WLODEK RABINOWICZ

DOES PRACTICAL DELIBERATION CROWD OUT SELF-PREDICTION?

ABSTRACT. It is a popular view that practical deliberation excludes foreknowledge of one's choice. Wolfgang Spohn and Isaac Levi have argued that not even a purely probabilistic self-prediction is available to the deliberator, if one takes subjective probabilities to be conceptually linked to betting rates. It makes no sense to have a betting rate for an option, for one's willingness to bet on the option depends on the net gain from the bet, in combination with the option's antecedent utility, rather than on the offered odds. And even apart from this consideration, assigning probabilities to the options among which one is choosing is futile since such probabilities could be of no possible use in choice. The paper subjects these arguments to critical examination and suggests that, appearances notwithstanding, practical deliberation need not crowd out self-prediction.

As is well known, Kant has argued for the existence of two fundamentally different perspectives on action: While an action can be seen as a natural event that falls under causality of nature, it can also be viewed from the fundamentally different perspective of freedom. While Kant does not quite say this, it might be tempting to argue that only from the former perspective the action is *predictable*, for only as natural events can actions be determinable from the past events via general laws. From the perspective of freedom, actions can be justified but they cannot be predicted.[1]

For Kant, this contrast between the two perspectives applies both to my current options *and* to the actions done by other persons, or by myself at other times. I can view each such action in two different ways. An alternative standpoint, which still is somewhat Kantian in spirit, would instead reserve the perspective of freedom primarily to the actions that are subject to my current deliberation and choice. On this view, the relevant distinction is between the *first-person* perspective of a practical deliberator and the *third-person* perspective of an observer.[2] While the observer can predict what I will do, I can't, insofar as I deliberate upon what is to be done. Deliberating in this way is incompatible with predicting the outcome of deliberation. To put it shortly, *deliberation crowds out prediction*.[3]

This claim allows for at least two interpretations:

Erkenntnis **57:** 91–122, 2002.

Weak Thesis: In a situation of choice, an agent does not assign extreme probabilities, one or zero, to options among which his choice is being made.

Strong Thesis: In a situation of choice, an agent does not assign any probabilities at all to options among which his choice is being made.[4]

As for the weak thesis, some of its variants instead deny that the agent, in a situation of choice, can be *certain* that he will, or will not, choose a certain option, or that he can *believe* this. On some other interpretations, again, what is being denied is the possibility of *knowledge* of the choice to be made. Among the proponents of the weak thesis in its various versions we find such philosophers and decision theorists as Shackle, Ginet, Pears, Goldman, Jeffrey, Schick and Levi.[5]

In the next section, I will shortly consider the theoretical significance of the predictability issue. Would it matter very much for our theories of practical rationality if the two theses were accepted? After this introductory discussion, I will focus on the *strong thesis*, whose foremost defenders have been Wolfgang Spohn and Isaac Levi (cf. Spohn (1977) and (1978), Levi (1989), (1991), and (1997), Introduction and chapters 2, 4 and 5). I will critically examine Spohn's and Levi's arguments and then, in the last section, I will adduce some positive reasons for rejecting the thesis in question.

As an aside, note that the two theses need not be interpreted so radically as to imply that a person simply cannot predict what he is about to choose. A weaker, and more plausible interpretation might be that these predictions are available to a person in his purely cognitive or doxastic capacity but not in his capacity of an agent or practical deliberator.[6] The agent cannot simultaneously see an action as an object of choice and as an object of prediction. But he can freely switch between these two perspectives. To put this in a fashionable terminology, our mind is *modular*, and it is reasonable to assume that the 'credence module' that is responsible for cognitive assessments is distinct from the module that is responsible for choices.[7] The two modules cooperate with each other to some extent. In particular, some input from the credence module is needed by the choice module to make a choice. Choice of an action often requires an assessment of the probability of its various possible consequences. But the two modules are to some extent mutually independent and separate. Therefore, it is conceivable that certain cognitive assessments need to be screened off from the choice module in order for the decision to be possible. A defender of the two theses might claim that this applies, in particular, to the predictions concerning those actions among which the choice is made.

Here is why such a modular interpretation of the two theses might be attractive. Suppose the agent faces a choice at t_1, with A being one of the options among which the choice is being made. If, as Levi suggests (cf. Levi 1997 (1991), pp. 80), there is nothing that hinders the agent from assigning probabilities to the options he will face in the *future*, as opposed to the options he faces at a given moment, we may suppose that at some earlier time t_0 he did assign a definite probability to his choice of A at t_1. However, at t_1, if the strong thesis is true, he can no longer assign any probability to A. How is this probability loss to be accounted for? Levi would say that the agent needs to 'contract' his probabilistic belief about A in order to make a choice. Note, however, that, at t_0, the agent might well have based his probability assignment to A on some reliable evidence. When t_1 comes, the same evidence may still be available to him and he may well remember how he arrived at his probability estimate at t_0. Still, at t_1, if the strong thesis holds, this probability estimate must be given up. How is this possible? Here, the modular version of the strong thesis might be of help.[8] According to it, the probability assignment to A may still be available to the subject in his purely doxastic capacity but not in his capacity of an agent or practical deliberator. The agent *qua* agent must abstain from assessing the probability of his options. Or, at least, this is what the defender of the strong thesis would want us to believe.

1. WHAT DOES IT MATTER?

It has been argued that the two theses, if accepted, would have far-reaching consequences for decision theory and for game theory.[9] Thus, for example, the standard game-theoretical assumption of *common knowledge of rationality* would have to go, for it implies that each player knows that he himself is rational (= acts rationally). Game theorists often assume that every game has a definite set of rational solutions, which can be identified by each player. Therefore, if the player knows he is rational, he must know he will do his part in one of these solutions. Consequently, if an option A does not belong to any such solution, the player must know he will not perform A. But this is incompatible with the weak version of the thesis.

Similarly, the assumption of the players having *common priors on their joint action space* will have to go, since it presupposes that each player has prior probabilities for all combinations of the players' actions, including his own. The probability of an action is just the sum of the probabilities of all possible action combinations that contain the action in question.

Consequently, the assumption of common priors entails that a player assigns probabilities to his own actions, in violation of the strong thesis. The assumption of common priors has been used by Aumann (1987), in his well-known argument for correlated equilibrium as an appropriate solution concept.

To take another example, *evidential decision theory*, such as the one proposed by Richard Jeffrey in *The Logic of Decision* (1983, 1st ed. 1965), defines the expected utility of an action A (or its 'desirability' in Jeffrey's terminology) in terms of the agent's conditional probabilities for various consequences given A. But $P(B/A)$, the conditional probability of a consequence B given an action A, is supposed to be defined as the ratio $P(B\&A)/P(A)$, which means that such conditional probability is ill-defined if the probability of the condition (action A) is not available. Thus, if the agent cannot assign probabilities to his actions, evidential decision theory would be in trouble. (This last example comes from Spohn (1978), while the first two are due to Levi (1997), ch. 5.)

Personally, I am not quite convinced by the suggestion that the acceptance of the two theses would have very dramatic consequences for theorising about games and decisions. To take the last example first, note that, according to the standard evidential decision theory, probabilities of actions play no role in choice. What does play such a role are conditional probabilities of various consequences given actions. Therefore, it is possible to argue that the agent might still have conditional probabilities for consequences given actions even though he has no probabilities for the actions themselves.[10] The agent's conditional and unconditional probability assignments may contain *gaps*.[11] It is enough to require that such a 'gappy' or partial assignment of conditional and unconditional probabilities is internally consistent, i.e., that it could in principle be extended to a complete assignment on which the conditional probabilities equal the ratios of the relevant unconditional probabilities. (In fact, even that requirement may be too strong if we want to allow for cases in which the agent has a conditional probability for a proposition B given a condition A, even though he assigns probability zero to A. Perhaps then we should only require that the agent's partial probability assignment P is extendable to a complete assignment P^* on which, for any A and B, if $P^*(A) > 0$, then $P^*(B/A) = P^*(B\&A)/P^*(A)$.)

Note, however, that Jeffrey's theory, as it was originally developed, did imply that the agent must have unconditional probabilities for options. To be compatible with the strong thesis, evidential decision theory should therefore be constructed in a more cautious way, to avoid this implication. In particular, the conditions on the agent's preference ordering on propositions, from which we determine his probability and utility assignments, must be considerably weakened to make room for probability gaps. Note also that conditional probability assignments may *indirectly* induce unconditional probabilities.[12] If the agent has definite conditional probabilities for the consequences given the actions *and* for the actions given the consequences, then these conditional probability assignments are jointly sufficient to determine his unconditional probabilities for actions. From the four conditional probabilities, $P(B/A)$, $P(B/\text{not-}A)$, $P(A/B)$, and $P(A/\text{not-}B)$, we can determine the unconditional probability of A, by solving two equations with two unknowns, $P(A)$ and $P(B)$:

$$P(A) = P(B) \times P(A/B) + (1 - P(B)) \times P(A/\text{not-}B)$$

$$P(B) = P(A) \times P(B/A) + (1 - P(A)) \times P(B/\text{not-}A))$$

On the other hand, an appropriately weakened evidential decision theory need not imply that all these conditional probabilities are defined. What it requires, for its applicability, is that the probabilities of consequences given actions are available, if the expected utility of the actions is to be defined. But the converse conditional probabilities (from consequences to actions) are not needed for the theory to apply. Still, it is important to recognize that detailed probabilistic information about various evidentiary relationships between actions and consequences may indirectly induce unconditional action probabilities. It is also important to recognize that Jeffrey's decision-theoretic framework, unlike the one developed by Savage, takes act propositions to be just a sub-class of propositions in general. This makes it difficult to treat acts in a special way.[13]

As to the first example, it has been argued by Schick (2000) that the assumption of common knowledge of rationality is unnecessary for game-theoretical proofs. To begin with, for most purposes it would be enough to assume common *belief* in rationality rather than outright knowledge. In particular, we need not require this belief to be veridical. The players make their choices on the basis of what they believe, whether or not these beliefs happen to be true. The efficacy of a belief, as far as choice is concerned, does not depend on its epistemic status. Replacing common knowledge with common belief is only the first step; by itself, it does not yet get us off the hook. Common belief in rationality implies that the agent believes that he himself is rational. But such a belief is incompatible with the weak thesis if we suppose that the agent has definite views as to what is rational for him to do. However, it seems that we could safely replace the assumption of common belief with a weaker requirement of *mutual* belief in rationality, without serious losses as far as game theory is concerned. The assumption of common belief in rationality requires everyone to believe that everyone is rational, that everyone believes that everyone is rational, etc. By contrast, the assumption of mutual belief in rationality only requires everyone to believe that *everyone else* is rational, that everyone believes that *everyone else* is rational, and so on.[14] Unlike common belief, mutual belief does not imply that any player believes that he himself is rational, which means that there is no danger of the weak thesis being violated. On the mutual-belief assumption, I believe that you believe that I am rational, but I need not believe that this belief of yours must be veridical.[15] Now, while reasoning about a game, it is often important for a player to ask whether *other* players are rational and whether they consider *him* to be rational. But he need not take a stand on his own rationality: To determine what action it is rational for him to choose, he need not assume that the action he will actually choose *will* be rational. The requirement of mutual belief in rationality seems therefore to be a satisfactory replacement for the requirement of common belief.

Extensive-form games are different: Unlike one-shot interactions, they do make a player's rationality an important consideration from his own point of view. To determine what action he should choose on a given occasion, the player may need to predict his future choices. And in making such predictions he may need to rely on his future rationality. (This applies, of course, not just to extensive-form games, but also to one-person sequential decision problems.) But the claim that deliberation crowds out prediction is by most of its defenders only meant to apply to the actions that are subject to the agents' *current* choice, and not to the actions he will decide on in the future. (For more on this point, see below.)

I have been arguing that the acceptance of the strong thesis need not have far-reaching effects on game theory and decision theory. Still, it might turn out that accepting it could create problems in some special areas. A case in point might be the assumption of the common priors on the players' action space. I do not know whether and how this assumption can be relaxed to avoid a clash with the strong thesis, if the game-theoretical results that are based on this assumption are to survive.[16] Which means that, my doubts notwithstanding, the strong thesis may after all have serious repercussions for some of our theories about practical rationality.

2. PROBABILITIES AND BETTING DISPOSITIONS

The common ground for Spohn's and Levi's arguments for the strong thesis is the assumption that the agent's probability assignments are his guides to action. As such, these assignments are related to his betting dispositions. In fact, the first three arguments for the strong thesis to be considered below assume that the agent who has a probability for a proposition is committed to the corresponding *betting rate* for the proposition in question. Clearly, some of us might want to reject a strict connection between probabilities and behavioural dispositions. While probability assignments normally are manifested in the agent's readiness to accept bets, the connection between the two might not be very tight. This would make it possible for some probability assignments not to have any direct behavioural manifestations in betting dispositions. Still, to understand their arguments, we should concede to the supporters of the strong thesis this connection between probabilities and betting rates as their point of departure. Let me therefore clarify this crucial notion of a betting rate before we go any further.

In what follows, we shall use the following notation for bets: $b^A_{C,S}$ is a bet on a proposition (or event) A that costs C to buy and pays S if won. (S and C are monetary amounts.) S is the *stake* of the bet (the prize to be won), while C is its *price*.[17] A bet shall be said to be *fair* if and only if the agent is prepared to take each side of the bet: buy it, if offered, or sell it, if asked. To pronounce a bet as fair, relative to a given agent, is thus to ascribe to the agent a certain betting disposition or a commitment to a certain betting behavior.

Note that the existence of a fair bet on a proposition, with a given stake, presupposes that there is some price C such that the agent is just as prepared to buy and to sell the bet for that price. We will be assuming that, for a given A and S, this C is *unique*. (We need this assumption, if we want to ascribe to the agent definite probabilities and at the same time to measure these probabilities in terms of fair bets; cf. below.) If the agent is prepared to buy the bet for a price C, then he should be prepared to buy it for a lower price. And if he is prepared to sell it for C, then he should be prepared to sell it for a higher price. Thus, we may think of C as the highest price the agent is prepared to pay for the bet and as the lowest price he is prepared to sell it for. The assumption behind the idea of a unique fair bet for any given A and S is that the agent's maximal buying price and his minimal selling price coincide.[18]

Another assumption we make is that, for various fair bets on A, with different stakes and prices, the *ratio* between the price and the stake remains constant. If the price is increased or decreased, then the price for a fair bet is increased or decreased in the same proportion. This simplifying assumption (which follows if we suppose that the agent is seeking to maximize his expected monetary payoff) is reasonable at least within a certain range, in which the monetary amounts S and C are not too high. We shall call this constant ratio the *betting rate* for A. I.e., the *betting rate* for A is the quotient C/S for a fair bet on A. The agent's *probability* for A, $P(A)$, can be identified with his betting rate for A. (Again, this identification will follow, if we take the agent to maximize his expected monetary payoff.) Probabilities can be measured by betting rates since the agent's betting dispositions are determined by his probability assignments. Example: If a bet on A with a stake $S = 200$ and a price $C = 90$ is fair for an agent, then his betting rate for A equals 90/200 = 0.45, which means that we can set his probability for A as equal to 0.45. The agent's betting rates are *coherent* if he is not vulnerable to a Dutch book, i.e., if no system of fair bets on various propositions would give him a positive loss whatever happens. As is well known, if probabilities are measurable by betting rates, the coherence of betting rates is equivalent to the assumption that the agent's probabilities obey the standard probability axioms.

If the monetary amounts S and C are not too large, the assumption of the constant ratio for all fair bets on a given proposition is not problematic, for in those cases we may assume that utility is proportional to money. If, however, we want the ratio C/S to be constant within a wider range, then S and C would need to be interpreted directly as amounts of *utility* rather than as monetary amounts. Given this interpretation,[19] which we shall for simplicity adopt in what follows, the expected utility of buying a fair bet $b^A_{C,S}$ for A is zero:

$$eu(\text{Buy } b^A_{C,S}) = P(A) \times (S - C) + (1 - P(A)) \times (-C) = C/S \times (S - C) - (1 - C/S) \times C = 0.$$

Similarly, of course, selling such a fair bet has an expected utility equal to zero:

$$eu(\text{Sell } b^A_{C,S}) = P(A) \times (-S + C) + (1 - P(A)) \times C = C/S \times (C - S) + (1 - C/S) \times C = 0.$$

Now, Spohn's and Levi's arguments for the strong thesis are meant to show that, *on pain of contradiction or incoherence, the agent cannot have betting rates for the actions that stand at his disposal*. Given the connection between probabilities and betting rates, this means that the agent cannot coherently assign probabilities to such actions. Thus, if the arguments hold, the strong thesis will be established.

3. ARGUMENT I (SPOHN)

Consider the following quote from Wolfgang Spohn:

The agent's readiness to accept a bet on an act does not depend on the betting odds [or betting rate, which amounts to the same[20]] but only on his gain. If the gain is high enough to put this act on the top of his preference order of acts, he will accept it, and if not, not. The stake of the agent is of no relevance whatsoever. (Spohn 1977, p. 115; comment in square quotes added. For the same argument, see also Spohn 1978, p. 73.)

The argument may not be crystal clear, but here is what I take it to be its plausible interpretation: The "gain", G, of the agent who accepts and wins a bet $b^A_{C,S}$ is supposed to be his net gain: the stake he wins minus the cost. I.e., $G = S - C$. Suppose, for simplicity, that the agent has a choice between two actions, A and B, where $eu(A)$ and $eu(B)$ are their expected utilities for the agent *disregarding any bets that he might place on the actions themselves*. Now, if he takes a bet on A with a net gain G, his expected utility of A will instead be $eu(A) + G$. The reason is obvious: If that bet is taken, then, if A is performed, the agent will receive G in addition to $eu(A)$.

Now, as far as I can see, it is implicitly assumed that the agent is certain of his own rationality, i.e., that he is certain he will perform the action with a maximal expected utility. On this assumption, he should take the bet on A if

$$eu(A) + G > eu(B).$$

For under these conditions, performing A will in fact maximise his expected utility (given that he bets on A) and thus the agent may rest assured that he will perform A if he takes the bet. This makes betting on A the right thing to do. Analogously, he should abstain from the bet on A if

$$eu(A) + G < eu(B).$$

But if the agent should take *any* bet on A, as long as the gain offered by that bet is sufficiently large, then it follows that the betting rate for A does not exist! For when we change the values of S and C, the size of the gain $G = S - C$ may well remain unchanged, or even increase, even though the quotient C/S decreases. Analogously, the quotient may remain unchanged, even though G decreases so much as to make the bet on A unattractive. Thus, the attractiveness of the bet does *not* depend on this quotient, but only on the value of G.

Here is an example. Suppose that $eu(A) < eu(B)$, but $eu(A) + G > eu(B)$ if and only if $G > 4$. That is, as long as $G > 4$, taking the bet on A is well-motivated. Let $C = 10$ and $S = 15$. Then, $G = 15 - 10 = 5$, which rationalises taking that bet on A. However, if C and S drop down to 4 and 6, respectively, the quotient C/S remains unchanged but G falls below 4, which makes the bet on A unattractive.

Thus, the attractiveness of the bet has nothing to do with the quotient C/S. The notions of a betting rate and of a fair bet are therefore inapplicable when we consider bets on one's own actions. Consequently, insofar as the agent's probabilities are manifested in his betting rates, the agent has no probabilities for his own actions. End of the argument.

To put this point in another way, suppose, for *reductio*, that the agent's betting rate for A equals x, for some x between 1 and 0. Now, if x is his betting rate, then we would expect him to decline buying an unfair bet $b^A_{C,S}$, in which $C/S > x$. However, this expectation is unjustified. As argued by Spohn, the agent may be assumed to buy the bet if only his net gain $G(= S - C)$ will be sufficiently high to put A on top of his preference ordering. Whether C/S is higher than x or not does not matter.[21]

Criticism: The argument, as I have re-constructed it, presupposes that the agent is *certain* that, if he takes the attractive bet on his action, he will perform the action in question. Otherwise, if there is some uncertainty on this account, the agent might well abstain from taking the bet even if the net gain G is high enough to make the sum $eu(A) + G$ larger than $eu(B)$. For he might not be assured of winning the bet in the first place!

Wouldn't it be enough for the agent to be certain that the action is under his direct control in order to be certain that he will perform the action if he bets on it?[22] The answer is that it wouldn't be enough. For the agent may believe to have such a direct control, i.e., believe that his decision to act would be fully efficacious in bringing about the action, but he may still be uncertain whether his bet on action A would be followed by the decision to perform A. Betting on an action is not the same as deciding to perform it.

Now, what about this assumption that the agent is certain that he will perform the action if he bets on it (and the net gain G is high enough)? The assumption is justified if the agent is certain of his own rationality. But, as we have seen above (in the introductory section), this kind of certainty about one's own rationality is really incompatible with the weak thesis (provided that the agent can work out which of his actions are rational). Spohn's argument for the strong thesis would therefore, on this reading, rest on the denial of the weak thesis, which is surely not his intention.

However, Spohn's main idea could be used to set up an argument that avoids this strong certainty assumption. Essentially the same point that Spohn has made could also be made in another way. The argument that follows shows how this can be done.

4. ARGUMENT II

Again, we assume that A and B are alternative actions that are available to the agent. Suppose, for *reductio*, that $P(A)$ is well-defined. Suppose also that

$$eu(A) < eu(B), \text{ but } eu(A) + G > eu(B).$$

In this version of the argument, we no longer assume that the agent is *certain* that he will take the action A if he bets on it. It is enough to suppose that he considers this to be relatively *probable*.[23]

If no bet on A is offered to the agent, the agent does not think it probable he will perform A. $P(A)$ is relatively low. But, if a bet on A is offered, with the net gain equal to G, $P(A)$ increases. The reason is obvious:

(i) The agent thinks it probable that

he will perform A *if* he takes the bet

(because $eu(A) + G > eu(B)$, and therefore $eu(A) + G > eu(B) - C$).[24]

At the same time,

(ii) The agent thinks it probable that

he will take the bet,

because of (i) together with the fact that the gain from the bet would move A to the top of his preference ordering (i.e., because $eu(A) + G > eu(B)$).

Thus, *the probability of an action depends on whether the bet is offered or not*. But then the probability of an action cannot be measured by the agent's betting rate, for the offer of a bet itself changes the agent's probabilities. The betting rate on A, if it exists, is a certain disposition to take bets on A if offered. But surely such a disposition cannot depend on whether a bet on A is offered or not.

Conclusion: If probabilities are to correspond to betting rates, the probabilities of actions under deliberation are not well-defined.

Criticism: This argument, if correct, proves too much – much more than it is meant to prove. Let me explain. Let A be an action that will be available to the agent at some point in *the future*. Suppose that the expected utility of A is relatively low. However, if the bet on A is offered, the promised gain G, if sufficiently large, makes A an attractive prospect. But then, by the same reasoning as in the argument above, the agent's probability for A increases as soon as the bet offer is made. For the agent expects to perform A *if* he takes the bet. Therefore, and because the gain from the bet, if added to the original expected utility of A, would make A an attractive prospect, he expects to take the bet. Which means that the offer of a bet on A increases the probability of A for the agent. Consequently, for *all* future actions, their probabilities cannot be defined if such probabilities are supposed to correspond to the agent's betting rates.

This point can also be made in another way. If a bet on a future A would now be offered to the agent, taking that bet would be a *current* option on his part. Thus, if Argument II is correct, the agent would have no probability for taking this bet. But A is supposed to be probabilistically dependent on that current action: The probability of A is higher on the assumption that the bet is taken than otherwise. Therefore, *if* the unconditional probability of A were defined, the agent would be able to simply *calculate* his probability for taking the bet, by the probability calculus:

$$P(\text{bet is taken}) = [P(A) - P(A/\text{the bet is not taken})]/[P(A/\text{the bet is taken}) - P(A/\text{the bet is not taken})],$$

provided that A is probabilistically dependent on taking the bet, i.e., provided that the denominator in the formula above does not equal zero.

Thus, the agent cannot assign any probability to his future action because this would mean that, if he now were offered a bet on that action, he would have a probability for taking that bet, contrary to the hypothesis.

Some philosophers might be prepared to bite the bullet and swallow the conclusion that the strong thesis should be extended to the agent's future actions. (Spohn (1978) seems to entertain such an extension.[25] And Schick (2000) suggests, with respect to the weak thesis, that the thesis in question applies to the options the agent expects to confront in the future.[26]) But many others will balk at this implication. It is one thing to deny the agent probability assignments to actions that are currently subject to his choice (or to the future events that would be influenced by these actions). It is quite another, and much more disputable claim to deny him probabilities for all the actions that are subject to his *future* choices (and for the events that would be influenced by these actions). As Levi, one of the main defenders of the strong thesis, puts it: "in [a] sequential choice problem [. . .], X [the agent] at the initial state might very well assign credal probabilities concerning how he will choose at stage 2 if he chooses A at stage 1 or if he chooses B at stage 1.[27] But at stage 1, X does not regard his options at stage 2 as available to him at stage 1. His choice at stage 1 is between A and B." (Levi 1997(1991), p. 80) According to Levi, the agent can take the same predictive attitude towards his own future choices ("towards his own choices at times other than the time at which deliberation is taking place") as he takes towards the choices of others (*ibid.*).[28]

Of course, *if* a future action A is probabilistically dependent on some action A' currently available to the agent, then the strong thesis will apply not just to A' but to A as well. As we have seen above, the strong thesis applies to all the events that are probabilistically dependent on the current options. However, some of the future actions may not be like this. Some of them may well be both probabilistically and causally independent of the options at hand. For some actions, nothing the agent can do at present can influence whether he will perform these actions in the future. Indeed, the agent may currently not confront any choice at all! Of course, *if* the agent were currently offered a bet on a future action, then he would be able to do something (accept the bet) that could influence his future performance. But that's beside the point: No bet offer need actually be available to the agent. It would be deeply worrying if the argument used to establish the strong thesis could be extended in such a way as to imply that predicting one's future actions is never possible.

What can we say about this difficulty? How can one stop Argument II from proving too much? A possible way out might be to assume a certain kind of *forgetfulness* on the part of the agent, with respect to the hypothetical bets that would test his probabilities. Thus, consider such a hypothetical betting offer, on some future action A. Suppose the offer comes together with the information that *the agent, if he takes the bet, will immediately forget about it*. (Still, if he takes it, the bet will be in force, but this will be revealed to him only *ex post facto*, after the action A has, or has not, taken place.) By taking such a bet, the agent would not acquire any additional future motive to perform A. But then the offer of the bet on A does not change the agent's probability for A: He realises that taking the bet will not change his future motives, at the time when A will be available for choice. Consequently, his probability for A *can* be measured by his betting quotient for fair bets of this kind. For this probability for A does not depend on whether the bet on A is offered or not.

But can't one use the same forgetfulness manoeuvre not only with respect to future actions but also in those cases when A is currently available for choice? The agent's probability for such an action A could, it seems, be elicited without inconsistency by asking what side bet on A he would now be prepared to accept, *if* he thought that this bet would be immediately forgotten by him and thus have no influence on his decision about A. On this interpretation, the forgetfulness manoeuvre is a very strong medicine: It allows us to avoid the untoward consequences of Argument II, but the price is that the argument itself becomes invalid! (The same applies, by the way, to Argument I as well. In that argument, as we remember, the suggestion was that the net gain G from the bet is the only thing the agent needs to care about, *provided* he is certain he will perform action A if he takes the side-bet and G is high enough. But to be certain of this he must believe that, in his decision on A, he will take the side bet into consideration.)

On the other hand, one might think that the forgetfulness manoeuvre as such is really implausible. If we need to assume that the agent forgets the relevant bets as soon as he takes them, then it is clear that the bets in question are pure *creatures of fiction*. But is it plausible to determine probabilities – our guides to real action – in terms of our willingness to accept such purely fictional bets?

At this point, we might think of the following reply from the defender of the strong thesis: Eliciting probabilities from the agent's dispositions to take very artificial bets is not really, he might say, an implausible procedure. From such dispositions we might be able to draw indirect conclusions about the agent's dispositions to real action. Note that even in the case of betting on *act-independent* events, the identification of probabilities with betting rates requires quite radical idealization assumptions. Receiving a bet offer on an act-independent event E would often decrease the agent's probability for E, simply because of the natural suspicion on the agent's part that the person who offers the bet might possess some private information in E's disfavour. Otherwise, why would she be offering the bet in the first place? When probabilities are identified with betting rates, however, real-life phenomena like this are idealized away. We postulate fictitious bet offers made by a bookmaker who is known to lack any hidden motives, perhaps offers generated by a random mechanism, in order to avoid the possibility that the bet offer itself would influence the agent's probabilities for the event on which the bet is to be made.

Therefore, the forgetfulness manoeuvre, while relying on pure fiction, is still acceptable as long as the fiction in question at least represents a logical possibility. But this means that the distinction between current and future choices can be upheld. The forgetfulness manoeuvre can only work for the bets on *future* actions. If the decision on an action instead is to be made *now*, simultaneously with the decision on the bet, there is simply no time to forget. The decision on the action and the decision on the bet are taken simultaneously. In this way, it seems, we can save Argument II and still avoid its untoward implications for the agent's probabilities for his future behaviour.

This reply, however, with its heavy stress on the issue of *simultaneity*, introduces a new consideration into our problem. If we take seriously the idea that the decision on A is to be made simultaneously with the decision on the bet, then it seems we must combine these two decision problems into *one*. It appears, therefore, that our options are not, on the one hand, to do A or to do B, and, on the other, to take the bet on A or to abstain. Rather, what we must choose between are the following four complex options:

take the bet on A & do A,

take the bet on A & do B;

abstain from the bet on A & do A,

abstain from the bet on A & do B.

This idea of *fusing two decision problems into one* lies behind Levi's argument for the strong thesis (cf. Levi (1989) and (1991)).

5. ARGUMENT III (LEVI)[29]

Levi starts with the weak thesis:

Claim 1: If an agent is certain that he won't perform an option, then this option is not feasible (i.e., it is not available for his choice).

According to Levi, an action must be doxastically possible to be a live option for the agent. If I am sure I will not perform A, then A is not a feasible option for me, even though I may still believe that doing A is in my causal power. In fact, A is not feasible under these circumstances even if I consider doing A to be preferable. As Levi puts it:

> Once the matter is settled in this respect, so that [performing an action] is no longer consistent with what [the agent] takes for granted, feasibility is also precluded. (1997 (1989), p. 28)

This restriction of feasible options to doxastically possible alternatives is clearly a controversial view.[30] But, in what follows, I am prepared to grant Levi that premise, for argument's sake.

In the second part of his argument, Levi provides a proof for the following claim:

Claim 2: If the agent assigns probabilities to options, then, on pain of incoherence, his probabilities for inadmissible (= irrational) options, as revealed by his betting dispositions, must be zero.

We shall consider and discuss his proof of that claim below. For now, let us note that Claims 1 and 2, taken together with the antecedent of Claim 2, imply that:

Only admissible options are feasible.[31]

This implication, however, cannot be accepted if the principles of rational choice are to have some normative bite as far as the agent is concerned. In this capacity, the principles in question are supposed to constrain the agent's choices by telling him not only what he should do but also what he should *not* do.

> When used for self-policing, the applicability of the principles should be *nonvacuous* in the sense that a nontrivial distinction may be made between feasible options which are admissible for choice and others which are not. (Levi 1997, pp. 25f, my emphasis)

Since one must allow the possibility of feasible but inadmissible options, Levi concludes that the antecedent of Claim 2 must be rejected. On pain of incoherence, the agent cannot assign probabilities to his feasible options. Which means that the strong thesis holds. Needless to say, this conclusion depends on Claim 2 being valid. Let us consider, therefore, how Levi argues for this crucial claim.

6. LEVI'S ARGUMENT FOR CLAIM 2[32]

For simplicity, let us suppose there are just two alternative options, A and B, with option B being inadmissible because of its lower expected utility:

$$eu(A) > eu(B).$$

Assume, for *reductio*, that $P(A)$ and $P(B)$ are well-defined. Let

$$P(A) = \text{the betting rate for } A = x.$$

Suppose that the agent is offered a fair bet b on A, with a positive stake S and a price C. Since b is fair, $C/S = x$. Since $1 \geq x \geq 0$ and $S > 0$, it follows that $S \geq C \geq 0$. Therefore, $G = S - C \geq 0$. Given the bet offer, the agent's decision problem has been altered. He has four options, instead of two:

$b\&A$ (taking bet b and performing A), $\neg b\&A$, $b\&B$ and $\neg b\&B$.

The expected utilities of these options are as follows:

$$eu(b\&A) = eu(A) + G,$$

$$eu(\neg b\&A) = eu(A),$$

$$eu(b\&B) = eu(A) - C,$$

$$eu(\neg b\&B) = eu(B).$$

Clearly, option $\neg b\&B$ dominates $b\&B$ (= taking the bet on A and performing B instead), if $C > 0$, or at least is as good as the latter option, if $C = 0$. Similarly, $b\&A$ dominates $\neg b\&A$, if $G > 0$, or at least is as good as that option, if $G = 0$. Thus, in what follows, we may safely concentrate on just $b\&A$ and $\neg b\&B$. At least one of them must be admissible. Among these two options, $b\&A$ and $\neg b\&B$, the former has a higher expected utility than the latter. Since $G \geq 0$ and $eu(A) > eu(B)$,

$$eu(b\&A) = eu(A) + G > eu(B) = eu(\neg b\&B).$$

Note that this inequality must hold even if the agent's net gain $G = S - C$ equals 0. But then it follows that the agent should be willing to accept the bet on A even if $S = C$, i.e., even if $S/C = 1$. Consequently, the (fair) betting rate x for A must equal 1, i.e., $P(A) = 1$. Which implies, on pain of incoherence, that $P(B) = 1 - P(A) = 0$. The inadmissible option has probability zero. Q.E.D.

Side Comment: If this argument for Claim 2 is correct, it proves more than Levi appears to claim. To see that, note that the conclusion above, that $P(A)$ if defined must equal 1, would follow even if $eu(A)$ did not exceed $eu(B)$. For the argument to go through, it is enough to require that $eu(A) \geq eu(B)$. For then, as long as $G \geq 0$, $eu(A) + G \geq eu(B)$. The agent should therefore be willing to accept the bet on A even if $S/C = 1$. Therefore, in the decision problems that contain *more than one admissible option*, Levi's proof of Claim 2 could be adjusted to generate the incoherent conclusion that the probability of each of these admissible options is 1 (if they are assigned any probabilities at all). Thus, assume that $eu(A) = eu(B)$ and suppose that the agent can, if he wishes, take a bet on A or a bet on B (but not both), with price C and stake S being the same in both cases. Now we have to consider six complex options rather than four: Performing A can be combined with taking the bet on A, or with taking the bet on B, or with refusing to take any of these bets. Similarly for B. By the same reasoning as in Levi's argument, we can then show that the agent should be willing to perform-A-and-take-the-bet-on-A and just as willing to perform-B-and-take-the-bet-on-B, even when the net gain $G = S - C$ for each of these bets is 0. But this would imply, if Levi is right, that the agent's betting rates for A and for B, if they existed, should both be equal to 1, which is incoherent. This line of thought would be enough to establish the strong thesis – relying on Claim 1 would not be necessary!

Criticism of Levi's argument: Levi argues that the agent is willing to bet on the admissible action A even if the net gain from that bet is zero and there is a theoretical danger of a positive loss in case A will not be performed. Thus, he argues, the agent's betting rate for A, *if* it exists, must be 1. But has he really shown this? What has been shown is that the agent is prepared to opt for a *combination* of actions: to do A *and* to take the unattractive bet on A. We also know that he is prepared to do one part in this combination, action A, even when he considers A on its own, abstracting from the bet. For we know that A is admissible. Furthermore, his willingness to opt for the combination is not weaker than his willingness to do A: The expected value of the combination is as high as the expected value of A. Does it then follow that he must be prepared to do the other part as well, i.e., to take the bet, *when that part is considered on its own*?

Surely, if the agent has any doubts as to whether he will perform A if he takes the bet, and if the cost of the bet is positive ($C > 0$), then he may well be wary of taking that bet on A if the most he can gain is nothing at all (i.e., if $G = 0$). Thus, being prepared to do a combination of the actions, and being prepared to do one part in this combination, does not imply that one is prepared to do the other part, considered on its own. Not even if the combination is no less valuable than the part he is prepared to perform. A whole may be at least as good as one of its parts, even if the other part is bad. A situation like this violates the separability principle for value: The parts do not make independent value contributions to the value of the whole. As a consequence, the value of the whole is not a monotonically increasing function of the (independent) values of the parts. But such violations of separability are to be expected. What we encounter in this case is a complementarity effect: The value of one component (the bet on A) increases in the presence of the other component (A itself).

Here is an analogy that should clarify this point. A motorist who is driving behind a slow-moving vehicle may be prepared to perform a complex action: to swerve to another lane and to accelerate, at the same time. He may also be prepared to swerve. Furthermore, he is more willing to swerve-and-accelerate than just to swerve. But if he has some doubts as to whether he *will* swerve if he accelerates, he is not prepared to accelerate. The complex action is admissible, but one of its components is not, if considered on its own.

Now, it seems that the fair betting rate for A can only be determined by considering what bet on A the agent is willing to take, period. What bet on A he is willing to take in *combination with* A is irrelevant for the specification of his betting rate for A. It seems, then, that Levi has not been successful in his argument for Claim 2.

7. DISCUSSION

At this point, I envisage the following retort from a defender of the strong thesis:

You want us to consider the agent's attitude to the bet on A, when that bet is considered on its own. This means we should assume that the agent makes *two independent but simultaneous choices*, one with respect to A and another with respect to the bet on A. One might wonder whether your demand is conceptually coherent. If the two choices are made simultaneously, by the same agent, then they may *have* to be seen as one choice. Bets on actions that are to be chosen at the same time when the bets are being made just cannot be considered on their own. This would by itself suffice to establish the strong thesis. If the existence of a betting rate requires that the relevant bet is considered on its own, independently of the event on which the bet is being made, it is conceptually incoherent to talk about betting rates for actions that are currently subject to choice. However, let us allow that perhaps we can keep the two choices apart, as you urge us to do. But then, as we have seen, we can establish the strong thesis via *Argument II*. According to that argument, the offer of a bet on an action would give the agent a reason to think he will accept the bet and thereupon perform the action. Thus, such a bet offer might well change his probability for the action, *if* such a probability exists. Consequently, such probability cannot exist if probabilities are essentially linked to betting rates (for such rates cannot be sensitive to whether the bet is offered or not). And remember that Argument II cannot be disarmed by the forgetfulness manoeuvre, for the simple reason that the agent has no time to forget if both choices are to be made simultaneously.

In my view, this reply is convincing as far as it goes: *If* we take all probabilities to be equal to betting rates, then we have to admit that the deliberating agent lacks probabilities for the options among which he makes his choice. His probability for an option cannot be an *ex ante* disposition to accept certain bets on that option. For the fact that a bet is being offered would give him new information with regard to the option in question. The probability cannot be a disposition to accept bets if offered, if the bet offer itself affects one's probability assignment. However, the conclusion I would like to draw from this argument is *not* that the strong thesis holds, but rather that the identification of probability assignments with betting rates must be put in question. Basically, probability is a measure of the strength of belief. Beliefs are our guides to action, and one form of action is betting. That much is true. Still, this does not mean that probabilities invariably can be measured in terms of betting rates: In those cases when bet offers themselves would influence our probabilities for the events on which the bets are made, probabilities no longer are translatable into betting dispositions. This does not mean, however, that probability estimates are impossible to make in cases like this. The correct conclusion is rather that the connection between probabilities and betting rates is not as tight as one might initially be tempted to think.

8. BETTING COMMITMENTS

Do we have to conclude that probabilities for one's current options must lack any connection at all to one's potential betting behaviour? This would be somewhat premature, I think. Certain kinds of betting arrangements that relate to the current options could still in some circumstances be proposed to the agent without influencing his motivations to act. Such offers would not provide him with new information with regard to the options in question. Consequently, his willingness to accept such betting arrangements could track his antecedent probabilities for options.

Here is how such a betting arrangement might be set up. Suppose the agent is offered an opportunity to make a *betting commitment* with respect to an option A, at a stake S and a price C. Accepting such a commitment means that the agent undertakes *to buy or to sell* a bet $b^A_{C,S}$, whichever the other party will require. The agent makes a commitment not knowing whether he will be required to sell or to buy the bet. As we remember, a bet is fair if the agent is just as willing to buy it as to sell it. Analogously, we might say that a betting commitment is fair if the agent is willing to accept it even if he is radically uncertain as to what will be required of him if he makes that commitment. Note that, given such radical uncertainty, the acceptance of the commitment does not provide the agent with any extra motivation to perform the option or to abstain. Thus, the offer of such a commitment does not provide him with any new information with regard to that option. His probability for the option is not affected.

Radical uncertainty about an event E is different from equal probability for E and for non-E. Instead, it can be seen as a state in which the agent has no determinate probability for E at all. We can represent the agent's uncertainty with regard to what will be required of him if he accepts the betting commitment as the set of probability distributions over {Buy $b^A_{C,S}$, Sell $b^A_{C,S}$}. In that set, the probability p of Buy $b^A_{C,S}$ varies from 1 to 0. Correspondingly, the probability $(1 - p)$ of Sell $b^A_{C,S}$ varies from 0 to 1. Now, when is a rational agent prepared to accept a betting commitment of this kind? What is the correct principle of choice under conditions of uncertainty? The issue is controversial, but consider what happens if we adopt a proposal that has been put forward in G,rdenfors and Sahlin (1982):

MMEU: In choice under uncertainty, the agent should *maximize the minimal expected utility*.

That is, on this proposal, the agent should determine each action's expected utility under every admissible probability distribution, and then choose the action whose minimal expected utility is as high as possible. In this sense, MMEU combines the expected utility maximization with the maximin idea of taking precautions against the worst case. (Gärdenfors and Sahlin assume a measure of epistemic reliability on the probability distributions that represent the agent's uncertainty and then take as admissible only those distributions whose epistemic reliability is high enough to be taken seriously, from the point of view of the decision maker. We can bypass this complication if we suppose that, in the envisaged situation, all the probability distributions over {Buy $b_{A,C,S}$, Sell $b_{A,C,S}$} are equally deserving to be taken seriously, and thus that all of them are admissible.)

The expected utility of *rejecting* the betting commitment is equal to 0 under all probability distributions. Consequently, given MMEU, the agent should be willing to *accept* the betting commitment if and only if the minimal expected utility of that commitment is not lower than 0. Now, for each admissible probability p, the expected utility of the betting commitment equals

$$p[eu(\text{Buy } b^A_{C,S})] + (1 - p)[eu(\text{Sell } b^A_{C,S})].$$

Letting $P(A)$ be the agent's probability for A, his expected utilities $eu(\text{Buy } b^A_{C,S})$ and $eu(\text{Sell } b^A_{C,S})$ equal, respectively, $P(A)S - C$ and $C - P(A)S$. Therefore, the expected utility of the betting commitment under a probability distribution p equals:

$$p(P(A)S - C) + (1 - p)(C - P(A)S).$$

For all values of p between 1 and 0, the value of this weighted sum lies between $P(A)S - C$ and $C - P(A)S$. Consequently, the minimal expected utility of the betting commitment is not lower than 0 if and only if none of these limiting values $P(A)S - C$ and $C - P(A)S$ is lower than 0:

(i) $P(A)S - C \geq 0,$

and

(ii) $C - P(A)S \geq 0.$

As is easy to see, (i) and (ii) hold if and only if $P(A)S = C$, i.e., assuming that $S \neq 0$, if and only if $P(A) = C/S$. Thus, the betting commitment with regard to A with a non-zero stake S and a price C is fair (= acceptable under conditions of radical uncertainty) if and only if the quotient between the price and the stake equals the agent's probability for A.[33]

To conclude: The probability for A determines the price/stake quotient in a fair betting commitment with regard to A. This means that the probabilities for one's current options can still be seen as being related to certain betting dispositions, namely to dispositions to undertake betting commitments under conditions of radical uncertainty.

Admittedly, the conditions of radical uncertainty are quite rare, to say the least. When the agent is offered a betting commitment, he might normally be expected to have at least some inkling as to what will be required of him if he accepts it. And as soon as he thinks it is more probable, or less probable, that he will be required to buy rather than to sell the bet on his option A, the offer of the commitment provides him with an extra motive to choose A, or to abstain. Thereby, the offer of a betting commitment will affect his probability for A. Thus, eliciting the agent's probabilities for his current options by offering him various opportunities for betting commitments is not a promising enterprise.

Despite this, we could insist that the agent does have such probability assignments even though their elicitation might prove practically impossible. But if he does assign probabilities to his current options, what possible practical role could these assignments play in his choice and deliberation? This query leads us to the final argument on behalf of the strong thesis.

9. "NO USEFUL ROLE"-ARGUMENT (SPOHN)

In addition to his betting argument, Spohn also has a much simpler argument for the strong thesis:

(...) probabilities of acts play no role in decision making. (.) The decision maker chooses the act he likes most – be its probability as it may. But if this is so, there is no sense in imputing probabilities for acts to the decision maker. (Spohn (1977), p. 115; cf. also Spohn (1978), p. 173f.)

In private communication, Spohn has suggested that it is this "no useful role"-argument that makes the strong thesis so intuitively plausible. If probabilities of options do not play any role in our choice, if we do not need them at all in order to make a decision, why should one impute them to the decision maker?

Criticism: I have three rejoinders to this "no useful role"-objection:

(i) Even if it were true that as a deliberators we have no use for the probabilities of the options among which we choose, Spohn would still need to show that such probabilities would be positively harmful. For, as we have seen, the probabilities for actions can be of use at an *earlier* stage, prior to deliberation. If at the onset of deliberation these probabilities no longer are present, we must assume that they must have been contracted. But why contract them if they don't do any harm? After all, the onset of deliberation does not provide us, by itself, with any new information that would undermine the evidence on which those probability assignments have originally been based.

In the introductory section, we have seen how such a probability loss could be understood, given a modular view of the mind: What happens in the alleged contraction is that the probability of an action, while still being accessible to the credence module, is somehow screened off from the module that makes a decision. But why should such a screening off be needed, in the first place, if the probabilities in question don't do any damage in the decision process?

(ii) In fact, they might do some good. According to Levi, I never deliberate whether to perform an option I am certain I am not going to choose. Now, something similar may apply to options with low probabilities: If I take an option to be very improbable, to begin with, then, in considering what to do, I might well allocate to it less time and effort in deliberation. This applies, in particular, in those cases when the option in question at the onset of deliberation does not appear to be attractive: Why bother giving a seemingly unattractive option any serious thought if the probability I will choose it is low anyway? But if this is how an efficient practical deliberation should be conducted, then it follows that probability assignments to options might have an important use as guideposts for deliberation.

(iii) Deliberation is an ongoing process in which we might incline to different choices at different stages. These changes in inclination might lead us to successively modify our probabilities for various options (assuming we do assign such probabilities), which in turn might modify our choice inclinations. This feedback process could, in particular, explain some instability phenomena in decision making. A case in point would be the well-known example of "Death in Damascus", in which my awareness of an inclination to perform a certain action, A, raises its probability and thereby raises the probability of the state of nature this action would be a reliable evidence for. If given that state, S, an alternative action, B, would be more profitable, then my inclination to act changes: I become more inclined to perform B rather than A, which increases my probability for B. But if the probability of B increases, this decreases the probability of S, which in turn makes A more attractive. And so on.[34] I face instability in my deliberation process.[35]

To conclude: (i) Betting arguments for the strong thesis are not convincing: While they manage to establish that the agent cannot have betting rates for options on which he deliberates, the connection between the agent's probability assignments and betting dispositions need not be very tight. (ii) Furthermore, the "no useful role"-argument can be met. Probabilities for one's current options need not be contracted: they make no harm in deliberation and could even be positively useful in allocation of time and effort. Also, their presence seems to account for some dynamic features of the deliberation process. None of these three considerations need to be decisive, if considered on its own.[36] But together they give us good grounds to question the strong thesis.[37] Deliberation is much less hostile to self-prediction than one might have thought.

APPENDIX: MUTUAL AND RECIPROCAL BELIEF IN RATIONALITY

The assumption of *common* belief in rationality requires everyone to believe (i) that everyone is rational, (ii) that everyone believes that everyone is rational, etc. By contrast, the assumption of *mutual* belief in rationality only requires everyone to believe (i) that *everyone else* is rational, (ii) that everyone believes that *everyone else* is rational, and so on. Here is how this notion of mutual belief in rationality can be made more precise. Let $B(x, \varphi)$ stand for the proposition that x believes that φ. $R(x)$ stands for the proposition that x is rational. Let x and y be variables that range over some fixed set of individuals (the set of players). Now, define recursively an infinite sequence of sentences:

Base clause:

$$(1) \qquad \forall x\, B(x, \forall y(y \neq x \rightarrow R(x))).$$

Recursion clause:

$$(n+1)\, \forall x B(x, (n)).$$

Mutual belief in rationality obtains iff all the sentences in this sequence are true.

The assumption of *reciprocal* belief in rationality (cf. Schick 2000) requires everyone to believe (i) that *everyone else* is rational, (ii) that everyone else believes that *everyone else* is rational, and so on. Reciprocal belief is less demanding than mutual belief in rationality since it does not, for example, entail that every agent believes that he himself believes that everyone else is rational. On the other hand, this difference between the two notions disappears given the assumption of *positive introspection*, according to which if any individual believes a proposition, then he believes that he believes that proposition. Then, as easily seen, reciprocal belief and mutual belief in rationality become equivalent concepts. For if the agent believes (i) that *everyone else* is rational, and (ii) that *everyone else* believes that *everyone else* is rational, then, by positive introspection, he also believes (a) that he believes that *everyone else* is rational, and (b) that *everyone else* believes that everyone else is rational. And the conjunction of the belief in (a) and (b) is equivalent with the belief that *everyone* believes that *everyone else* is rational. All the consecutive clauses in the definition of mutual belief can be proved in this way.

Should one prefer the concept of reciprocal belief, here is how this concept can be formalized. Let $x_1, x_2, \ldots$ be variables that range over a fixed set of individuals. Define recursively a sequence of *open* formulas:

Base clause:

$$(\mathrm{i}) \qquad B(x_1, \forall x_0(x_0 \neq x_1 \rightarrow R(x_0))).$$

Recursion clause:

$$(n+1)\, B(x_{n+1}, \forall x_n(x_n \neq x_{n+1} \rightarrow (n))).$$

Then, on the basis of the formulas in this sequence, define the sequence of their universal generalisations: $\forall x_1$ (i), $\forall x_2$ (ii), etc. *Reciprocal belief in rationality* obtains iff all the sentences in the latter sequence are true.

ACKNOWLEDGEMENTS

Earlier versions of this paper were presented at conferences and seminars in Leipzig, Bowling Green, St Andrews, Uppsala and Lund. I am indebted to the participants of these events and to several other colleagues for their stimulating comments and discussion. In particular, I would like to thank Luc Bovens, Michael Bratman, Krister Bykvist, Erik Carlson, David Copp, Christoph Fehige, Ghijs Donselaar, Wolfram Hinzen, Jeff Horty, Magnus Jiborn, Marion Ledwig, Sten Lindström, Georg Meggle, Howard Sobel, Rudolf Schüssler, Pirmin Stekeler-Weithofer, Bruno Verbeek, and, especially, Isaac Levi and Wolfgang Spohn. I am grateful to the anonymous referees for their helpful suggestions. My work on this paper has been supported by a grant from The Bank of Sweden Tercentenary Foundation.

NOTES

[1] Kant does not put the contrast between the two perspectives in quite these terms. As far as the "natural" perspective is concerned, his focus is on causal explanations rather than predictions. Still, explanation and prediction are two sides of the same coin and Kant does sometimes connect the opposition between the two perspectives with the issue of predictability. Cf. the following passage in which he envisages something like the point of view of Laplace's demon: "[Even] if it were possible for us to have so deep an insight into a man's character as shown both in inner and in outer actions, that every, even the least, incentive to these actions and all external occasions which affect them were so known to us that his future conduct could be *predicted* with as great a certainty as the occurrence of a solar or lunar eclipse, we could nevertheless still assert [from another point of view] that the man is free." (Kant 1956 (1788), pp. 99f in the standard pagination; my italics)

[2] Christine Korsgaard, in her paper on Kant's conception of freedom, suggests that even for Kant the perspective of freedom has its roots in practical deliberation: "The standpoint from which you adopt the belief in freedom is that of the deliberating agent. (...) Thus it is primarily your own freedom that you are licensed to believe in, and, as a consequence, it is primarily yourself that you hold imputable." (Korsgaard 1996, p. 174)

[3] This slogan has been coined by Isaac Levi. See Levi (1997), p. IX and 81. Cf. also ibid. p. 32 for another formulation of the slogan: "To be an agent crowds out being a predictor".

[4] Two clarifications: (i) The strong thesis only concerns unconditional probabilities for actions. It does not exclude that the deliberating agent might assign conditional probabilities to his acts, of the form: On a condition B, the probability of an action A equals k. In symbols, $P(A/B) = k$. (ii) If the strong thesis holds, it extends to all events that are probabilistically dependent on the agent's current options. (Cf. Spohn 1978, p. 74; Levi 1997, p. 83.) If the agent were to assign an unconditional probability to such an event, E, then his probability for his current option A could be calculated from his unconditional probability for E together with his conditional probabilities for E given A and given not-A. To be more precise, for each such option A, the probability calculus implies that

$$P(A) = [P(E) - P(E/\text{non-}A)]/[P(E/A) - P(E/\text{non-}A)].$$

E's probabilistic dependence on *A* guarantees that the denumerator in the right-hand side of this equation differs from 0.

[5] Cf. Shackle (1958), pp. 21–27, Ginet (1962), Pears (1968), Goldman (1970), ch. 6, Jeffrey (1977) and (1983), p. 85, Schick (1979) and (2000), and Levi (1986), sect. 4.3, (1989) and (1991).

[6] A similar suggestion in fact is made by Spohn, in his (1978), section 2.5.

[7] A modular conception of the mind is hugely influential among cognitive scientists. David Chalmers' useful bibliography of the philosophy of psychology (http://www.u.arizona.edu/~chalmers/biblio/5.html) lists nearly thirty books and papers on the subject, and his list is by no means complete.

[8] Howard Sobel's comment, in private communication: "The modular theory – common sense in fashionable terminology? When we are deliberating what to do, we are not wondering what we will do. But if someone interrupted, and asked, we could switch to wondering with them, and find that we had an opinion with which to start. Did it suddenly materialize, in response to the question, and with the switch? No one would say so, except for the sake of maintaining an argument."

[9] Cf. Levi 1997 (1991), p. 82: "This view [= the strong thesis] has profound implications for many topics". He proceeds to give examples from game theory and theories of sequential choice.

[10] In fact, this way out has been suggested by Spohn himself, in his (1978), p. 77.

[11] Nowadays, Jeffrey is fully prepared to allow for such gaps in the probability assignment. As he puts it in Jeffrey 1996, p. 7, "[p]robabilistic judgment is not generally a matter of assigning definite probabilities to all propositions in a Boolean algebra [...] probabilistic judgment may assign values to some propositions, none to others." One point should be noted, though: Since in Jeffrey's framework probabilities are derived (up to certain transformations) from the preference ranking of propositions, it is the latter that should be allowed to be incomplete if one wants to make room for serious gaps in the probability assignment. (In fact, since even a complete preference ranking does not *uniquely* determine the probability assignment, there is always room in this approach for some rudimentary gaps in probabilities.) I am indebted to the referees for this reminder.

[12] This point has been stressed by Levi in private communication. For the same point, cf. Jeffrey 1996, p.8f.

[13] I am indebted for this reminder to an anonymous referee.

[14] Schick 2000 works, in fact, with an even weaker notion of *reciprocal* belief. Reciprocal belief in rationality requires everyone to believe that *everyone else* is rational, that *everyone else* believes that *everyone else* is rational, and so on. Reciprocal belief is less demanding than mutual belief in rationality since it does not, for example, entail that every agent believes that he himself believes that everyone else is rational. On the other hand, this difference between the two notions disappears given the standard assumption of *positive introspection*. That is, suppose it holds for every individual that, if he believes something, then he believes that he believes this. Then, as easily seen, reciprocal belief and mutual belief in rationality become equivalent concepts. For an explication of both these notions, see Appendix.

[15] In contrast, mutual knowledge of rationality *would* entail such a self-assurance on each player's part. If I know that you know that I am rational, then I know myself that I am rational.

[16] Another case in point, in which the strong thesis might get us into trouble, are sequential decision problems with *imperfect recall*. In some problems of this kind, at the moment of

choice, an "absentminded" agent is unable to tell whether this is the first occasion of choice or whether he has been in exactly similar choice situations before. And he believes he is (will be or has been) equally absentminded on the other occasions. In order to determine what he should do on a given occasion of choice, he needs to form an opinion as to the probable consequences of his current options. For this purpose, he might need to make a probabilistic estimate of his behaviour on the potential future occasions of choice, at which he will have exactly the same beliefs and preferences as he has now and will follow exactly the same decision procedures. But then the estimate he makes of his behaviour on other similar occasions applies just as well to his current occasion of choice. For he recognises that the factors that determine his choice are the same on each occasion. For a discussion of an example of this kind, the case of "The Absentminded Driver", see Aumann, Hart and Perry (1997) and Rabinowicz (2001). The example is due to Piccione and Rubinstein (1997).

[17] Alternatively, for bets that are costless, one can think of C as the amount the bettor has to pay if he loses (i.e., if A turns out to be false). If he wins, he collects a certain amount G. The stake S is then the sum of C and G.

[18] If the maximal buying price were *higher* than the minimal selling price, then, for a given A and S, there would be a whole range of bets on A with stake S that the agent is prepared to cover on each side. If, on the other hand, the maximal buying price were *lower* than the minimal selling price, then no bet on A with stake S would be such that the agent is prepared to take each of its sides. We exclude both these possibilities by assuming, for a given A and S, the existence of a unique C such a bet on A with stake S and price C is fair.

[19] Note that, on this interpretation, we need to assume that utility is measured on a ratio scale if the ratio C/S is to be a meaningful quantity.

[20] While the betting rate for A is defined as the quotient C/S in a fair bet $b^A_{C,S}$, the betting odds are defined as S/C.

[21] In this formulation of the argument it becomes clear that the argument can only work against the assumption that the agent's probability for A is lower than 1. For if his betting rate x is 1, then we cannot find S and C such that $C/S > x$ and the net gain $S - C$ is non-negative.

[22] This question has been posed to me by an anonymous referee.

[23] But if he considers it probable that he will perform the action if he takes the bet, then he must assign a probability to the conditional *If I were to bet on A, I would do A*. In the case under consideration, we may suppose that this probability of a conditional coincides with the conditional probability of A given the bet on A. Can we ascribe to the agent such a conditional probability for A in the course of the argument to the effect that $P(A)$ is not well-defined? Yes, I think so, for two reasons. (i) As pointed out above, the strong thesis only concerns unconditional probabilities for actions. (ii) The argument to be presented is *reductive* in form. We assume that the agent does assign probabilities to his acts and then show that this assumption cannot be upheld.

[24] The value $eu(B) - C$ is what the agent would receive if he placed the bet on A and performed B instead.

[25] However, in his later work, Spohn changes his mind on this issue. In Spohn (1999, pp. 44–45), he postulates that in the case of sequential decision making, the decision maker can ascribe subjective probabilities to his future, but not to his present actions. I am indebted to Marion Ledwig for pointing this out.

[26] Schick thinks that a future action is something I can "agonize" about even beforehand. In this sense, a future action is already now an option for me, something I can deliberate upon. As such, it falls within the scope of the weak thesis.

[27] There is here a misprint in Levi's text that I have corrected.

[28] As noted above, Schick (2000) thinks otherwise. In that paper, Schick wants to establish that the agent's inability to predict his own choices does not require any dramatic changes in decision theory and game theory. However, if we deprive the agent the ability to predict his *future* choices, the implications for sequential decision making and for treatment of games in extensive form would be very serious indeed. In particular, contrary to what Schick suggests, using backward induction would in some cases become impossible. Here is a simple example that illustrates this. Consider a simple game between two players in which player X, who moves first, has a choice between going up or down. If he goes down, the game terminates, and each player receives the payoff of 9. If X instead goes up, Y will have a choice between up and down. If Y goes down, the game terminates, and each player receives the payoff of 0. If Y instead goes up, X will have one last choice between up and down. If he goes down, each player again receives 0 but if he goes up, each player ends up with 10. The backward-induction solution is up-up-up, where the argument for X going up in the first choice node assumes not only that he expects Y, in the second choice node, to count on X going up in the last node, but also that he himself counts on this. I.e., it is assumed that X, in the first node, predicts that he will go up in the last node (if the game will reach that far). If he has doubts on this score, he might well decide to opt for the safe solution and go down in his first move. Thus, self-predicting one's future choices is sometimes necessary for backward induction reasoning.

[29] Cf. Levi 1997, p. 32 and pp. 76–80.

[30] It would be less controversial to say that an action A is not an option for the agent if he not only believes he will not do it but also believes that he wouldn't do it even if he subjected that action to deliberation. For under these circumstances, deliberating on A is seen as *pointless* by the agent. However, merely believing that one won't do A, if motivated by the belief that one will not deliberate on A, is fully compatible with the belief that one might or even would do A, if one gave that action a serious consideration. Cf. Carlson (2000).

[31] To derive this conclusion we assume, for simplicity's sake, that an agent who assigns probability zero to an option is thereby certain he will not perform it. Levi actually does not make this assumption. As he points out, in the decision problems that involve infinitely many equiprobable alternatives, each of the alternatives is assigned probability zero, but they may still be considered to be doxastically possible. However, in the cases we are interested in, this special problem does not arise.

[32] This is a reconstruction of the reasoning presented by Levi in his 1997, on p. 32 and in a footnote 5 on pp. 76f.

[33] The reader should note, however, that other possible principles for choice under uncertainty might not yield this result. This applies, in particular, to Levi's well-known choice rule, which enjoins the agent to identify the actions that under some admissible probability distributions maximize expected utility and then, among these actions, to choose the one whose security level (= the worst possible outcome) is maximal (cf. Levi 1980, 1986). In the case at hand, Levi's rule would recommend the agent to abstain from making the betting commitment. If $P(A) = S/C$, making that commitment has the same expected utility as abstaining, on all probability distributions over {Buy $b^A_{C,S}$, Sell $b^A_{C,S}$}. But if $C > 0$ or

$S > C$, the worst possible outcome of making the commitment (i.e., $-C$, or $C - S$, as the case may be) is worse than the worst possible outcome of abstaining.

[34] In the "Death in Damascus"-example, due to Gibbard and Harper (1978), Death is awaiting the agent in one of two localities, a or b. The agent, who would like to avoid his appointment with Death, is in a severe quandary: should he flee to a (option A) or should he stay in b (option B)? We assume that no other option is available. The agent believes that Death is a good predictor of human choices. Therefore, if he inclines to go to a, his probability for A increases, which in turn increases his probability that Death awaits him in a. This gives him a reason to stay in b instead and thus increases his probability for B. But then it becomes more probable that Death awaits him in b, which gives him a reason to escape to a. And so on.

[35] It might be mentioned that the instability cases are not the only ones in which the decision is influenced by the process of updating probabilities for options. Thus, to give another example, Ellery Eells exploits this updating process in his defence of evidential decision theory. Taking this process into consideration, he argues (cf. Eells 1984), allows us to show how evidential decision theory can deal with the Newcomb problem. In that problem, taking just one box is seen to have a high "value as news" (i.e., as a symptom of the favourable state of nature). Therefore, the probability of this option initially increases in the course of deliberation, which increases the probability of the favourable state. As a result, however, the option's value as news tends to decrease. The option's evidentiary significance as the signal of the favourable state is gradually screened off by the incoming introspective information concerning the agent's probabilities for options. This in turn has consequences for the decision: the option's value as news eventually converges to its expected value as cause. Or so Eells argues. For an interesting criticism of Eells' dynamic approach to the maximization of evidentiary expected utility, see Sobel (1994), chapter 2, sections 5.3–5.5.

[36] Wolfgang Spohn's comment (in private communication): "[I]n objections (ii) and (iii), decision theory is taken as a theory of the deliberational process. But this is something entirely different. I would think, then, that decision theory is entirely inadequate [for that task] and that there is no adequate theory [for that task] at all."

[37] Howard Sobel's comment (in private communication): "About the possible use of views concerning likelihoods of our doing this or that, about the possible use of such views when we are deliberating what to do, perhaps the main thing to say is that ordinary deliberation is forward-looking. It is for what to do, not exactly now, but, maybe, tomorrow. The likelihoods of carrying out plans/decisions/choices/intentions made now are relevant." But this particular point, about the importance of predictions concerning implementation, would be admitted even by the proponents of the strong thesis. It is only the choice itself that they take to be unpredictable, not its implementation. Levi, for one, argues that, for an action to be an option fit for deliberation, the agent must be certain that he will perform the action if he makes that choice.

REFERENCES

Aumann, R. J., S. Hart, and M. Perry: 1997, 'The Absent-Minded Driver', *Games and Economic Behavior* **20**, 102–120.

Aumann, R. J.: 1987, 'Correlated Equilibrium as an Expression of Bayesian Rationality', *Econometrica* **55**, 67–96.

Carlson, E.: 2000, 'Deliberation, Foreknowledge, and Morality as a Guide to Action', draft.

Eells, E.: 1984, 'Newcomb's Many Solutions', *Theory and Decision* **16**, 59–105.

Gärdenfors, P. and N.-E. Sahlin: 1982, 'Unreliable Probabilities, Risk-Taking and Decision-Making', *Synthese* **53**, 36–86; re-printed in Gärdenfors and Sahlin (eds.), *Decision, Probability and Utility – Selected Readings*, University Press, Cambridge 1988, pp. 313–334.

Gibbard, A. and W. Harper: 1978, 'Counterfactuals and Two Kinds of Expected Utility', in C. A. Hooker, J. J. Leach and E. F. McClennen (eds.), *Foundations and Applications of Decision Theory*, Vol. 1: *Theoretical Foundations*, University of Western Ontario Series in Philosophy of Science, 13, Reidel, Dordrecht and Boston, pp. 125–162.

Ginet, C.: 1962, 'Can the Will Be Caused?', *Philosophical Review* **71**, 49–55.

Goldman, A.: 1970, *A Theory of Human Action*, Prentice-Hall, Englewood Cliffs.

Jeffrey, R.: 1977, 'A Note on Probability Kinematics', *Erkenntnis* **11**, 135–141.

Jeffrey, R.: 1983, *Logic of Decision*, 2nd ed., Univ. of Chicago Press, Chicago and London.

Jeffrey, R.: 1996, 'Decision Kinematics', in K. Arrow, E. Colombatto, M. Perlman and C. Schmidt (eds.), *The Rational Foundations of Economic Behaviour*, IEA conference vol. No. 114, Macmillan Press and St. Martin's Press, London and New York, pp. 3–19.

Kant, I.: 1956 (1788), *Critique of Practical Reason*, transl. by Lewis White Beck, The Bobbs-Merrill Comp., Indianapolis – New York.

Korsgaard, C. M.: 1996, 'Morality as Freedom', in Korsgaard (ed.), *Creating the Kingdom of Ends*, Cambridge University Press, pp. 159-87. Reprinted from Y. Yovel (ed.), *Kant's Practical Philosophy Reconsidered*, Kluwer, Dordrecht 1989, pp. 23–48.

Levi, I.: 1980, *The Enterprise of Knowledge*, The MIT Press, Cambridge.

Levi, I.: 1986, *Hard Choices*, Cambridge University Press, Cambridge.

Levi, I.: 1989, 'Rationality, Prediction, and Autonomous Choice', *Canadian Journal of Philosophy* **19** (suppl.), 339–363, re-printed in Levi (1997), pp. 19–39.

Levi, I.: 1991, 'Consequentialism and Sequential Choice', in M. Bacharach and S.Hurley (eds.), *Foundations of Decision Theory*, Blackwell, Oxford, pp. 92–122; re-published in a slightly different form in Levi (1997), pp. 70–101.

Levi, I.: 1997, *The Covenant of Reason*, Cambridge University Press, Cambridge.

Pears, D.: 1968, 'Predicting and Deciding', in P. F. Strawson (ed.), *Studies in the Philosophy of Thought and Action*, Oxford, London.

Piccione, M. and A. Rubinstein: 1997, 'On the Interpretation of Decision Problems with Imperfect Recall', *Games and Economic Behavior* **20**, 3–24.

Rabinowicz, W.: 2001, 'Remarks on the Absentminded Driver', in W. Rabinowicz (ed.), *Value and Choice* Vol. 2, *Lund Philosophy Reports*, pp. 194–207; a revised version forthcoming in *Studia Logica* 2002.

Shackle, G. L. S.: 1958, *Time in Economics*, North Holland, Amsterdam.

Schick, F.: 1979. 'Self-Knowledge, Uncertainty and Choice', *British Journal for the Philosophy of Science* **30**, 232–252; re-printed in P. Gärdenfors and N.-E. Sahlin (eds.), *Decision, Probability and Utility*, Cambridge Univ. Press, Cambridge 1988, pp. 270–286.

Schick, F.: 2000. 'Surprise, Self-Knowledge and Commonality', *Journal of Philosophy* **97**, 440–453; originally published in B. Hansson, S. Halldén, W. Rabinowicz and N.-E. Sahlin (eds.), *Spinning Ideas – An Electronic Festschrift for Peter Gärdenfors*, 1999, http://www.lucs.lu.se/spinning/

Sobel, J. H.: 1994, *Taking Chances*, Cambridge University Press, Cambridge.

Spohn, W.: 1977, 'Where Luce and Krantz Do Really Generalize Savage's Decision Model', *Erkenntnis* **11**, 113–134.

Spohn, W.: 1978, *Grundlagen der Entscheidungstheorie*, Scriptor, Kronberg/Ts.
Spohn, W.: 1999, *Strategic Rationality*, Forschungsbericht der DFG-Forschergruppe 'Logik in der Philosophie', University of Konstanz.

Department of Practical Philosophy
Lund University
Box 117
221 00 Lund
Sweden

Accepted: 4 February, 2002