

RATIONALITY AND PARADOX IN DECISION MAKING

James Dreier and Roberto Serrano

Brown University Faculty Bulletin, Fall 2001, pp. 30-33

Department of Philosophy and Department of Economics

1 The Philosopher's View

Decisions. In what we'll call *ordinary decision problems* a person faces a choice alone. Typically, the outcome of the choice depends partly on the act he chooses, and partly on the *state of the world*. For example, suppose you are on vacation and you have to decide whether to go to the beach or visit a museum. The outcome will depend partly on which you choose, and partly on the weather. And the weather (the state of the world) is uncertain, though you may have a judgment about probabilities (there may be, for example, a 40% chance of rain).

Dominance. In some fortunate cases, there is a *dominant* action. A dominant action is one that, from your point of view, is better than any other action in some states and never worse than any other action in any state. For instance, it may be that you prefer the museum to the beach in case it rains, and you are indifferent between them in case it is sunny. Then your choice is easy: you should go to the museum because it is dominant.

Probabilistic independence of states. Adopting the dominant action is reliably rational only when states are *probabilistically independent of acts*. To see this, consider this example from Shakespeare's play King Henry V. The scene is just before the battle of Agincourt, at which the British know that they are terribly outnumbered. One of Henry's men wishes aloud that just a few thousand knights could join them in the morning. Henry disagrees. His argument: Either we will win tomorrow, or we will lose. If we win, how much better to win having been so hopelessly outnumbered! If we lose, how ignominious it would be to lose with a multitude on our side! Then in either case, it is better that we have few soldiers. What is wrong with

Henry's Argument? Apparently what is wrong is that the State (win or lose) depends probabilistically on "act" (whether we have few or many). It is very much more likely that we will win given that we have many, than given that we have few. Assuming that Henry's judgment of probabilities and values of outcomes is roughly like our own, he should conclude that it would be better to have many. Decision theory puts it this way: the *expected utility* of having many is greater than the expected utility of having few. The theory says that the *rational action* is the one with the greater expected utility.

Causal decision theory. More intuitively, we might say that having more men can influence the outcome. In general we expect to exert *causal influence* on the world by our actions. Decision theory represents this causal influence by weighting the utilities of the outcomes by conditional probabilities. Sometimes the representation is inadequate. Getting theories to capture adequately the intuitive ideas that drive them is one of the callings of philosophers.

Newcomb's Paradox. A rich and powerful (and possibly superhuman) Predictor shows you two boxes, and offers you a choice. You may either take just one box, an opaque one whose contents you don't know, or both, the opaque one plus a transparent one in which he has placed \$1000. The Predictor has predicted your choice. If he predicted that you would take both, he has put nothing in the opaque box. If he thought you'd take just the opaque box, he has put \$1,000,000 therein. He seems to be a very good predictor. You have watched thousands of other humans play this game, and the Predictor has been correct nearly every time. Because the Predictor is so accurate, you think it is very likely that he will predict your choice correctly. So the probability that there is (already) a million dollars in the opaque box given that you choose just one box, is very high (say, .99), and the probability that the box is empty given that you take both boxes is also high (also .99). That does not mean you think you can cause money to be in the box by your choice. You cannot, and you know it. Rather, you think that your choice is very good evidence of what the Predictor has

already done.

If we let your utility follow the dollar amounts, then the expected utility of taking one box is very close to a million (because the conditional probability that he predicted you'd take one is nearly one). But the expected utility of taking both boxes will be much lower: only about eleven thousand, because the conditional probability that he predicted you'd take both will be almost one, so you will almost certainly get a thousand dollars (and there is a one percent chance that you will get a million plus the thousand). Classic decision theory, therefore, tells you to take just one box.

Most people think this is a foolish thing to do. The money is already in the box, or not in it, and nothing you can do will change that! Most people think that because your choice is causally independent of the state, you should adopt the dominant strategy (which is to take both boxes). A new version of decision theory replaces conditional probabilities with causal conditionals. To most people, this new version gets the "right answers" where the classic version fails.

2 The Economist's View

Newcomb's problem. A paradox? In Newcomb's problem, the agent's probabilistic assessments are not independent of his actions, which some people argue may give rise to a paradox. But I think there is no paradox. If the decision maker is convinced that the Predictor is infallible or that he can affect the fact that there is money in the opaque box with his action, he should choose only the opaque box. Given this bizarre beliefs, this is what decision theory would recommend, but note that in this case choosing the two boxes is not dominant. On the other hand, if his beliefs take into consideration that the money either *is* or *is not* in the opaque box regardless of what action he takes, he should choose both boxes. Furthermore, classical decision theory endorses this recommendation, whatever probability he assigns to the money being in the opaque box, he has a dominant action.

From decisions to games. Most of the time economic agents are involved in decision problems, where it is important to understand that there are other agents making their own decisions, and that the final outcome is a result of the actions of all of them. Such situations are called *games* and are studied by *game theory*.

The prisoners' dilemma and dominance. The first game anyone encounters in a game theory book is the prisoners' dilemma. Two people are arrested, suspect of having committed a crime. The police lock them in separate cells, and each of them receives the following set of instructions: "There is another prisoner in a nearby cell, but you will not be able to communicate with him. You can either "cooperate with the other suspect" by keeping your mouth shut, or "defect" by accusing him of the crime. If both of you defect, we will take it as evidence of at least one of you lying and we will manage to send each of you to jail for 25 years. If only one defects, he will be put back in the street immediately, while the other will be sent to jail for 30 years. Finally, if both cooperate, we will not be able to press charges against either of you and you will get out after 72 hours." In a game, we say that an action is *dominant* for an agent if he, *regardless of the actions chosen by others*, always prefers its consequences to those associated to other actions. The dominant action is to "defect," which yields a terrible outcome for the prisoners, thereby contradicting Adam Smith's "invisible hand" principle: the rational pursuit of individual satisfaction interferes here with the interests of society as a whole. Note how it would not be sound to use the "Newcomb's paradox" logic in the analysis, i.e., to make the assumption that each prisoner believes the other to very likely cooperate if he cooperates, and believes that the other will very likely defect if he defects. Remember that these two prisoners are making their decisions independently.

Equilibrium in the expanded battle of the sexes. Another paradox? A rational boyfriend and his rational girlfriend are trying to coordinate their plans to go out in the evening. Unfortunatley, they have no

way to communicate beforehand. Each of them can choose to show up to the local football game (her preferred outcome) or to a ballet performance (his preferred outcome). Of course, what they would really dislike is to end up going to different places. This is more complicated than the prisoners' dilemma, because neither player has a dominant action. We now need to speak of an *equilibrium*, whose logic provides two predictions: two systems of *consistent expectations and actions* can be constructed ("football-football" and "ballet-ballet"). Thus, rationality alone ceases to give us a clear-cut prediction. Moreover, if a third action is added for both players (staying home), the new prediction (the unique equilibrium of the new game) may be "home-home" and leave both worse off. This would never happen in an ordinary decision problem, where adding one action can never hurt the decision maker.

King Solomon's dilemma: from game theory to implementation theory. The above two games raise the question of whether something can be done to influence the final outcome in a desirable way. Implementation theory, instead of analyzing a given fixed game, studies how the rules of the game can be changed so that, when played by rational agents, a socially desirable outcome obtains. One can, for example, solve King Solomon's dilemma, i.e., give the baby to the true mother, by eliciting the true information from both women. And this can be done without cheating! The solution reported in the Bible is not sound, because Solomon changed the rules after the women were playing: while he had promised the baby to the only woman that claimed her, he ended up doing exactly the opposite after he threatened to kill the baby and one of the women ceased to claim her.