Locally rational decision-making

Richard M. Shiffrin
Psychology Department, Indiana University, Bloomington, IN 47405.
shiffrin@indiana.edu
http://www.cogs.indiana.edu/people/homepages/shiffrin.html

Abstract: Colman shows that normative theories of rational decision-making fail to produce rational decisions in simple interactive games. I suggest that well-formed theories are possible in local settings, keeping in mind that a good part of each game is the generation of a rational approach appropriate for that game. The key is rationality defined in terms of the game, not individual decisions.

Colman gives an intriguing, interesting, and at times amusing account of the failures of normative theories of rational decision-making. He suggests moving toward a "psychological" game theory that would be "primarily descriptive or positive rather than normative," and adds "a collection of tentative and ad hoc suggestions" (target article, sect. 8). I suggest that a well-formed psychological theory of rational decision-making may well be possible in local contexts (of a scope and generality large enough to be interesting). The approach is rooted in the thought that rationality itself is a psychological rather than axiomatic concept, justifying the need to reinvent it (or at least restrict it) for different settings.

I propose that all the decision-makers in a social/interactive game are faced with a dual task: They must decide (quite possibly without any communication) what theory of rational decision-making applies in that situation, and given that, whether a jointly rational solution exists, and what it is. The first of these tasks is not typically made explicit, but is a necessary consequence of the current lack of a general (axiomatic) theory of rational decision-making.

It will suffice for this commentary to consider the Centipede game (Colman's Fig. 5). This is a good exemplar of a social/interaction game without communication (except through the choices made), and with the goal for each player to maximize individual utility (not beat the other player). I assume that both players know that both players are rational, and not subject to the sundry "irrational" forces that lead human decision-makers to their choices. I also assume that each player knows his or her own mapping of monetary payoffs onto subjective utility, but does not know the mapping for the other player, other than the shared knowledge that a larger payoff (in monetary amount, say) corresponds to a larger utility. Note that this assumption (in most cases) eliminates the possibility that a rational strategy will involve a probabilistic mixture. Assuming that player A's mixture of choices affects player B's mixture of outcomes, player A generally cannot know whether the utility to B of a given mixture exceeds that for some other fixed or mixed payoff.

Therefore, the players at the outset of a game will both consider the same finite set of strategies S_j, where a given strategy consists of the ordered set of decisions $\langle D(1_A), D(2_B), D(3_A), D(4_B), \ldots D(N)\rangle$, where $D(I)$ is one of the choices allowed that player by the sequence of previous choices in that strategy. A game utility U_j is associated with each strategy: $\langle U_{jA}, U_{jB}\rangle$. Each player's goal is to reach a strategy that will maximize his or her personal U_j, in the knowledge that both players are rational and both have this goal.

In a Centipede game with many trials (say, 20), backward induction seems to lead to the "irrational" decision to stop (defect) on trial 1, even though both players can gain lots of money by playing (cooperating) for many trials. Of course, backward induction is flawed when used here in the usual way: Player A defects on, say, trial 15 in the certainty that Player B will defect on trial 16. But trial 15 could not have been reached unless B had been cooperating on all previous choices, so certainty is not possible. Thus, by cooperating on the first trial, the player eliminates backward induction as a basis for reasoning, and allows cooperation to emerge as a rational strategy. Yet, the forces in favor of defecting grow over trials, until backward induction seems to regain its force on the penultimate choice (e.g., trial 19 of 20, or 3 of 4).

Consider, therefore, a two-trial version of Colman's Centipede game. Both players at the outset consider the three possible strategies: ⟨stop⟩, ⟨play, stop⟩, ⟨play, play⟩, with associated payoffs of ⟨0,0⟩, ⟨−1,10⟩, ⟨9, 9⟩. The players look for a rational solution, in the hope that one exists (they share the knowledge that some games may not have a rational solution). So each player reasons: Which of the three strategies could be rational? Player B might like ⟨play, stop⟩, but both players could not decide this strategy was rational. If it were, A would stop on trial 1 (forcing a better outcome). Therefore, both players know ⟨play, stop⟩ could not be a rational strategy. Of the two remaining strategies, both players have little trouble seeing ⟨play, play⟩ as the rational choice, given that ⟨9, 9⟩ is preferred to ⟨0,0⟩.

This solution is "selfish," not predicated on maximizing joint return. It derives from the shared knowledge of playing a two-trial social game: In a one-trial game even a rational, cooperative decision-maker would clearly defect. Rationality is defined in terms of the entire game and total payoffs, not the payoff on a given trial. This approach could perhaps be considered a kind of generalization of the "Stackelberg reasoning" discussed by Colman, but is even more closely related to "dependency equilibria" discussed by Spohn (2001). It can be generalized and formalized (though not in this commentary). I note only that it gives justification for cooperative choices in simultaneous-choice games, such as the Prisoner's Dilemma (and sequential-play extensions of those games).

Perhaps the chief objection to this approach involves the perception that accepted causal precepts are violated: What is to stop B from defecting once trial 2 is reached? This issue is reminiscent of that obtaining in Newcomb's paradox (Nozick 1969), or the "toxin" puzzle (Kavka 1983), but in those cases a defense of a seemingly irrational later choice depends on uncertainty concerning an earlier causal event (I say "seemingly" because I am quite certain a Newcomb's chooser should take "one" and the "toxin" should be imbibed). The present case is more troublesome, because the first choice is known when the last choice is made. I nonetheless defend cooperation with the primary argument that rationality ought to be, and in fact must be, defined in terms of the entire game and not an individual decision within that game.

"Was you ever bit by a dead bee?" – Evolutionary games and dominated strategies

Karl Sigmund
Institut für Mathematik, Universität Wien, 1090 Vienna, Austria.
karl.sigmund@univie.ac.at **http://mailbox.univie.ac.at/karl.sigmund/**

Abstract: On top of the puzzles mentioned by Colman comes the puzzle of why rationality has bewitched classical game theory for so long. Not the smallest merit of evolutionary game theory is that it views rationality as a limiting case, at best. But some problems only become more pressing.

Aficionados of Humphrey Bogart will recognize this title's question as being a running gag from the film "To Have and Have Not." Apparently, if you step barefoot on a dead bee, you are likely to get hurt. The assumption that human behavior is rational died a long time ago, for reasons Colman summarizes very well, but it has failed to be buried properly. And if you carelessly tread on it, you will learn about its sting.

The question is, of course, why one should tread on it in the first place. There seems no reason ever to come close. The hypothesis that humans act rationally has been empirically refuted not only in the context of interactive decisions, but also for individual decision-making, where, in a way, it is even more striking. Indeed,

https://doi.org/10.1017/S0140525X03440053 Published online by Cambridge University Press