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/S0140525X0345005X Published online by Cambridge University Press

some of the observed interactive behavior can be explained in rational terms, if the utility function is modified by a term depending on the payoff difference between the player and the coplayers (see Fehr & Schmidt 1999). But this device, a "fix" that resembles the modifications of epicycles in the Ptolemaic model of celestial mechanics, cannot explain deviations from rationality in individual decision-making as evidenced, for instance, by the paradoxes of Allais (see, e.g., Allais & Hagen 1979) or Ellsberg (1961).

The founding fathers of game theory had little knowledge of such experiments. But it seems difficult to understand why, to our day, after all the work by Tversky, Kahnemann (see e.g., Kahnemann & Tversky 1979), and many others, full rationality can still be termed "not such a bad assumption." Not every scientific idealization deserves as much respect as that of a perfect gas! Clinging to human rationality, in the face of evidence, must be a way of protecting faith in the existence of a "unique rational solution" for every game – a supernatural claim.

Game theory is the conceptual tool for analyzing social interactions in terms of methodological individualism. That it should be used in any normative sense smacks of dogmatism. Game theory is a branch of mathematics, and in this sense is not more "normative" or "descriptive" than algebra. It helps to analyze the logical consequences of certain assumptions. The assumption of fully rational agents is just one of many alternatives. Its prominent role is caused by force of habit alone. Almost two hundred years ago, mathematicians rejected the creed in a unique set of geometrical axioms. Why should there be a unique set of postulates for game theory?

The rationality axiom is obviously not needed in game-theoretical analyses dealing with the chemical warfare between bacterial mutants, the mating behavior of male lizards, or the economical solidarity between students (see Fehr & Gächter 2000; Kerr et al. 2002; Sinervo & Lively 1996). Even the term "bounded rationality" seems ill-advised in such contexts, implying to lay-persons that rationality is the norm that bacteria, lizards, and undergraduates fail to achieve.

In applications to real-life situations (as opposed to philosophical puzzles), game theory can do just as well without the postulate of rationality, and Occam's razor demands, therefore, to get rid of it. That it held out for so long is mostly due to historical contingency.

An illustration of historical contingency at work is the fact that John Nash, in his Ph.D. thesis, explicitly stated that his equilibrium notion could be motivated, not only by an appeal to rational players, but also by what he called the "mass action" approach. Oddly, this section was deleted in the published version from 1950 (see Weibull 1995). It seems that a reviewer had discarded it. Nash's mass action approach was resuscitated decades later in evolutionary game theory: Thinking in terms of populations came naturally to evolutionary biologists. No longer do the players have to be rational; all they need is some propensity for adopting successful strategies. This can be due to learning, to imitation, to infection, or to inheritance (see, e.g., Gintis 2000; Hofbauer & Sigmund 1998; Weibull 1995).

But, and here comes the sting, getting rid of the rationality axiom as a foundational postulate does not get rid of all problems. Evolutionary games lead, in many cases, back to the puzzles described by Colman. It only places them in the context of natural science. Whenever successful strategies spread, dominated strategies will get eliminated, defection will evolve in the Prisoner's Dilemma game, and selfishness will be just as self-defeating as it is between rational players bent on out-smarting their equally rational coplayers.

This is the dead bee's revenge. Far from explaining it away, evolutionary game theory emphasizes the urgency of the paradox. There are societies out there – not only in philosophical mind games – that display cooperation, although it is a dominated strategy. Opting for the evolutionary approach is beneficial, nevertheless, because it opens up so many testable solutions to the puzzles. Consider, for example, the Ultimatum game. Here an experimenter offers ten dollars to a pair of test persons, provided they keep to the following rules: A toss of the coin decides who of the two is the "Proposer" and must decide which part of the ten dollars to offer to the coplayer. If the "Responder" accepts the offer, this is how the money is split between the two players. If the "Responder" rejects the offer, the experimenter pockets the money. In each case, the game is over, and all go their separate ways – no haggling, and no further rounds.

In real experiments, small offers get rejected by most Responders, and most Proposers offer a substantial share. This blatantly contradicts the usual rationality assumptions, whereby Proposers ought to offer the minimal amount and Responders ought to accept it. Numerical simulations of evolving populations of players yield the same prediction. But, whereas the rationality axiom just leads to an impasse, the evolutionary approach suggests ways out. If one assumes, for example, (a) that players usually interact only within their neighborhood (rather than with a randomly chosen member of a large, well-mixed crowd); or (b) that there is always some small percentage of players who would never offer, as Proposers, less than they would accept as Responders; or (c) that players occasionally offer less if they learn, somehow, that their coplayer is likely to swallow it; then offers coming close to reality will evolve (see Nowak et al. 2000; Page & Nowak 2002; Page et al. 2002). None of these three hypotheses need be right; but all allow for testable predictions. Game theory is not only a tool for philosophical debates, but – rid of the straitjacket of rationality – it is an instrument for every social science.

Irrationality, suboptimality, and the evolutionary context

Mark Steer and Innes Cuthill

School of Biological Sciences, University of Bristol, Bristol, BS8 1UG, United Kingdom. **mark.steer@bristol.ac.uk** **i.cuthill@bristol.ac.uk**

Abstract: We propose that a direct analogy can be made between optimal behaviour in animals and rational behaviour in humans, and that lessons learned by the study of the former can be applied to the latter. Furthermore, we suggest that, to understand human decisions, rationality must be considered within an evolutionary framework.

We believe that Colman raises valuable and interesting points about the nature of rational choice in humans. Nonetheless, we would like to make the important point that behaviour considered to be irrational within the confines of an experimental situation may nonetheless be rational within a wider context. We believe there are illuminating parallels between the study of the adaptive value of behaviour (in terms of individual optimality or evolutionary stability) and that of rationality in decision-making. Just as a rational decision is one that maximizes some measure of utility, so to a behavioural ecologist, an optimal decision is one that maximizes Darwinian fitness given certain constraints. Thus, we believe that the appropriate research program to understand the rationality (or otherwise) of decision-making in humans should be analogous to that needed to understand the adaptive value of behaviour in the face of evolution by natural selection. These issues are of broad concern, not just confined to game-theoretic situations.

Imagine, for example, an investigation into the foraging behaviour of a bird in an aviary. It has a choice between foraging in two locations. At location A, situated deep in a bush, the bird experiences a low rate of food intake; at the more open location B, the rate of intake is much higher. Contrary to the predictions of a simple model of energetic intake maximization, the bird prefers to feed at A. Why?

Although the experimenters appreciate that the bird is in no danger of predation, it doesn't necessarily follow that the bird does. Foraging in the open may be deemed too risky, even though

https://doi.org/10.1017/S0140525X0345005X Published online by Cambridge University Press