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/S0140525X03460056 Published online by Cambridge University Press

the rate of gain is higher. It is only when we start to take these considerations into account, and include them in the model, that the bird's behaviour begins to make sense (see Houston & McNamara 1989; 1999; McNamara 1996 for further discussion). Of course, we must test our new assumptions independently before accepting our revised hypothesis.

So is the bird making optimal, or suboptimal, foraging decisions? This depends on what information we can expect the bird to have about its current situation. If it can perceive that it is in no danger of predation, then to carry on exhibiting antipredator behaviour may indeed be considered suboptimal. However, if the bird is incapable of perceiving the decreased threat, either through a lack of the cues it would use in nature or an inflexible decision rule (that nonetheless works well under natural conditions), then the behaviour may be optimal within the context for which the rule evolved. The information that the bird bases its decisions upon depends not only on its perception of the current environment, but also on its own developmental and evolutionary history. Indeed, to an evolutionary game theorist, it is appropriate to ask why a decision rule that may be susceptible to developmental experience (e.g., learning) has evolved as opposed to a relatively inflexible, innate, rule.

We suggest that a direct comparison can be made between our investigation of the apparent suboptimality of the bird foraging under cover and how one should investigate the apparent irrationality of, for example, an altruistic person in a one-shot game. Although the behaviour of the person in question might seem to be irrational within the confines of the experimental set-up, we have to consider not only whether the person perceives (or even *can* perceive) the game in the way the experimenters have conceived it, but also whether it makes evolutionary sense for that person to do so.

We would go further than simply drawing analogies between the study of optimality in animal decision-making and of rationality in human behaviour. We believe that a full understanding of human decision-making, even if the rules are the result of conscious reasoning rather than inflexible preprogrammed strategies, requires an evolutionary perspective. As behavioural ecologists, our belief is that natural selection should have equipped humans, just as other animals, with rules that maximize fitness in the appropriate environment. Therefore, from our point of view, the most rational decision for a human to make should be that which maximizes fitness. If indeed human decisions are constrained in some way by evolutionary influences, then considering our evolutionary past could be instrumental in understanding why seemingly irrational decisions are made in certain circumstances.

Not only might human behaviour be less flexible than we imagine, the subject may perceive that the game is being played with a wider range of people than the experimenter has planned: the experimenters, other players, and observers. For example, individuals behave differently if they are informed that they are playing another person, than if they are told they are playing against a computer (Baker & Rachlin 2002).

So we return to our main point. Whether our bird is behaving optimally or not depends to some degree on whether it is reasonable to expect that bird to know that predation risk is zero. Similarly, whether a person's choices in a game can be considered (ir)rational depends on whether we expect that person to have understood the precise experimental paradigm and then acted accordingly. The importance of looking at the wider context when considering responses to games is paramount to understanding how people might react. When individuals make seemingly irrational choices in a predictable fashion, there are three possibilities among which we need to distinguish: First, subjects don't possess the capabilities to fully assess the conditions of the situation, precipitating fully rational behaviour as far as the subject is concerned. Or, subjects might fully understand the conditions of the game, however, because of a developmental or evolutionary hangup, they perform the (evolutionarily) right behaviour in the wrong context. The behaviour is thus locally irrational, but rational within a wider framework. Or, third, all the conditions of the game are well understood by a subject, and the decisions truly are irrational (and thus maladaptive?).

Bridging psychology and game theory yields interdependence theory

Paul A. M. Van Lange and Marcello Gallucci

Department of Social Psychology, Free University, 1081 BT Amsterdam, The Netherlands. **pam.van.lange@psy.vu.nl** **m.gallucci@psy.vu.nl**

Abstract: This commentary focuses on the parts of psychological game theory dealing with preference, as illustrated by team reasoning, and supports the conclusion that these theoretical notions do not contribute above and beyond existing theory in understanding social interaction. In particular, psychology and games are already bridged by a comprehensive, formal, and inherently psychological theory, interdependence theory (Kelley & Thibaut 1978; Kelley et al. 2003), which has been demonstrated to account for a wide variety of social interaction phenomena.

Understanding social interaction phenomena is obviously a key issue in the social and behavioral sciences. It is, therefore, surprising that several important theories, such as rational choice theory, game theory, and complementary forms of decision theory, tend to focus on the individual level of behavior, as if there is no mutual dependence between individuals. In other words, these theories tend to neglect the conceptual importance of *interdependence* in understanding social interaction – the realization that important psychological processes, including deviations of rationality and self-interest, are ultimately rooted in the dimensions of interdependence underlying interaction situations.

In the target article, Colman proposes psychological game theory as an alternative to "orthodox theory" in accounting for deviations of rationality and self-interest, thereby extending orthodox theory in at least two important respects. First, it addresses *interpersonal psychology* by emphasizing the role of beliefs and expectations relevant to a situation and the actions and beliefs of the interaction partner. Second, it addresses a *collective level of analysis,* emphasizing, for example, the importance of collective preferences in settings of interdependence.

Although we are in agreement with both extensions, we do not think that these extensions make a very novel contribution to existing social psychological theory. Moreover, we share the critical view expressed by Colman that psychological game theory "amounts to nothing more than a collection of tentative and ad hoc suggestions for solving the heterogeneous problems that have been highlighted in earlier sections" (sect. 8). Notwithstanding their heuristic value, the suggestions would have been considerably less tentative and less ad hoc, if Colman had discussed psychological game theory in relation to existing social psychological theories. In particular, we argue that the parts of psychological game theory dealing with preference, illustrated mainly with team reasoning, are already well-understood in terms of interdependence theory (Kelley & Thibaut 1978; Kelley et al. 2003; Rusbult & Van Lange 2003), which in many ways can be conceptualized as an inherently psychological, yet formal, theory of social interaction. As we will see, after a brief discussion of interdependence theory, it is not clear whether the part of psychological game theory dealing with preference – a part that is very essential to game theory – contributes above and beyond interdependence theory in understanding social interaction.

Interdependence theory. Interdependence theory may be characterized by at least three qualities. First, using games and related conceptual tools, interdependence theory provides a taxonomy of interaction situations, which can be analyzed in terms of several dimensions, such as degree and mutuality of dependence, basis for dependence, corresponding versus conflicting interest, temporal structure, and information availability (Kelley et al. 2003;

https://doi.org/10.1017/S0140525X03460056 Published online by Cambridge University Press