In November 1945, a conference of mathematicians was held at the Museum of Science and Industry in Chicago. A robot was displayed in the entrance hall, inviting the visitors to play a game of tic-tac-toe. Needless to say, regardless of who made the first move, every game ended either in a draw or in a win for the robot. Many were impressed. Today, an exhibit of this sort would be unthinkable, except, possibly in a children's section.

The term "game," in the context of interacting actors with usually different goals, was introduced by von Neumann and Morgenstern (1944) in their seminal treatise *Theory of Games and Economic Behavior.* It will surprise many that von Neumann did not recognize chess as a "game" in the sense that he used the term.

"Chess is not a game," von Neumann told Jacob Bronowski, who worked with him during World War II (Poundstone 1992, p. 6). He meant that there is a "correct" way to play the game – although no one presently knows what it is – and that the game should therefore be trivial, in much the same sense as tic-tac-toe is trivial to players aware of a "correct" strategy.

It turned out that the inspiration for game theory was not chess or any parlor game, which can be shown to have one or more "correct" ways to play it, but poker instead, where it is not possible to guess with certainty the choice of strategy of one's opponent(s).

According to Luce and von Winterfeldt (1994), "[M]ost people in real situations attempt to behave in accord with the most basic (conventional) rationality principles, although they are likely to fail in more complex situations." The "situations" are not mentioned, but one can surmise that "utility" is (perhaps tacitly) represented in them by a linear function of some "good" (money, survivors), and that expected utilities are either given (EU) or subjectively assumed (SEU).

This observation tends to imply that normative (prescriptive) decision theory has a positive role to play along with recent empirical descriptive approaches, which seek to gain understanding of how people actually make decisions in a great variety of situations. Yet, already in the late eighteenth century, maximization of expected utility was shown to lead to absurd results in the so-called Petersburg Paradox. A fair coin is thrown. The gambler wins $n2^{n-1}$ rubles if "heads" appears n times before the first "tails." The gambler's expected gain is infinite. Daniel Bernoulli (1738/1954), to whom the invention of the game is attributed, modified the rule, whereby expected utility increased logarithmically rather than linearly with n. This still made the expected gain, and thus a "rational" maximum stake, enormous, and hence unacceptable to any "rational" player. Indeed, as long as expected gain increases monotonically with n, a rule can be devised to make the expected gain enormous. No "rational" gambler can be expected to pay anywhere near it for the privilege of playing the game once.

Passing from gambling to two-or-more-person games, we encounter similar difficulties with prescriptive decision theory. Especially impressive are paradoxes resulting from backward induction. Consider the Prisoner's Dilemma game played a large known number of times. In a single play, defection by both players is a minimax outcome, which, according to von Neumann, is the only rational one. In a long sequence of plays, however, one might suppose that repeated cooperation (CC) might emerge, as each player forestalls the other's "revenge" for defection. Nevertheless, the last "rational" outcome ought to be double defection (DD), because no retaliation can follow. Given this conclusion, the next to the last play also ought to be (DD), and so on down to the first play.

When Flood and Dresher, discoverers of the Prisoner's Dilemma game (Poundstone 1992), reported to von Neumann that in a long sequence of plays of the game, the outcome was not at all a solid string of DD's, as predicted by the minimax theorem, the great mathematician did not take the result of the admittedly informal experiment seriously. Subsequent experiments, however, showed that, especially in long repeated plays, substantial runs of CC are a rule rather than an exception. Even in single plays by total strangers, frequent CC outcomes have been observed (Rapoport 1988).

Colman cites backward induction in "Centipede," a weirdly designed multi-move game in which both players could win fabulous sums if they tacitly agreed to cooperate after the first play. Nevertheless, backward induction would dictate stopping after the first play, whereby both would receive zero. In contrast, backward induction in R. Selten's "Chain Store" game prescribes CC throughout. The inventor of the game writes that, in the role of the chain store, he would not play as prescribed and presumably would get more money (Selten 1978). Luce and Raiffa (1957) also preferred to violate the backward induction prescription in finitely repeated Prisoner's Dilemma, thus avoiding the only minimax equilibrium of this game.

It turns out that these frequent failures of rational choice theory to prescribe acceptable actions in gambling or game-like situations can be traced to two often tacitly implied assumptions, namely, rejection of so called "evidential" decision theory (Joyce 1999) and independence of decisions of individual players.

Suppose a believer in predestination (Calvinist, Presbyterian) is asked why, if his ultimate abode is fixed, he leads a sober and chaste life. Why doesn't he drink, gamble, chase women, and so on while he can? He might answer, "Since God is just, I can assume that I am among the saved, because I live as I live." He considers his frugal and chaste life as "evidence" that he has been saved and he cherishes this feeling (Joyce 1999).

Asked why he bothers to vote in a general election, seeing that his vote can't possibly make a difference in the result, Herr Kant replies, "I vote, because I would like everyone to vote and because it makes me feel that I have done my duty as a citizen." In the wilderness, Dr. Z has one dose of a life-saving medicine. If given to Mr. X, it will save his life with a probability of 0.9; if given to Mr. Y it has a probability of 0.95. Maximization of expected utility demands that she give the medicine to Mr. Y. But Dr. Z tosses a fair coin to decide. She doesn't want to "play God."

The concept of rationality in classical prescriptive decision theory has three weak spots: individualism, decision independence, and the minimax equilibrium dogma. "Individualism" in this context means "egoism." To avoid the pejorative connotation, Wicksteed (1933) called it "non-tuism." Decision independence is dropped in evidential decision theory. "However I decide, so will my co-players, since there is no reason to suppose that they think not as I do." This perspective dethrones the Nash equilibrium from its role as a *sine qua non* condition of rational choice.

In spite of his generous appreciation of game-theoretic contributions to decision theory, Colman effectively pronounces the end of prescriptive theory founded on the orthodox paradigm, and discusses the promising dawn of inductive experimental-psychological approaches.

Why not go all the way

Richard Schuster

Department of Psychology, University of Haifa, Haifa 31905, Israel.
schuster@psy.haifa.ac.il

Abstract: "Psychology Game Theory" grafts social-process explanations onto classical game theory to explain deviations from instrumental rationality caused by the social properties of cooperation. This leads to confusion between cooperation as a social or individual behavior, and between *ultimate* and *proximate* explanations. If game theory models explain the existence of cooperation, different models are needed for understanding the proximate social processes that underlie cooperation in the real world.

Colman's provocative paper reminds me of a familiar scenario in science. A popular doctrine is under stress but refuses to die. Instead, it is amended again and again in a vain attempt to forge an accommodation with a new reality. A good example is the assumption that individual self-interest, which can explain the evolution of cooperation, must also underlie the *behavior* of cooperating in the real world. Colman characterizes this assumption (from Hume) as "instrumental rationality." The stress comes from

https://doi.org/10.1017/S0140525X03430057 Published online by Cambridge University Press

the emerging reality (cf. Palameta & Brown 1999) that humans, in a variety of interactive decision games, prefer to cooperate more than predicted from instrumental rationality alone. Colman's proposal is to explain the cooperation bias with "Psychological game theory," a set of "formal principles of nonrational reasoning" linked to the social context of cooperation in which the actions and motives of other participants may not be completely known. This commentary will suggest that Colman's allegiance to the basic paradigm of game theory does not allow his theory to go far enough when addressing the social reality of cooperation under free-ranging conditions.

The root of the problem seems to be the familiar mistake of confounding *ultimate* versus *proximate* explanations. If game theoretical models are designed to provide an ultimate explanation for the existence of cooperation, such models only create obscurity and confusion when they are stretched to incorporate what happens when live subjects are used in the kinds of experiments spawned by game theory. Setting aside the evolution of cooperation and altruism at the level of groups (e.g., Sober & Wilson 1998), natural selection at the level of individuals is necessarily selfish. But as Colman realizes (see also Chase 1980; Dennett 1995), classical game theory was never meant to be about behavior. Evolution is a process that guarantees maximization of individual fitness precisely because it is mindless and therefore "nonrational," leading to higher payoffs in the absence of intentional choosing. It is this nonrational aspect of cooperation that is grist for the mill of game theory, whether in behavioral ecology (e.g., Mesterton-Gibbons & Dugatkin 1992), the social sciences (e.g., Luce & Raiffa 1957), or economics (e.g., Arrow 1963).

When humans or animals actually choose whether or not to cooperate, they are no longer "mindless," in the sense that their choices are no longer immune from the influence of proximate psychological processes evoked by the presence and behaviors of others (Boesch & Boesch 1989; Roberts 1997; Schuster 2001; 2002; Schuster et al. 1993). The existence of such processes invites the question of how best to study them. By grafting psychological game theory onto classical game theory, the result is a peculiar kind of hybrid explanation that is part social and part individual, part proximate and part ultimate. On the one hand, cooperation is retained as an individual behavior, resting on "a bedrock of *methodological individualism*" (cf. target article, sect. 8.1, emphasis Colman's). This is obvious from the basic design of laboratory experiments in which social interaction is deliberately minimized or totally absent. Instead, anonymous players are physically isolated in separate cubicles and individually compensated according to how all players have chosen between strategic options.

On the other hand, humans playing such games show a persistent bias towards cooperating, despite the impoverished social conditions. The experimental conditions leave the experimenter with few social variables to manipulate. And the theoretician is left with the less-than-enviable task of speculating about why decision games with human subjects might generate too much cooperation. One proposal is to suggest traits such as "group-identity" or "social value orientation," which can vary across subjects. Colman's proposal is a set of unobservable intervening variables, such as "focal points," "collective preferences," and "team reasoning." Still another possibility is to speculate that humans, despite the physical isolation, are nevertheless influenced by the possibility that outcomes also depend on others (e.g., Forsythe et al. 1994). A player might then opt for cooperation because it is intrinsically reinforcing (Frank 1988; Schuster 2001; 2002; Sober & Wilson 1998), or merely to avoid the embarrassment of meeting opponents who were hurt by defection. At the end of the day, all such explanations remain plausible, but highly speculative, with little likelihood that they could be disentangled using the nonsocial experimental conditions associated with game theory.

To better understand the social dimensions of cooperation, and their impact on instrumental rationality, one alternative is to study examples such as team sports, military strategies, and collaborative business ventures under free-ranging conditions, where cooperation is intrinsically social (see Dugatkin 1997 for a review of animal examples). The social properties of cooperation can then be separately analyzed *before, during* and *after* engaging in a bout of cooperation (Schuster 2001; 2002). Instead of the anonymity among partners favored by game theorists, real-life "players" are usually familiar from group membership and/or prior encounters. Cooperators are therefore likely to have preferred partners (Dugatkin 1995). With interaction unrestricted, social influences also impact on how cooperation is performed, including the ability to coordinate actions, use signals, and develop a division-of-labor based on different and complementary roles (Boesch & Boesch 1989; Schuster et al. 1993; Stander 1992). There is also the possibility that the cooperators themselves are affected by working together in ways that affect the incentive to cooperate (Schuster 2001; 2002). Finally, social influences can affect how outcomes are allocated following an act of cooperation, because of factors such as sex, age, aggression, or status, which influence priority of access (e.g., Boesch & Boesch 1989; Noë 1990). In humans, allocation can also follow from bargaining or prior arrangements. Instrumental rationality is also violated when cooperation persists even though individual behavior might be more profitable (e.g., Packer et al. 1990).

A second alternative is to incorporate free-ranging conditions into experimental models of cooperation. One example is a simple model in which pairs of laboratory rats (*Rattus norvegicus*) are rewarded with a saccharine solution for coordinating back-and-forth shuttling within a shared chamber in which social interaction is unrestricted (Schuster 2001; 2002). Two experiments are relevant to instrumental rationality. In one, the option of competing over outcome allocation was modeled by reinforcing cooperative shuttles with intermittent presentations of either one or two cups of saccharine solution. Although pairs varied in how the single outcomes were allocated, from near "sharing" to strong dominance by one partner, there was *no* relationship between outcome disparity and the level of performance (Schuster et al., in preparation). The second experiment measured choice between entering one chamber, where the rats shuttled alone, or another chamber in which they always cooperated with the same partner. This situation is analogous to the game theory choice between cooperation and noncooperation, with the important difference being that the cooperation option, as in real life, was also a social option. Even though there was no advantage from choosing either chamber, because outcomes and the proportions of reinforced shuttles were matched, cooperation was strongly preferred (Schuster 2001; 2002; Schuster et al., in preparation).

By incorporating the social dimensions of cooperation into experimental models, it becomes possible to clarify the differences between ultimate explanations based on game theory or proximate explanations based on social processes. This clarification would not be needed if predictions from the two kinds of explanations were congruent. But congruence seems more likely when behavior is *not* social, so that the abstract prediction of maximizing fitness or expected utility can be better matched by the influence of feedback from individual experience (e.g., Staddon 1983). The cooperation bias shows that the advantage from using game-theory models, eliminating the "noise" generated by uncontrollable social interactions, can no longer be defended. The same methods are also inadequate for analyzing why the preference for cooperation deviates from instrumental rationality.

So why not go all the way and explain cooperation under free-ranging conditions by jettisoning the game-theory approach, in both method and theory, in favor of alternative models that are more faithful to the reality of cooperation in the real world? Then we could begin to study what is impossible with Colman's approach, namely, the experimental analysis of deviations from predictions based on instrumental rationality, and the likely identification of, as yet undiscovered, sources of reinforcement that underlie cooperation as we know it.

https://doi.org/10.1017/S0140525X03430057 Published online by Cambridge University Press