namics can be rationalized as a form of social or cultural learning: BGS emphasize the role of aspirations, but evolution toward fair outcomes is also consistent with imitation (Björnerstedt & Weibull 1996). It is tempting, then, to interpret the second BGS outcome as a Falk and Fischbacher (2000) "fairness equilibrium."

All of this said, we share most of Colman's concerns with standard game theoretic arguments, and suspect that psychological game theorists, both old and new, will have much to contribute to the literature.

ACKNOWLEDGMENTS
We thank Corinna Noelke and Carolyn Craven for their comments on a previous draft.

## To have and to eat cake: The biscriptive role of game-theoretic explanations of human choice behavior

William D. Casebeer[a] and James E. Parco[b]
[a]*Department of Philosophy, United States Air Force Academy, Colorado Springs, CO 80840;* [b]*American Embassy, Tel Aviv, 63903 Israel.*
**william.casebeer@usafa.af.mil james.parco@usafa.af.mil**
**http://www.usafa.af.mil/dfpfa/CVs/Casebeer.html**
**http://parco.usafa.biz**

**Abstract:** Game-theoretic explanations of behavior need supplementation to be descriptive; behavior has multiple causes, only some governed by traditional rationality. An evolutionarily informed theory of action countenances overlapping causal domains: neurobiological, psychological, and rational. Colman's discussion is insufficient because he neither evaluates learning models nor qualifies under what conditions his propositions hold. Still, inability to incorporate emotions in axiomatic models highlights the need for a comprehensive theory of functional rationality.

The power and beauty of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* (1944) and Luce and Raiffa's *Games and Decisions* (1957) lie in their mathematical coherence and axiomatic treatment of human behavior. Once rational agents could be described mathematically, game theory provided a far-reaching normative model of behavior requiring an assumption of common knowledge of rationality. This assumption (in addition to the often unstated requirement that a player fully understand the game situation) is subsumed under the phrase "the theory assumes rational players" (Luce & Raiffa 1957). But we know that, descriptively speaking, this is not always the case. The literature has clearly shown that not only are these (mathematically required) assumptions often too strong to be met in practice, but also that the "rational actor theory" (hereafter RAT) is underspecified in that it cannot effectively accommodate emotions. But does this constitute a failure of RAT? We think not.

Nevertheless, we agree with Colman's larger point that we need a "psychological game theory," or rather, a neurobiologically informed theory of decision-making. This is not because of the spectacular failure of game theoretic assumptions in any particular experiment, but rather stems from an ecumenical and fully naturalizable worldview about the causes of, and norms governing, human behavior. Choice-driven behavior is a function of multiple, highly distributed brain subsystems that include affect and emotion. For example, in the domain of moral judgment, good moral cognition is driven by a variety of brain structures, only some involved in ratiocination as traditionally construed (Casebeer & Churchland 2003). Even the most ardent RAT enthusiast recognizes that if your *explanandum* is all human behavior, your *explanans* will be more comprehensive than adverting to RAT alone.

Thus, we question the usefulness of Colman's ad hoc refinements for prescriptions of behavior in interactive decision-making, primarily because he has neither (1) qualified his theory as to when and under what conditions it applies, nor (2) provided an account for learning in games (beyond simple Stackelberg reasoning). For example, Colman uses the two-player centipede game as a primary domain in which he justifies his theory. However, recent evidence experimentally investigating three-player centipede games (Parco et al. 2002) directly contradicts it. Parco et al. extended the McKelvey and Palfrey (1992) study to three players using small incentives (10 cents for stopping the game at the first node, and $25.60 for continuing the game all the way to the end) and obtained similar results, soundly rejecting the normative equilibrium solution derived by backward induction. However, when the payoffs of the game were increased by a factor of 50 (and each player thus had the opportunity to earn $7,680), the results were markedly different. Although initial behavior of both the low-pay and high-pay conditions mirrored that of the McKelvey and Palfrey study, over the course of play for 60 trials, behavior in the high-pay treatment converged toward the Nash equilibrium and could be well accounted for using an adaptive reinforcement-based learning model. Furthermore, as noted by McKelvey and Palfrey (1992) and later by Fey et al. (1996), in all of the centipede experiments that were conducted up until then, there were learning effects in the direction of equilibrium play. Colman's oversight of the extant learning in games literature and his brief account for the dynamics of play through Stackelberg reasoning is insufficient. Learning in games manifests itself in a variety of processes quite different from simple Stackelberg reasoning (see Camerer & Ho 1999; Erev & Roth, 1998). For example, Rapoport et al. (2002) document almost "magical" convergence to the mixed-strategy equilibrium over 70 trials without common knowledge or between-trial feedback provided to subjects. Neither traditional game theory nor Colman's model can account for such data.

Generally speaking, Colman does little to improve prescriptions for human behavior both within and outside of the subset of games he has described; his paper is really a call for more theory than a theory proper. RAT's difficulty in dealing with emotions serves as proof-of-concept that we need a more comprehensive theory. Humans are evolved creatures with multiple causes of behavior, and the brain structures that subserve "rational" thought are, on an evolutionary timescale, relatively recent arrivals compared to the midbrain and limbic systems, which are the neural mechanisms of affect and emotion. Ultimately, our goal should be to formulate an explanation of human behavior that leverages RAT in the multiple domains where it is successful, but that also enlightens (in a principled way) as to when and why RAT fails. This more comprehensive explanation will be a neurobiological cum psychological cum rational theory of human behavior.

The problems game-theoretic treatments have in dealing with the role of emotions in decision-making serve to underscore our point. There are at least two strategies "friends of RAT" can pursue: (1) attempt to include emotions in the subjective utility function (meaning you must have a mathematically rigorous theory of the emotions; this is problematic), or (2) abandon RAT's claim to be discussing proximate human psychology and, instead, talk about how emotions fit in system-wide considerations about long-term strategic utility (Frank 1988). The latter approach has been most successful, although it leaves RAT in the position of being a distal explanatory mechanism. The proximate causes of behavior in this story will be locally arational or possibly irrational (hence the concerns with emotions). How would "new wave RAT" deal with this? One contender for a meta-theory of rationality that can accommodate the explanatory successes of RAT, yet can also cope with their failure in certain domains, is a functional conception of rationality. The norms that govern action are reasonable, and reason-giving for creatures that wish to be rational, insofar as such norms allow us to function appropriately given our evolutionary history and our current environment of action (Casebeer 2003).

We acknowledge that RAT will require supplementation if it is to fully realize its biscriptive explanatory role of predicting human action and providing us with a normative yardstick for it. Utility theory must incorporate neurobiological and psychological deter-

https://doi.org/10.1017/S0140525X03280055 Published online by Cambridge University Press

minants, as well as the rational, if game theory is to become as descriptively appealing as it is normatively.

## Experience and decisions

Edmund Fantino and Stephanie Stolarz-Fantino

*Department of Psychology, University of California–San Diego, La Jolla, CA 92093-0109.* **efantino@ucsd.edu** **sfantino@psy.ucsd.edu**

**Abstract:** Game-theoretic rationality is not generally observed in human behavior. One important reason is that subjects do not perceive the tasks in the same way as the experimenters do. Moreover, the rich history of cooperation that participants bring into the laboratory affects the decisions they make.

Colman reviews many instances of game playing in which human players behave much more cooperatively and receive larger payoffs than permitted by conceptions of strict rationality. Specifically, he points out that although "Game-theoretic rationality requires rational players to defect in one-shot social dilemmas" (sect. 6.11), experimental evidence shows widespread cooperation. We agree that strict rationality does not accurately portray or predict human behavior in interactive decision-making situations. Particularly problematic are predictions made on the basis of backward induction. The Chain-store and Centipede games are good examples. In each case, backward induction makes it appear that the likely last move is inevitable, rather than one of a number of possible outcomes, as it must appear to the participant. In any case, it is unlikely that participants would reason backwards from the conclusion, even if such reasoning made sense. For example, Stolarz-Fantino et al. (2003) found that students were more likely to demonstrate the conjunction effect (in which the conjunction of two statements is judged more likely than at least one of the component statements) when the conjunction was judged before the components, than when it was judged after them. Further, if people easily reasoned backward from likely end-states, they should be more adept at demonstrating self-control (preferring a larger, delayed reward to a smaller, more immediate reward) than in fact they are (see discussion in Logue 1988).

Colman proposes "Psychological game theory" as a general approach that can be argued to account for these deviations. We agree that this is a promising approach, although it is a fairly broad and nonspecific approach as presented in the target article. We would add a component to Psychological game theory that appears to be relevant to the types of problems discussed: the pre-experimental behavioral history of the game participants. We are studying various types of irrational and nonoptimal behavior in the laboratory (e.g., Case et al. 1999; Fantino 1998a; 1998b; Fantino & Stolarz-Fantino 2002a; Goodie & Fantino 1995; 1996; 1999; Stolarz-Fantino et al. 1996; 2003) and are finding a pronounced effect of past history on decision-making (a conclusion also supported by Goltz' research on the sunk-cost effect, e.g., Goltz 1993; 1999). One example will suffice.

A case of illogical decision-making is base-rate neglect, first developed by Kahneman and Tversky (1973) and discussed often in this journal (e.g., Koehler 1996). Base-rate neglect refers to a robust phenomenon in which people ignore or undervalue background information in favor of case-specific information. Although many studies have reported such neglect, most have used a single "paper-and-pencil" question with no special care taken to insure attentive and motivated subjects. Goodie and Fantino wondered if base-rate neglect would occur in a behavioral task in which subjects were motivated and in which they were exposed to repeated trials. We employed a matching-to-sample procedure (MTS), which allowed us to mimic the base-rate problem quite precisely (Goodie & Fantino 1995; 1996; 1999; Stolarz-Fantino & Fantino 1990). The sample in the MTS task was either a blue or green light. After sample termination, two comparison stimuli appeared: these were always a blue and a green light. Subjects were instructed to choose either. We could present subjects with repeated trials rapidly (from 150 to 400 trials in less than a one-hour session, depending on the experiment) and could readily manipulate the probability of reinforcement for selecting either color after a blue sample and after a green sample. Consider the following condition (from Goodie & Fantino 1995): Following either a blue sample or a green sample, selection of the blue comparison stimulus is rewarded on 67% of trials, and selection of the green comparison stimulus is rewarded on 33% of trials; thus, in this situation the sample has no informative or predictive function. If participants responded optimally, they should have come to always select blue, regardless of the color of the sample; instead they focused on sample accuracy. Thus, after a green sample, instead of always choosing blue (for reward on 67% of trials) they chose the (matching) green comparison stimulus on 56% of trials (for a 48% rate of reward). This continued for several hundred trials. In contrast, Hartl and Fantino (1996) found that pigeons performed optimally, ignoring the sample stimulus when it served no predictive function. They did not neglect base-rate information.

What accounts for pigeons' and people's differing responses to this simple task? We have speculated that people have acquired strategies for dealing with matching problems that are misapplied in our MTS problem (e.g., Stolarz-Fantino & Fantino 1995). For example, from early childhood, we learn to match like shapes and colors at home, in school, and at play (e.g., in picture books and in playing with blocks and puzzles). Perhaps, this learned tendency to match accounts for base-rate neglect in our MTS procedure. If so, Goodie and Fantino (1996) reasoned that base-rate neglect would be eliminated by using sample and comparison stimuli unrelated to one another (line orientation and color). In this case, base-rate neglect was indeed eliminated. To further assess the learning hypothesis, Goodie and Fantino (1996) next introduced an MTS task in which the sample and comparison stimuli were physically different but related by an extensive history. The samples were the words "blue" and "green"; the comparison stimuli were the colors blue and green. A robust base-rate neglect was reinstated. Ongoing research in our laboratory is showing that pigeons with sufficient matching experience (where matching is required for reward) can be induced to commit base-rate neglect. These and other studies have led us to conclude that base-rate neglect results from preexisting learned associations.

How might learned associations account for nonoptimal decisions in the Prisoner's Dilemma Game (PDG)? Rationality theory argues that the selfish response is optimal. But we have been taught since childhood to be unselfish and cooperative. For many of us, these behaviors have been rewarded with praise throughout our lives (see the discussion of altruism in Fantino & Stolarz-Fantino 2002b; Rachlin 2002). Moreover, actual deeds of unselfish and cooperative behavior are often reciprocated. Why then should these behaviors not "intrude" on the decisions subjects make in the laboratory? Viewed from this perspective, there is nothing surprising about the kinds of behavior displayed in PDG. Indeed, such behavior is variable (many subjects cooperate, many defect), as one would expect from the variable behavioral histories of the participants.

## A critique of team and Stackelberg reasoning

Herbert Gintis

*Emeritus Professor of Economics, University of Massachusetts, Northampton, MA 01060; External Faculty, Santa Fe Institute, Santa Fe, NM.*
**hgintis@comcast.net** **http://www-unix.oit.umass.edu/~gintis**

**Abstract:** Colman's critique of classical game theory is correct, but it is well known. Colman's proposed mechanisms are not plausible. Insufficient reason does what "team reasoning" is supposed to handle, and it applies to a broader set of coordination games. There is little evidence ruling out more traditional alternatives to Stackelberg reasoning, and the latter is implausible when applied to coordination games in general.

https://doi.org/10.1017/S0140525X03280055 Published online by Cambridge University Press