Game theory need not abandon individual maximization

John Monterosso[a] and George Ainslie[b]

[a]*Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA 90024;* [b]*Department of Psychiatry, Coatesville VA Medical Center, Coatesville, PA 19320.* **jmont@ucla.edu** **george.ainslie@med.va.gov**

Abstract: Colman proposes that the domain of interpersonal choice requires an alternative and nonindividualistic conception of rationality. However, the anomalies he catalogues can be accounted for with less radical departures from orthodox rational choice theory. In particular, we emphasize the need for descriptive and prescriptive rationality to incorporate recursive interplay between one's own choices and one's expectation regarding others' choices.

Colman proposes that an alternative conception of rationality is required to account for human interaction, and he provides some suggestions in this direction. What he specifically sees the need to give up is "methodological individualism" – the premise that "rational play in games can be deduced, in principle, from one-person rationality considerations" (Binmore 1994a, quoted in target article, sect. 4.1, para. 1). We think the anomalies he catalogues can be accounted for without abandoning this foundational principle of deterministic behavioral science.

First, the prevailing payoffs in experimental games are not the same as the specified payoffs. Social interactions are rife with invisible contingencies that are impossible to bring under full experimental control. Human beings are fundamentally social creatures, which entails the presence of powerful interpersonal motivations, too numerous to list. Otherwise, anomalous play, such as rejecting a low offer in a one-shot ultimatum game or cooperating in a one-round prisoner's dilemma game, is sensible if we allow that the dollars offered do not exhaust the prevailing payoffs. Colman discusses this type of proposal (Camerer's "behavioral game theory," Rabin's fairness equilibrium), but he concludes it is not enough to account for all the phenomena he presents. We agree, and furthermore, do not think that the subjects' many motives, beyond maximizing the specified matrix outcomes, are orderly enough to inspire any useful addition to game theory (such as adding X points to particular cells); discrepancies between the specified payoffs and the prevailing payoffs will always be noise in the experiment, the friction that distorts the ideal physics lab.

However, permitting the free use of probability estimates of other's choices should be enough to let "methodological individualism" both describe and prescribe rationality to the extent that subjects *are* motivated by the specified matrices of the game. Of particular importance in explaining otherwise anomalous play is the subject's use of her own inclinations and behavior as test cases that inform her expectations regarding what others will do. In the kinds of situations game theorists care about, it is neither descriptively tenable *nor prescriptively effective* to require individuals to finalize their assessments of what others will do prior to considering what they will do. Instead, we think that a rational individual is simultaneously engaging in both computing expectation of what the other player will be motivated to do and contemplating what she herself should do, and each process informs the other. In the absence of specific information about one's counterpart, what better basis is there to predict her behavior than via one's own response to the situation?

Colman describes such a recursive process in characterizing one attempt to develop a game-theoretic rationale for the Pareto-dominant H-H solution in the Hi-Lo game. In this account, Player I assumes by default that Player II's strategies are equally probable. She thus concludes she should choose H, because the probability-weighted sum is higher. But this, Colman adds, violates rational choice theory. "By the transparency of reason, Player I 's intention to choose *H* would be common knowledge and would induce Player II to choose the best reply, namely *H*, with *certainty*, contradicting Player I's initial assumption" [i.e., of equal probability of moves] (sect. 5.6, para. 4). While such recursion may violate game theory's constraints, we think it is descriptively accurate, prescriptively rational, and it does not entail abandoning methodological individualism.

The recursion between someone's own perceived choices and their expectations about the choices of others is easier to see in a *discoordination* variant of the Hi-Lo game, in which players get to keep their choice (in some monetary unit) if and only if they chose differently from each other. With no a priori expectation regarding what Player II will choose, Player I's first-order inclination is to choose the high amount, following the same logic as above. But seeing the similarity of her counterpart's predicament, she may expect her to have thought the same way, giving her the second-order inclination that she must go for the lower amount to get anything. But then again, if she thinks her counterpart is a similarly sophisticated sort, she might get the feeling that her counterpart went through the same thought process, thus giving her the third-order inclination that maybe she *should* therefore go for *H*. The more similar she thinks her counterpart to be to herself, the more dizzying the potential for iteration, and the less likely there will be a probable solution.

The recursive prediction model has the advantage that it also provides intertemporal bargaining within the individual person. In situations that involve resisting temptation, individuals cannot be certain of their own future choices. The need to choose in the present, with an eye to the precedent this choice will set for the future (e.g., whether or not I am sticking to my diet), places people in a situation analogous to a repeated prisoner's dilemma (PD) game, as we have argued elsewhere (Ainslie 2001, pp. 90–104; Ainslie & Monterosso 2003; Monterosso et al. 2002). Briefly, the danger that future selves will see past violations of a resolution as a reason to violate it, in turn, is similar to the danger that one player's defection will cause the other(s) to defect. But, in this bargaining, a person may propose a choice to herself ("I'll have an ice cream"), then put herself in the shoes of her future self to evaluate it retrospectively ("I'll have gone off my diet"), then revise her current choice in light of this evaluation ("I'll have a muffin instead"), and evaluate this ("no"), and propose again ("a *bran* muffin") at some length before making a single concrete choice. Choices may turn out to divide along salient features, just as in the Hi-Lo game, not because of their intrinsic payoff, but because they make intertemporal cooperation more likely. Intertemporal bargaining theory predicts the emergence of both positive and negative features that have been ascribed to willpower. It generates an internal version of Adam Smith's "unseen hand" without assuming an innate faculty of self-control.

Second-order indeterminacy

Marco Perugini

Department of Psychology, University of Essex, Colchester, CO4 3SQ United Kingdom. **mperug@essex.ac.uk**
http://privatewww.essex.ac.uk/~mperug

Abstract: Psychological game theory, as defined by Colman, is meant to offer a series of solution concepts that should reduce the indeterminacy of orthodox game theory when applied to a series of situations. My main criticism is that, actually, they introduce a second-order indeterminacy problem rather than offering a viable solution. The reason is that the proposed solution concepts are under-specified in their definition and in their scope.

Colman looks at game theory from a psychological perspective. In the first part of his article, he convincingly argues about the limitations of orthodox game theory, especially when applied to social interactions. The examples are well chosen and the case is well built. This is an important contribution that might help us to focus, once and for all, on these important issues. However, Colman's suggestion of psychological game theory as a way forward to overcome the severe limitations of orthodox game theory in explaining social interactions is not entirely convincing. The spirit behind this attempt should be praised, yet psychological game theory as defined and exemplified by Colman does not offer a truly viable solution. The key problem is that the suggested solutions are theoretically under-specified, quite limited in scope, and lead to a second-order indeterminacy.

To illustrate my point I will focus on the concept of "team reasoning." What is so special about team reasoning that cannot be said about other ways of reasoning? For example, one might define "altruistic reasoning," "individualistic reasoning," "fairness reasoning," "reciprocity reasoning," and so on, in the same kind of holistic way as the definition is offered for "team reasoning." It is easy to find examples of games that can be solved using some of these concepts; although they can be solved promptly also via "team reasoning," the intuition is that it would not necessarily be the best solution concept. By best solution concept I mean a concept that is intuitively compelling and likely to be empirically supported with actual behavioral data.

I will present two examples of games. For the first example, let's consider all modified coordination games for two players with asymmetrical payoffs. Let's consider this asymmetric coordination game with the following payoffs and choices (Fig. 1):

As for every coordination game, a standard analysis would show two Nash equilibria (*H*, *H* and *L*, *L*), and the issue would be how to select one of the two. Applying a team reasoning would single out *H*, *H* as the best equilibrium. Would this be a compelling solution? I doubt it. If I were Player I, I would think twice before choosing *H*. By applying "fairness reasoning" or "Reciprocity reasoning," I could anticipate that Player II would like *L*, *L* much more than *H*, *H* (or, put differently, dislike much more the inequality of payoffs resulting from *H*, *H*). I would therefore anticipate that the other player would play *L*, and as a consequence I would decide to play *L*. On the other hand, if I were to apply "altruistic reasoning" or "individualistic reasoning," for opposite reasons I should come to the conclusion that Player II will play *H*, and hence so would I. The problem is threefold: First, we can list a series of reasoning concepts besides "team reasoning"; second, psychological game theory, as defined by Colman, would offer no tools to select among these different reasoning concepts; and third, the solution concept which would be the best for a player, depends on his expectations about the other player's type.

The second example is perhaps even more intriguing.[1] The Ultimatum Game (UG) is a well-known paradigm that has been the subject of several studies in experimental economics and in social psychology. The UG is a very simple game whereby two players bargain over a given monetary endowment. The first player proposes a division of the endowment and the second player can either accept or refuse it. If she refuses it, both players end up with nothing. Orthodox game theory predicts that the first player will propose a small amount for the second player (e.g., 99% for self vs. 1% for other) and the second player will accept the proposal. However, several experimental studies have found systematic deviations from these predictions (e.g., Guth 1995; Thaler 1988). It is well established that a consistent portion of second players would reject low offers (e.g., 25% or lower) even though this means that both players end up with nothing. What about team reasoning? A team-reasoning second player should never reject any offer, because from the perspective of a second player the strategy that maximizes the joint payoff is to accept any offer. In fact, for every offer, the alternative would be to reject it, which is always dominated in terms of joint payoffs, given that it implies no payoff for both players. Therefore, a team reasoning second player would be equally likely to accept a 1/99 or a 50/50 split. The intriguing conclusion is that a team-reasoning player often will behave exactly as dictated by orthodox game theory, even in those situations where our intuition would suggest we do otherwise.

Equally problematic are those cases where team reasoning offers different predictions from orthodox game theory. Take social dilemmas. Of course, social dilemmas can be solved by using team reasoning, but this is equally true for several of the nonstandard solution concepts that I have sketched previously. I wonder how well a team reasoning concept would fare when compared with other nonstandard solution concepts across a comprehensive range of social dilemmas. To sum up, I am not convinced that team reasoning can be a good solution to much more than the specific example of the Hi-Lo matching game with symmetrical payoffs illustrated by Colman. But then, why should it not be named "matching reasoning" instead?

These examples illustrate my main problem with Colman's suggestions: Concepts such as team reasoning must be defined more precisely, which ultimately means that it will be necessary to specify the payoffs involved, how they are transformed, and under which conditions each solution concept primarily applies. The preceding examples have made clear that an important parameter is the symmetry of the payoffs for the players: Everything else being equal, the more asymmetrical the payoffs, the less likely is that team reasoning can offer a compelling solution for all players. But this implies that the team reasoning concept should specify what level of asymmetry is acceptable to the players, which ultimately means to specify some function of weighting the payoffs involved. Only in this way can the solution concepts pass more stringent theoretical and empirical tests. The alternative would be to have a storage bin full of loose ad-hoc reasoning concepts that can be used post-hoc for different situations, but without any rule that specifies when and why they should be adopted. In other words, ironically, the lack of a reason for choosing, which was the main point behind many of Colman's sharp criticisms on the indeterminacy of orthodox game theory, will strike back with a vengeance. Without specifying the concepts more precisely – given that they can explain or predict only some interactions and not others, and that alternative nonstandard concepts can be compellingly applied in several circumstances – we will be left without any reason why to apply a given nonstandard psychological solution concept in the first place.

		I: H	I: L
II	H	100, 10	0, 0
	L	0, 0	9, 9

Figure 1 (Perugini). Example of a coordination game with asymmetric payoffs.

ACKNOWLEDGMENT
Preparation of this commentary was supported by RPF grant DGPC40 from the University of Essex.

NOTE
1. I owe this example to Tim Rakow.

Chance, utility, rationality, strategy, equilibrium

Anatol Rapoport
Department of Psychology, University of Toronto, Toronto, Ontario M5S 3G3, Canada. **anatol.rapoport@utoronto.ca**

Abstract: Almost anyone seriously interested in decision theory will name John von Neumann's (1928) Minimax Theorem as its foundation, whereas Utility and Rationality are imagined to be the twin towers on which the theory rests. Yet, experimental results and real-life observations seldom support that expectation. Over two centuries ago, Hume (1739–40/1978) put his finger on the discrepancy. "Reason," he wrote "is, and ought to be the slave of passions, and can never pretend to any other office than to serve and obey them." In other words, effective means to reach specific goals can be prescribed, but not the goals. A wide range of experimental results and daily life behavior support this dictum.