Available online at www.sciencedirect.com

SCIENCE DIRECT

Theoretical Population Biology 69 (2006) 339–348

Theoretical Population Biology

www.elsevier.com/locate/tpb

Game theory and human evolution: A critique of some recent interpretations of experimental games

Edward H. Hagen*, Peter Hammerstein*

Institute for Theoretical Biology, Humboldt University, Invalidenstr. 43, 10115 Berlin, Germany

Received 25 July 2005
Available online 3 February 2006

Abstract

Economists and psychologists have been testing Nash equilibrium predictions of game theory models of human behavior. In many instances, humans do not conform to the predictions. These results are of great interest to biologists because they also raise questions about well-known ESS models of cooperation. Cooperation in certain one-shot, anonymous interactions, and a willingness to punish others at a net cost to oneself are some of the most intriguing deviations from standard theory. One proposed explanation for these results that is receiving increasing attention invokes the cultural group selection of 'other regarding' social norms. We critically review this explanation. We conclude that experimental results reveal limits in two implicit models of cognitive structure commonly employed by economists and evolutionary biologists.

Keywords: Altruism; Strong reciprocity; Behavioral economics; Ultimatum game; Cultural group selection

I once saw a woman hit by a car...and she was lying in the middle of the road. I knew that at that moment I would have risked my life if necessary to help her, whereas if I had merely read about the accident or heard about it, it could not have meant too much. Stanley Kubrick

1. Introduction

In biology, the concept of the ESS was developed to explain behavior that evolved by frequency-dependent selection (Maynard Smith and Price, 1973). In economics, the closely related concept of the Nash equilibrium was developed to analyze decisions in the context of strategic interaction. Economists have been testing Nash equilibrium predictions of game theory models of human behavior using a variety of experiments (e.g., Camerer et al., 2004). In many cases, humans do not conform to these predictions, casting some doubt on well-known ESS models as well. This poses a problem to both economists and biologists, who have recently begun to collaborate theoretically—developing population models—and empirically studying experimental economic games (Hammerstein and Hagen, 2005). Here we describe and critique one pioneering explanation for some of the discrepancies between game equilibrium predictions and experimental results. This explanation is based on models of the cultural evolution of cooperation that include an element of cultural group selection, resulting in prosocial (other regarding) utility functions. Our critique highlights the differences between economists' implicit view of the human brain as a utility maximizer (subject to constraints) with evolutionary biologists' implicit view of the brain as an assemblage of evolved, specialized mechanisms. These dissimilar perspectives make it difficult to interpret the results of the experimental games and the extent to which they support cultural group selection. Successfully explaining the game results will require integrating aspects of the cognitive and other social sciences into theoretical population biology, an important challenge for the future development of evolutionary game theory.

*Corresponding authors.
E-mail addresses: e.hagen@biologie.hu-berlin.de (E.H. Hagen), p.hammerstein@biologie.hu-berlin.de (P. Hammerstein).

0040-5809/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2005.09.005

2. Strong reciprocity and the ultimatum game

Strong reciprocity is a predisposition to cooperate with others and to punish those who violate the norms of cooperation, at personal cost, even when these costs would not be repaid (Gintis et al., 2003; Gintis, 2000). A number of scholars in anthropology and economics have argued that the results of certain economics experiments provide convincing evidence for strong reciprocity. Strong reciprocity, they believe, must have evolved by some sort of genetic or, more plausibly, cultural group selection, and is therefore an adaptation (e.g., Boyd et al., 2003; Fehr and Henrich, 2003; Fehr et al., 2002; Henrich et al., 2005).

The experiments cited in support of strong reciprocity typically involve a game, such as the ultimatum game, played for real money. In the ultimatum game, a proposer offers a responder a fraction s of a fixed amount of money put up by the experimenter. If the responder accepts s, he gets s and the proposer keeps the remainder. If the responder rejects s, both get 0. A rational, profit-maximizing responder would accept any s, no matter how small. However, in one-shot experiments with complete anonymity conducted in numerous cultures, proposers routinely offer much more than 0, and in a few societies many responders reject even relatively generous offers (Henrich et al., 2005).

We will mostly limit our discussion to the ultimatum game because results are available for a wide range of societies, and the cross-cultural pattern of altruism that was discovered in a study of 15 cultures (Henrich et al., 2005) has been highlighted as supporting strong reciprocity. We will occasionally mention other games, however, such as the public goods game and dictator game. In the public goods game, the experimenter multiplies voluntary contributions to a 'pot' by a factor >1; the pot is then equally distributed to all players regardless of their contribution level. In the dictator game, player 1 can offer any fraction of an endowment to player 2, keeping the rest for himself. Player 2 has no option to refuse the offer.

3. Rationality and inequity aversion

Economics has long assumed that (1) individuals are rational decision-makers who seek to maximize their utility, where (2) utility is defined in terms of individual benefit, such as monetary profit. This model of decision-making was motivated as much by its analytic tractability and intuitive appeal as it was by empirical facts (Tversky and Kahneman, 1986).

Because players in the ultimatum game often do not maximize their individual monetary profit, one or both of the above assumptions must be incorrect. Strong reciprocity theorists have chosen to retain (1) while modifying (2).

To explain the altruism in the ultimatum game and other examples of strong reciprocity, Fehr and Schmidt (1999) propose a utility function where for a given individual payoff, utility is maximized if inequity is minimized. If $x = (x_1, \dots, x_n)$ is the vector of monetary payoffs, the 'inequity aversion' utility function of player $i \in \{1, \dots, n\}$ is given by:

$$U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\}, \quad (1)$$

where $\beta_i \leqslant \alpha_i$ and $0 \leqslant \beta_i < 1$. The second term is the utility loss from disadvantageous inequality, and the third term the loss from advantageous inequality. Given a monetary payoff x_i, player i's utility is maximized at $x_j = x_i$. After empirically calibrating α_i and β_i, this function can account for players' performance in the ultimatum and similar games (but see Fehr and Gächter, 2005 for evidence that inequity aversion cannot be the sole motive for punishment).

How would such an 'other regarding' utility function evolve? Perhaps, Fehr et al. (2002) argue, by cultural group selection.

4. Cultural group selection

Here we describe one sophisticated theory for the evolution of the 'other regarding' utilities that, in the view of strong reciprocity scholars, could explain the anomalous game results while preserving the rational actor model. This explanation was inspired by the observation that standard evolutionary theories of cooperation could not account for a singularly human style of cooperation.

4.1. The evolution of human cooperation in large groups

Humans cooperate in groups ranging in size from dyads to millions. Even in small scale, foraging societies, group sizes exceeding 100 are not unknown, especially during seasonal aggregations. Further, although average residential group size is around 25, cooperation among groups is common, creating fairly large corporate entities (Kelly, 1995). Kin selection and reciprocal altruism might explain cooperation in families and among friends, but they do not explain close cooperation in moderately large groups. Relatedness decreases exponentially with each genealogical remove, so, under plausible assumptions about the composition of human families and human mating systems, the cost/benefit ratio of helping would have to decrease rapidly with increasing group size, which is unlikely. Models of n-player prisoner's dilemma games similarly show that TIT-FOR-TAT type strategies, in which a single defector forces everyone to defect, cannot easily evolve in even moderately sized groups (e.g., Boyd and Richerson, 1988).

Under certain conditions, genetic group selection could select for altruism in large groups. To be effective, genetic group selection usually requires relatively high levels of between-group genetic variance relative to within-group genetic variance. Ethnographic (Murdock, 1967) and genetic (e.g., Seielstad et al., 1998) evidence for humans suggests that women often migrate between groups, however, making genetic group selection a weak evolutionary force.

4.2. The cost of acquiring information

Using cultural evolutionary, and gene-culture co-evolutionary frameworks (e.g., Cavalli-Sforza and Feldman, 1973, 1981; Feldman and Cavalli-Sforza, 1976; Boyd and Richerson, 1985; Feldman and Zhivotovsky, 1992; Durham, 1991), Boyd, Richerson, and their students, have grounded the evolution of large-scale human cooperation in the cost of acquiring information (Boyd and Richerson, 1985; Henrich and Boyd, 1998; Richerson and Boyd, 1998). If learning about varying physical or social environments is costly, it can pay to substitute social learning—acquiring information from others—for individual learning—acquiring information oneself. In *conformist transmission*, for example, individuals are biased to adopt common behaviors (Boyd and Richerson, 1985; Henrich and Boyd, 1998). In *prestige-biased transmission*, individuals are biased to adopt the behaviors of successful individuals (e.g., Henrich and Gil-White, 2001).

Boyd and Richerson (1985) argued out that once a social learning strategy like conformist transmission evolved by (genetic) natural selection, it could bootstrap what they termed *cultural group selection*. If groups competed with one another, and if group success co-varied with culturally acquired norms, such as group-altruism norms, then (non-genetic) selection would favor groups with more effective norms relative to those with less effective norms, and the more effective norms would spread in the population. In particular, individually costly but group-beneficial norms—strong reciprocity—could spread.

Boyd and Richerson's argument hinges on the proposal that migration between groups would not hinder cultural group selection as severely as it does genetic group selection. Immigrants arriving with different norms would tend to conform to the norms of the community they are joining. This conformism reduces within-group variance in norms, and maintains between-group variance. Ethnographic evidence for group extinction rates suggests that under this type of cultural group selection it would take centuries or millennia for a group-beneficial trait to go to fixation (Soltis et al., 1995). History tells us, however, that cultural evolution can be rapid.

Prestige-biased group selection is a much faster cultural evolutionary mechanism. If different groups find themselves at different cooperative equilibria with different average payoffs for group members, and if individuals have a genetically evolved bias to acquire the norms of successful individuals, then individuals in groups at equilibria with lower payoffs will copy individuals in groups with higher payoffs, rapidly spreading the norms of the more successful groups (Boyd and Richerson, 2002; Henrich and Boyd, 2001).

A somewhat different group selection model of large scale cooperation exploits the fact that, unlike altruistic cooperation, the cost of altruistic punishment decreases as the number of defectors decreases because there is less need to punish (Boyd et al., 2003). Selection against altruistic punishers within groups is therefore weak when punishment is common. As long as migration rates are not too high, payoff biased imitation maintains variation among groups in the frequency of defectors, because in groups in which punishers are common, defectors achieve a low payoff and are unlikely to be imitated, whereas in groups in which punishers are rare, defectors achieve a high payoff and are widely imitated. Group selection then maintains the correlated traits of altruistic punishment and within-group cooperation. For plausible parameter values, this model can support cooperation in groups of at least 100 individuals.

Strong reciprocity theorists argue that the generous offers and costly punishment seen in the ultimatum and other games are evidence for group-beneficial norms that probably first evolved by some type of cultural group selection, followed by genetic evolution that produced prosocial motives and enforcement mechanisms.

4.3. Strategic non-conformism

The research program of Boyd and Richerson and their colleagues has produced many genuinely important insights, and there is intriguing ethnographic evidence in support of it that we do not discuss here. We nonetheless have some concerns. Principally, many of their models rely heavily on conformism. There is little doubt that conformism is a significant aspect of human psychology, but it is not clear how vulnerable models of cultural group selection are to *strategic non-conformism*. An immigrant could, for example, copy individually beneficial behaviors, like what to eat in her new environment, but refuse to conform to individually costly, but group beneficial norms, like helping her mother-in-law. This strategic non-conformism would undermine cultural group selection. Henrich (2004) makes several reasonable verbal arguments that strategic non-conformism would be difficult, and perhaps impossible, to evolve. In variable social environments where information acquisition is difficult and costly, how would immigrants know which norms to copy, and which to ignore? These verbal arguments need to be translated into formal models, if possible, to assess the sensitivity of the conformism story to strategic non-conformism.

5. The mismatch hypothesis

The cultural group selection hypothesis for strong reciprocity preserves economists' assumption that humans are rational actors by instead modifying the utility function. It could be, however, that rationality is violated. If humans did not evolve to interact with others in the anonymous, one-shot encounters of the ultimatum game, there is no reason to expect that players would be adapted to successfully maximize monetary utility under these conditions (e.g., Trivers, 2004). We term this the *mismatch hypothesis*.

Mismatch phenomena are well documented. The neurotoxins of many spiders and snakes, for example, were a real danger to ancestral humans, as well as our distant primate ancestors. Now, however, they kill less than 20 people a year in the US (virtually all of whom were owners of dangerous spiders and snakes), whereas automobile accidents kill about 40–50,000 people a year (National Safety Council, 2000). Yet decades of research have shown that fear of spiders and snakes is more readily learned than fear of contemporary dangers like automobiles, guns, and electric outlets (Öhman and Mineka, 2001), a clear example of a mismatch. More generally, mismatches are successfully exploited by many large industries, including advertising and entertainment.

There are at least two distinct versions of the mismatch hypothesis for game play. In one, players misapprehend the situation, but nonetheless act consistently with that misapprehension. Players might be cooperating in one-shot interactions with strangers because, for example, they are using psychological machinery that evolved for reciprocal altruism, where cooperating on the first round with strangers is often a good strategy. And they might be punishing in one-shot encounters because, for example, they are using psychological machinery that evolved to defend individual reputation via retaliation. The psychological machinery is wrongly activated, but, once activated, it functions as designed. We term this variant of the mismatch hypothesis the *misapprehension* hypothesis.

Alternatively, it could be that games are so odd that players' psychological machinery simply malfunctions, and there is little rhyme or reason to game play. We term this variant of the mismatch hypothesis the *malfunction* hypothesis.

5.1. Assessing arguments against the mismatch hypothesis

In order to strengthen their case for humans as rational actors with a culturally group-selected utility function that includes an element of altruism, strong reciprocity scholars have provided several counter-arguments to the mismatch hypothesis.

5.1.1. Are humans adapted to one-shot encounters?

Fehr and Henrich (2003) argue that in the ethnographic record there are many examples of transient, one-shot encounters with distant groups, especially in harsh climate conditions or long-range raids. If one-shot encounters are not uncommon ethnographically, humans might be adapted to reason about them. There is then potentially no mismatch between conditions in one-shot experimental games and the social circumstances that humans are adapted to. They also point out that a positive probability of a repeated encounter does not, in and of itself, imply cooperation.

Although their point is well-taken that, a posteriori, one-shot encounters might have been more common in ancestral environments than is usually acknowledged, they have not made a convincing argument that upon encountering a stranger, the a priori probability of a future interaction was zero, or very close to zero. If one group raids a distant group, this might well increase the probability that the distant group would conduct a counter-raid. Or if two distant groups come together during a drought and get along well, this might create a desire for an alliance or trade relationship.

If humans were adapted to one-shot encounters, two additional questions are raised. First, would simply telling players in the game that interactions are one-shot trigger the psychological mechanisms that (putatively) evolved to deal with one-shot interactions? This concern is especially acute for the many studies, including those in the 15-cultures study, in which players were told, or correctly infer, that the other players were from their school, town, or village. Such information should undercut one-shot adapted players' certainty that the encounter in the game was truly a one-shot encounter. Second, what strategic response would strong reciprocity theorists predict for one-shots? Their model is based on the advantages of within-group altruism for between-group competition. A one-shot encounter is, almost by definition, not an encounter with a member of one's in-group. It is therefore not clear that strong reciprocity predicts altruism towards others in one-shot encounters. If it does not, altruism in one-shot games is not evidence for group selected strong reciprocity.

5.1.2. Are humans adapted to anonymous interactions?

Equally important, with the possible exception of murders, ambushes or certain rituals, we do not believe that there were any circumstances in ancestral environments where personal interactions with others were anonymous. Information about personal interactions, even one-shot interactions, could have been communicated to others. For virtually every personal interaction in ancestral human environments, reputational effects were probably impossible to ignore. The anonymity of the ultimatum game should indicate to players that there will be no reputational consequences to game play. If features of the ultimatum game, such as the proposer–responder paradigm or the exchange of money, suggest to players that a personal interaction is taking place, however, then psychological machinery that evolved to take reputation into account might nonetheless be activated.

Fehr and colleagues respond that there is clear evidence from the games themselves that players do respond differently when there is anonymity vs. when reputation building is possible. In post-game interviews, players also affirm that they believed the game to be anonymous. This is a key point that undermines the simple mismatch hypothesis. In our opinion, however, it is not a very strong argument against it. Young men certainly know that there is no chance they will encounter the attractive woman appearing in the *Playboy* centerfold, and they would truthfully and accurately affirm that they knew that the centerfold was just a picture in a magazine that they could not interact with, yet they might still become sexually aroused by the photograph.

6. Disentangling the debate about cognition from the debate about altruism

Although we think the arguments against the mismatch hypothesis are unconvincing, this does not mean that the hypothesis can explain the results from the 15 cultures study or other economics games. Players respond to many features of various games in seemingly rational ways (e.g., by usually accepting generous offers). Rationality per se might not be the key issue. We believe, in fact, that differing implicit models of human cognition are currently more central to the debate over strong reciprocity than rationality. Until a debate about the nature of cognition has been made more explicit in the debate about strong reciprocity, it will be difficult to make progress elucidating game play.

6.1. Models of cognition in economics and evolutionary biology

In economic theory, humans maximize utility, subject to informational and cognitive constraints. This implies that humans have a generalized 'optimizing engine' in their head that operates in real, or ontogenetic, time over a set of complete, transitive preferences that either evolved by natural selection, or were acquired culturally or by individual learning. In both their daily life, as well as in the experimental games we discuss here, people have the optimizing engine at their disposal to reason about novel situations. If agents play TIT-FOR-TAT, for example, it is because their optimizing engines have computed, or learned, that this is the way to maximize a 'self-regarding' utility in dyadic exchange.

To strong reciprocity theorists, many of whom are economists, the economic games are simple, posing little problem to the optimizing engine implicitly posited by their model. Naïve players should easily understand one-shot, anonymous interactions, so their choices cannot be dismissed as irrational. Hence, there is a fairly straight line of inference from game play to a utility function such as Eq. (1).

The strong reciprocity theorists have a sophisticated understanding of evolutionary theory. They are fully conversant with current debates and theories about human evolution and the evolution of cognition, and have themselves made important contributions to the debate. They see the utility maximization model as a useful approximation of poorly understood, evolved cognitive processes. Nonetheless, the difference between their implicit model and that of many evolutionary biologists, particularly those studying animal behavior, leads to different interpretations of game results.

In most evolutionary biological models of behavior, the brain of the agent does not contain a generalized optimizing engine that maximizes utility. Instead, natural selection is the 'optimizing engine,' and 'utility maximization' is external to the individual, occurring over evolutionary time in a population under selection. This implies that, as a product of natural selection, the brain of the agent contains a specialized mechanism that evolved to solve fitness-relevant problems in a specific domain—the optimizing strategy for the domain has been built into the mechanism (strategy parameters can, of course, be set by individual learning or acquired culturally). If the agent evolved to solve problems in two or more unrelated domains, its brain would comprise multiple evolved mechanisms. When an evolutionary biologist invokes an optimizing engine, it is almost always for a circumscribed domain like foraging.

Emphasis therefore shifts to the patterns of informational cues that indicate the particular strategic context. If agents play TIT-FOR-TAT, it is because their brains have detected informational patterns indicating an opportunity for reciprocal exchange. An innate, behavior-generating mechanism that is specialized for TIT-FOR-TAT is then triggered, delivering benefits to others under certain conditions.

In this model, mechanisms will usually perform well *if* they are activated in the appropriate environment. But that is only likely to happen when there is enough information to reliably determine the strategic context.

In the evolutionary biologists' picture, the more-or-less modular structure of cognition also means that multiple, and perhaps even opposing, mechanisms might be activated in evolutionarily novel environments. Just as some parts of a young man's brain know that the woman in the centerfold is ink on paper, and other parts do not, some parts of the ultimatum game players' brains could know that the game is anonymous, but other parts might not. No single utility is maximized. Hence, in experimental economics games that deliberately restrict informational cues, there is no straight line of inference from game play to a utility function. Irrationality is a real possibility.

6.2. The entanglement of cognitive models and human altruism

Essentialism is the belief that entities are defined by a set of necessary properties—*essences*. Although essentialism is controversial, human psychology probably has an essentialist bent (e.g., Barrett, 2001) that influences the reasoning of human scientists. The differing essences in economists' and biologists' implicit models of cognition can thus lead to very different intuitions about human 'selfishness.'

What is the essential nature of the agents in the economists' and evolutionary biologists' models? In the economists' model, preferences, especially evolved preferences, play the role of essences. As the mechanical product of the optimizing engine, strategies are of secondary importance. Agents that cooperate because they have optimized self-regarding preferences are still essentially selfish. Cooperative agents could be genuinely unselfish only if their preferences were 'other regarding.' Because 'other regarding' preferences can clearly evolve by group selection, this evolutionary mechanism seems necessary to explain the genuine altruism that we all experience within ourselves and see in others.

In the evolutionary biologists' implicit model, strategies that have been optimized by the external process of natural selection play the role of essences. The essence of an agent with a conditionally cooperative strategy like TIT-FOR-TAT is that it is conditionally cooperative, full stop. In this model, when conditions of cooperation are met, agents are genuinely unselfish. Group-selected cooperation is no more genuinely altruistic than is individually-selected cooperation because in most individual and group selection models of the evolution of cooperation agents are only *conditionally* altruistic: agents deliver benefits to certain categories of individuals (e.g., reciprocal partners or in-group members, respectively), and only under certain conditions (e.g., repeated interactions or between-group competition, respectively).

6.3. *The disentanglement*

In our view, the foregoing debate about implicit cognitive models needs to be disentangled from the debate about strong reciprocity and the evolution of group beneficial norms. Suppose, for example, that players had no ability to understand one-shot, anonymous interactions. They could nonetheless be interpreting the ultimatum game as a group-beneficial institution, and exhibiting group-selected altruistic behavior.

Conversely, assume players had a superb ability to understand one-shot, anonymous interactions. Assume further that they possessed culturally group selected norms that dictated altruistic behavior towards members of one's in-group. Ironically, this does not mean that they would play the ultimatum game as strong reciprocators because the conditionality inherent in most group selection scenarios for the evolution of strong reciprocity (cooperate with in-group members only) is not an explicit feature of the ultimatum game.

Student participants in experimental economics games are often told, or correctly infer, that fellow players are students at the same university. But do students view one another as cooperative members of the same in-group, competitors for grades, strangers, or some other social category? To our knowledge, no experimental economics study has adequately explored how Western players in the one-shot, anonymous ultimatum game conceive of their partners when group identity is not an explicit feature of the game. Yet when group identities are incorporated into the game, they often have a strong impact on game play. In an East African group that encouraged cooperation within, but not between, villages, for example, the adjusted mean offer for partners from the same village was more than three times that for partners from different villages (Paciotti and Hadley, 2003). Group identities are also known to have an important impact on players in other types of games. Kollock (1998), for example, found that individuals who were members of different groups played a Prisoner's Dilemma game as a Prisoner's Dilemma ($DC > CC > DD > CD$), whereas players who were members of the same group played the Prisoner's Dilemma as an Assurance Game ($CC > DC > DD > CD$). In the Assurance Game, players who can trust their partners to cooperate do the best.

Putting aside our concern that categorizing fellow players as in-group members might suggest a repeated interaction, high offers to in-group members in the one-shot ultimatum game is consistent with cultural group selection; lower but non-zero offers to out-group members or strangers in the one-shot game is perhaps more consistent with, e.g., something like a misfiring TIT-FOR-TAT strategy. Without knowing how players categorize their game partners, it is difficult to infer from the cross-cultural pattern of altruism in the ultimatum game that players are acting as group-selected strong reciprocators.

Cultural group selection explanations of strong reciprocity do not require the assumption of rationality. The rationality assumption *is* required, however, to interpret the ultimatum and other game results as strong evidence of cultural group selection. The fairly straight line of inference from game play to utility function seen in the economists' model of cognition does not exist for the evolutionary biologist, who needs to understand much more about the informational cues available to the players, and the strategies they are triggering. The debate about strong reciprocity will therefore be difficult to resolve until the differences in the underlying models of cognition are reconciled.

7. Five interesting facts about experimental economics games

We will now sketch a framework for interpreting experimental economics games results based on the following five facts. These facts highlight the important distinction between the explicit and implicit features of experimental games. By 'explicit' we mean the oral and written instructions to the players about the game, including any assurances of anonymity. By 'implicit' we mean all other information that might potentially influence play.

7.1. *Explicit features*

First, and most importantly, explicit features of games, such as opportunities for reputation building and punishment, can dramatically influence game play in ways predicted by rational choice theory (e.g., Fehr and Gächter, 2000). This is strong evidence that the structure of the games is not completely opaque to the players.

7.2. Implicit features

Second, relatively subtle cues that are independent of the formal structure of an experimental game can nonetheless have a significant impact on players' contributions. Kurzban (2001), for example, found that in a public goods game, brief, oblique eye contact or light taps on a shoulder or arm before each round significantly increased contributions by males (but not females), sometimes by more than 70%. He speculates that in ancestral environments these behaviors might have served as psychophysical cues of cooperation that would have been particularly important for men in hunting and warfare. In another example of the importance of implicit features, Haley and Fessler (2005) found that in a dictator game played on computer workstations, players using workstations with stylized eye-spots as part of the desktop background image were 55% more generous than players using workstations whose desktop background image did not contain eyespots. Haley and Fessler conjecture that eyespots enhanced players' (perhaps unconscious) perception that their decisions would have reputational consequences. These examples raise the possibility that players in experimental economics games are using a wide variety of implicit cues that are not controlled by the experimenter to determine the extent to which they should cooperate in the game. If so, some of the altruism found in anonymous, one-shot games could be caused by such cues.

7.3. The importance of culture

Third, there is increasing evidence that culture plays an important role in the explanation of game results. The most intriguing finding from cross-cultural studies is that game play in many societies seems to reflect local social and economic institutions (Henrich et al., 2005). In the New Guinean societies of Au and Gnau, for example, accepting a gift creates a strong obligation to reciprocate, often in ways that the receiver finds onerous. If the receiver fails to reciprocate, he finds himself in a subordinate social position. Large gifts are consequently often refused. Perhaps not coincidentally, in these societies and unlike many other societies, large, 'hyperfair' offers exceeding 50% were frequently rejected in the ultimatum game (Tracer, 2003). This again illustrates that players draw upon information that is not specified by the formal structure of the game or the experimenter's instructions.

7.4. Emotions

Fourth, rightly or wrongly, emotions ('hot' cognition) are often seen as fundamentally different from the rationality assumed to underly economic decision-making ('cold' cognition). Yet hot cognition plays a central role in the decisions of players in economics games, especially when it comes to punishing free-riders. In a public goods game with a punishment option, for example, the pattern of emotional responses to free riding was consistent with the hypothesis that strong negative emotions trigger the willingness to punish (Fehr and Gächter, 2000, 2002).

7.5. Variation in game play

Finally, we are struck by the substantial individual variation in game play even within cultures. Across a number of different Western societies, there are large fractions of both free-riding and strongly reciprocating players (Fehr et al., 2002). Fischbacher et al. (2001), for example, found that in a public goods game in Switzerland, about 50% of players could be classified as conditionally cooperative with a self-serving bias; about 30% as purely selfish; and about 14% as conditionally cooperative when others contributed smaller amounts, but increasingly selfish when others contributed larger amounts. Kurzban and Houser (2005) similarly found in a public goods game in the US that 63% of players could be classified as reciprocators, 20% as free-riders, and 13% as cooperators (i.e., consistently generous).

8. Framing: The ESS meets cognitive science

On the basis of these five facts, we offer a tentative approach to game results that employs the concept of semantic *frames*, also commonly referred to as *schemata*, or *scripts*. A frame is a knowledge structure or conceptual abstraction used to interpret a complex reality or experience, and guide behavior. It would be difficult, for example, to interpret the sentences "May I have your order?" and "Check please," and respond properly, without a 'restaurant frame.' The concept was first developed in social psychology (Bartlett, 1932) and then spread to linguistics (Fillmore, 1968; Lakoff, 1970), cognitive anthropology (e.g., Bateson, 1972), artificial intelligence (e.g., Minsky, 1975), cognitive psychology (e.g., Tversky and Kahneman, 1981), and economics (Heiner, 1983; Elliot and Hayward, 1998).

Framing effects, where the same facts are interpreted differently depending on the framing, are well documented. In the famous Asian Disease experiment of Tversky and Kahneman (1981), for example, subjects' preferences for disease control programs reversed if the outcomes were framed in terms of number of lives lost instead of number of lives saved. Framing effects have also been documented in experimental economics games. Contributions to a public goods game were higher when the game was framed as community social event, for instance, than when it was framed as an economic investment (Pillutla and Chen, 1999; Ross and Ward, 1996). See Elliot and Hayward

(1998) for a review of framing effects in experimental economics.

Economics experiments, however, are typically bereft of explicit contextualizing cues or framing. Players are given the rules, the payoffs, and are usually tested to confirm that these are understood. In most cases, they are deliberately not given any explicit information about the broader context of the game. Are the players allies or enemies? Friends or competitors? Members of one's in-group or an out-group? Is the game meant to represent an economic transaction, a ritualized competition like a chess game, or a test of intelligence, personality, or skill? Despite the lack of explicit contextualization or framing, the games are nonetheless rich in social cues. As Haley and Fessler (2005), commenting on the anonymity of these games, point out:

> Participants frequently come face-to-face with experimenters and other participants, providing complex stimuli likely to influence intuitive judgments as to whether one's actions are observable. Moreover, shared language, similar styles of dress, and familiar comportment and patterns of interaction indicate to participants that those around them are members of the same social group, a fact which, in ancestral populations, would have corresponded to an increased likelihood of, and greater consequences associated with, future interaction. If natural selection shaped the mind to attend to a variety of sources of information about the consequences of social behavior, then manipulation of these and similar factors should affect levels of prosociality.

Given a lack of explicit framing combined with a rich stream of social and other cues (such as the use of money), players must provide their own framing of the game. The individual variability often seen in game play suggests to us that players might be using a variety of frames to guide their choices. Sometimes the frame might be a pan-human social frame, such as reciprocal altruism, indirect reciprocity, or status competition, that conceivably evolved by natural selection. Other times it might be a social or economic institution that was acquired culturally or by individual learning. Among the Orma, an East African pastoralist group, for example, the public goods game was seen to be very similar to the *harambee*, a local social institution for organizing cooperative investments in public goods like roads or schools. The Orma, in fact, began referring to the public goods game as the *harambee* game (Henrich et al., 2005), essentially confirming the framing idea in this particular case. Alternatively, the variation in game play might be evidence of a polymorphism (e.g., Kurzban and Houser, 2005).

Strong reciprocity theorists agree that differences in framing could explain differences in game play. Comparing the framing hypothesis with an alternative, dispositional hypothesis, Henrich et al. (2005) propose that "individual differences result from the differing ways that individuals frame a given situation, not from generalized dispositional differences."

In both the economists' and evolutionary biologists' implicit models of cognition, framing is required to determine the context the agent finds herself in. Strong reciprocity theorists, however, lean towards frames that only determine which preferences or norms are to be applied. The optimizing engine is then used to maximize utility in real or ontogenetic time. Evolutionary biologists studying animal behavior, in contrast, lean towards frames that have the appropriate evolved strategy attached. Framing, in this view, is a way for people to *avoid* the difficulties of optimizing by applying pre-established cognitive templates to problems (cf. Simon, 1957; Gigerenzer and Todd, 1999). This perspective assumes that, unlike mathematicians, most people do not, and cannot, see these games as abstract structures that can be logically analyzed.

9. Concluding remarks

We agree with strong reciprocity theorists that most people do genuinely care about the welfare of others, and that this explains much about the experimental game results. We disagree, however, that results to date provide convincing evidence that genuine care for others is a group-selected utility or norm. Our disagreement stems from the fact that little is currently known about the frames players are using to interpret the experimental economics games cited in support of a group-selected strong reciprocity.

To test rational choice theory, it was necessary to construct experiments that adhered as closely as possible to the formal structure of the theory. Now that it is clear, however, that rational choice theory cannot be applied to human decision-making in a straightforward manner, there is scant reason to continue using the spare, highly abstract structure of most experimental economics games. The lack of explicit framing or contextualization merely allows players to interpret the games in idiosyncratic ways that are often opaque to the experimenter.

The clear impact of explicit features of games on game play, and the fact that players often do well, is evidence that humans are, in some sense, rational actors that understand the games. The equally clear impact of 'implicit' features, and the variability of play in games with few contextualizing cues is evidence that much about player behavior in these games is currently not understood. In our view, the next step is to empirically determine how participants in the various games categorize their fellow players—are they in-group members, strangers, or some other social category? Players' framing of the games must then be investigated. The important role in game play of both strong, universal emotions like anger, and of cultural institutions like the *harambee*, indicates that gene-culture co-evolution theory (e.g., Boyd and Richerson, 1985; Feldman and Cavalli-Sforza, 1976; Durham, 1991) is a promising framework in which to conduct such studies.

Fruitful collaboration between economists and biologists will require frank discussion of their respective implicit cognitive models. The issue is whether the human brain itself optimizes, in which case it is relatively easy to infer preferences from play in simple games, or whether natural selection optimizes, in which case inferring strategies from behavior requires knowing how players have categorized their game partners and interpreted the game.

In this review, bounded rationality was not the key issue. The extent to which rationality is bounded, however, is an important question that appears to undermine both the economists' and the evolutionary biologists' implicit cognitive models. In our view, the typical economists' model is a problematic idealization because it posits generalized computational abilities and preference consistencies that most people do not seem to possess (e.g., Gigerenzer and Todd, 1999; Rieskamp et al., in press). Typical models in evolutionary biology, on the other hand, are also unrealistic because they do not explain humans' obvious talent for dealing successfully with novel situations. A synthesis of the two ideas is clearly needed. Human cognition is still a profound mystery that will require the combined efforts of all fields of biology and the social sciences to unravel.

Despite our reservations about the conclusions that can currently be drawn from experimental economics regarding cultural group selection and strong reciprocity, strong reciprocity theorists have developed an impressive body of theory with which researchers in many disciplines must grapple. They have also astutely spotlighted one of the most pressing challenges to theoretical biology: integrating population biology with the cognitive and social sciences.

Acknowledgments

We thank Rob Boyd, Marc Feldman, Dan Fessler, Herb Gintis, Nicole Hess, and two anonymous referees for many useful comments. Support provided by Deutsche Forschungsgemeinschaft SFB 618 and ECAgents Contract No. 001940.1.

References

Barrett, H.C., 2001. On the functional origins of essentialism. Mind and Society 3 (2), 1–30.
Bartlett, F.C., 1932. Remembering: A Study in Experimental and Social Psychology. Cambridge University Press, Cambridge.
Bateson, G., 1972. Steps to an Ecology of Mind. Ballantine, New York.
Boyd, R., Richerson, P.J., 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.
Boyd, R., Richerson, P.J., 1988. The evolution of reciprocity in sizable groups. J. Theor. Biol. 132, 337–356.
Boyd, R., Richerson, P.J., 2002. Group beneficial norms spread rapidly in a structured population. J. Theor. Biol. 215, 287–296.
Boyd, R., Gintis, H., Bowles, S., Richerson, P.J., 2003. The evolution of altruistic punishment. Proceedings of the National Academy of Sciences USA 100, 3531–3535.
Camerer, C., Loewenstein, G., Rabin, M., 2004. Advances in Behavioural Economics. Princeton University Press, Princeton.
Cavalli-Sforza, L.L., Feldman, M.W., 1973. Models for cultural inheritance. I. Group mean and within group variation. Theor. Popul. Biol. 4, 42–55.
Cavalli-Sforza, L.L., Feldman, M.W., 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton.
Durham, W.H., 1991. Coevolution: Genes, Culture, and Human Diversity. Stanford University Press, Stanford.
Elliot, C.S., Hayward, D.M., 1998. The expanding definition of framing and its particular impact on economic experimentation. J. Socio-Econom. 27 (2), 229–243.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. Amer. Econom. Rev. 90, 980–994.
Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415, 137–140.
Fehr, E., Gächter, S., 2005. Human behaviour: egalitarian motive and altruistic punishment (reply). Nature 433, E1–E2.
Fehr, E., Henrich, J., 2003. Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In: Hammerstein, P. (Ed.), Genetic and Cultural Evolution of Cooperation. MIT Press, New York, pp. 55–82.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. Q. J. Econ. 114, 817–868.
Fehr, E., Fischbacher, U., Gächter, S., 2002. Strong reciprocity, human cooperation and the enforcement of social norms. Hum. Nat. 13, 1–25.
Feldman, M.W., Cavalli-Sforza, L.L., 1976. Cultural and biological evolutionary processes selection for a trait under complex transmission. Theor. Popul. Biol. 9, 238–259.
Feldman, M.W., Zhivotovsky, L.A., 1992. Gene-culture coevolution: toward a general theory of vertical transmission. Proceedings of the National Academy of Sciences USA 89, 11935–11938.
Fillmore, C.J., 1968. The case for case. In: Bach, E., Harms, R. (Eds.), Universals in Linguistic Theory. Holt, Rinehart, and Winston, New York, pp. 1–88.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Econ. Lett. 71, 397–404.
Gigerenzer, G., Todd, P., 1999. Simple Heuristics that Make Us Smart. Oxford University Press, Oxford.
Gintis, H., 2000. Strong reciprocity and human sociality. J. Theor. Biol. 206, 169–179.
Gintis, H., Bowles, S., Boyd, R., Fehr, E., 2003. Explaining altruistic behavior in humans. Evol. Hum. Behav. 24, 153–172.
Haley, K.J., Fessler, D.M.T., 2005. Nobody's watching? Subtle cues affect generosity in an anonymous economic game. Evol. Hum. Behav. 26, 245–256.
Hammerstein, P., Hagen, E.H., 2005. The second wave of evolutionary economics in biology. Trends Ecol Evol. 20, 604–609.
Heiner, R.A., 1983. The origin of predictable behavior. Am. Econ. Rev. 73 (4), 560–595.
Henrich, J., 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. J. Econ. Behav. Organ. 53, 3–35.
Henrich, J., Boyd, R., 1998. The evolution of conformist transmission and the emergence of between-group differences. Evol. Hum. Behav. 19, 215–242.
Henrich, J., Boyd, R., 2001. Why people punish defectors: conformist transmission stabilizes costly enforcement of norms in cooperative dilemmas. J. Theor. Biol. 208, 79–89.
Henrich, J., Gil-White, F., 2001. The evolution of prestige: freely conferred status as a mechanism for enhancing the benefits of cultural transmission. Evol. Hum. Behav. 22, 1–32.
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, G., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N.S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F.W., Patton, J.Q., Tracer, D., 2005. 'Economic man' in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behav. Brain Sci. 28, 795–815.
Kelly, R.L., 1995. The Foraging Spectrum. Smithsonian Institution Press, Washington.
Kollock, P., 1998. Social dilemmas: the anatomy of cooperation. Annu. Rev. Sociol. 24, 183–214.

Kurzban, R., 2001. The social psychophysics of cooperation: nonverbal communication in a public goods game. J. Nonverbal Behav. 25, 241–259.

Kurzban, R., Houser, D., 2005. Experiments investigating cooperative types in humans: a complement to evolutionary theory and simulations. Proceedings of the National Academy of Sciences, USA 102, 1803–1807.

Lakoff, G., 1970. Irregularity in Syntax. Holt, Rinehart, and Winston, New York.

Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. Nature 246, 15–18.

Minsky, M., 1975. A framework for representing knowledge. In: The Psychology of Computer Vision. McGraw-Hill, New York, pp. 211–277.

Murdock, G.P., 1967. Ethnographic Atlas. University of Pittsburgh Press, Pittsburgh.

National Safety Council, 2000. What are the odds of dying?

Öhman, A., Mineka, S., 2001. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. Psychol. Rev. 108, 483–522.

Paciotti, B., Hadley, C., 2003. The ultimatum game in southwestern tanzania: ethnic variation and institutional scope. Curr. Anthropol. 44, 427–432.

Pillutla, M.M., Chen, X.P., 1999. Social norms and cooperation in social dilemmas: the effects of context and feedback. Organ. Behav. Hum. Dec. Process. 78, 81–103.

Richerson, P.J., Boyd, R., 1998. The evolution of human ultra-sociality. In: Eibl-Eibisfeldt, I., Salter, F. (Eds.), Ideology, Warfare, and Indoctrinability. Berghan Books, pp. 71–95.

Rieskamp, J., Busemeyer, J.R., Mellers, B.A. Extending the bounds of rationality: a review of research on preferential choice. J. Econ. Literature, in press.

Ross, L., Ward, A., 1996. Naive realism: implications for social conflict and misunderstanding. In: Brown, T., Reed, E., Turiel, E.-E. (Eds.), Values and Knowledge. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 103–135.

Seielstad, M.T., Minch, E., Cavalli-Sforza, L.L., 1998. Genetic evidence for a higher female migration rate in humans. Nat. Genet. 20, 278–280.

Simon, H., 1957. A Behavioral Model of Rational Choice. Rand Corporation, Santa Monica, Calif.

Soltis, J., Boyd, R., Richerson, P.J., 1995. Can group-functional behaviors evolve by cultural group selection? An empirical test. Curr. Anthropol. 63, 473–494.

Tracer, D.P., 2003. Selfishness and fairness in economic and evolutionary perspective: an experimental economic study in papua new guinea. Curr. Anthropol. 44, 432–438.

Trivers, R., 2004. Mutual benefits at all levels of life. Science 304, 964–965.

Tversky, A., Kahneman, D., 1981. The framing of decisions and the psychology of choice. Science 211, 453–458.

Tversky, A., Kahneman, D., 1986. Rational choice and the framing of decisions. J. Bus. 59 (4), S251–S278.