Attractiveness is one of these factors. Apart from being desirable in its own right, the halo effect causes many assessments of another, such as their intelligence and character, to be estimated more highly. Thus, Solnick and Schweitzer (1999) found that more was expected of attractive faces. Joergensen and Hancock (2001) reported an Ultimatum Game where proposers saw a picture of the responder. Offers were higher to faces rated as attractive, echoing results from Solnick and Schweitzer (1999), but with a stronger effect for attractive women. The correlation between rated attractiveness and offer level was 0.83. However, the effect of attractiveness was transient, it disappeared in a second round of the game following information about who had refused low offers. Hancock and Ross (2002) investigated the Dictator Game with similar results: The correlation between offer levels within the game and independently rated attractiveness was 0.91.

These experiments use anonymous faces, but what effect might the perception that someone is related to you have? Hamilton's theory of kin selection (Hamilton 1964) suggests that people should be favorably disposed toward relatives. Any gene promoting altruistic behavior can influence its own success by benefiting those most likely to share a copy of itself. Thus, a gene causing altruism will be favored if the benefit (b) to the recipient multiplied by the relatedness[1] (r) between the altruist and recipient is greater than the cost (c) to the altruist.

Given this logic, cues of relatedness between individuals may change the payoffs attributed to different behaviors. DeBruine (2002) explored behavior in a two-person, two-node sequential trust game (after Eckel & Wilson 1998b, and related to the Centipede game described by Colman). The first player can decide either not to trust the second, in which case both get a sure payoff of $3, or to trust the second player with the decision. The second player can decide between selfish behavior, keeping $5 and giving $2 to player one, or unselfish behavior, allocating $4 to each. Given no information about the second player, the first player's expected payoff is $3 for either choice, so a rational player should be indifferent. However, if the other player is a relative with relatedness of 0.5,[2] then the structure of the game is changed to a choice between a sure payoff of $4.50 ($3 to self plus 0.5 times $3 to the other) and a risky payoff with an expected value of $5.25 ($3 to self plus 0.5 times $4.50 to the other). In addition, assessment of the second player's trustworthiness may bias the expected payoff of trusting.

DeBruine (2002) digitally morphed images of the other player to manipulate one possible cue of relatedness, facial resemblance. Players in this trust game chose the riskier option more often when the image of the opponent had been morphed to resemble the first player than when the image had been morphed to resemble an unknown person. This is consistent with an increase in either the expected payoff of trusting or the expected probability of trustworthy behavior. Analysis of the responses indicated that independently rated attractiveness of the second player did not influence behavior in this situation, although current research by DeBruine indicates that resemblance to self increases the attractiveness of faces (also see Penton-Voak et al. 1999).

In an unpublished study, DeBruine randomized the computer-generated second players' responses in the trust game. Players were less likely to trust opponents in games immediately after they had been cheated than in games after the opponent was unselfish. This echoes findings by Eckel and Wilson (1998a) that the previous opponent's response influences the current choice, even when the current opponent is a different person. Within the sets of games played after either a selfish or unselfish response, players were still more likely to trust faces morphed to resemble themselves.

In any social situation, people evaluate others. We have shown that even a static photograph of another player can cause significant differences to behavior in simple games. A playing partner's attractiveness may introduce noneconomic motivations to the game or change the player's predictions of the partner's behavior. The perception that someone may be related to you introduces further complications, because of the shift of possible inclusive fitness payoffs. The inclusion of such factors will broaden the scope of a psychological game theory.

ACKNOWLEDGMENTS

We wish to thank M. Daly, P. Joergensen, K. Ross, I. Penton-Voak, C. Taylor, and M. Wilson.

NOTES

1. Relatedness refers not to the total proportion of genes shared, but to those shared by identical descent. In this case, r is the probability that the recipient shares a gene for altruism with the altruist.

2. The conclusion holds for any $r > 0$ with this particular game payoff structure.

Rational belief and social interaction

Daniel M. Hausman

Department of Philosophy, University of Wisconsin-Madison, Madison, WI 53706-1474. **dhausman@wisc.edu**
http://philosophy.wisc.edu/hausman

Abstract: Game theory poses problems for modeling rational belief, but it does not need a new theory of rationality. Experimental results that suggest otherwise often reveal difficulties in testing game theory, rather than mistakes or paradoxes. Even though the puzzles Colman discusses show no inadequacy in the standard theory of rationality, they show that improved models of belief are needed.

The theory of rational choice takes choice to be rational when it tracks preferences. Preferences are rational when they are, in a precise sense, consistent. When there is risk or uncertainty, preferences and hence choices depend on beliefs; and neither preference nor choice is rational unless belief is rational. Rational beliefs must conform to the calculus of probabilities. When they do, and preferences satisfy relevant consistency and technical axioms, then preferences can be represented by expected utilities, and choice is rational if and only if it maximizes expected utility.

Expected utility maximization is defined whenever rational beliefs and preferences are defined. The fact that an interaction is strategic by itself causes no problem. If I am playing the pure coordination game in Colman's Figure 1 and believe that the other player will play Tails, then I should choose Tails, too.

But suppose I have no beliefs about what strategies other players will choose other than those that I can deduce from (1) beliefs about the other players' preferences over the payoffs, (2) beliefs about the other players' beliefs concerning both the game and its players, and (3) beliefs about the other players' rationality. All of these beliefs of mine have to be rational – that is, they have to be consistent with the calculus of probabilities – but this constraint permits many different sets of beliefs to count as rational. One way of developing game theory that greatly narrows the set of rational beliefs has been to assume that the players are all perfectly rational, that they all share the same subjective prior probabilities, that they have complete knowledge of the extensive form of the game, and that all of this is common knowledge. Call this "the total rationality representation" (TRR). TRR is not required by the standard theory of rational belief, and the fact that it leads to surprising and sometimes arguably paradoxical results is no indictment of the standard theory.

Colman also objects that game theory employing TRR may be uninformative. In the Hi-Lo Matching game of Figure 2, the theory fails to predict and recommend that players choose strategy H. As Colman correctly points out, if TRR requires common prior point probabilities, then no argument can be given for the rationality of playing H. But the remedy here is just a mild relaxation of the idealizations. If one does not require that the players have point priors, then Player I can believe that the probability that Player II will play *H* is not less than one-half, and also believe that

https://doi.org/10.1017/S0140525X03330055 Published online by Cambridge University Press

	II		
I		C	D
	C	\$3,\$3	\$1,\$4
	D	\$4,\$1	\$2,\$2

Figure 1 (Hausman). A prisoner's dilemma game form

Player II believes the same of Player I. Player I can then reason that Player II will definitely play *H*, update his or her subject probability accordingly, and play *H*. The problem lies with one idealized development of the standard view of rational belief, not with the view itself.

Many of the purported paradoxes Colman discusses yield to similar, though more complicated treatment. But some of the purported paradoxes are not paradoxes at all, and some of the apparent experimental disconfirmations are dubious. Consider first the standard single-shot prisoner's dilemma (PDG). Mutual defection is the uniquely rational outcome. Colman takes this to be paradoxical and to show that rationality is self-defeating, on the grounds that mutual cooperation is better for both players. In addition, he cites evidence showing that many experimental subjects, in fact, cooperate.

Although rationality is indeed *collectively* self-defeating in a PDG, there is no paradox or problem with the theory of rationality, and the apparently disconfirming data Colman cites are questionable. Consider the following game form (Fig. 1), which represents a PDG if the two players care only about their own monetary payoffs.

Mutual cooperators do better than mutual defectors. But the benefit comes from the choice the other player makes, not from one's own choice. (Remember this is a simultaneous play one-shot game in which I and II choose independently.) Unlike the finite iterated prisoner's dilemma or the centipede game, mutual cooperators cannot taunt mutual defectors, "If you're so rational, how come you ain't rich?" because the defectors can reply, "Because I wasn't lucky enough to be playing against a fool."

In addition, the apparently disconfirming experimental evidence is dubious, because cooperating subjects facing a game form like the one in Figure 1 might not be playing a PDG. To know what game they are playing one needs to know their preferences. For example, unless II prefers the outcome where II gets \$4 and I gets \$1 to the actual outcome of \$3 each, II was not playing a PDG. For those who do not have these preferences, the interaction depicted in Figure 1 is not a prisoner's dilemma. Similar remarks apply to the tetrapod in Colman's Figure 5. If the numbers represented dollars, many people would prefer the outcome where both get \$18 and player I's trust is rewarded, to the outcome where II gets \$19 and I gets \$8. The numbers in Figure 5 are, of course, supposed to represent utilities rather than dollars, but the common view, that the recommendation to play down on the first move is absurd, may reflect a common refusal to believe that these numbers correctly represent the preferences.

A great deal remains to be done to figure out how to represent rational beliefs. Wonderful controversy still rages. But one should not thereby conclude, as Colman does, that "the conception of rationality on which it [game theory] rests appears to be internally deficient" (target article, sect. 9.2). His essay does not address the treatment of rational preference, and the problems Colman explores concerning rational belief show, at most, the limitations of specific modeling choices, rather than a deficiency in basic concepts.[1]

NOTE

1. I do not, in fact, think that the standard theory of rationality is unproblematic (see e.g., my 1992 book, Chs. 2, 12, 13), but the difficulties I see are independent of those that Colman alleges.

The limits of individualism are not the limits of rationality

Susan Hurley

PAIS, University of Warwick, Coventry CV4 7AL, United Kingdom.
susan.hurley@warwick.ac.uk **www.warwick.ac.uk/staff/S.L.Hurley**

Abstract: Individualism fixes the unit of rational agency at the individual, creating problems exemplified in Hi-Lo and Prisoner's Dilemma (PD) games. But instrumental evaluation of consequences does not require a fixed individual unit. Units of agency can overlap, and the question of which unit should operate arises. Assuming a fixed individual unit is hard to justify: It is natural, and can be rational, to act as part of a group rather than as an individual. More attention should be paid to how units of agency are formed and selected: Are the local processes local or nonlocal? Do they presuppose the ability to understand other minds?

I disagree with little that Colman says about the limitations of orthodox rational choice theory, but wonder why he doesn't say more to challenge individualism as their source, and why he omits references to trailblazers such as Regan (1980) and Howard (1988).

In 1989, I argued that Hi-Lo and Prisoner's Dilemma games (PDs) exemplify the limits of individual rationality. In Hi-Lo, individuals have the same goals, yet individual rationality fails to guarantee them the best available outcome. In PDs, individuals have different goals, and individual rationality guarantees an outcome worse for all than another available outcome. These problems stem not from nature of individuals' goals, or the instrumental character of rationality, but from individualism about rationality, which holds the unit of rational agency exogenously fixed at the individual (cf. Hurley 1989).

Activity by a given unit of agency has consequences, calculated against a background of what occurs outside that unit, and can be evaluated instrumentally. Such consequentialist evaluation does not require the unit whose activity is evaluated to be fixed at the individual. Larger units of agency can subsume smaller ones, and consequentialist evaluation can apply to different units, with different results. We can think of individuals as composed of persons-at-times (or in other ways, involving multiple personalities); similarly, we can think of collective agents as composed of persons. In both cases, lower-level rationality (or irrationality) may coexist with, or even explain, higher-level irrationality (or rationality). For example, we understand from social dilemmas and social choice theory how a group can behave irrationally as a unit, although the agents composing it are individually rational. Intrapersonal analogues of social dilemmas may explain some forms of individual irrationality. Conversely, agents can behave irrationally as individuals, yet their actions fit together so that the group they compose behaves rationally (Hutchins 1995, pp. 235ff).

Individualism requires the individual to do the individual act available that will have the best expected consequences, given what other individuals are expected to do. Given others' expected acts, an individual agent has certain possible outcomes within her causal power. The best of these may not be very good, and it may be indeterminate what others are expected to do. But a group of individuals acting as a collective agent can have different possible outcomes within its causal power, given what agents outside the group are expected to do. A collective agent may be able to bring about an outcome better than any that the individual agent can bring about – better for that individual, inter alia. If so, the issue is not just what a particular unit of agency should do, given others'

https://doi.org/10.1017/S0140525X03330055 Published online by Cambridge University Press