Available online at www.sciencedirect.com
ScienceDirect
Acta Psychologica 128 (2008) 402–404
acta psychologica
www.elsevier.com/locate/actpsy

Nash equilibrium, team reasoning and cognitive hierarchy theory ☆

Robert Sugden *

School of Economics, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Received 24 September 2007; received in revised form 14 December 2007; accepted 17 December 2007
Available online 1 February 2008

Abstract

This paper comments on two experiments, carried out by Colman, Pulford and Rose, which investigate the prevalence of team reasoning. I argue that because the first experiment uses 'decomposable' games, it cannot discriminate between team-reasoning and the conceptually distinct 'prosocial' orientation. In the second experiment, Colman et al. find more support for the team reasoning hypothesis than for the rival hypothesis that subjects choose Nash equilibrium strategies. I suggest that the most credible explanation of the data is that some subjects are team reasoners while others act according to cognitive hierarchy theory.

PsycINFO classification: 2340; 3020

Keywords: Nash equilibrium; Team reasoning; Cognitive hierarchy theory; Social value orientation

The theory of team reasoning is a recent introduction to game theory, although it has a longer history in philosophy. It offers explanations of coordination and cooperation that are radically different from those of conventional analysis. Although some writers on team reasoning have drawn on ideas of group identity from social psychology (e.g. Bacharach, 2006, chap. 2), the theory has so far been relatively little used in psychology. Colman, Pulford and Rose provide a very useful summary of the theory, and report two experiments designed to investigate the prevalence of team reasoning.

Colman et al. emphasise a crucial feature of the theory that conventionally trained economists often have difficulty in grasping – that team reasoning can not be represented simply by 'getting the payoffs right' (as game theorists say). When a group of individuals act on team reasoning, each member of the group chooses her own component of the *combination* of actions that maximises the value of some collective objective. In general, this mode of reasoning cannot be represented by postulating that each individual seeks to maximise some 'payoff' to herself – even if payoffs are allowed to represent group-oriented concerns. Because of this feature, team reasoning cannot be reduced to any of the standard 'social value orientations' used in social psychology. Colman et al.'s Experiment 1 is designed to discriminate between team reasoning and other social value orientations. Experiment 2 pits the theory of team reasoning against the hypothesis, often proposed by game theorists, that if a unique Nash equilibrium exists, players conform to it.

This commentary is mainly concerned with the interpretation of the results of these two experiments. I will raise questions about some of the interpretations offered by Colman et al., particularly in relation to Experiment 1. I will agree with their main conclusion that their results provide considerable support for the theory of team reasoning. However, I will argue that some of the patterns in the data may be explained by a very different form of game theory.

I begin by considering the theoretical principles underlying the design of Experiment 1, focusing on two-player games. Consider any such game, and let x_1 and x_2 denote

☆ This work was supported by the Economic and Social Research Council (award no. RES 051 27 0146).
* Tel.: +44 1603 593423.
E-mail address: r.sugden@uea.ac.uk

0001-6918/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2007.12.009

the 'material' payoffs (say, measured in money units) to Players 1 and 2. In conventional game theory, each player is modelled as maximising a utility function. A standard approach would be to assume that Player 1's utility is given by some function $u_1 = u_1(x_1, x_2)$; similarly, Player 2's is given by $u_2 = u_2(x_2, x_1)$. Given this assumption, we can think of the social value orientation of (say) Player 1 as a property of his utility function. Colman et al. characterise a range of orientations. An *individually rational* orientation is represented by $u_1 = x_1$, an *altruistic* orientation by $u_1 = x_2$, a *prosocial* orientation by $u_1 = x_1 + x_2$, a *competitive* orientation by $u_1 = x_1 - x_2$, and an *equality-seeking* orientation by $u_1 = \min(x_1 - x_2, x_2 - x_1)$. It is useful to distinguish between those *linear* utility functions that are special cases of $u_1 = a_1x_1 + b_1x_2$ (where a_1 and b_1 are parameters, which may be positive, zero or negative) and the others; of the functions just listed, all except the equality-seeking one are linear.

This distinction is significant in relation to the *decomposable games* used in Experiment 1. Consider a game in which each player chooses between the strategies Left and Right. Suppose that Player 1 chooses strategy i and Player 2 chooses strategy j (i, j = Left, Right). In a decomposable game, the payoff to Player 1 is the sum of an *own-component* v_{1i} determined by his own strategy (independently of Player 2's) and an *other-component* w_{1j} determined by Player 2's strategy (independently of Player 1's). Correspondingly, Player 2's payoff is the sum of an own-component v_{2j} determined by her own strategy and an other-component w_{2i} determined by Player 1's strategy. If Player 1's utility function has the linear form $u_1 = a_1x_1 + b_1x_2$, he maximises utility, irrespective of Player 2's decision, by maximising $a_1v_{1i} + b_1w_{2i}$. Thus, *given the assumption that utility is linear*, each player's strategy choice (i.e. his choice of his own-component and of his opponent's other-component) reveals his social value orientation. In Experiment 1, the altruistic, prosocial, individually rational and competitive orientations imply the choice of options A, C, D and E, respectively. (The equality-seeking orientation does not have similarly straightforward implications, because its utility function is non-linear: the optimal choice of an equality-seeking player depends on what the other player is expected to do.)

In a decomposable game, strategy choices cannot be used to discriminate between the prosocial orientation and team reasoning. If Player 1 has the prosocial orientation, he chooses whichever strategy i maximises $v_{1i} + w_{2i}$. If he is a team reasoner, he chooses his component of the combination of strategies which maximises $x_1 + x_2$; necessarily, this component maximises $v_{1i} + w_{2i}$. Thus, to the extent that the results of Experiment 1 can be interpreted as supporting the team reasoning hypothesis, they are equally supportive of the hypothesis of prosocially oriented individual-reasoning. To discriminate between these hypotheses, one needs a non-decomposable game, such as the Hi–Lo game that Colman et al. discuss in Section 1, or the games used in Experiment 2.

Experiment 2 tests two alternative but not jointly exhaustive hypotheses. One is that of team reasoning, on the (natural) assumption that the objective of the two players as a team is to maximise the sum of their payoffs: this implies the choice of strategy C in each game. The other hypothesis is that each player has the individually rational orientation (let us call this part of the hypothesis *self-interest*) and that, in a game with a unique Nash equilibrium, each player chooses his component of that equilibrium (*Nash choice*). This implies the choice of strategy E in each game. The results of Experiment 2 are consistent with the assumption that the majority of the subjects act on team reasoning. They provide very little support for the hypothesis of self-interest plus Nash choice.

It is not clear whether this should be read as evidence against individual rationality. As applied to one-shot games (that is, games that are played once only, as in Experiment 2), Nash choice has a peculiar theoretical status. Applications of game theory often assume that Nash equilibrium strategies will be played. There is a distinguished tradition of theoretical work which presupposes that every game has a unique rational solution (e.g. Harsanyi & Selten, 1988). Given that presupposition, one can prove the rationality of Nash choice. (It is a theorem that, under normal assumptions, *if* a game has a unique solution which is common knowledge among rational players, *then* the solution must be a Nash equilibrium.) However, the presupposition is ungrounded. It is now widely accepted that the standard assumptions of rationality and common knowledge do *not*, in general, imply Nash choice in one-shot games; these assumptions imply only the much weaker property of 'rationalisability' (Bacharach, 1987; Bernheim, 1984; Pearce, 1984).

Given the ambiguous status of the Nash choice assumption, it is more useful to focus on cases in which each player has a strictly dominant strategy. In such cases, it is clear that individual rationality requires that the corresponding Nash equilibrium is chosen. In Experiment 2, Game 5 is the only game in which E is a dominant strategy. This game allows us to test team reasoning against self-interest without assuming Nash choice. The results for this game (54% of subjects chose C, 46% chose D) suggest that subjects are divided more or less equally between the two orientations.

One way of trying to organise the data for all five games is to use *level-n theory* or *cognitive hierarchy theory* (Camerer, Ho, & Chong, 2004; Stahl & Wilson, 1995) rather than postulating Nash choice. This is a theory of bounded individual rationality. Each player is modelled as having a particular *level* of reasoning. Players at level 0 simply choose between strategies at random. A level 1 player maximises expected payoff, given the belief that her opponent is at level 0. A level 2 player maximises expected payoff, given the belief that her opponent is at level 0 or level 1; and so on. Clearly, players at level 1 or above never choose dominated strategies. This approach has often proved quite successful in organising experimental data; typically, levels 1 and 2 are the most frequent.