

Love is not enough: Other-regarding preferences cannot explain payoff dominance in game theory

DOI: 10.1017/S0140525X07000659

Andrew M. Colman

*School of Psychology, University of Leicester, Leicester LE1 7RH,
United Kingdom.*

amc@le.ac.uk

<http://www.le.ac.uk/home/amc>

Abstract: Even if game theory is broadened to encompass other-regarding preferences, it cannot adequately model all aspects of interactive decision making. Payoff dominance is an example of a phenomenon that can be adequately modeled only by departing radically from standard assumptions of decision theory and game theory – either the unit of agency or the nature of rationality.

Gintis rests his attempt to unify the behavioral sciences on a claim that “if decision theory and game theory are broadened to encompass other-regarding preferences, they become capable of modeling all aspects of decision making” (Abstract). This claim seems unsustainable in relation to many aspects of both individual and interactive decision making, but I shall confine my comments to just one, namely the payoff-dominance phenomenon. The simplest illustration of it is the Hi-Lo matching game depicted in Figure 1.

		II	
		H	L
I	H	2, 2	0, 0
	L	0, 0	1, 1

Figure 1 (Colman). Payoff matrix of the Hi-Lo game.

Player I chooses between rows H and L, and Player II independently chooses between columns H and L. The pair of numbers in the cell where the chosen row and column intersect are the payoffs to Player I and Player II, respectively. The Hi-Lo game is a pure coordination game, because the players' interests coincide exactly and they are motivated to match each other's strategy choices. This payoff structure might apply to an incident in a football game, for example, when Player I can pass the ball either left or right for Player II to shoot for goal, and Player II can move either left or right to intercept it. If the chances of scoring are better if both choose left than if both choose right, and zero if they make non-matching choices, then their problem can be modeled as a Hi-Lo game (Bacharach 2006, pp. 124–27; Sugden 2005). Many other dyadic interactions have this simple strategic structure, and payoff dominance is also a property of more complicated games.

In game theory, payoffs represent utilities, but for the purposes of the argument that follows, we may interpret them simply as monetary payoffs – dollars, let us say. A fundamental assumption of orthodox game theory is that players are rational, in the sense of invariably acting to maximize their own (individual) expected payoffs, relative to their knowledge and beliefs at the time. This merely formalizes the notion that decision makers try to do the best for themselves in any circumstances that arise.

In the Hi-Lo game, it is obvious that rational players should choose H, and experimental evidence confirms that that is what (almost) everyone does in practice (Gold & Sugden, in press; Mehta et al. 1994). The HH outcome is in Nash equilibrium, because each player's strategy is a best reply to the co-player's; and this equilibrium is payoff dominant, in the sense that it yields both players a strictly higher payoff than the LL equilibrium, where strategies are also best replies to each other. Nevertheless, it is strange but true that game theory provides no justification for choosing H (Bacharach 2006, Ch. 1; Casajus 2001; Colman 2003a; Cooper et al. 1990; Crawford & Haller 1990; Harsanyi & Selten 1988; Hollis 1998; Janssen 2001). A player has no reason to choose H in the absence of a reason to expect the co-player to choose H, but the symmetry of the game means that the co-player faces the same dilemma, having no reason to choose H without a reason to expect the co-player to choose it. This generates an infinite regress that spirals endlessly through loops of "I expect my co-player to expect me to expect..." without providing either player with any rational justification for choosing H.

Other-regarding preferences provide no help in solving this problem, notwithstanding Gintis's claim. The usual way of modeling other-regarding preferences, although Gintis does not spell this out, is by transforming the payoffs of any player who is influenced by a co-player's payoffs, using a weighted linear function of the player's and the co-player's payoffs. This technique was introduced by Edgeworth (1881/1967, pp. 101–102) and has been adopted by more recent researchers, such as Rabin (1993) and Van Lange (1999). It alters the strategic structure of the well-known Prisoner's Dilemma game radically, providing a reason for cooperating where there was none before, but it leaves the Hi-Lo game totally unchanged. For example, suppose that both players attach equal weight to their own and their co-player's payoffs, then Player I's payoff for joint H choices is transformed from 2 to $(2 + 2)/2 = 2$, but this is exactly the same as before.

The transformed, other-regarding payoff is identical to the untransformed, self-regarding payoff; and the same applies to all other payoffs of the game. This game is unchanged by other-regarding payoff transformation, and other-regarding preferences cannot solve the payoff-dominance problem in other games.

This is just one illustration of the fact that game theory cannot model all aspects of strategic decision making, even if it is broadened to encompass other-regarding preferences. The payoff-dominance phenomenon, illustrated by the Hi-Lo game, cannot be modeled within the framework of orthodox game theory (Colman 2003a; 2003b). The only valid solutions, as far as I am aware, involve either abandoning the assumption of individual agency that is fundamental to both decision theory and game theory (Bacharach 1999; 2006; Sugden 1993b; 2005) or assuming that players use a form of evidential reasoning that violates orthodox assumptions of rational decision making (Colman & Bacharach 1997; Colman & Stirk 1998).

It is worth commenting that any evolutionary game-theoretic model that operates by adaptive learning in a non-rational process of mindless trial and error would tend to converge on the payoff-dominant equilibrium in a game such as Hi-Lo, although this cannot explain why human players choose it in a one-shot game. But the version of evolutionary game theory favored by Gintis incorporates a rational actor "BPC" model in which the brain, as a decision-making organ, follows the standard principles of rationality. Gintis believes this to be a basic insight that is surprisingly "missing from psychology," and he devotes the whole of section 9 of his target article to defending it against its critics.

I must comment, finally, on Gintis's surprising assertion that the Parsons-Shils general theory of action was "the last serious attempt at developing an analytical framework for the unification of the behavioral sciences" (Note 2). There have been other attempts, of which the theory of operant conditioning (Ferster & Skinner 1957) is surely the most comprehensive, successful, and enduring (Dragoi & Staddon 1999), and it even underpins the Pavlov strategy of evolutionary games (Nowak & Sigmund 1993).

ACKNOWLEDGMENT

The preparation of this commentary was supported, in part, by an Auber Bequest Award from the Royal Society of Edinburgh, Scotland.

The place of ethics in a unified behavioral science

DOI: 10.1017/S0140525X07000660

Peter Danielson

W. Maurice Young Centre for Applied Ethics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada.

pad@ethics.ubc.ca http://www.ethics.ubc.ca/people/danielson/

Abstract: Behavioral science, unified in the way Gintis proposes, should affect ethics, which also finds itself in "disarray," in three ways. First, it raises the standards. Second, it removes the easy targets of economic and sociobiological selfishness. Third, it provides methods, in particular the close coupling of theory and experiments, to construct a better ethics.

The target article proposes to unify behavioral science around evolutionary game theory. Although Gintis makes no explicit reference to ethics (except, perhaps, as part of philosophy), it is clear that concerns central to ethics – accounting for and, we hope, justifying prosocial attitudes – are also central to his proposal. On Gintis's account, *unified* behavioral science (UBS) is quite friendly to ethics. It is centered on choice, gives a central role to the normative ideal of rationality, and makes a case for moralized preferences as a product of evolution. Here I argue that a unified behavioral science should lead to, if not include, a unified science of ethics. In particular, I expect the change