

Testing Theories of Other-regarding Behavior

A Sequence of Four Laboratory Studies*

Gary E Bolton, Jordi Brandts, Elena Katok, Axel Ockenfels and Rami Zwick

Revised August 2002

1. Introduction

Behavior that Hoffman, McCabe, Shachat and Smith (1994) have labeled “other-regarding” often seems incompatible with the standard assumption of payoff maximizing when rationality is common knowledge. Other-regarding behavior includes altruistic action, action motivated out of fairness, and cooperative behavior in the face of incentives to free ride.

These 'anomalies' have given rise to a diverse set of theories. Some explain other-regarding behavior in terms of strategic reputation building, by modifying standard optimization models by introducing a small amount of incomplete information about rationality (e.g., Kreps et al.'s, 1982, tit-for-tat explanation of the finitely repeated prisoner's dilemma), or by taking explicit account of the larger supergame (e.g., Fudenberg and Maskin's, 1986, treatment of the folk theorem). Other theories explicitly introduce bounded rationality (e.g., Sugden's, 1984, reciprocity principle for public good games, Selten's, 1987, theory of equal division payoff bounds for coalition games, and Roth and Erev's, 1995, reinforcement learning model). Both of these classes of theory, strategic reputation building and bounded rationality, are usually based, if only implicitly, on the premise that own material payoff is the sole motivator of behavior. A third class of theory introduces additional motivations, usually by way of interdependent preferences (e.g. Becker's, 1974, theory of social interactions, Andreoni's, 1989, theory of warm

* Correspondence. Bolton: 310 Beam, Penn State University, University Park, PA 16802, USA; geb3@psu.edu. Brandts: Institut d'Anàlisi Econòmica (CSIC), Campus UAB, 08193 Bellaterra, Barcelona, Spain; Jordi.Brandts@uab.es. Katok: 509H BAB, Smeal College of Business, Penn State University, State College, PA 16870, USA; ekatok@psu.edu. Ockenfels: Max Planck Institute for Research into Economic Systems, Strategic Interaction Unit, Kahlaische Straße 10, D-07745 Jena, Germany; ockenfels@mpiew-jena.mpg.de. Zwick: Department of Marketing, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; mkzwick@ust.hk.

glow giving, Bolton's, 1991, comparative bargaining model, Rabin's, 1993, fairness equilibrium, and the inequity aversion models by Fehr and Schmidt, 1999, and Bolton and Ockenfels, 2000).¹

This paper describes a sequence of four experiments. The hypotheses tested are all informed by one or more of the aforementioned types of models. Each experiment builds on the findings of the previous experiment in the sequence. At the same time, each experiment involves a different game, thereby allowing us to study the phenomenon from a variety of perspectives.

While some of the models we examine get important features of the data correct, none is completely satisfactory. What emerges from this sequence of experiments is a set of empirical regularities that should provide some useful guidance to the construction of more accurate models.

2. Bolton and Zwick (1995): Reputation building versus self-centered fairness in an ultimatum game

At first blush, strategic reputation building might not seem to apply to much of the other-regarding behavior we observe in the lab, the reason being that lab games are typically played among anonymous subjects interacting for a single play of the game. Under such conditions, there is little chance for a subject to build a reputation with other subjects. But on reflection, one might conjecture that lab subjects act to build reputations with the person who does usually observe their actions: the experimenter. So for example, subjects might reject a small share of the payoff to avoid looking greedy and unprincipled in the eyes of the experimenter; by rejecting, the subject establishes a reputation with the experimenter that might be profitable to the subject in the future.² An alternative explanation, implied by models such as Bolton and Ockenfels (2000) and Rabin (1993), is that subjects reject money because they care about fairness independent of strategic considerations (such as reputation building). Both of these models treat fairness as a matter of individual preference; they imply that those subjects who reject money in the ultimatum game have an aversion to being treated unfairly, much as some people have an aversion to risk.

Bolton and Zwick (1995) report on an experiment that concerns the simplified version of the ultimatum game, called cardinal ultimatum, illustrated in the top panel of Figure 1. Player A

¹ Work by Hoffman et al (1998) takes an evolutionary psychology approach to these questions. The work we report here has less to say about these sorts of model.

² Reputation building with the experimenter was first put forth as an explanation by Hoffman et al. (1994) to explain gift giving in the dictator game.

chooses between offering Player B an equal division of the payoff pie, or an unequal division favoring Player A. If Player B rejects the offer then the game ends with both players receiving no money. The payoffs in the unequal division were rotated across rounds as indicated on the horizontal axis of the bottom panel in the figure (each player had a different partner each round). In the standard perfect equilibrium for this game, bargainers settle on the unequal division.

Figure 1 here.

In the base condition of the experiment, known as cardinal ultimatum, the game is played with experimenter observation; that is, the experimenter gathers information in such a way that he can accurately attribute individual actions to individual subjects. Each of the other two conditions of the experiment features a variation on the base condition. In the experimenter-subject anonymity condition, cardinal ultimatum is played in such a way that the experimenter does not have the information necessary to attribute individual actions to individual subjects (see Bolton and Zwick, 1995, for a description of how this is done). In the impunity condition, the game is modified as shown in the top panel of the figure, so that rejecting an unequal split leaves Player A's share of the payoff intact; the game is played with experimenter observation, as in the base condition.³

Relative to the base condition, the experimenter-subject anonymity condition diminishes the strategic reputation building motive for rejecting the unequal split without changing the unfairness aversion motive, whereas the impunity condition leaves the reputation building motive for rejecting intact but eliminates the unfairness aversion motive, since rejecting in the impunity condition does not avert the unfairness (and arguably makes it worse). Hence the strategic reputation building hypothesis predicts more perfect equilibrium play in the experimenter-subject anonymity condition than in either the cardinal ultimatum or impunity conditions, with no difference for the latter two. In contrast, the unfairness aversion hypothesis predicts the same amount of perfect equilibrium play in cardinal ultimatum and experimenter-subject conditions, but more in the impunity condition.

The bottom panel in Figure 1 summarizes the main results of the experiment. There is a dramatic increase in perfect equilibrium play in the impunity condition over cardinal ultimatum or the experiment-subject anonymity condition. There is little difference between the latter two

conditions. Fong and Bolton (1997) analyze the data using a Bayesian bioassay technique. Given the assumptions of the Bayesian model, the analysis identifies some statistical evidence that perfect equilibria are higher in experimenter-subject anonymity than in cardinal ultimatum, but the difference is quite small.

A substantial portion of the deviation from perfect equilibrium in cardinal ultimatum stems from Player B rejections of an offer of the unequal division; the rejection rate ranges from 7 to 100 percent depending on how unequal the division is. In contrast, no Player B rejects the unequal division in the impunity condition. Hence the results of the experiment imply that unfairness aversion is a much stronger motive for rejecting money in the ultimatum game than is building a reputation with the experimenter. Also, Players A in cardinal ultimatum and experimenter-subject anonymity deviate from the unequal division (offer the equal division) about 40 to 50 percent of the time, whereas Players A in impunity deviate only twice in the first two rounds of play and never after that. It appears then that Players A act strategically in the sense that they are far more likely to ask for the unequal division when there is a negligible chance of it being turned down than when the chance is non-negligible.

3. Bolton, Katok and Zwick (1998): The nature of giving behavior in dictator games

The dictator game differs from the ultimatum game solely in that the second mover cannot reject what the first mover offers. If money is the only motive then the dictator should keep the entire pie for himself. But in fact dictator game investigations (e.g., Forsythe, Horowitz, Savin, and Sefton, 1994; and Hoffman, McCabe, Shachat and Smith, 1994) report a wide dispersion of dictator game giving. Some dictators do leave nothing, but others give away as much as 50 percent of the pie. The modal amount left is sometimes as high as 30 percent. What explains this behavior?⁴ Bolton, Katok and Zwick (1998) provides some clues.

The experiment began as a 2x2 design and was later extended in response to the initial results. The treatment variables were the number of division choices per game (2 or 6) and the number of games each dictator participated in (1 or 10). Here we focus on two conditions with 6

³ Relative to the original paper, some of the experimental conditions and hypotheses have been renamed to fit the present exposition. The changes are purely semantic.

⁴ The preferences in Bolton (1991) imply that a person is concerned exclusively with fairness for himself. The preferences in Rabin (1993) imply that a person gives to another only when he expects the other to give to him. Neither of these characterizations allows giving in the dictator game. However, more recent models by Fehr and

division choices, one condition featuring 1 game, and another featuring 10 games.⁵ In the 1 game condition, the game involved a \$10 pie, and the dictator could leave \$0 to \$5 to the recipient in \$1 increments. In the 10 games treatments, each game involved a \$1 pie, and the dictator could leave \$0, 0.1, 0.2, 0.3, 0.4 or 0.5 in each game. So in all cases, the total amount available for division by a single dictator was \$10, and up to half of this amount could be given away in gifts.

Figure 2 here.

Figure 2 presents individual dictator giving to 10 recipients. Here, only the 15 dictators (out of 25) who are willing to give a positive total amount are considered. As can be seen from the figure, when distributing a gift among several recipients, individual dictators show little tendency toward equal treatment. Of the 15 subjects who are willing to give a positive total amount, only two subjects gave the same amount to each recipient. In most other cases, the distribution of total gifts appears to be arbitrary.

The haphazard nature of the distribution is puzzling. It is perhaps then all the more striking that we find very consistent behavior when we look at the total sum allocated. The bottom panel of Figure 2 compares frequency distributions for total giving in the 10 games \$1 pie and the 1 game \$10 pie treatments. While most dictators seem to distribute gifts among multiple recipients in an arbitrary manner, the distribution of *total* gifts is the same whether dividing a pie of \$10 with one other or dividing 10 pies of \$1 with 10 others. This is what Selten and Ockenfels (1998) call the ‘fixed total sacrifice effect’.

Bolton, Katok and Zwick propose a hypothetical decision procedure (a pre-model model if you will) that posits that dictator giving arises from a concern on the part of the dictator for a fair distribution between self and the group. In the first step of the decision procedure, dictators decide how much to keep for themselves and how much to give in gifts to the group. Only after deciding the total gift to the group do they decide how to distribute the gift across the group. In the procedure, concerns from fair distribution are posited to originate from personal and social rules that effectively constrain self-interested behavior – although within these constraints dictators behave in a self-interested manner (they act first to secure what they consider to be their

Schmidt (1999) and Bolton and Ockenfels (2000) capture both ultimatum and dictator game behavior along with behavior observed in other games.

⁵ The experiment included several other treatments, including one designed to look for an experimenter-subject anonymity effect. No effect of this sort was evident. Hoffman et al. (1994) find such an effect in their study. Bolton, Katok and Zwick provide some discussion towards reconciling these two results.

own fair share).⁶ What purpose these rules might ultimately serve, whether it be to improve others' welfare (altruism) or some other purpose, is not clear from the data in this study.

4. Selten and Ockenfels (1998) and Ockenfels and Weimann (1999): The fixed total sacrifice effect in the solidarity game

The solidarity game was designed by Selten and Ockenfels (1998) to examine the fixed total sacrifice effect in greater detail. The solidarity game is a one-shot three person game in which each player independently wins a fixed positive pecuniary amount with probability 2/3 and zero with probability 1/3. Before the dice are rolled, each player is asked how much he is willing to give to a loser if there is only one loser in the group, and how much he is willing to give to each loser if there are two. Let x_1 denote the conditional gift for the case of exactly one loser, and x_2 the conditional gifts given to each of two losers. Of course, no positive gifts should be handed over by payoff maximizing subjects.

The types of conditional gift behavior that were observed are classified in Table 1. The classification is concerned with the question of whether the conditional gifts are positive and how they relate to each other. The corresponding relative frequencies of subjects in the western and eastern part of Germany are based on the data collected by Selten and Ockenfels (1998) and Ockenfels and Weimann (1999), in solidarity experiments which are conducted shortly after the reunification of Germany in Bonn (western Germany) and Magdeburg (eastern Germany), respectively, both under conditions of experimenter-subject anonymity.

Table 1 here.

In both studies, a majority of subjects choose positive conditional gifts. Among the gift givers, a large majority exhibit fixed total sacrifice behavior, $x_1 = 2x_2 > 0$. The same sort of hypothetical decision procedure as suggested in Bolton et al. fits this data. But in the context of the solidarity game, Selten and Ockenfels (1998) are able to say more: They formally show that fixed total sacrifice behavior cannot easily be interpreted as the result of utility maximization of altruistic subjects (see the paper for the details).

⁶ This line of reasoning is consistent with Bolton and Ockenfels' (2000) model.

It is natural to wonder whether the rules that govern giving behavior are influenced by culture. Ockenfels and Weimann (1999) examine solidarity game, and repeated public good game experiments in eastern and western Germany in search of potential cross-culture effects. Since the study was conducted in two parts of one nation, it avoided currency, language and related effects which usually cause methodological problems in multinational experimental settings (see Roth et al., 1991). In both subject pools individual behavior exhibits a great deal of heterogeneity. Moreover, as illustrated in Table 1, eastern subjects behave on average in a significantly more self-interested manner than do western subjects. In particular, eastern subjects choose a zero conditional gift in the solidarity game more frequently than do western subjects. Analogously, Ockenfels and Weimann observe that eastern subjects contribute dramatically less than western subjects in all rounds of a repeated linear public good game. A more detailed data analysis suggests the hypothesis that behavioral differences are due to differing social norms that in turn may have resulted from sharply differing economic and social histories in the two parts of Germany.

Despite the differences across cultures, however, the same *qualitative* features of behavior are observed in both subject pools. In the solidarity game, both the distributions of subjects across non-selfish behavioral types of conditional gift behavior and the distributions of quantitative gift levels of non-selfish subjects are very similar in the two parts of Germany (Roth et al., 1991, make similar observations for multinational ultimatum bargaining experiments).

5. Bolton, Brandts, and Ockenfels (1998): Distribution versus intentions in a 2-person dilemma game

Is the other-regarding motive behind rejecting in an ultimatum game or contributing in a dilemma game (ex., the public goods game by Ockenfels and Weimann discussed in the last section) fundamentally the same as, or fundamentally different than, the giving behavior we see in the dictator game? Rabin's (1993) model suggests they are fundamentally different. In Rabin's model, other-regarding behavior is motivated by a desire to reward or punish the intentions behind another's actions towards ones self. So rejecting in an ultimatum game punishes the poor intentions behind the first mover's poor offer, and contributing rewards the good intentions of those who contribute to ones self. Giving in the dictator game cannot be about the same thing since the recipient can take no action, and hence there are no intentions to reward

or punish. Models by Bolton and Ockenfels (1998 and 2000) and Fehr and Schmidt (1999), on the other hand, imply that the motive is fundamentally the same for all three types of games: By these models, players act to choose the distribution of payoffs they prefer most among those that are feasible independent of the intentions of other players.

Contrary to Rabin's model, we observe giving in the dictator game, which would seem to suggest that distribution is at least part of the motive in ultimatum and dilemma games. But it is also plausible that there is an additional motive, like the one posited by Rabin's model, involved in other-regarding behavior in dilemma and ultimatum games.

Bolton, Brandts and Ockenfels (1998) analyze the role of distributional preferences and of intentionality in explaining behavior in two simple sequential dilemma games: treatments I and III in Figure 3. The key insight behind the design is that, in games of this type, the motivation of a second mover who chooses a certain response to a given action by a first mover can not be assessed correctly without additional information. What needs to be known is what the second mover would choose if faced with the same opportunity set as in the C(oooperation) choice of treatment I and as in the D(efection) choice in treatment III, but without the first mover having had any part in bringing that opportunity set about.

Figure 3 here.

The appropriate control is provided by treatment II, shown in Figure 3, in which the second mover can select a distribution of payoffs for himself and the other player in a "free choice" manner. Choices in treatment II give a measure of what level of contributions is motivated by purely distributional concerns. The relevance of intentionality can be assessed by comparing second movers' choices across treatments. If second movers' responses to C in treatment I were more generous than their choices in treatment II, one could say that second movers' rewarding the first mover's good intentions behind the action. If second movers were more generous in treatment III than in treatment II, then one could speak of punishing bad intentions.

The lower part of Figure 3 compares the cumulative distributions of the second mover's decisions for the three treatments. There are no significant differences between the three distributions. The order of the results of treatments I and II is reversed relative to what corresponds to a reward of good intentions: contributions are lower in treatment I than in

treatment II. Although there is no significant difference between treatments II and III, the ordering of contributions is consistent with the punishment of bad intentions. A quantitative comparison, however, shows that the contribution level in treatment II is equal to about nine times the small decrease in contributions that could possibly be attributed to a punishment of bad intentions.⁷

As a whole, we interpret the above results as evidence that behavior in these games can basically be explained on distributional grounds and that the role of intentionality is secondary. At the same time our results suggests that the tendency to punish bad intentions may be a somewhat stronger force than the propensity to reward good intentions, but a firmer conclusion requires additional research.

6. Summary: Regularities for theory building

To summarize, three types of regularities of potential importance to theory building emerge from these experiments:

First, there is a strong strategic element to other-regarding behavior. A successful model will therefore be a strategic model. Unfairness averse second movers reject ultimatum bargaining offers only when doing so makes the resulting division less unequal. And the first mover's proposal is more generous when there is a potential penalty (rejection) for not being generous.

Second, the empirical manifestation of other-regarding behavior – altruism, fairness, and reciprocity – differs in important respects from philosophical and other introspective notions of these concepts. While we do observe a concern for fairness, this concern is in an important sense self-centered (people appear more concerned with whether the agreement is fair to self than fair to others; see also Bolton and Ockenfels on this issue). It is difficult for the standard theory of altruism to account for the nature of giving that we observe; specifically, altruism does not easily account for the fixed total sacrifice effect. The Bolton, Brandts and Ockenfels experiment suggests that instead of being distinctly different motives, altruism, fairness and reciprocity are facets of a common pattern of behavior.

Third, among the factors that do influence other-regarding behavior there is considerable variance in economic significance. Since no model can usefully capture the influence of all

⁷ Bolton, Brandts and Katok (2000) present evidence that confirms the comparison of treatments I and II above. Bolton, Brandts and Ockenfels (2001) and Bolton and Ockenfels (2001) discuss the role of intentionality in variants of the ultimatum game.

factors that influence human behavior, our work has something to say about which factors should be the focus of attention. We found that reputation building with the experimenter does not account for nearly as much of the rejection behavior in ultimatum games as does unfairness aversion. We found that preferences for distribution, independent of considerations of intentions, are sufficient to explain contributions in a simple dilemma game. The solidarity game and a linear public goods game proved sensitive to east German/west German cultural differences, but not so much so that it changes the qualitative features of giving.

References

- Andreoni, James (1989), "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, 97, 1447-1458.
- Becker, Gary S. (1974), "A Theory of Social Interactions," *Journal of Political Economy*, 82, 1063-1093.
- Bolton, Gary E (1991), "A Comparative Model of Bargaining: Theory and Evidence," *American Economic Review*, 81, 1096-1136.
- Bolton, Gary E, Jordi Brandts and Elena Katok (2000), "How strategy sensitive are contributions? A test of six hypotheses in a two-person dilemma game," *Economic Theory*, 15, 367-387.
- Bolton, Gary E, Jordi Brandts and Axel Ockenfels (1998), "Measuring Motivations in the Reciprocal Responses Observed in a Dilemma Game," *Experimental Economics*, 1, 207-219.
- Bolton, Gary E, Jordi Brandts and Axel Ockenfels (2001), "Fair Procedures: Evidence from Games Involving Lotteries," working paper, Max Planck Institute, Jena.
- Bolton, Gary E, Elena Katok and Rami Zwick (1998), "Dictator Game Giving: Rules of Fairness versus Acts of Kindness," *International Journal of Game Theory*, 27, 269-299.
- Bolton, Gary E and Axel Ockenfels (2000), "ERC: A Theory of Equity, Reciprocity and Competition," *American Economic Review*, 90, 166-193.
- Bolton, Gary E and Axel Ockenfels (1998), "Strategy and Equity: An ERC-analysis of the Güth-van Damme game," *Journal of Mathematical Psychology*, 42, 215-226.
- Bolton, Gary E and Axel Ockenfels (this volume), "Self-centered Fairness in Games with More than Two Players."

- Bolton, Gary E, and Axel Ockenfels (2001), "A Stress Test of Fairness Measures in Models of Social Utility," working paper, Penn State University.
- Bolton, Gary E and Rami Zwick (1995), "Anonymity versus Punishment in Ultimatum Bargaining," *Games and Economic Behavior*, 10, 95-121
- Fehr, Ernst and Klaus Schmidt (1999), "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 817-868.
- Fudenberg, Drew, and Eric Maskin (1986): "The Folk Theorem in Repeated Games with Discounting or with Incomplete Information," *Econometrica*, 54, 533-56.
- Fong, Duncan and Gary E Bolton (1997), "Analyzing ultimatum bargaining: a Bayesian approach to the comparison of two potency curves under shape constraints," *Journal of Business and Economics Statistics*, 15, 335-344.
- Forsythe, Robert, Joel L. Horowitz, N. E. Savin and Martin Sefton (1994), "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 6, 347-369.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat and Vernon L. Smith (1994), "Preferences, Property Rights and Anonymity in Bargaining Games," *Games and Economic Behavior*, 7, 346-380.
- Hoffman, Elizabeth, Kevin McCabe and Vernon L. Smith (1998), "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry*, 36, 335-352.
- Kreps, David M., Paul Milgrom, John Roberts and Robert Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27, 245-252.
- Ockenfels, Axel, and Joachim Weimann (1999): "Types and Patterns – An Experimental East-West-German Comparison of Cooperation and Solidarity," *Journal of Public Economics*, 71, 275-287.
- Rabin, Matthew (1993), "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281-1302.
- Roth, Alvin E. (1995), "Bargaining Experiments," in J. Kagel and A. E. Roth (eds.), *The Handbook of Experimental Economics*, Princeton: Princeton University Press, 253-348.

- Roth, Alvin E. and Ido Erev (1995), "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior*, 8, 164-212.
- Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir (1991), "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo," *American Economic Review*, 81, 1068-1095.
- Selten, Reinhard (1987), "Equity and Coalition Bargaining in Experimental Three-person Games," in A. E. Roth (ed.), *Laboratory Experimentation in Economics: Six Points of View*, Cambridge: Cambridge University Press, p. 42-98.
- Selten, Reinhard and Axel Ockenfels (1998), "An Experimental Solidarity Game," *Journal of Economic Behavior and Organization*, 34, 517-539.
- Sugden, Robert (1984), "Reciprocity: The Supply of Public Goods through Voluntary Contributions," *Economic Journal*, 94, 772-87.