Chapter 59

SELF-CENTERED FAIRNESS IN GAMES WITH MORE THAN TWO PLAYERS

GARY E. BOLTON

Smeal College of Business, 310 Beam, Penn State University, University Park, PA 16802, USA
e-mail: geb3@psu.edu

AXEL OCKENFELS

Max Planck Institute for Research into Economic Systems, Strategic Interaction Unit, Kahlaische Str. 10, D-07745 Jena, Germany
e-mail: ockenfels@mpiew-jena.mpg.de

1. Introduction

What we now label as 'fair behavior' in the lab often differs from philosophical notions of the concept. Establishing a clear understanding of the empirical nature of fairness is important if we are to gauge the impact fairness has on economic and political institutions.

Most experiments that aim to reveal the nature of fairness examine two-player games. Among these, the ultimatum game and the dictator game are probably the most prominent. In the ultimatum game, a proposer offers a division of a monetary cake, which the responder can either accept or reject; rejection leaves both players with a zero payoff. The dictator game differs from the ultimatum game only in that the responder has no choice but to accept. In both games, we observe outcomes that deviate from standard perfect equilibrium in the direction of the *egalitarian solution*, the outcome in which both subjects receive half the cake. In ultimatum games, the egalitarian solution is typically the modal outcome and in dictator games it is typically the second most frequent outcome (cf. Roth, 1995).

Recent experiments show that adding players to ultimatum and dictator games in certain ways tends to invalidate the attraction of the egalitarian solution. In some of these games, egalitarianism does not seem to guide individual behavior at all, even though people still deviate from standard predictions of selfish payoff maximization. In other games, behavior conforms nearly perfectly to selfish payoff maximization, even though the corresponding outcomes are far from egalitarian. Hence games with more than two players challenge our understanding of the two-player games.

In the next section, we describe a class of preferences that characterize 'self-centered fairness.' In Section 3, we outline how these preferences can reconcile seemingly incompatible behavior from selected games. Except for the number of players, all of these games are closely related to either the two-person ultimatum or dictator game.

Handbook of Experimental Economics Results, Volume 1

DOI: 10.1016/S1574-0722(07)00059-5

2. Sketch of ERC Preferences

'ERC' is short for equity, reciprocity and competition, the patterns of behavior that ERC preferences attempt to reconcile. Bolton and Ockenfels (1998, 2000) provide a full account of the model. An earlier model, Bolton's (1991) comparative model, posited preferences consistent with ultimatum game behavior. According to this model a player cares not only about her material or *absolute payoff*, but also about her share of the cake or *relative payoff*. Ultimatum game second movers turn down money – sacrifice own pecuniary payoff – in order to gain relative payoff (all receiving nothing is an equal split). Beyond ultimatum games, the comparative model is consistent with a variety of phenomena observed in laboratory play of two-period alternating offer bargaining games. But the comparative model fails with the dictator game. The reason for the failure is straightforward: Comparative model utility functions are non-decreasing in both own pecuniary and relative payoffs. Consequently, no one should be inclined to give a gift to another.

ERC posits that people behave in a self-centered way, but different from the comparative model, as well as from the standard assumption of selfish behavior. Specifically, in an n-player game, player i maximizes the expected value of her *motivation function*:

$$v_i = v_i(y_i, \sigma_i)$$

own absolute payoff   own relative payoff

where

$$\sigma_i = \sigma_i(y_i, c, n) = \begin{cases} y_i/c, & \text{if } c > 0, \\ 1/n, & \text{if } c = 0, \end{cases} \quad \text{and} \quad c = \sum_{j=1}^{n} y_j$$

is the total payoff distributed.

Preferences over the relative payoff are interpreted as preferences over the fairness of the payoff allocation. Note that players care only about own absolute and own relative payoffs, but that the payoff (or welfare) of other individuals does not enter directly into the motivation function. In this sense, the concern for fairness is self-centered.[1]

[1] Self-centered fairness must be distinguished from the "self-serving bias" in assessments of fairness. We define the former as a property of individual motivations while the latter describes a tendency of human beings "[...] to conflate what is fair with what benefits oneself"; see Babcock and Loewenstein (1997, p. 110) and the references cited therein for empirical studies of the self-serving bias. Fehr and Schmidt (1999) propose a model of "self-centered inequity aversion." The model differs from our notion of self-centered fairness in that absolute payoffs of all other players directly enter into the utility function. In the domain of 'advantageous inequality,' individuals are assumed to have standard altruistic preferences.

We assume the motivation function increases in the absolute payoff y_i, and decreases as the relative payoff σ_i moves away from the *social reference share* $1/n$, the average proportion of the cake.[2] Furthermore, we assume that each player i prefers a distribution in which i receives more than the social reference share to an allocation in which all players receive nothing.[3]

The assumptions imply that, *fixing the absolute payoff*, all players prefer their relative payoff to be the social reference share; in this sense the egalitarian solution is of collective prominence. In most situations, however, the relative payoff is not easily separated from the absolute payoff, and in fact the two are usually positively correlated (for instance, if the cake size is fixed). For this reason, some, perhaps most people will strive for more than their social reference share – how much more depends on the marginal rate of substitution between absolute and relative payoffs.

We obtain a visual representation by observing that $c\sigma_i = y_i$, and writing $v_i(y_i, \sigma_i)$ as $v_i(c\sigma_i, \sigma_i)$. Holding c fixed, a typical motivation function looks like Figure 1. The peak of the function at 0.75 indicates that, as a dictator in the dictator game, i keeps 75 percent of the cake, and gives the rest to the other player. The function taking the value zero at 0.2 indicates that, as a second mover in the ultimatum game, i is indifferent between accepting 20 percent of the cake or rejecting, in which case $v_i(0, 0.5) = 0$.

The marginal rate of substitution between absolute and relative payoffs will differ from person-to-person, and so too the ultimatum and dictator game thresholds. There are, however, some general bounds: The model implies that all dictators will choose to keep at least the social reference share. The rejection threshold of a responder in the ultimatum game is always an offer between half the cake and nothing. Hence when making an offer to a randomly chosen responder, the probability of acceptance increases as the offer increases from zero to half the cake. Therefore the proposer in the ultimatum game should never offer more than half the cake. These bounds are in basic accord with what we see in the lab (cf. Roth, 1995).[4]

[2] In Bolton and Ockenfels (2000) we demonstrate that our simple definition of the social reference share organizes a wide range of behaviors observed in experimental fairness games. But, of course, the definition can be modified for games that allow for competing fairness norms. For instance, what is perceived as fair may depend on the payoff menu (see Bolton and Ockenfels, 2001, and the references cited therein) as well as on the fairness of the procedure by which the allocation is reached (see Bolton, Brandts and Ockenfels, 2001).

[3] Our statement of the assumptions is rough. See Bolton and Ockenfels (2000) for a detailed depiction.

[4] In Bolton and Ockenfels (2000) we model the heterogeneity in marginal rates of substitution in terms of the proportion a person takes when in the role of dictator in the dictator game, together with the rejection threshold when in the role of ultimatum game responder. The model is then solved by applying perfect Bayesian equilibrium under the assumption that individual motivation functions are private information. We derive predictions for a variety of experimental games, including variations on dictator and ultimatum games, the gift-exchange game, the prisoner's dilemma, Bertrand markets, and the guessing game.

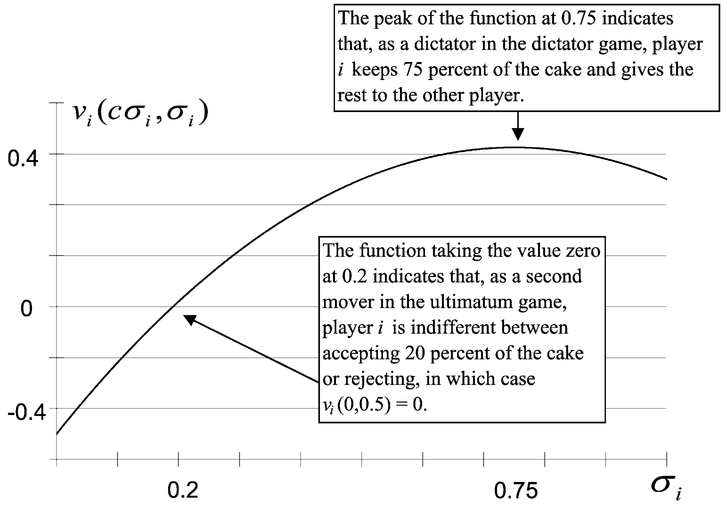

Figure 1. A typical motivation function $v_i(y_i, \sigma_i)$, y_i = 'absolute payoff,' σ_i = 'relative' payoff. We obtain a visual representation by holding the cake size c fixed, observing that $c\sigma_i = y_i$, and writing $v_i(y_i, \sigma_i)$ as $v_i(c\sigma_i, \sigma_i)$. The motivation function in the figure provides an example of ERC preferences. These are based on the assumption that people are motivated by their own absolute (pecuniary) payoff as well as their own relative share of the payoff cake. The payoffs of other individuals do not enter directly into the motivation function. In this sense, the concern for fairness is self-centered.

3. Evidence in Games with More Than Two Players

3.1. The Güth–van Damme Bargaining Game

Güth and van Damme (1998) present a three-person bargaining game, which is closely related to the standard two-person ultimatum game. The only difference is a third player, the 'dummy.' The proposer proposes a division of 120 points among herself, the responder and the dummy. A minimal amount, 5 points, must be allocated to each player, but otherwise the proposer is free to allocate as she chooses. The responder either accepts or rejects the proposal. If accepted, the money is distributed accordingly. If rejected, all receive nothing. The dummy has no say in the negotiation, and no choice but to accept any agreement set by the other two.

The game was played under three information conditions. We restrict attention to two. In the *full information* condition, the responder knows the full proposal at the time of accepting or rejecting. In the *irrelevant information* condition, the responder knows only the dummy's share.[5]

[5] In the information condition not described, the responder knows only his own allocation. Güth and van Damme (1998) examined all three information conditions in two modes. In the constant mode, all games had

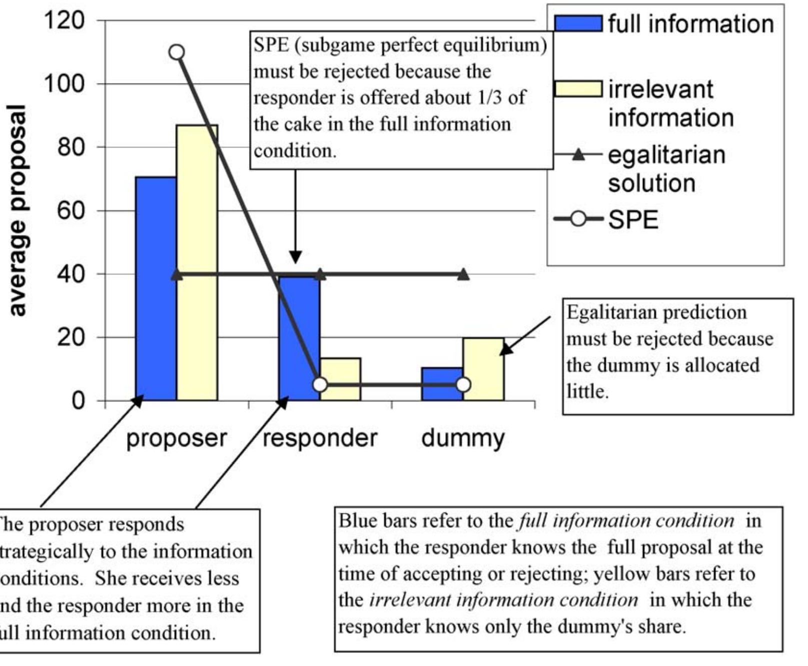

Figure 2. Average proposals in the Güth–van Damme game: The proposer proposes a division of 120 points among herself, the responder and the dummy. The responder either accepts or rejects the proposal. If rejected, all receive nothing. The dummy has no say in the negotiation. ERC preferences, like those described in Figure 1, can organize the proposer's behavior. First, since the risk of rejection is decreasing in the offer to the responder, proposers offer more to the responder than predicted by the standard subgame perfect equilibrium. Second, since ERC-players are self-centered, and do not strive for egalitarian allocations, the dummy – who has no veto-power – is allocated little (see Table 1 for more on this point). Third, since there is no evidence in the data that the responders could ascertain their own share from what is offered to the dummy, ERC-proposers strategically exploit the incomplete information condition.

The subgame perfect equilibrium under the standard self-interest assumption (henceforth SPE) is invariant to the information condition: every feasible proposal gives each bargainer a positive amount, so the responder always makes more money accepting than rejecting. The proposer should therefore ask for the maximum allowable. The egalitarian solution is also invariant to the information condition: each bargainer gets a one-third share. Figure 2 shows the average proposals.

The figure illustrates that neither SPE nor egalitarianism fits the data particularly well. In both information conditions, the proposer receives more than the responder and the dummy combined but still significantly less than in the SPE allocation. While the

the same information condition, and in the cycle mode, games were rotated through all conditions. We confine ourselves to the constant mode. See Bolton and Ockenfels' (1998) analysis.

Table 1
Rejection behavior in the full information condition in the Güth–van Damme game. Self-centered ERC-responders care neither about the absolute nor the relative payoffs of the other individual players, which explains why the responders reject only when they personally receive less than the social reference share (40) regardless of the dummy's payoff. Second, offering something to the responder decreases the risk of rejection and improves the proposer's relative position while giving to the dummy can only improve the relative position. Consequently, responders are 'served first' by ERC-proposers. Third, from the point of view of self-centered players, the three-person Güth–van Damme game creates more room to agree on a distribution of relative payoffs than a standard two-person ultimatum game. Therefore, ERC predicts a smaller rejection rate in the three-person game than in the two-person game

Number of proposals in the full information condition in the Güth–van Damme game	Dummy's payoff = 5 (minimum value allowed)		Dummy's payoff > 5	
	Responder's payoff <40	Responder's payoff ⩾40	Responder's payoff <40	Responder's payoff ⩾40
Accepted	23	18	0	26
Rejected[a]	5[b]	0[c]	0[d]	0[d]

[a]The overall rejection rate (7%) is remarkably small.

[b]Rejections only occur if the offer to the responder is smaller than the 'social reference share' 1/3 of 120 = 40.

[c]In particular, none of the proposals that give the minimum value to the dummy are rejected if the responder's payoff is at least 40.

[d]The proposer is 'served first': The dummy is never allocated more than the minimum as long as the proposer is offered less than the 'social reference share.'

responder is, on average, offered about the social reference payoff (1/3 of 120) in the full information condition, the dummy is allocated little in both conditions.[6] In contrast with what we observe in the dictator game, there is nothing here to make us think that the proposer cares about the dummy's payoff.

Table 1 suggests that the responder too cares little about the dummy's payoff. Rejections only occur if the offer to the responder is smaller than the social reference share. Note in particular that if the responder's allocation is at least 40, none of the proposals that give the minimum value to the dummy are rejected.

ERC preferences can organize the behavior of both the proposer and the responder. ERC-players care neither about the absolute nor about the relative payoffs of other individuals, which immediately explains why the responder rejects only when he personally receives less than the social reference.

Offering something to the responder has, from the proposer's point of view, two positive effects: First, the risk of rejection decreases, at least up to the point that the offer is equal to the social reference share. Second, giving something may improve the proposer's relative position (see Figure 1). In contrast, giving to the dummy can only

[6] Keep in mind that the rules of the game require that a minimum of 5 be allocated to each player.

improve the relative position. Therefore, ERC predicts that the responder is 'served first,' at least up to the social reference share. Table 1 is in line with this prediction: only in those cases in which the responder receives at least the social reference share is the dummy allocated more than the minimum.

Figure 2 and Table 1 capture two other phenomena.

First, observe from Figure 2 that, compared to the full information condition, the proposer receives more, and the responder less, in the irrelevant information condition. It turns out that the data shows no evidence of signaling behavior in the irrelevant information condition; that is, no evidence that the responder could ascertain her own share from what is offered the dummy. As we would expect if the responder cares only about her own share of the payoff, proposers strategically exploit the situation by allocating smaller amounts for the responder in the irrelevant information condition than they would in the full information condition.[7]

Second, observe in Table 1 that the overall rejection rate is small, 7 percent, compared to rates in the two-person ultimatum game, which typically run in the neighborhood of 15 to 20 percent (see Roth, 1995). ERC offers an explanation. The underlying idea is that the three-person game creates more room to agree on a distribution of relative payoffs between the proposer and the responder than a two-person game: A proposer in the three-person game can demand up to $2/3$ of the cake for herself with no risk of rejection, but not generally so in the two-person game. It can be shown that (risk-averse) proposers who want more than $2/3$ of the cake will use some of the extra resources in the three-person game to lower the probability of rejection relative to the risk they would assume in the two-person game (cf. Bolton and Ockenfels, 1998).

3.2. Market Game

Roth et al. (1991) report an experiment on the 'market game' similar to the ultimatum game: Nine proposers simultaneously submit offers on the cake. One responder is given the opportunity to accept or reject the highest offer. In the case of a tie, one offer is chosen at random. All SPE imply a highly unequal payoff distribution: the responder receives virtually the entire cake, namely 995 or 1000. Roth et al. ran the market game in each of four countries. In each country, there were four markets, each market iterated for 10 rounds. Figure 3 shows the minimum of the four winning offers per round by country. We see that, after a few rounds, outcomes are remarkably consistent with the SPE prediction, with the best offer rising to the SPE offer no later than round seven.

Competing in this market game is consistent with ERC preferences. For simplicity, let us restrict attention to pure strategies. Suppose that all proposers but i are offering 995. If i offers less than 995 his offer will not get to the responder and he receives

[7] That proposers exploit their private information about the share offered to the responder has also been observed in the two-person ultimatum game; cf. Mitzkewitz and Nagel (1993), Kagel, Kim, and Moser (1996), and Rapoport, Sundali, and Potter (1992).

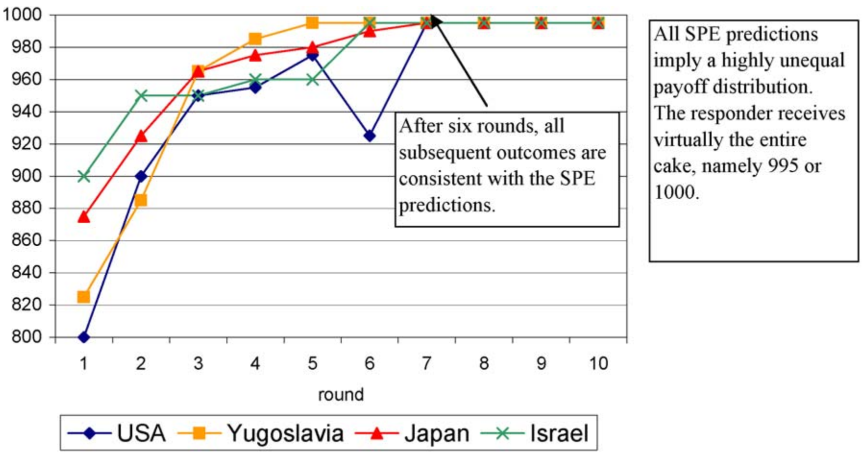

Figure 3. Winning offers in the market game (Roth et al., 1991): Nine proposers simultaneously submit offers on the cake (=1000). One responder is given the opportunity to accept or reject the highest offer. The game was run in each of four countries. In each country there were four markets, each market iterated for 10 rounds. We show the minimum of the four winning offers per round by country. Typical market environments interact with ERC-preferences in a way that aligns absolute and relative motives: Self-centered proposers can improve both their absolute and relative position by overbidding offers that are smaller than SPE-offers. Consequently, competing in the market game is consistent with self-centered fairness.

nothing; if he offers the entire cake, 1000, it comes to the same: $v_i(0, 0)$, a very poor payoff. If i offers 995, then he has a $1/9$ chance of receiving $v_i(5, 5/1000)$, which is better. So all proposers offering 995 is stable.[8] Suppose, however, that everybody else is offering something less than 995, say 980. Now if i offers just a bit more, 985, the offer goes to the responder for sure, and i receives $v_i(15, 15/1000)$ instead of a $1/9$ chance of $v_i(20, 20/1000)$. So long as the difference between $v_i(15, 15/1000)$ and $v_i(20, 20/1000)$ is small, i has an incentive to overbid the others, and hence a winning offer of less than 995 cannot be stable.

3.3. *The Fixed Total Sacrifice Effect*

Bolton, Katok, and Zwick (1998) study a dictator game in which the dictator divides the cake among himself and 10 recipients. They found no difference between the distribution of the total gift that dictators leave the multiple recipients and the distribution for the game where there is but one recipient.

Selten and Ockenfels (1998) observe the same sort of fixed total sacrifice effect in the solidarity game: Each player in a three-person group independently throws dice

[8] By our assumptions, the responder will accept the offer with probability one. A positive acceptance probability, however, would be enough. A similar argument shows that a winning offer of 1000 can also be stable.

to determine whether they (individually) win a fixed monetary sum. Before the dice are rolled, each player announces how much he wishes to compensate the losers, for both the case where there is one loser, and for the case where there are two. Most subjects give the same total (positive) amount independent of the number of losers. Selten and Ockenfels formally demonstrate that the behavioral pattern is not easy to justify if subjects have standard altruistic preferences (a class that includes egalitarian preferences).[9]

4. Summary

The concept of self-centered fairness, as embodied in ERC preferences, organizes the data from the two-person ultimatum and dictator game, as well as the Güth–van Damme game, the market game, the dictator game with multiple recipients, and the fixed total sacrifice effect as observed by Bolton, Katok, and Zwick (1998).

There are also some clear limitations to the approach. Specifically, a more comprehensive decision theory need incorporate learning and bounded rationality:

Subjects in the market game require experience before playing SPE. Likewise, Güth and van Damme (1998) report an experience effect for their game. While the dynamics in both games are consistent with the idea that behavior converges towards equilibrium play of self-centered subjects, the dynamics themselves cannot be explained in the framework of a static model.

The solidarity game by Selten and Ockenfels (1998) poses a different challenge. There, dictator behavior together with the expectations they report sometimes imply that a recipient may receive more than the dictator which contradicts ERC-predictions. Selten and Ockenfels (1998) propose a model of boundedly rational decision making in order to capture the phenomenon.

An interesting theme for future empirical and theoretical research, particularly in games with more than two players, is the identification of the 'appropriate' reference group. While this task is rather obvious in the simple games considered here, this need not to be true in more complex social environments. A nice example of an environment with competing reference groups can be found in European parliaments. The income distribution of the Europe-parliamentarians in Brussels is highly unequal. Since Europe-parliamentarians of all European countries basically have the same tasks and work load, this was regarded as unfair, and a proposal was made to equalize incomes. However, German European-parliamentarians (among others) objected to the proposal, on the grounds that, under the proposal, they would earn much less than their colleagues in the German parliament in Bonn. Likewise, a Spanish European-parliamentarian voted

[9] The "embedding effect" in the contingent valuation literature is a related phenomenon. It suggests that the willingness-to-pay does not vary substantially with the number of projects to be valued. Kahneman and Knetsch (1992) and Desvousges et al. (1993) find empirical support. See Hanemann (1994) for a skeptical assessment.

against the proposal, stating that he would earn much more than his colleagues in Spain. Due to the conflicting reference groups and given the differing incomes of the parliamentarians across nations, it is far from obvious what constitutes a 'fair' income distribution within the European parliament.

References

Babcock, Lind, Loewenstein, George (1997). "Explaining bargaining impasse: The role of self-serving biases". Journal of Economic Perspectives 11, 109–126.

Bolton, Gary E. (1991). "A comparative model of bargaining: Theory and evidence". American Economic Review 81, 1096–1136.

Bolton, Gary E., Ockenfels, Axel (1998). "Strategy and equity: An ERC-analysis of the Güth–van Damme game". Journal of Mathematical Psychology 42, 215–226.

Bolton, Gary E., Ockenfels, Axel (2000). "ERC: A theory of equity, reciprocity and competition". American Economic Review 90 (1), 166–193.

Bolton, Gary E., Ockenfels, Axel (2001). "A stress test of fairness measures in models of social utility". Working paper, Penn State University.

Bolton, Gary E., Brandts, Jordi, Ockenfels, Axel (2001). "Fair procedures: Evidence from games involving lotteries". Working paper, Max Planck Institute, Jena.

Bolton, Gary E., Katok, Elena, Zwick, Rami (1998). "Dictator game giving: Rules of fairness versus acts of kindness". International Journal of Game Theory 27, 269–299.

Desvousges, W.H., Johnson, F.R., Dunford, R.W., Boyle, K.J., Hudson, S.P., Wilson, K.N. (1993). "Measuring natural resource damages using contingent valuation: Tests of validity and reliability". In: Hausman, J. (Ed.), Contingent Valuation: A Critical Assessment. North-Holland Press, Amsterdam, pp. 91–164.

Fehr, Ernst, Schmidt, Klaus (1999). "A theory of fairness, competition and cooperation". Quarterly Journal of Economics 114, 817–868.

Güth, Werner, van Damme, Eric (1998). "Information, strategic behavior and fairness in ultimatum bargaining: An experimental study". Journal of Mathematical Psychology 42, 227–247.

Hanemann, W. Michael (1994). "Valuing the environment through contingent valuation". Journal of Economic Perspectives 8, 19–43.

Kagel, John, Kim, Chung, Moser, Donald (1996). "Fairness in ultimatum games with asymmetric information and asymmetric payoffs". Games and Economic Behavior 13, 100–110.

Kahneman, Daniel, Knetsch, Jack L. (1992). "Valuing public goods: The purchase of moral satisfaction". Journal of Environmental Economics and Management 22, 57–70.

Mitzkewitz, Michael, Nagel, Rosemarie (1993). "Envy, greed and anticipation in ultimatum games with incomplete information". International Journal of Game Theory 22, 171–198.

Rapoport, Amnon, Sundali, James A., Potter, Richard E. (1992). "Ultimatum games with incomplete information: Effects of the variability of the pie size". Mimeo, University of Arizona.

Roth, Alvin E. (1995). "Bargaining experiments". In: Kagel, J., Roth, A.E. (Eds.), Handbook of Experimental Economics. Princeton University Press, Princeton.

Roth, Alvin E., Prasnikar, Vesna, Okuno-Fujiwara, Masahiro, Zamir, Shmuel (1991). "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo". American Economic Review 81, 1068–1095.

Selten, Reinhard, Ockenfels, Axel (1998). "An experimental solidarity game". Journal of Economic Behavior and Organization 34, 517–539.