Theories of Fairness: Experimental Evidence

Jae Bong Chang
Research Fellow

KOREA RURAL ECONOMIC INSTITUTE

TABLE OF CONTENTS

I Introduction

Many influential economists, including Adam Smith (1759), Gary Becker (1974), Kenneth Arrow (1981), Paul Samuelson (1993) and Amartya Sen (1995), pointed out that people often do care for the well-being of others and that this may have important economic consequences. However, most economists still routinely assume that material self-interest is the sole motivation of all people. This practice contrasts sharply with a large body of evidence gathered by experimental economists and psychologists during the last two decades. This evidence indicates that a substantial percentage of the people are strongly motivated by other-regarding preferences and that concerns for the well-being of others, for fairness and for reciprocity, cannot be ignored in social interactions. One purpose of this chapter is to review this evidence, suggest how it can be best interpreted, and how it should be modeled. We take up this task in Section 2, where we describe the most important experiments that have radically changed the views of many experimental economists over the last two decades. Section 2 also describes recent neuroeconomic experiments that combine the tools of experimental economics with non-invasive brain imaging methods of modern neuroscience to better understand how the brain generates other-regarding behaviour.[1)]

1) Readers who are interested in the role of reciprocity and altruism at the workplace and, more generally, in cooperative endeavours, should consult the excellent handbook chapters by Putterman and Rotemberg. Kolm, provides an interesting discussion of the concept of reciprocity that differs from the preference based theories dealt with in our chapter.

In hindsight, it is ironic that experiments have proven to be critical for the discovery and the understanding of other-regarding preferences because experimental economists were firmly convinced for several decades that other-regarding motives only had limited impact. They believed that the self-interest assumption provides a good description for most people's behaviour. At best, other-regarding behaviour was viewed as a temporary deviation from the strong forces of self-interest. Vernon Smith discovered in the 1950s that experimental markets quickly converge to the competitive equilibrium if subjects trade a homogeneous good and all aspects of the good are fully contractible (Smith 1962). Hundreds of experiments have since confirmed the remarkable convergence properties of experimental markets (see Davis and Holt 1993, for example). The equilibrium in these experiments is computed assuming that all players are exclusively self-interested. Therefore, the quick convergence to equilibrium was interpreted as a confirmation of the self-interest hypothesis.

However, the bargaining and cooperation experiments described in Section 2 below illustrate that this conclusion was premature because a large percentage of the subjects in these experiments - some of which involve fully representative subject pools for whole countries - exhibit other regarding behaviour that the self-interest hypothesis cannot rationalize in any reasonable way. Subjects in these experiments have to make simple decisions in situations where the self-interested choice is salient and easy to understand. Thus, if they deviate from the self-interested choice, we can conclude that they exhibit some form of other-regarding preference. Given this evidence, the real question is no longer whether many people have other-regarding preferences, but under which conditions these preferences have important economic and social effects and what the best way to describe and model these preferences is.

However, the evidence from competitive market experiments remains. How can we reconcile the fact that the self-interest model predicts behaviour in competitive experimental markets with fully contractible goods very well while it completely fails in the simple experiments described in Section 2 below? Some of the recently developed models of other-regarding preferences that are described and discussed in some detail in Section 3 provide a solution to this puzzle; they show

that competition may completely remove the impact of other-regarding preferences. Thus, the fact that we do not observe other-regarding behaviour in certain competitive markets does not mean that other-regarding preferences are absent. Instead, rational individuals will not express their other-regarding preferences in these markets because the market makes the achievement of other-regarding goals impossible or infinitely costly. However, a large amount of economic activity takes place outside competitive markets - in markets with a small number of traders, in markets with informational frictions, in firms and organizations, and under contracts which are neither completely specified nor enforceable. Models based on the self-interest assumption frequently make very misleading predictions in these environments, while models of other-regarding preferences predict much better. These models thus provide fresh and experimentally confirmed insights into important phenomena like the persistence of non-competitive wage premiums, the incompleteness of contracts and the absence of explicit incentive schemes, the allocation of property rights, the conditions for successful collective action, and the optimal design of institutions.

One of the exciting aspects of this development is that the newly developed theories of other-regarding preferences were tested in a new wave of experiments, sometimes before they were even published. This led to important insights into the power and the limits of different models which will be discussed in Section 4. These experiments also show that it is possible to discriminate between different motivational assumptions, answering one important objection to this research program. There has always been a strong convention in economics of not explaining puzzling observations by changing assumptions on preferences. Changing preferences is said to open Pandora's Box because everything can be explained by assuming the "right" preferences. We believe that this convention made sense in the past when economists did not have the tools to examine the nature of preferences in a scientifically rigorous way. However, due to the development of experimental techniques these tools are now available. In fact, one purpose of this paper is to show that the past decade has yielded both progress on and fascinating new insights into the nature of other regarding preferences.

While many people are strongly concerned about others' well-being, fair-

ness, and reciprocity, we consider it equally important to stress that the available experimental evidence suggests that there are also many subjects who behave quite selfishly even when they are given a chance to affect other people's well-being at a relatively small cost. One of the exciting insights of some of the newly developed theoretical models is that the interaction between fair and selfish individuals is key to understanding the observed behaviour in strategic settings. These models explain why almost all people behave as if they are completely selfish in some strategic settings, while the same people will behave as if driven by fairness in others.

We describe several examples that show the economic importance of other-regarding preferences in different settings in the final part of the paper, Section 5. Among other things, we provide evidence indicating that other-regarding preferences are decisive for explaining collective action and multi-lateral cooperation. We present, in particular, recent evidence showing that if individuals can choose between an institution allowing mutual punishment of non-cooperative behaviour or one which rules out mutual punishment, they converge to a behavioral equilibrium in which the selfish and the other-regarding types unanimously prefer the punishment institution. Moreover, punishment of free riders actually occurs and drives the behaviour in the punishment institution towards a state in which full cooperation und no punishment occurs. The threat of punishment alone suffices to generate full cooperation. This experiment constitutes a powerful example suggesting that other-regarding preferences have shaped many of our cooperative institutions. In addition, we document that other-regarding preferences have deep effects on outcomes in markets with moral hazard problems, while the interaction between selfish and fair-minded subjects in markets with fully contractible goods generates outcomes that are close to the competitive prediction. Finally, we report how other-regarding preferences influence voting behaviour in taxation games. These examples, although important, provide only a glimpse into the full range of possibilities how other-regarding preferences shape social and economic interactions including, perhaps, some of our most fundamental institutions. The examples also show that the main reason why other-regarding preferences are important lies in the fact that even a minority of other-regarding people may gen-

erate powerful cooperation incentives for selfish people.

To set the stage for the discussion of the following sections we give an informal and intuitive definition of several types of other-regarding preferences that received a lot of attention in the recent literature that tries to explain behavior in economic experiments. In Section 3 we define these preferences in a formal and more rigorous way. The theoretical literature on other-regarding preferences has focused on three departures from the standard self-interest model. In addition to the material resources allocated to him a person may also care about: (i) The material resources allocated to other agents in a relevant reference group. (ii) The fairness of the behavior of relevant reference agents. (iii) The "type" of the reference agents, i.e. whether the agents have selfish, altruistic, spiteful, or fair minded preferences.

Consider first the case where the utility function of an individual also depends on the material resources that other agents in a relevant reference group receive. A typical example is altruism. Altruism is a form of unconditional kindness; that is, a favor given does not emerge as a response to a favor received (Andreoni 1989; Andreoni and Miller 2002; Cox, Sadiraj and Sadiraj 2001, Charness and Rabin 2002). In technical terms, altruism means that the first derivate of the utility function of an individual with respect to the material resources received by any other agent is always strictly positive. Thus, an altruist is willing to sacrifice own resources in order to improve the well being of others. The opposite case is envy or spitefulness. A spiteful person always values the material payoff of relevant reference agents negatively. Such a person is, therefore, always willing to decrease the material payoff of a reference agent at a personal cost to himself (Bolton 1991, Kirchsteiger 1994, Mui Vai-Lam 1995) irrespective of both the payoff distribution and the reference agent's fair or unfair behavior. Therefore, spiteful preferences represent the antisocial version of other-regarding preferences. A conditional form of altruism and/or envy is inequity aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000, Charness and Rabin 2002). An individual is inequity averse if, in addition to his material self-interest, his utility increases if the allocation of material payoffs becomes more equitable. Thus, an inequity averse person may value additional material resources allocated to a reference agent pos-

itively or negatively, depending on whether the allocation becomes more or less equitable. Obviously, the definition of equity is very important in these models. In the context of experimental games equity is usually defined as equality of monetary payoffs. However, departures from equality have been defined differently. They can be measured in terms of the income differences between the individual and all relevant reference agents, or in terms of the difference between the individual and the least well-off in his reference group, or in terms of the individual's relative share of the overall surplus.

The case where preferences depend on the fair or unfair behavior of other agents has also received much attention in the literature and is often called reciprocity. A reciprocal individual, as we define it here, responds to actions he perceives to be kind in a kind manner, and to actions he perceives to be hostile in a hostile manner (Rabin 1993, Segal and Sobel 2004, Dufwenberg and Kirchsteiger 2004, Falk and Fischbacher 2005). Thus, preferences do not only depend on material payoffs but also on intentions, i.e. on beliefs about why an agent has chosen a certain action. This cannot be modeled by using conventional game theory but requires the tools of psychological game theory (Geanakoplos, Pearce and Stacchetti, 1989).

Finally, preferences may depend on the type of opponent (Levine 1998). According to type-based reciprocity, an individual behaves kindly towards a "good" person (i.e, a person with kind or altruistic preferences) and hostilely towards a "bad" person (i.e. a person with unkind or spiteful preferences). Note that it is the "type" of a person and not the intention" of his action that affects preferences in this case. Therefore, type-based reciprocity can be modeled using conventional game theory.

It is important to emphasize that it is not the expectation of future material benefits that drives reciprocity. Reciprocal behavior as defined above differs fundamentally from "cooperative" or "retaliatory" behaviour in repeated interactions that is motivated by future material benefits. Therefore, reciprocal behaviour in one-shot interactions is often called "strong reciprocity" in contrast to "weak reciprocity" that is motivated by long-term self-interest in repeated interactions. (Gintis 2000; Fehr and Fischbacher 2003).

Readers who are mainly interested in the experimental evidence that documents the existence of other-regarding preferences should first consult Section 2 and then Section 4 of this chapter. In Section 2 we present a list of simple experiments that indicate the existence and the prevailing patterns of other-regarding preferences. In Section 4 we discuss the most recent evidence in the light of the newly developed models of other-regarding preferences. Readers who are mainly interested in the different models of other-regarding preferences and how they perform relative to the available evidence can directly jump to Section 3 and Section 4. Finally those readers who are mainly interested in the economic impact of other-regarding preferences may directly jump to Section 5.

II Empirical Foundations of Other-regarding Preferences

1. Other-regarding Behaviour in Simple Experiments

In the introduction, we referred to the previously held belief of many experimental economists in the validity of the self-interest hypothesis. This "commitment" to the self-interest hypothesis slowly weakened in the 1980s, when experimental economists started studying bilateral bargaining games and interactions in small groups in controlled laboratory settings (see e.g. Roth, Malouf and Murningham 1981, Güth, Schmittberger and Schwarze 1982). One of the important experimental games that eventually led many people to realize that the self-interest hypothesis is problematic was the so-called Ultimatum Game by Güth, Schmittberger and Schwarze (1982). In addition, games like the Dictator Game, the Power to Take Game, the Third Party Punishment Game, the Gift Exchange Game and the Trust Game played an important role in weakening the exclusive reliance on the self-interest hypothesis. All these games share the feature of simplicity, enabling the experimental subjects to understand them and therefore making inferences about subjects' motives more convincing. In fact, in all these games one player has a strictly dominant strategy if he is self-interested and this selfish strategy is salient and easy to understand in all cases. Therefore, if this player does not choose his or her selfish strategy, we can infer that he deliberately did not do so, i.e., we can make inferences about his motives.

In the Ultimatum Game, a pair of subjects has to agree on the division of a fixed sum of money. Person A, the proposer, can make one proposal of how to

divide the amount. Person B, the Responder, can accept or reject the proposed division. In case of rejection, both receive nothing; in case of acceptance, the proposal is implemented. Under the standard assumptions that (i) both the proposer and the responder are rational and care only about how much money they get and (ii) that the Proposer knows that the Responder is rational and selfish, the subgame perfect equilibrium prescribes a rather extreme outcome: the Responder accepts any positive amount of money and, hence, the Proposer gives the Responder the smallest money unit, ε, and keeps the rest to himself.

A robust result in the ultimatum game, across hundreds of experiments, is that the vast majority of the offers to the Responder are between 40 and 50 percent of the available surplus. Moreover, proposals offering the Responder less than 20 percent of the surplus are rejected with probability 0.4 to 0.6. In addition, the probability of rejection is decreasing in the size of the offer (see, e.g., Güth, Schmittberger and Schwarze, 1982; Camerer and Thaler, 1995; Roth, 1995, Camerer 2003 and the references therein). Apparently, many Responders do not behave in a self-interest maximizing manner. In general, the motive indicated for the rejection of positive, yet "low", offers is that subjects view them as unfair. A further robust result is that many Proposers seem to anticipate that low offers will be rejected with a high probability. A comparison of the results of dictator games and ultimatum games suggests this. The Responder's option to reject is removed in a dictator game; the Responder must accept any proposal. Forsythe et al. (1994) were the first to compare the offers in ultimatum and dictator games. Self-interested proposers should allocate nothing to the Recipient in the dictator game. In experiments, Proposers typically dictate allocations that assign the Recipient on average between 10 and 25 percent of the surplus, with modal allocations at 50 percent and zero. These allocations are much less than Proposers' offers in ultimatum games, although most players do offer something. Comparing dictator with bilateral ultimatum games shows that fear of rejection is part of the explanation for Proposers' generous offers, because they do offer less when rejection is precluded. But many subjects offer something in the dictator game, so fear of rejection is not the entire explanation. The considerably lower offers in the dictator game suggest that many Proposers apply backwards induction. This interpretation

is also supported by the surprising observation of Roth, Prasnikar, Okuno-Fujiwara and Zamir, 1991, who showed that the modal offer in the ultimatum game tends to maximize the Proposer's expected income.[2)]

The "power to take game", invented by Bosman and van Winden (2002), is another tool that has proven useful in understanding punishment behaviour. Both the Proposer and the Responder are endowed with some income in this game. Subjects may have earned this income, as in Bosman and van Winden (2002), or the experimenter may have allocated the money to the subjects as in Bosman, Sutter and van Winden (2005). The Proposer can set a take or "theft" rate $t \in [0,1]$ which is the fraction of the Responder's endowment that will be transferred to the Proposer. The Responder is then informed of the take rate and can destroy part or all of his income. Thus, if the Responder destroys his or her whole income nothing is transferred to the Proposer. If the Responder destroys only a fraction d, $d \in [0,1]$, of his income, the Proposer receives a share of $t(1 - d)$ of the Responder's pre-destruction income. In contrast to the ultimatum game, the power to take game allows the punishment behaviour to vary continuously with the take rate. The evidence indicates that the destruction rate is roughly $d = 0.5$ for take rates around $t = 0.8$, regardless of whether the initial endowment was earned through effort or exogenously allocated by the experimenter. However, the destruction rate is higher for lower take rates if the initial endowment is given to the subjects without effort, whereas the destruction rate is higher for takes rates above 0.8 if the endowment was earned through effort. This indicates that the way the initial endowment is allocated to the subjects matters because it seems to affect their feelings of entitlement. Hoffman, McCabe and Smith (1996) also reported that feelings of entitlement may be important for punishment behaviour in the context of the ultimatum game.

---

2) Suleiman (1996) reports the results of ultimatum games with varying degrees of veto power. In these games a rejection meant that λ percent of the cake was destroyed. For example, if $\lambda = 0.8$, and the Proposer offered a 9:1 division of \$10, a rejection implied that the Proposer received \$1.8 while the Responder received \$0.2. Suleiman reports that Proposers' offers are strongly increasing in λ.

The Responders' feelings may be hurt if he or she receives an unfairly low offer in the ultimatum game. Thus, pride or motives to retain self-respect may drive a rejection. Therefore, the question arises whether people would also be willing to punish violations of social or moral norms if they themselves are not the victim of the norm violation. A game that is particularly suited to examine this question is the so-called third party punishment Game (Fehr and Fischbacher 2004). The three players in this game are denoted A, B, and C. A and B play a simple dictator game. Player A, the Proposer, receives an endowment of S tokens of which he can transfer any amount to player B, the Recipient. B has no endowment and no choice to make. Player C has an endowment of S/2 tokens and observes player A's transfer. Player C can then assign punishment points to player A. Player C incurs costs of 1 token and player A is charged 3 tokens for each punishment point player C assigns to player A. Since punishment is costly, a self-interested player C will never punish. However, if there is a sharing norm, player C may well punish player A if A gives too little.

In fact, in the experiments conducted by Fehr and Fischbacher (2004), where S = 100, player A was rarely punished if he transferred 50 or more tokens to player B. If he transferred less than 50 tokens, roughly 60 percent of players C punished A and the less A transferred, the stronger was the punishment. If nothing was transferred, A received on average 14 punishment points, reducing A's income by 42 tokens. Thus, if nothing was transferred player A earned (on average) more money in this setting than if he transferred the fair amount of 50. However, if player C was himself the recipient in another dictator game unrelated to that played between A and B, C punished more. All transfer levels below 50 were on average punished so strongly in this case that it was no longer in player A's self-interest to transfer less than 50. It seems that if C is himself a recipient, he is more able to empathize with B if B receives little and thus increase the punishment imposed on A. Finally, if third party punishment is compared to second party punishment (i.e. if B can punish A), it turns out that second party punishment is significantly stronger than is third party punishment. Note that this does not necessarily mean that third party punishment is less effective in sustaining social norms because third parties are often more numerous than second parties.

Dictator games measure pure altruism. Interesting companion games are the trust game (Berg, Dickhaut and McCabe 1995) and the gift exchange game (Fehr, Kirchsteiger and Riedl 1993). In a trust game, both an Investor and a Trustee receive an amount of money S from the experimenter. The Investor can send between zero and S to the Trustee. The experimenter then triples the amount sent, which we term y, so that the Trustee has S + 3y. The Trustee is then free to return anything between zero and S + 3y to the Investor. The Investor's payoff is S − y + z and that of the Trustee is S + 3y − z where z denotes the final transfer from the Trustee to the Investor. The trust game is essentially a dictator game in which the Trustee dictates an allocation, with the difference, however, that the Investor's initial investment determines the amount to be shared.

In theory, self−interested Trustees will keep everything and repay z = 0. Self−interested Investors who anticipate this should transfer nothing, i.e., y = 0. In experiments in several developed countries, Investors typically invest about half the maximum on average, although there is substantial variation across subjects. Trustees tend to repay roughly y so that trust is not or only slightly profitable. The amount Trustees repay increases on average with y if the change in the Investors' transfer is sufficiently high; the Trustees do not necessarily pay back more if the increase in y is modest.

In the gift exchange game, there is again a Proposer and a Responder. The Proposer offers an amount of money w $\in [\underline{w},\overline{w}]$, $\underline{w} \geq 0$, which can be interpreted as a wage payment, to the Responder. The Responder can accept or reject w. In case of a rejection, both players receive zero payoff; in case of acceptance, the Responder has to make a costly "effort" choice e $\in [\underline{e},\overline{e}]$, $\underline{e} > 0$. A higher effort level increases the Proposer's monetary payoff but is costly to the Responder. A selfish Responder will always choose the lowest feasible effort level and will, in equilibrium, never reject any w. Therefore, if the Proposer is selfish and anticipates the Responder's selfishness the sub−game perfect proposal is the lowest feasible wage level $\underline{w}$. The main difference between the gift exchange game and the trust game is that in the trust game it is the first mover's action that increases the available surplus, while in the gift exchange game it is the second mover who can increase the

surplus.

The gift exchange game captures a principal-agent relation with highly incomplete contracts in a stylized way. Several authors have conducted variants of the gift exchange game.[3] All of these studies report that the mean effort is, in general, positively related to the offered wage which is consistent with the interpretation that the Responders, on average, reward generous wage offers with generous effort choices. However, as in the case of the ultimatum and the trust game, there are considerable individual differences among the Responders. While a sizeable share of Responders (frequently roughly 40 percent, sometimes more than 50 percent) typically exhibit a reciprocal effort pattern, a substantial fraction of Responders also always make purely selfish effort choices or choices which seem to deviate randomly from the self-interested action. Despite the presence of selfish Responders, the relation between average effort and wages can be sufficiently steep to render a high wage policy profitable which may induce Proposers to pay wages far above . Evidence for this interpretation comes from Fehr, Kirchsteiger and Riedl (1998), who embedded the gift exchange game into an experimental market.[4] In addition, there was a control condition where the experimenter exo-

---

3) See, e. g., Fehr, Kirchsteiger and Riedl (1993, 1998), Charness (1996, 2000), Fehr and Falk, (1999), Gächter and Falk (1999), Falk, Gächter and Kovacs (1999), Hannan, Kagel and Moser (1999), Brandts and Charness (2004) and Fehr, Klein and Schmidt (2004).

4) When interpreting the results of gift exchange games it is important to stress that - depending on the concrete form of the proposer's payoff function - gift exchange is more or less likely to be profitable for the proposer. In Fehr, Kirchsteiger and Riedl (1993, 1998), the proposer's payoff function is given by $xP = (v - w)e$ and effort is in the interval [0.1, 1]. With this payoff function the proposer cannot make losses and paying a high wage is less costly if the agent chooses a low effort level. In contrast, in Fehr, Klein and Schmidt (2004) the payoff function used is $xP = ve - w$ which makes it more risky for the principal to offer a high wage. Indeed, while paying high wages was profitable for the principal in the experiments of Fehr, Kirchsteiger and Riedl, it did not pay off in Fehr, Klein and Schmidt. This difference in performance is predicted by the theory of inequity aversion by Fehr and Schmidt (1999) that is discussed in more detail in section 3. For a further discussion of gift exchange games in com-

genously fixed the effort level. Note that the Responders can no longer reward generous wages with high effort levels in the control condition. It turns out that the average wage is substantially reduced when the effort is exogenously fixed.

The facts observed in the games mentioned above are now well established and there is little disagreement about them. However, questions remain about which factors determine and change the behavior in these games. For example, a routine question in discussions is whether a rise in the stake level will eventually induce subjects to behave in a self-interested manner. Several papers examine this question (Hoffman McCabe and Smith 1995, Fehr and Tougareva 1995, Slonim and Roth 1998, Cameron 1999); the surprising answer is that relatively large increases in the monetary stakes did little or nothing to change behavior. Hoffman, McCabe and Smith (1995) could not detect any effect of the stake level in the ultimatum game. Cameron (1999) conducted ultimatum games in Indonesia and subjects in the high stake condition could earn the equivalent of three months' income in this experiment. She observed no effect of the stake level on Proposers' behavior and a slight reduction of the rejection probability when stakes were high. Slonim and Roth (1998) conducted ultimatum games in Slovakia. They found a small interaction effect between experience and the stake level; the Responders in the high-stake condition (with a 10-fold increase in the stake level relative to the low stake condition) reject somewhat less frequently in the final period of a series of one-shot ultimatum games. Fehr and Tougareva (1995) conducted gift exchange games (embedded in a competitive experimental market) in Moscow. They did not observe an interaction effect between stake levels and experience. The subjects earned, on average, the equivalent amount of the income of one week in one of their conditions, while they earned the equivalent of a ten weeks' income in another condition. Despite this large difference in the stake size, neither the Proposers' nor the Responders' behavior shows significant differences across conditions.

Of course, it is still possible that there may be a shift towards more selfish behavior in the presence of extremely high stakes. However, the vast majority of

---

petitive environments see also section 5.3.

economic decisions for most people involve stake levels well below three months' income. Thus, even if other-regarding preferences played no role at all at stake levels above that size, these preferences would still play a major role in many economically important domains.

Another important question is to what degree the behavior of students is respresentative for the general population. All the experiments mentioned above were predominantly conducted with students as experimental subjects. Two representative data sets recently addressed this question - one from Germany (Fehr et al. 2002) and one from the Netherlands (Bellemare and Kröger 2003). In both cases, the authors conducted (modified) trust games and in both cases, certain demographic variables affected how the game is played, but these effects do not change the general pattern observed in the experiments with students. In particular, the trustees' back transfers are increasing in the investors' transfer and a large share (79 percent in the Fehr et al. study) of the trustees pays back money. Likewise, 83 percent of the investors transfer positive amounts; roughly 60 percent of them transfer 50% or more of their endowment. Moreover, the Proposers' and Responders' behavior remains constant, regardless of whether the players' endowment in the trust game is €10 or €100.

Among the demographic variables, age seems to be important. Both studies find that people above the age of 60 give less than middle-aged individuals when in the role of an investor. However, both studies also find that the elderly tend to give back more, ceteris paribus, when in the role of a trustee. Fehr et al. also report that subjects who experienced a divorce from their partner during the last year and people who favor none of the parliamentary parties in Germany (i.e. those who feel that they are not represented by the major political parties) pay back significantly less when in the role of a trustee. Furthermore, people who report that they are in good health give back significantly more. The most important result these studies provide, however, is that only very few individual level demographic variables seem to matter for behaviour. This suggests that it is possible to detect meaningful behavioral patterns with student subject pools that are representative for a more general subject pool, at least for the trust game.

To what extent does culture affect behaviour in these experiments? We de-

fine culture in terms of subjects' preferences and their beliefs about others' behavior. For example, in the context of the ultimatum game cultural differences may be reflected in different rejection rates for the same offer or in different beliefs about the rejection rate. In the past, many researchers took subjects' nationality as a proxy for culture. Nationality may be a very imperfect measure for culture in modern nations, however, because different cultures may coexist within the same country. Cohen and Nisbett (1994) provide evidence, for example, indicating that individuals who grew up in the American South have a culture of honour whereas Northerners do not have such a culture. Having said this, comparing subjects' behaviour across different continents may nevertheless yield interesting insights. Roth et al conducted ultimatum games in Japan, Israel, Slovenia, and the USA. Their results indicate somewhat lower rejection rates and lower offers in Japan and Israel compared to the US and Slovenia. Whereas the modal offers remain at 50% of the surplus throughout a ten period experiment with randomly assigned partners in the latter two countries, the modal offer converges to 40% in Israel and to two modes in Japan at 40% and 45%, respectively. The relatively low offers in Israel are also associated with relatively low rejection rates, indicating that a lower proposal in Israel was a rational choice for a self-interested proposer.

Buchan, Croson and Dawes conducted trust games in China, Japan, South Korea, and the USA. They find significant differences in investors' and in trustees' behaviour across countries. American and Chinese Investors transfer significantly more money than do their Japanese and Korean counterparts. Moreover, Chinese and Korean trustees send back a significantly higher proportion of their money than do American and Japanese subjects. Thus, Chinese subjects exhibit relatively high levels of trust (as indicator by investors' behaviour) and reciprocation (as indicated by trustees' behaviour) whereas Japanese subjects show relatively little trust and little reciprocation. The picture is more mixed for US and Korean subjects. Americans show a relatively high level of trust but a low level of reciprocation, whereas the Koreans show little trust but exhibit high levels of reciprocation.

The study by Henrich et al. (2001) documented the perhaps largest differences across cultures. This study reports the results of ultimatum game experi-

ments conducted in 15 small scale societies located in 5 different continents. The subjects in the cross cultural studies previously discussed were university students; one could therefore argue that, despite national differences, they all share much in common. They probably all have above-average skills, probably stem from higher income families and, perhaps most importantly, share an academic learning environment. This provides a sharp contrast to the Henrich et al study, where subjects come from vastly different cultures. For example, the Ache from Paraguay practice extreme forms of egalitarianism in which big game is shared equally among the tribe members. Others, like the Au and the Gnau from Papua New Guinea obey norms of competitive gift giving: accepting gifts, even unsolicited ones, obliges one to reciprocate at some future time to be determined by the giver. Acceptance of gifts also establishes a subordinate position between the giver and the receiver. Therefore, large gifts are frequently rejected in this society because of the fear associated with the unspecific commitments.

Henrich et al. observe vastly different proposer behaviour across cultures. For example, among the Machiguenga, who live in Peru, the average offer is only 26%, among the Gnau it is 38%, among the Ache it is 51%, while it even reaches 58% among the Lamelara, who are whale hunters on an Island in the Pacific Ocean. Likewise, there are also strong differences regarding rejection rates across several cultures. However, since most offers were around 50% in several societies, few rejections are observed, rendering the analysis of rejection behaviour impossible in these societies. Similar to the two representative studies in Germany and the Netherlands, only few, if any, individual level variables predict individual behaviour in the experiment. Two group level variables, however, explain a large share of the cross cultural variation in behaviour: the more the resources in a society are acquired through market trading and the higher the potential payoffs to group cooperation that are associated with the environment in which the society lives, the higher are the offers in the ultimatum game. For example, groups of 20 and more individuals have to cooperate in order to catch a whale and after the catch, they have to solve a difficult distribution problem: who gets which part of the whale. The Lamaleras have developed an extremely elaborate set of norms that determine in detail who gets what (Alvard 2004). These elaborate cooperation and

distribution practices may well spill over to the experimental context and induce subjects to make egalitarian offers. In contrast to the Lamelara, the Machiguenga in Peru exhibit little cooperation in production outside narrow family boundaries (Henrich and Smith 2004). They are also at the lower end of the spectrum with regard to market integration. It seems plausible that the absence of cooperation norms manifests itself in low offers in the ultimatum game. A third piece of telling evidence comes from the competitive gift giving societies in Papua New Guinea. Among the Au and the Gnau, a significant number of proposers offered more than 50% of the surplus, only to have these offers rejected in many cases. Thus, deeply seated social norms again seem to affect behaviour in the experiment.

2. Other-regarding Preferences or Irrational Behaviour

While there is now little disagreement regarding the facts reported above, there is still some disagreement about their interpretation. In Section 3, we will describe several recently developed theories of altruism, fairness, and reciprocity that maintain the rationality assumption but change the assumption of purely self-ish preferences. Although opinions about the relative importance of different mo-tives behind other-regarding behaviour differ somewhat (see section 4), it is prob-ably fair to say that most experimental researchers believe that some form of oth-er-regarding preferences exists. However, some interpret the behaviour in these games as elementary forms of bounded rationality. For example, Roth and Erev (1995) and Binmore, Gale and Samuelson (1995) try to explain the presence of fair offers and rejections of low offers in the ultimatum game with learning models that are based on purely pecuniary preferences, which assume that the rejection of low offers is not very costly for the Responders who therefore only learn very slowly not to reject such offers. The rejection of offers, however, is quite costly for the Proposers, who thus quickly realize that low offers are not profitable. Moreover, since Proposers quickly learn to make fair offers, the pressure on the Responders to learn to accept low offers is greatly reduced. This gives rise to very

slow convergence to the subgame perfect equilibrium - if there is convergence at all. The simulations of Roth and Erev and Binmore, Gale and Samuelson show that it often takes thousands of iterations until play comes close to the standard prediction.

In our view, there can be little doubt that learning processes are important in real life as well as in laboratory experiments. There are numerous examples where subjects' behaviour changes over time and it seems clear that learning models are prime candidates for explaining such dynamic patterns. We believe, however, that attempts to explain the basic facts in simple games, such as the ultimatum game, the third party punishment game, or the trust game, in terms of learning models that assume completely selfish preferences are misplaced. The Responders' decisions, in particular, are so simple in these games that it is difficult to believe that they make systematic mistakes and reject money or reward generous offers, even though their true preferences would require them not to do so. Moreover, the above cited evidence from Roth et al. (1991) Forsythe et al (1995), Suleiman (1996) and Fehr, Kirchsteiger and Riedl (1998) suggests that many Proposers anticipate Responders' actions surprisingly well. Thus, at least in these simple two-stage games, many Proposers seem to be quite rational and forward looking.

It is also sometimes argued that the behaviour in these games is due to a social norm (see, Binmore 1998, for example). In real life, so the argument goes, experimental subjects make the bulk of their decisions in repeated interactions. It is well known that the rejection of unfair offers or the rewarding of generous offers in repeated interactions can be sustained as an equilibrium among purely self-interested agents. According to this argument, subjects' behaviour is adapted to repeated interactions and they tend to apply behavioral rules that are appropriate in the context of repeated interactions erroneously to laboratory one-shot games.

We believe that this argument is half right and half wrong. The evidence from the cross-cultural experiments in 15 different small scale societies strongly suggests that social norms of cooperation and sharing have an impact on game playing behaviour. Indeed, the very fact that the behaviour in the experiment cap-

tures relevant aspects of real life behaviour is the main reason why such experiments are interesting; if they did not tell us something about how people behave in real life, the external validity of the experiments could be called into question. However, the fact that social norms affect subjects' behaviour in the experiment does not at all mean that they are inappropriately applying repeated game heuristics when they play one-shot games. In fact, the evidence suggests that subjects are well aware of the difference between one-shot interactions and repeated interactions where their reputation is at stake. Subjects in the experiments by Andreoni and Miller (1993), Engelmann and Fischbacher (2002), Gächter and Falk (2002), Fehr and Fischbacher (2003), Seinen and Schram (2000) exhibit much more cooperative behaviour or punish much more if the probability of repeatedly meeting the same subject increases or if they can acquire a reputation.

Fehr and Fischbacher (2003), for example, conducted a series of ten ultimatum games in two different conditions. Subjects played against a different opponent in each of the ten iterations of the game in both conditions. The proposers knew nothing about the past behaviour of their current responders in each iteration of the baseline condition. Thus, the responders could not build up a reputation for being "tough" in this condition. In contrast, the proposers knew the full history of their current responders' behaviour in the reputation condition, i.e., the responders could build up a reputation for being "tough". A reputation for rejecting low offers is, of course, valuable in the reputation condition because it increases the likelihood of receiving high offers from the proposers in future periods.

Therefore, if the responders understand that there is a pecuniary payoff from rejecting low offers in the reputation condition, one should generally observe higher acceptance thresholds in this condition. This is the prediction of an approach that assumes that subjects are rational and not only care for their own material payoff but also have a preference for punishing unfair offers: only the punishment motive plays a role in the baseline condition, while the punishment motive and the self interest motive influence rejection behaviour in the reputation condition. If, in contrast, subjects do not understand the logic of reputation formation and apply the same habits or cognitive heuristics to both conditions, there should be no observable systematic differences in responder behaviour across

conditions. Since the subjects participated in both conditions, it was possible to observe behavioral changes at the individual level. It turns out that the vast majority (slightly more than 80 percent, N = 72) of the responders increase their acceptance thresholds in the reputation condition relative to the baseline condition.[5] Moreover, the changes in rejection behaviour occur almost instantaneously when subjects move from the baseline condition to the reputation condition or vice versa. Thus, the data refutes the hypothesis that subjects do not understand the strategic differences between one-shot play and repeated play.

Therefore, instead of assuming that simple decisions that deviate systematically from self-interest reflect merely a form of erroneous application of rules of thumb, it seems more reasonable to assume that the prevailing social norms affect subjects' preferences. After all, the elaborate cooperation and distribution norms practiced by the whale hunters in Indonesia, or the gift giving norms among the Au and the Gnau in Papua New Guinea have been in place for decades if not centuries. They represent deep seated social practices that are likely to affect subjects' preferences. As these social practices are rather stable, the associated preferences inherit this stability. If a subject rejects a low offer in an anonymous one-shot ultimatum game because he or she is upset by the offer, the subject's emotional reaction to the situation probably drives the behaviour. Anger, after all, is a basic emotion and the prevailing fairness norms are likely to be reflected in the emotional response to a greedy offer. Recent papers by Fehr and Gächter (2002), Bosman and van Winden (2002) and Ben-Shakhar, Bornstein, Hopfensitz and van Winden (2004) provide evidence for the involvement of anger in punishment behaviour.

---

5) The remaining subjects, with one exception, exhibit no significant change in the acceptance threshold. Only one out of 70 subjects exhibits a significant decrease in the threshold relative to the baseline. Note that if a subject places a very high value on fairness, the acceptance threshold may already be very high in the baseline condition so that there is little reason to change the threshold in the reputation condition. Identical thresholds across conditions are, therefore, also compatible with a social preference approach. Only a decrease in the acceptance threshold is incompatible with theories of social preferences.

The view that emotions are important determinants of other-regarding behaviors, however, does not imply that these behaviors are irrational. If I feel bad if I let a greedy Proposer go unpunished, and if punishing him makes me feel good, I simply have a taste for punishing a greedy proposer. From a choice theoretic viewpoint, this taste does not differ from my taste for chocolate or lobster. In fact, there is strong experimental evidence suggesting that the demand for punishment increases if its price decreases (Eckel et al., Andreoni et al. in QJE, Putterman et al., Carpenter et al.). In addition, evidence from dictator games (Andreoni 2002) also shows that most subjects' preferences for giving in a dictator game obey the generalized axiom of revealed preferences, implying that the preferences can be represented by a utility function. Finally, Andreoni, Castillo and Petrie (2003) have shown that the responder's behaviour in a modified ultimatum game, in which the responder could shrink the available pie continuously, can be represented by convex fairness preferences.

The above arguments suggest that there is no reason for treating other-regarding preferences differently than other types of preferences. This means that we can apply the standard tools of economics and game theory to this area, enabling us to explain a great deal of behaviour in the games described above. For example, why do in Forsythe et al. (1995) the Proposers give so much less in the DG compared to the UG? Why do the Proposers in the control condition with exogenously fixed effort (Fehr, Kirchsteiger and Riedl 1998) make such low wage offers? Why do subjects punish less if the price of punishing is higher? Why do subjects reject higher offers if they can gain a reputation for being a tough bargainer compared to a situation where no reputation can be acquired? All these questions can be answered if one assumes that subjects are rational and care both for their own and others' payoffs. The problem with the alternative approach, which invokes some form of bounded rationality, is that at least so far it cannot explain these important behavioral variations across different games.

Most of the experiments that we consider in the rest of this paper are fairly simple. Therefore, we restrict attention in the following to approaches that maintain the assumption of rationality and ignore the potential role of learning.[6)]

6) There are a few models that combine other regarding preferences and learning, e.g. Cooper and Stockman (1999) and Costa-Gomes and Zauner (1999).

III Theories of Other-regarding Preferences

The experimental evidence sketched in Section 2 has provoked several theoretical attempts to explain the observed behaviour across different experiments within the rational choice framework. Three different approaches can be distinguished:

1. Models of "social preferences" assume that a player's utility function not only depends on his own material payoff, but may also be a function of the allocation of resources within his reference group, i.e. a player may also be concerned about the material resources other people receive. Furthermore, several models assume that people differ. Some people seem to be quite strongly concerned about how they compare to other people, while others seem to be mainly self-interested. Given these social preferences, all agents are assumed to behave rationally, meaning that the well known concepts of traditional utility and game theory can be applied to analyze optimal behavior and to characterize equilibrium outcomes in experimental games.
2. Models of "interdependent preferences" assume that people are concerned about their opponent's "type". Suppose that each player may be either a selfish type or a (conditionally) altruistic type. If an altruistic player knows that he interacts with another altruistic player, his preferences are altruistic and he is willing to be generous. If however, he knows that he deals with a selfish opponent, his preferences become selfish, too. Thus, whether player 1's preferences are altruistic or selfish depend on player

2's preferences and vice versa.

3. The third class of models deals with "intention based reciprocity". This approach assumes that a player cares about his opponent's intentions. If he feels that the opponent wanted to treat him kindly, he wants to return the favor and be nice to his opponent as well. If he feels that his opponent has hostile intentions, he wants to hurt his opponent. Thus, a player's interpretation of his opponent's behavior is crucial in this approach. Note that it is not the "type" of a player but rather his intention that is kind or hostile. Thus, in a given situation there may be an equilibrium in which a player has kind intentions, but there may also be a second equilibrium in which he has hostile intentions. Traditional game theory cannot capture this phenomenon; the framework of psychological game theory is needed.

Almost all models of these three approaches start out by making some fairly specific assumptions about the players' utility functions. Alternatively, one could start from a general preference relation and ask which axioms are necessary and sufficient to generate utility functions with certain properties. Axiomatic approaches are discussed at the end of this section.

Before we discuss the different approaches in detail, a word of caution is required. Many of the models under consideration here use terms such as "fairness", "equity", "altruism" or "reciprocity" that have been debated for a long time by moral philosophers and economists and that can be interpreted in different ways. Furthermore, some of these models are not entirely clear about what the domain of the theory is and what they want to achieve. In this section we will interpret all of these theories very restrictively. First of all, we view them as purely positive theories that try to explain actual human behavior. Thus, we disregard any normative implications the theories may have. Second, we view these models as first attempts to explain the outcomes of economic experiments. Typically, subjects enter these experiments as equals, they interact anonymously, and the physical outcome of the experiment is an allocation of monetary payoffs. Thus, for the experiments it is fairly straightforward to give a precise (and hopefully uncon-

troversial) definition of "altruistic preferences", "equitable allocation", "fair behavior" and the like. Of course, the theories discussed here do have implications for human behavior outside the laboratory as well. In some situations these implications may be very straightforward, but in general there are many important questions that have to be answered before the models can be applied to the "real world". This is a very important next step of this research agenda, but it will not be discussed here.

1. Social Preferences

Classical utility theory assumes that a decision maker has preferences over allocations of material outcomes (e.g. goods) and that these preferences satisfy some "rationality" or "consistency" requirements, such as completeness and transitivity. However, this fairly general framework is often interpreted much more narrowly in applications, by implicitly assuming that the decision maker only cares about one aspect of an allocation, namely the material resources that are allocated to her. Models of social preferences assume, in contrast, that the decision maker may also care about the material resources allocated to others.

Somewhat more formally, let $\{1,2,...,N\}$ denote a set of individuals and $x=(x1,x2,...,xN)$ denote an allocation of physical resources out of some set X of feasible allocations. For concreteness we assume in the following that xi denotes the monetary payoff of person i. The self-interest hypothesis says that the utility of individual i only depends on xi. We will say that individual i has social preferences if for any given xi person i's utility is affected by variations of xj, $j \neq i$. Of course, simply assuming that the utility of individual i may be any function of the total allocation is often too general because it yields very few empirically testable restrictions on observed behavior.[7] In the following we will discuss several models

7) One implication, however, is that if a decision maker can choose between two allocations then his decision should be independent on how the two allocations have been

of social preferences, each of which assumes that an individual's preferences depend on xj, j≠i, in a different way.

3.1. Altruism

A person is altruistic if the first partial derivatives of u(x1,...,xN) with respect to x1,...,xN are strictly positive, i.e., if her utility increases with the well being of other people. The hypothesis that (some) people are altruistic has a long tradition in economics and has been used to explain charitable donations and the voluntary provision of public goods.

Clearly, the simplest game for eliciting altruistic preferences is the dictator game (DG). Andreoni and Miller (2002) conducted a series of DG experiments in which one agent could allocate "tokens" between herself and another agent for a series of different budgets. The tokens were exchanged into money at different rates for the two agents and the different budgets. Let Ui(x1,x2) denote subject i's utility function representing her preferences over monetary allocations (x1,x2).

In a first step, Andreoni and Miller check for violations of the General Axiom of Revealed Preference (GARP) and find that almost all subjects behaved consistently and passed this basic rationality check. Thus, their preferences can be described by (quasi-concave) utility functions. Then Andreoni and Miller classify the subjects into three main groups. They find that about 30 percent of the subjects give tokens to the other party in a fashion that equalizes the monetary payoffs between players. The behavior of 20 percent of the subjects can be explained by a utility function in which x1 and x2 are perfect substitutes, i.e., these subjects seem to have maximized the (weighted) sum of the monetary payoffs. However, almost 50 percent of the subjects behaved "selfishly" and did not give any significant amounts to the other party. In a different experiment, they find that a

---

generated. This prediction is refuted by some experiments on variants of the ultimatum game, where the proposer either could or could not influence the allocation of resources. See e.g. Falk, Fehr and Fischbacher (2004) and Blount (1995) and the discussion in Sections 3.2 and 3.3 below.

sizeable minority (23 percent) of the subjects behaved spitefully by reducing their opponent's payoff if the opponent was better off then they were. Thus, they seem to have preferences that are non-monotonic in the monetary payoff of their opponent. Andreoni and Miller (2002, p.750) conclude that many individuals seem to have other-regarding preferences and that the individual choice behavior of subjects in dictator games is consistent with rationality. However, individuals are heterogeneous, and only a minority of subjects can be described as unconditional altruists who have a utility function that is always strictly increasing in the payoff of their opponent.[8)]

3.2. Relative Income and Envy

An alternative hypothesis is that subjects are not only concerned about the absolute amount of money they receive but also about their relative standing compared to others. The importance of relative income for a person's well being, of envy and jealousy, and of conspicuous consumption has long been recognized by economists and goes back at least to Veblen (1922).[9)]9 Bolton (1991) formalized this idea in the context of an experimental bargaining game between two players. He assumes that $Ui(xi,xj)=ui(xi,xi/xj)$, where $u(\cdot,\cdot)$ is strictly increasing in its first argument and where the partial derivative with respect to xi/xj is strictly positive for $xi<xj$ and equal to 0 for $xi \geq xj$. Thus, agent i suffers if she gets less than player j, but she does not care about player j if she is better off herself. Note that this utility function implies that $\partial Ui/\partial xj \leq 0$, just the opposite of altruism. Hence, while this utility function is consistent with the behavior in the bargaining games considered by Bolton, it neither explains generosity in dictator games and

8) Another, more specific model of heterogeneous altruistic preferences has been developed by Cox, Sadiraj and Sadiraj (2001). They assume that the marginal rate of substitution between own income and the income of the opponent depends on whose income is higher.

9) See e.g. Kolm (1995) for a detailed discussion and formalization of "envy" in economics.

kind behaviour of responders in trust games and gift exchange games nor voluntary contributions in public good games. The same problem arises in the envy-approach of Kirchsteiger (1994).

3.3. Inequity Aversion

The preceding approaches assume that utility is either monotonically increasing or monotonically decreasing in the well being of other players. Fehr and Schmidt (1999) assume that a player is altruistic towards other players if their material payoffs are below an equitable benchmark, but she feels envy when the other players' material payoffs exceed this level.[10] For most economic experiments it seems natural to assume that an equitable allocation is an equal monetary payoff for all players. Thus, inequity aversion reduces to inequality aversion in these games. Fehr and Schmidt consider the simplest utility function capturing this idea.

$$U_i(x_1, x_2, \cdots, x_N) = x_i - \frac{\alpha_i}{N-1}\sum_{j \neq i}\max x_j - x_i, 0 - \frac{\beta_i}{N-1}\sum_{j \neq i}\max x_i - x_j, 0$$

with $0 \leq \beta i \leq \alpha i$ and $\beta i \leq 1$. Note that $\partial Ui/\partial xj \geq 0$ if and only if $xi \geq xj$. Note also that the disutility from inequality is larger if another person is better off than player i than if another person is worse off ($\alpha i \geq \beta i$).

This utility function can rationalize positive and negative actions towards other players. It is consistent with generosity in dictator games and kind behaviour of responders in trust games and gift exchange games, and at the same time with the rejection of low offers in ultimatum games. It can explain voluntary contributions in public good games and at the same time costly punishments of free-riders.

A second important ingredient of this model is the assumption that individuals are heterogeneous. If all people were alike, it would be difficult to explain why we observe that people sometimes resist "unfair" outcomes or manage

---

10) Daughety (1994) and Fehr, Kirchsteiger and Riedl (1998) also assume that a player values the payoff of reference agents positively, if she is relatively better off, while she values the others' payoff negatively, if she is relatively worse off.

to cooperate even though it is a dominant strategy for a selfish person not to do so, while fairness concerns or the desire to cooperate do not seem to have much of an effect in other environments. Fehr and Schmidt show that the interaction of the distribution of types with the strategic environment explains why very unequal outcomes are obtained in some situations while very egalitarian outcomes prevail in others. For example, even a population that consists only of very fair types (high α's and β's) cannot prevent very uneven outcomes in certain competitive environments (see, e.g., the ultimatum game with proposer competition in Section 5.3) because none of the inequity averse players can enforce a more equitable outcome through her own actions. In contrast, a small fraction of inequity averse players in a public good game with punishment is sufficient to credibly threaten that free riders will be punished, inducing selfish players to contribute to the public good.

Fehr and Schmidt choose a distribution for αand β that is consistent with the experimental evidence of the ultimatum game. Keeping this distribution fixed, they show that their model yields surprisingly accurate predictions across many bargaining, market and social dilemma games.[11]

Bolton and Ockenfels (2000) independently developed a similar model of inequity aversion. They also show that their model can explain a wide variety of seemingly puzzling evidence such as generosity in dictator, gift exchange and trust games and rejections in the ultimatum game. In their model, the utility function is given by

$U_i = U_i(x_i, \sigma_i)$

where

---

11) One drawback of the piece-wise linear utility function employed by Fehr and Schmidt is that it implies corner solutions for some games where interior solutions are frequently observed. For example, a decision maker in the dictator game with a Fehr-Schmidt utility function would either give nothing (if her $\beta<0.5$) or share the pie equally (if $\beta>0.5$). Giving away a fraction that is strictly in between 0 and 0.5 is optimal only in the non-generic case where $\beta=0.5$. This problem can be avoided, at the cost of tractability, by assuming non-linear inequity aversion.

$$\sigma_i = \begin{cases} \dfrac{x_i}{\sum_{j=1}^{N} x_j} & \text{if } \sum_{j=1}^{N} x_j \neq 0 \\ \dfrac{1}{N} & \text{if } \sum_{j=1}^{N} x_j = 0 \end{cases}$$

For any given σi, the utility function is assumed to be weakly increasing and concave in player i's own material payoff xi. Furthermore, for any given xi, the utility function is strictly concave in player i's share of total income, σi, and obtains a maximum at σi=1/N.[12)]

Fehr-Schmidt and Bolton-Ockenfels often yield qualitatively similar results for two-player games, while some interesting differences arise with more than two players. Fehr and Schmidt assume that a player compares herself to each of her opponents separately in this case. This implies that her behavior towards an opponent depends on the income difference towards this person. In contrast, Bolton and Ockenfels assume that the decision maker is not concerned about each individual opponent but only about the average income of all players. Thus, whether ∂Ui/∂xj is positive or negative in the Bolton-Ockenfels model does not depend on j's relative position towards i, but rather on how well i does compared to the

12) This specification of the utility function has the disadvantage that it is not independent of a shift in payoffs. Consider, for example, a dictator game in which the dictator has to divide X Dollars. Note that this is a constant sum game because x1+x2 ≡ X. If we reduce the sum of payoffs by X, i.e., if the dictator can take away money from her opponent or give to him out of her own pocket, then x1+x2 = 0 for any decision of the dictator and thus we always have σ1=σ2=1/2. Therefore, the theory makes the implausible prediction that, in contrast to the game where x1+x2 = X > 0, all dictators should take as much money from their opponent as possible. Camerer (2003, p. 111) notes a related problem. Suppose that the ultimatum game is modified as follows: If the Responder rejects a proposal, the monetary payoffs are 10 percent of the original offer. In this case the relative shares are the same no matter whether the Responder accepts or rejects. Hence, Bolton and Ockenfels predict that the responder will always accept any offer, no matter how unequal it is. These problems do not arise in Fehr and Schmidt's model of inequity aversion.

average. If xi is below the average, then i would like to reduce j's income even if j has a much lower income than i herself. On the other hand, if i is doing better than the average, then she is prepared to give to j even if j is much better off than i.[13)]

2. Models of Intention based Reciprocity

The models considered so far do not allow for the possibility that players care about their opponents' intentions. I may be happy to be kind to my opponent if I believe that he intends to be kind to me - independent of what he actually does. In order to evaluate my opponent's intentions, I not only have to form beliefs about what he is going to do, but also about why he is going to do it. But in order interpret his behavior, I have to form beliefs about which actions my opponent believes I will take. Thus, for a given action of my opponent, it makes a difference for my utility payoff whether I believe that he takes this action because he believes that I will be kind to him or because he believes that I am going to hurt him. Traditional game theory cannot capture this, as it assumes that outcomes (and not beliefs) determine payoffs. However, Geanakoplos, Pearce and Stacchetti (1989) developed the concept of "psychological game theory" that generalizes traditional game theory by allowing for the possibility that payoffs are a function of players' beliefs. All models discussed in this subsection are based on psychological game theory.

13) See Camerer (2003, Section 2.8.5) and Section 4.1 for a more extensive comparison of these two approaches.

IV Discriminating between Theories of Other-regarding Preferences

Most theories discussed in Section 3 were developed during the last few years and the evidence to discriminate between these theories is still limited. As we will show, however, the available data do exhibit some clear qualitative regularities which give a first indication of the advantages and disadvantages of the different approaches.

1. Who are the Relevant Reference Actors?

All theories of other-regarding preferences are based on the idea that actors compare themselves with a set of reference actors or take these actors' payoffs directly into account. To whom do people compare themselves? Who are the relevant reference actors whose payoff is taken into account? There is no ambiguity about who the relevant reference actor is in bilateral interactions; the answer is less clear, however, in multi-person interactions. Most of the theories applicable in the n-person context assume that players make comparisons with all other n-1 players in the game. The only exemption is the theory of Bolton and Ockenfels (BO). They assume that players compare themselves only with the "average" player in the game and do not care about inequities between the other players. In this regard, the BO approach is inspired by the data of Selten and Ockenfels (1998) and Güth and van Damme (1998), which seem to suggest that actors do not care for inequities among the other reference agents. It would greatly simplify matters

if this aspect of the BO theory were correct.

One problem with this aspect of the BO approach is that it disenables the theory to explain punishment in the Third-Party Punishment Game. Recall that there are three players, A, B, and C in the third party punishment game. Player A is endowed with some surplus S and must decide how much of S to give to B, who has no endowment. Player B is just a dummy player and has no decision power. Player C is endowed with S/2 and can spend this money on the punish-ment of A after he observes how much A gave to B. For any money unit player C spends on punishment the payoff of player A is reduced by 3 units. Note that the total surplus available in this game is (3/2)S. Therefore, without punishment, player C is certain to get her fair share (S/2) of the total surplus, implying that the BO model predicts that C will never punish. In contrast to this prediction, roughly 60 percent of the C players punished in this game. This indicates that many players do care about inequities among other players. Further support for this hypothesis comes from Charness and Rabin (2002) who offered player C the choice between the payoff allocations (575,575,575) and (900,300,600). Because both allocations give player C the fair share of 1/3 of the surplus, the BO model predicts that player C will choose the second allocation which gives him a higher absolute payoff. However, 54 percent of the subjects preferred the first allocation. Note that the self-interest hypothesis also predicts the second allocation, so one cannot conclude that the other 46 percent of the subjects have BO-preferences. A recent paper by Zizzo and Oswald (2000) also strongly suggests that subjects care about the inequities among the set of references agents.

It is important to note that theories of other-regarding preferences, in which subjects have multiple reference agents, do not necessarily imply that the subjects take actions in favour of all other reference agents, even if all other reference agents have the same weight in their utility function. To illustrate this, consider the following three-person UG (Güth and van Damme 1998). This game includes a Proposer, a Responder who can reject or accept the proposal, and a passive Receiver who can do nothing but collect the amount of money allocated to him. The Proposer proposes an allocation (x1,x2,x3) where x1 is the Proposer's payoff, x2 the Responder's payoff and x3 the Receiver's payoff. If the Responder rejects,

all three players get nothing, otherwise the proposed allocation is implemented.

It turns out that the Proposers allocate substantial fractions of the surplus to the Responder in this game but little or nothing to the Receiver. Moreover, Güth and van Damme (p. 230) report that "there is not a single rejection that can clearly be attributed to a low share for the dummy (i.e., the Receiver, FS)". BO take this as evidence in favour of their approach because the Proposer and the Responder apparently do not take the Receiver's interest into account. However, this conclusion is premature because it is easy to show that approaches with multiple reference agents are fully consistent with the Güth and van Damme data. The point can be demonstrated in the context of the Fehr-Schmidt model. Assume for simplicity that the Proposer makes an offer of $x1=x2=x$ while the Receiver gets $x3<x$. It is easy to show that a Responder with FS-preferences will never (!) reject such an allocation even if $x3 = 0$ and even if he is very fair-minded, i.e., has a high β-coefficient. To see this note that the utility of the Responder if he accepts is given by $U2 = x - (\beta/2)(x - x3)$ which is positive for all $\beta \leq 1$, and thus higher than the rejection payoff of zero. A similar calculation shows that it takes implausibly high β-values to induce a Proposer to take the interests of the Receiver into account.[14)]

The above arguments suggest that the "average" player in a game is not an empirically relevant reference agent. This is particularly important for all games in which subjects may want to punish a particular individual for unfair or morally inappropriate behaviour. In all these cases, a model, in which the differences (or the ratio) between a player's own payoff and the group's average payoff is the driving force of the punishment, is not able to predict which individual will be punished. A player who just wants to reduce the difference between his payoff

---

14) The Proposer's utility is given by $U1 = x1 - (\beta/2)[(x1 - x2) + (x1 - x3)]$. If we normalize the surplus to one and take into account that $x1 + x2 + x3 = 1$, $U1 = (\beta/2) + (3/2)x1[(2/3) - \beta]$. Thus, the marginal utility of x1 is positive unless β exceeds 2/3. This means that Proposers with $\beta < 2/3$ will give the Responders just enough to prevent rejection and, since the Responders neglect the interests of the Receivers, nothing to the Receivers.

and the group's average payoff does not care about the target of the punishment. Any punishment that reduces this difference, even if it is targeted on cooperative or norm abiding individuals, is equally desirable from the perspective of such a player (see also Falk, Fehr and Fischbacher 2001).

In general, however, very little is known about the outcome of social comparison processes in games. Therefore, our empirical knowledge about what makes a player a relevant reference agent is very limited. The assumption that all players in a game are relevant reference agents to each other should only be taken as a first approximation and may not be true in some games. It seems reasonable to assume that player A is a relevant reference agent for player B if A can affect B's payoff in a salient way. However, there neither seems to be much theoretical work on this question nor persuasive empirical evidence beyond such general statements. Thus, the question "who are the relevant reference agents" is clearly an important unsolved problem.

2. Equality versus Efficiency

Many models of other-regarding preferences are based on the definition of a fair or equitable outcome to which people compare the available payoff allocations. In experimental games, the equality of material payoffs is a natural first approximation for the relevant reference outcome. The quasi-maximin theory of Charness and Rabin assumes instead that subjects care for the total surplus ("efficiency") accruing to the group. A natural way to study whether there are subjects who want to maximize the total surplus is to construct experiments in which the predictions of both theories of inequality aversion (BO and FS) are in conflict with surplus maximization. This has been done by Andreoni and Miller (2000), Bolle and Kritikos (1998), Andreoni and Vesterlund (forthcoming), Charness and Rabin (2002), Cox (2000) and Güth, Kliemt and Ockenfels (2000). Except for the Güth et al. paper, these papers indicate that a non-negligible fraction of the subjects in dictator game situations is willing to give up some of their

own money in order to increase total surplus, even if this implies that they generate inequality that is to their disadvantage. Andreoni and Miller and Andreoni and Vesterlund, for example, conducted dictator games with varying prices for transferring money to the Receiver. In some conditions, the Allocator had to give up less than a dollar to give the Receiver a dollar, in some conditions the exchange ratio was 1:1, and in some other conditions the Allocator had to give up more than one dollar. In the usual dictator games, the exchange ratio is 1:1 and there are virtually no cases in which an Allocator transfers more than 50 percent of the surplus. In contrast, in dictator games with an exchange ratio of 1:3 (or 1:2) a non-negligible number of allocators transfer in such a way that they end up with less money than the Receiver. This contradicts the models of Bolton and Ockenfels (2000), of Fehr and Schmidt (1999), and of Falk and Fischbacher (2005) because other-regarding subjects never take actions that give the other party more than they get in these models. It is, however, consistent with altruistic preferences or quasi-maximin preferences.

What is the relative importance of this kind of behavior? Andreoni and Vesterlund are able to classify subjects in three distinct classes. They report that 44 % of their subjects (N=141) are completely selfish, 35 percent exhibit egalitarian preferences, i.e. they tend to equalize payoffs, and 21 percent of the subjects can be classified as surplus maximizers. Charness and Rabin report similar results with regard to the fraction of egalitarian subjects in a simple Dictator Game where the Allocator had to choose between (own, other) allocations of (400, 400) and (400, 750). 31 percent of the subjects preferred the egalitarian and 69 percent the surplus maximizing allocation. Among the 69 percent there may, however, also be many selfish subjects who no longer choose the surplus-maximizing allocation when this decreases their payoff only slightly. This is suggested by the game where the Allocator had to choose between (400, 400) and (375, 750). Here only 49 percent of surplus-maximizing choices were observed. Charness and Rabin also present questionnaire evidence indicating that when the income disparities are greater the egalitarian motive gains weight at the cost of the surplus maximization motive. When the Allocator faces a choice between (400, 400) and (400, 2000), 62 percent prefer the egalitarian allocation.

More recently, Engelmann and Strobel (2004) argued that "efficiency" is an important motive that clearly dominates the desire for equality in 3 player dictator games. For example, the Allocator (who was always player B) could choose between 3 different payoff allocations in one of their games: (14, 4, 5), (11, 4, 6) and (8, 4, 7). Thus B's material payoff was the same in each of the three allocations, but he could redistribute income from the rich person to the poor person. Redistribution has a high efficiency cost in this game because it reduces the rich person's income by 3 units and increases the poor person's income by only 1 unit. Maximin preferences and selfish preferences cannot play a role in this game because the Allocator receives the lowest payoff regardless of the allocation chosen. This game allows, therefore, for a clean examination of how important the equality motive is relative to the "efficiency" motive. Engelmann and Strobel report that 60% of their subjects (N = 30) chose the first allocation, i.e., the one with the highest surplus and the highest inequality, and only 33% chose the most egalitarian allocation (8, 4, 7).

However, only students of economics and business administration, which we call for brevity "economists", participated in the Engelmann and Strobel study. These students learn from the very beginning of their studies that surplus maximization is normatively desirable. Therefore, Fehr, Naef and Schmidt (2004) replicated this game with N = 458 subjects to examine potential subject pool biases. They find a robust subject pool bias indicating that non-economists (N = 291) chose the most egalitarian allocation with the lowest surplus in 51% of the cases whereas economists' probability to choose this allocation was only 26% (N = 167). Likewise, the non-economists chose the least egalitarian allocation with the maximal surplus in only 28% of the cases, whereas the economists chose it in 56% of the cases. This result is also important with regard to the interpretation of the results of Charness and Rabin, who also have disproportionately many economists in their subject pool.

Since the evidence in favour of preferences for surplus maximization comes exclusively from dictator games, it is important to ask whether these preferences are likely to play a role in "strategic situations". We define strategic situations to be those in which the potential gift recipients are also capable of affecting the gift

givers' material payoffs. This question is important because the dictator game is different from many economically important games and real life situations, because one player is rarely at the complete mercy of another player in economic interactions. It may well be that in situations where both players have some power to affect the outcome, the surplus maximization motive is less important than in dictator games or is easily dominated by other considerations. The gift-exchange experiments by Fehr, Kirchsteiger and Riedl (1993, 1998) are telling in this regard because they embed a situation that is like a DG into an environment with competitive and strategic elements.

These experiments exhibit a competitive element because the gift exchange game is embedded into a competitive experimental market. The experiments also exhibit a strategic element because the Proposers are wage setters and have to take the Responders' likely effort responses into account. Yet, once the Responder has accepted a wage offer, the experiments are similar to a dictator game because, for a given wage, the Responder essentially determines the income distribution and the total surplus by his choice of the effort level. The gift exchange experiments are an ideal environment for checking the robustness of the surplus maximization motive because an increase in the effort cost by one unit increases the total surplus by five units on average. Therefore, the maximal feasible effort level is, in general, also the surplus maximizing effort level. If surplus maximization is a robust motive, capable of overturning preferences for equality or reciprocity, one would expect that many Responders choose effort levels that give the Proposer a higher monetary payoff than the Responder.[15] Moreover, surplus maximization also means that we should not observe a positive correlation between effort and wages because, for a given wage, the maximum feasible effort always maximizes the total surplus.[16]

---

15) The Responders' effort level may, of course, also be affected by the intentions of the Proposer. For example, paying a high wage may signal fair intentions which may increase the effort level. Yet, since this tends to raise effort levels, we would have even stronger evidence against the surplus-maximization hypothesis, if we observe little or no effort choices that give the Proposer a higher payoff than the Responder.

However, the data supports neither of these implications. Effort levels that give the Proposer a higher payoff than the Responder are virtually non-existent. In the overwhelming majority of the cases, effort is substantially below the maximally feasible level and the Proposer earns a higher payoff than the Responder in less than two percent of the cases.[17] Moreover, almost all subjects who regularly chose non-minimal effort levels exhibited a reciprocal effort-wage relation. A related result was observed by Güth, Kliemt and Ockenfels (2002) who also conducted experiments in which dictators face a trade-off between equality and surplus maximization. They report that equality concerns dominate surplus maximization concerns in the sense that dictators never perform transfers such that they earn less than the recipient, even if such transfers would be surplus enhancing. These results are in sharp contrast to the 49 percent of the Allocators in Charness and Rabin who preferred the (375, 750) allocation over the (400, 400) allocation. One reason for the difference across studies is perhaps the fact that it was much cheaper to increase the surplus in the Charness-Rabin example. While the surplus increases in the gift exchange experiments on average by five units, if the Responder sacrifices one payoff unit, the surplus increases by 14 units per payoff unit sacrificed in the Charness-Rabin case. This suggests that surplus maximization only gives rise to a violation of the equality constraint if surplus increases are extremely cheap. A second reason for the behavioural difference may be that when both players have some power to affect the outcome, the motive to increase the surplus is quickly crowded out by other considerations. This reason is quite plausible insofar as the outcomes in dictator games themselves are notoriously non-robust.

While the experimental results on ultimatum games are fairly robust, the dictator game seems to be a rather fragile situation in which minor factors can have large effects. Cox (2004), e.g., reports that 100 percent of all subjects transferred positive amounts in his dictator games.[18] This result contrasts sharply with

---

16) There are degenerate cases in which this is not true.

17) The total number of effort choices is N = 480 in these experiments, i.e., the results are not an artefact of a low number of observations.

many other games, including the games in Charness and Rabin and many other dictator games. To indicate the other extreme, Eichenberger and Oberholzer (1998), Hoffman, McCabe, Shachat and Smith (1994) and List and Cherry (2000) report on dictator games with extremely low transfers.[19] Likewise, in the impunity Game of Bolton and Zwick (1995), which is very close but not identical to a dictator game, the vast majority of Proposers did not shy away from making very unfair offers. The impunity Game differs from the dictator game only insofar as the Responder can reject an offer; however, the rejection destroys only the Responder's but not the Proposer's payoff. The notorious non-robustness of outcomes in situations resembling the dictator game indicates that one should be very careful in generalizing the results found in these situations to other games. Testing theories of other-regarding preferences in dictator games is a bit like testing the laws of gravity with a table tennis ball. In both situations, minor unobserved distortions can have large effects. Therefore, we believe that it is necessary to show that the same motivational forces that are inferred from dictator games are also behaviorally relevant in economically more important games. One way to do this is to apply the theories that were constructed on the basis of dictator game experiments to predict outcomes in other games. With the exemption of Andreoni and Miller (2002) this has not yet been done.

Andreoni and Miller (2002) estimate utility functions based on the results of their dictator game experiments and use them to predict cooperative behavior in a standard public goods game. They predict behaviour in period one of these games, where cooperation is often quite high, rather well. However, their predictions differ greatly from final period outcomes, where cooperation is typically very low. In

---

18) In Cox's experiment, both players had an endowment of 10 and the Allocator could transfer his endowment to the Receiver, where the experimenter tripled the transferred amount. The Receiver made no choice.

19) In Eichenberger and Oberholzer (1998), almost 90 percent of the subjects gave nothing. In Hoffman et al. (1992) 64 percent gave nothing and 19 percent gave between 1 and 10 percent. In List and Cherry subjects earned their endowment in a quiz. Then they played the DG. Roughly 90 percent of the Allocators transferred nothing to the Receivers.

our view, the low cooperation rates in the final period of repeated public good games constitutes a strong challenge for models that rely exclusively on altruistic or surplus-maximizing preferences. Why should a subject with a stable preference for others' payoffs or for those of the whole group contribute much less in the final period compared to the first period? Models of inequity aversion and intention based or type based reciprocity models provide a plausible explanation for this behaviour. All of these models predict that fair subjects make their cooperation contingent on the cooperation of others. Thus, if the fair subjects realize that there are sufficiently many selfish decisions in the course of a public goods experiment, they cease to cooperate as well (see also section 5 below).

3. Maximin Preferences

The papers by Charness and Rabin (2002) and by Engelmann and Strobel (2004) show that a substantial percentage of the Allocators in multi person dictator games care for the material payoff of the least well-off group member. The relevance of the maximin motive in these games is, for example, illustrated by the dictator game taken from Engelmann and Strobel (2004), in which player B is the dictator who can choose among the following three allocations: (11, 12, 2), (8, 12, 3) and (5, 12, 4). Both surplus maximization as well as the theories by Bolton and Ockenfels and Fehr and Schmidt predict that B will choose the first allocation in this game, whereas a player with maximin preferences chooses the third allocation. In fact, 53% of the players chose the third and only 27% chose the first allocation, indicating the importance of the maximin motive in these games. This game also shows, however, that nonlinear forms of inequity aversion may come close to maximin preferences. This is, for example, the case if the marginal disutility from advantageous inequality strongly increases in the amount of inequality. In this case also an inequity averse player may prefer the third allocation.

Although the maximin motive plays a prominent role in multi person dicta-

tor games, there are several papers that cast doubt on the relevance of this motive in strategic games. A salient example is the three-person experiment of Güth and van Damme (1998) that combines an ultimatum and a dictator game. Recall from Section 4.1 that the Proposer has to make a proposal (x,y,z) on how to allocate a given sum of money between himself and players two and three in this game. Then the Responder has to decide whether to accept or reject the proposal. If he accepts, the proposal is implemented, otherwise all players get zero. Player 3 remains inactive and cannot affect the final outcome. Güth and van Damme report that the Proposer allocates only marginal amounts to the passive Receiver and the Responder's rejection behaviour is seemingly unaffected by the low amounts allocated to the passive Receiver. These observations contradict maximin preferences while they are consistent with the linear Fehr and Schmidt model and the model by Bolton and Ockenfels (see Bolton and Ockenfels 2000 and Section 4.1).

Frechette, Kagel and Lehrer (2003) provide another striking example of the neglect of the weak player's interests in strategic interactions. One player in a group of five can make a proposal on how to allocate a fixed sum of money among the five players in their experiments. Then the players vote on the proposal under the majority rule, i.e., the support of 3 players is sufficient to implement the proposal. In 65% of the cases, the proposals implied that two of the five players received a zero payoff, completely neglecting the interests of members that are not part of the winning coalition. Moreover, such proposals received the support of the majority in most cases. Thus, maximin preferences seem do play little role in this environment.

Finally, the experiments by Okada and Riedl (2005) also indicate that maximin preferences are of little importance in strategic games. In their three person experiments, a Proposer could propose an allocation (x, y) to one Responder or an allocation (x, y, z) to two Responders. If he proposes forming a three person coalition, i.e., making an offer to two Responders, the total amount to be distributed among the three players is 3000 points whereas if he only proposes a two person coalition, the total amount to be distributed is an element of the set {1200, 2100, 2500, 2800}. However, both Responders have to accept the proposal (x, y, z) in the case of a three person coalition, whereas only a single Responder has to

accept the proposal (x, y) in the case of the two person coalition. If one of the Responders rejects a proposal, all players receive zero. If only the two person coalition is proposed, the third player automatically receives a payoff of zero. Therefore, Proposers with maximin preferences that dominate their self-interest will always propose a three person coalition with x=y=z, regardless of the amount available for the two person coalition. In the case of quasi maximin preferences in the sense of Charness and Rabin (2002) the "efficiency" motive puts even more weight on this proposal because the grand coalition produces a larger surplus.

Okada and Riedl report that 90% of the Proposer's went for the two-person coalition when the total amount available for the two person coalition is 2500 or 2800. If the available amount for the small coalition is only 2100 still about 40% of the Proposers went for the two person coalition. The grand coalition is favoured by almost all proposers only in those cases when the small coalition became very inefficient because the available amount shrank to 1200. These regularities in Proposers' behaviour are predicted by the Fehr and Schmidt and the Bolton and Ockenfels model of inequity aversion.

Given the evidence from the above mentioned papers, it remains to be shown that maximin preferences play a role in strategic games. It seems that dictator games put players in a different frame of mind than strategic games, where the players can mutually affect each others' payoffs. Players in strategic games seem to be much more willing to neglect weak players' interests and to demand fairness or equity mainly for themselves, whereas the dictators seem to care a lot for the interests of the worst-off players in dictator games. This insight may also help in determining when the maximin motive plays a role in naturally occurring environments. In a competitive environment or in an environment where the players view each other as agents behaving strategically, the maximin motive is likely to be not important. However, the maximin motive may be more or even highly relevant in the context of charitable giving or in the context of referenda or elections with a large number of people, where strategic voting is unlikely to occur.

4. Summary

Although most models of other-regarding preferences discussed in Section 3 are just a few years old, the discussion in this section shows that there is already a fair amount of evidence that sheds light on the merits and the weaknesses of the different models. This indicates a quick and healthy interaction between experimental research and the development of new theories. The initial experimental results discussed in Section 2 gave rise to a number of new theories which, in turn, have again been quickly subjected to careful and rigorous empirical testing. Although these tests have not yet led to conclusive results regarding the relative importance of the different motives many important and interesting insights have been obtained. In our view the main results can be summarized as follows:

1) The average payoff in the group is an empirically invalid reference standard for explaining punishment individual behaviour. Approaches that rely on this comparison standard cannot explain important aspects of punishment behaviour. Evidence from the Third Party Punishment Game and other games indicates that many subjects compare themselves with other people in the group and not just to the group as a whole or to the group average.
2) Pure revenge as captured by intention based and type based reciprocity models is an important motive for punishment behaviour. Since pure equity models do not capture this motive they cannot explain a significant amount of punishment behaviour. While the inequality of the payoffs also is a significant determinant of payoff reducing behaviour, the revenge motive seems to be more important in bilateral interactions as illustrated in those experiments where responses to a computerized first-mover choice are compared to the responses to human first mover choices.
3) In the domain of kind behaviour, the motives captured by intention or type based models of reciprocity seem to be less important than in the domain of payoff-reducing behaviour. Several studies indicate that inequity aversion or maximin preferences play a more important role here.
4) In dictator games, a significant share of the subjects prefers allocations with a

higher group payoff and a higher inequality within the group over allocations with a lower group payoff and a lower inequality. However, this motive only dominates among economists, while the clear majority of non-economists is willing to sacrifice substantial amounts of the group payoff in order to ensure more equality within the group. Moreover, the relative importance of the motive to increase the group's payoff has yet to be determined for strategic games.

5) In multi person dictator games, a large share of the subjects cares for the least well-off player's material payoff. However, evidence from several strategic games casts doubt on the relevance of this motive in strategic interactions.

6) Some recent papers report that a substantial share of the subjects has indicated a preference for honesty.

Which model of other-regarding preferences does best in the light of the data, and which should be used in applications to economically important phenomena? We believe that it is too early to give a conclusive answer to these questions. There is a large amount of heterogeneity at the individual level and any model has difficulties in explaining the full diversity of the experimental observations. The above summary provides, however, some guidance for applied research. In addition to the summary statements above, we believe that the most important heterogeneity in strategic games is the one between purely selfish subjects and subjects with a preference for fairness or reciprocity.

Within the class of inequity aversion models, the evidence suggests that the Fehr and Schmidt model outperforms or does at least as well as the Bolton and Ockenfels model in almost all games considered in this paper. In particular, the experiments discussed in Section 4.1 indicate that people do not compare themselves with the group as a whole but rather with other individuals in the group. The group average is less compelling as a yardstick for measuring equity than are differences in individual payoffs. However, the Fehr and Schmidt model clearly does not recognize the full heterogeneity within the class of fair-minded individuals. Section 4.4 makes it clear that an important part of payoff-reducing behaviour is not driven by the desire to reduce payoff-differences, but by the desire to reduce the payoff of those who take unfair actions or reveal themselves as

unfair types. The model therefore cannot explain punishing behaviour in situations where payoff differences cannot be changed by punishing others. Fairness models exclusively based on intentions (Rabin 1993, Dufwenberg and Kirchsteiger 2004) can, in principle, account for this type of punishment. However, these models have other undesirable features, including multiple, and very counterintuitive, equilibria in many games and a very high degree of complexity due to the use of psychological game theory. The same has to be said about the intention based theory of Charness and Rabin (2002). It is also worthwhile to point out that intention based reciprocity models cannot explain punishment in the third party punishment game because they are based on bilateral notions of reciprocity. The third party was not treated in an unkind way in this game and will therefore never punish. Falk and Fischbacher (1999) do not share these problems of pure intention models. This is due to the fact that they incorporate equity as a global reference standard. Their model shares however, the complexity costs of psychological game theory.

Even though none of the available theories of other-regarding preferences takes the full complexity of motives at the individual level into account, some theories may allow for better approximations than others, depending on the problem at hand. If, for example, actors' intentions constitute a salient dimension of an economic problem, consideration of some form of intention based reciprocity might be advisable, despite the complexity costs involved. Or, to give another example, a type based reciprocity model in the spirit of Levine (1998) may provide a plausible explanation for third party punishment. The essence of third party punishment is that the punisher is not directly hurt but nevertheless punishes a norm violation. While bilateral notions of reciprocity are unable to explain this kind of punishment type based models provide a natural explanation because norm violations are type revealing. However, the most important message of the evidence presented in Section 2 clearly is that there are many important economic problems where the self-interest theory is unambiguously, and in a quantitatively important way, refuted. Therefore, in our view, it is certainly not advisable to only consider the self-interest model, but to combine the self-interest assumption with the other-regarding motive that is likely to be most important in the problem at hand.

V Economic Applications: Fairness and Food Choice

1. Introduction

A wealth of evidence indicates that people are often motivated by altruism, reciprocity, and concerns for inequity (e.g., see Andreoni and Miller, 2002; Güth et al., 1982; Fehr et al., 1993). Such findings have led to the development of models which incorporate other-regarding preferences into the traditional economic framework. Such models assume that in addition to their own payoffs, people also care about others' payoffs. Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), for examples, have proposed models of self-centered inequality aversion in which people are motivated by their own payoff and their payoff relative to the payoffs of others.[20] Other models suggest people also exhibit preferences for efficiency and maximizing minimum payoffs (Charness and Rabin, 2002).

Although such "fairness" models have recently gained much attention, there have only been a few attempts to determine the relative explanatory power of these theories (e.g., see Bereby-Meyer and Niederle, 2005; Engelmann and Strobel,

20) The word "inequality" is used in many ways throughout the literature. Although the term often refers to disparities in distribution of income or wealth, in this paper we adopt the definition used in the experimental/behavioral economics literature, in which "inequality" refers to the distribution of relative payoffs or game outcomes (see Fehr and Schmidt (1999) or Bolton and Ockenfels (2000)).

2004; Fehr et al., 2006; Bolton and Ockenfels, 2006). Such comparisons have primarily been limited to abstract, experimental games devoid of naturally occurring field context. Furthermore, the vast majority of such experiments have been conducted with a convenient sample of student subjects. It is unclear whether and to what extent other-regarding behavior will hold up when the decision context is moved to a more natural setting or when money allocations are no longer anonymous (e.g., see the findings in List, 2006).

We consider the explanatory power of several competing models of other-regarding preferences in a field context for which all people have experience: food choice. Recent years have witnessed pronounced differentiation of food products, ranging from organic to "eco-friendly" to "hormone free", and to genetically modified (GM) food products, and a number of academic studies have analyzed such issues. For example, Kolodinsky (2008) examined consumers' preferences for rBST-free and organic milk and Costa-Font et al. (2008) reviewed a number of studies related to consumer preferences for GM foods. The reasons for growth in food product differentiation are multi-faceted, but a factor that may have some explanatory is altruism or other-regarding preferences. For examples, Sunding (2003) argued that people are motivated to purchase some products because of an "associated public good," and Umberger et al. (2009) showed that about a fifth of the willingness to pay premium for "natural, regionally produced" beef was explained by people's concerns for societal benefits such as supporting local agriculture and environmental benefits.

The recent growth in the organic food market, in particular, has been partially attributed to people's concerns about inequity in the food supply chain and motivations to support small farmers. Indeed, one of key principles of organic agriculture is the concern for fairness, which emphasizes the relationships to between all parties in the food chain - farmers, processors, distributors, traders and consumers (e.g., see Alrøe et al., 2006; Danish Research Center for Organic Farming, 2000; International Federation of Organic Agriculture Movements, 2008; Padel et al., 2009). To what extent are people's preferences for organic food driven by concerns for "fairness" and distribution of outcomes versus, say, environmental and food safety concerns? Unfortunately, existing literature provides scant evidence

on this issue. Despite the arguments by some that "fairness" and support for small farms is a key principle of organic agriculture (Padel et al., 2009), we are aware of only one previous empirical study investigating a link between fairness motivations and consumer demand for organic food; Lusk and Briggeman (2008) found that stated willingness-to-pay a premium for organic food was positively correlated with the food value of "fairness."

In this paper, we seek to determine whether food choice is influenced by concerns for "fairness" and to determine the extent to which the fairness models proposed in the literature explain food purchasing behavior. Moreover, we elicit consumers' perceptions about the distribution of benefits resulting from the sales of non-organic and organic food and determine the extent to which preferences for the distribution of benefits can explain preferences for organic food. Our results indicate that the fairness models proposed in the literature do not exhibit much explanatory power unless modified in nontrivial ways. Whereas the proposed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) assume people dislike inequality as it relates to others' payoffs relative to their own payoff, our data suggests that in the context of food people dislike inequality as it relates to others' payoffs relative to the payoffs to small farmers. Finally, we find that preferences for distribution of benefits, along with measured beliefs about the relative distribution on benefits accruing to producers of organic and conventional foods, are significant factors explaining consumer willingness to pay a premium for organic food.

2. Background

2.1. Fairness Models

In an attempt to study the predictions of game theory, psychologists and experimental economists constructed simple bargaining games, and when the results failed to match the Nash equilibrium predictions, the assumptions of ration-

ality and pure self-interest were called into question. In one of the earlier studies on the issue, Güth et al. (1982) devised the so-called Ultimatum game, where a proposer makes an offer to split a sum of money with an anonymous responder. If the responder accepts the offer, the pie is divided as suggested by the proposer; if the responder rejects the offer, neither party gets any money. Güth et al. (1982) found a tendency for proposers to offer half the pie, which contradicted the game theoretic prediction in which the proposer would offer the smallest possible share of the pie and responder would accept. As it turns out, the results were quite robust; in the hundreds of replications of the Ultimatum game, proposers typically offer about 40-50% of the allocation and responders typically reject offers less than 20% of the pie (Camerer, 2003).

Because a variety of factors could explain such high offers in the Ultimatum game, researchers sought to isolate the role of other-regarding preferences by studying so-called Dictator and gift-exchange games. The Dictator game, introduced by Forsythe et al. (1994), is like the Ultimatum game but it strips the responder of the right to reject; the proposer simply dictates the share of the pie the responder receives. A strictly selfish proposer would offer none of the available surplus, but results suggest that proposers transfer, on average, about between 20 to 25% of the pie (Forsythe et al., 1994).

In response to these findings, several theoretical models were developed to describe the apparent observed pro-social behavior observed in different types of experiments (e.g., see Fehr and Schimdt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Erlei, 2008; Konow, 2000). These models distinguish between two classes or types of pro-social preferences: distributive and reciprocal preferences. Distributive preferences refer to the fact that people have preferences over the final allocation of consequences or outcomes, and include preferences such as equity, altruism, and efficiency, whereas reciprocal preferences refer to the fact that people are concerned about others' types or intentions, meaning that an agent's response depend upon other's actions. In our application, and in the remainder of the paper, we focus on distributive preferences.

Among these social preference models, perhaps the most popular and widely discussed is Fehr and Schmidt's (1999) theory of inequality aversion, which as-

sumes people derive disutility from unequal outcomes. Their model is two dimen-sional in the sense that one parameter characterizes disutility from others being better off than self (disadvantageous inequality) and another parameter character-izes disutility from others being worse off than self (advantageous inequality). It is commonly assumed that people are more averse to disadvantageous inequality than to advantageous inequality. Bolton and Ockenfels (2000) proposed another model which assumed people are averse to inequality. In their model, people de-rive disutility when their payoff diverges from the average payoff to all parties.

A prominent difference between the Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) models relates to the reference point against which people compare themselves. People are assumed to comparisons their own payoff with all the other people's payoff separately in Fehr and Schmidt (1999), but in Bolton and Ockenfels (2000) disutility is caused by differences between one's own payoff and the average payoff irrespective of distribution of payoffs. Thus, a person with Fehr and Schmidt (1999) preferences might favor income or payoff redistribution even if their own payoff is unaltered, but a person with Bolton and Ockenfels (2000) preferences is unconcerned about redistribution so long as they receive the average payoff.

In addition to inequality concerns, other models of distributive preferences suggest people have preferences for the total surplus generated by a game, i.e., "efficiency." For example, Charness and Rabin (2002) proposed a quasi-maximin preference model, in which it was assumed that people care about aggregate sur-plus and always prefer a higher sum of payoff for all subjects. In support of the model, Andreoni and Miller (2002) and Andreoni and Vesterlund (2001) show that people are willing to give up some of their payoff in order to increase or maximize the total payoff.

Engelmann and Strobel (2004) conducted some simple distributional experi-ments in which people indicated their preference between several choice options that differed in terms of the payouts going to self and two other anonymous individuals. Engelmann and Strobel (2004) sought to distinguish between the in-equality aversion models of Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) and the efficiency and maximin concerns proposed by Charness and Rabin

(2002). Although some support was found for all the models, Engelmann and Strobel (2004) concluded that efficiency concerns exhibited the most explanatory power.

To our knowledge, there has been no previous attempt to determine whether people's food choices are affected by concerns for inequality, efficiency, or mini-mum payoffs. This paper represents an attempt to apply what has been learned in abstract experimental games, and what has been theorized as a result, to the problems of food demand.

2.2. Organic Foods

With growing popularity of organic foods, a great deal of attention has been devoted to uncover consumers' motivations for choosing organic over con-ventional food. Environmental and health concerns are often considered to be among the most important factors influencing the purchase of organic foods (e.g., Durham and Andrade, 2005; Gracia and de Magistris, 2008; Huang, 1996; Torjusen et al., 2001; Verhoef, 2005). Other empirical studies have found that other quality attributes such as taste, freshness, appearance and animal welfare significantly affect consumers' decisions of whether to buy organic food products (e.g., Kuhar and Juvancic, 2005; Harper and Makatouni, 2002; Makatouni, 2002; Milock et al., 2004).

In addition to these more tangible product attributes, others have argued that social aspects of organic farming, such as supporting local agriculture, played a role in explaining why people prefer organic over conventional foods (e.g., Padel and Foster, 2005). In particular, with the industrialization of agriculture, and the rise of so-called "factory farms," many have express concern about the dis-tribution of outcomes that have resulted. In particular, the concept of fairness has been outlined as one of the four core principles of organic agriculture by the International Federation of Organic Agriculture Movements (e.g., Alrøe et al., 2006; Padel et al., 2009). Moreover, recent findings by Lusk and Briggeman (2009) have indicated that food value of "fairness" was positively correlated with consumers' stated willingness to pay premiums for organic food.

This study seeks to identify whether and to what extent concerns for "fairness" can explain people's preferences for organic food using the conceptual framework discussed in the previous sub-section.

3. Methods and Procedures

3.1. Survey Design

A mail survey was developed and sent to a random sample of 2,000 consumers in the United States in April 2007.[21] The survey contained a stated-preference experiment in which people were asked to respond to a series of purchase intention questions. In particular, respondents were asked to indicate how likely they were to purchase 12 loaves of bread that differed by price and the amount of profit from the purchase going to the following participants in the food marketing channel: small farmers (defined as farming less than 500 acres), large farmers (defined as farming 500 acres or more), agribusiness processors (such as wheat millers and bakers), and supermarkets.[22] For each loaf of bread, people

21) The mailing list was purchased from a reputable survey research company which randomly selected names from the white pages of the telephone directory. In designing the survey, we followed the guidance and suggestions in Dillman (2000). The survey questionnaire was mailed out with a personalized cover letter, and the mailing included a prepaid return envelope. One week after the survey was send, a reminder postcard was mailed out.

22) To define a "small" farm, a variety of concepts could have been used such as agricultural scale, type of organization (e.g., family farms versus corporation), agricultural sales, level of resource endowments, etc. For example, the USDA-ERS uses $250,000 in agricultural sales as the cutoff between small and large farms as does the National Commission on Small Farms. We were reluctant to use this definition, however, because people might have confused gross sales with net income. In this paper, we adopt a common, size-based approach and set the criteria for small farms as an operation farming less than 500 acres. According to the most recent Census of

were asked to indicate how likely they were to buy the loaf on a scale of 0 to 10, where 0 was defined as "definitely would not buy," 5 was defined as "equal chance of buying and not buying," and 10 was defined as "definitely would buy." Across the 12 bread options, prices were varied among the values of $1.99, $2.99, and $3.99 and the profits accruing to each of the participants in the food marketing channel was varied among the values of $0.01, $0.07, and $0.15. Thus, the experimental design consisted of price being varied at three levels and the profits accruing to the four participants in the production of bread (small farmers, large farmers, agribusiness processors, and supermarkets) each being varied at three levels, creating 35=243 possible types of bread. Each survey contained 12 descriptions of bread that were randomly selected from the full set of 243 (note: no two surveys were alike; each survey had a different set of 12 bread options). An example of two bread options presented to one respondent is shown in figure 1.

Figure 1. Example Survey Questions

Product	Definitely Would Not Buy				Equal Chance of Buying and Not Buying						Definitely Would Buy
Price of bread loaf: **$2.99** Profit to small farmers: **$0.01** Profit to large farmers: **$0.15** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10
Price of bread loaf: **$1.99** Profit to small farmers: **$0.15** Profit to large farmers: **$0.01** Profit to agribusinesses: **$0.01** Profit to grocery store: **$0.15**	0	1	2	3	4	5	6	7	8	9	10

These stated preference questions were designed to determine people's preferences for profits accruing to different participants in the food supply chain.

Agriculture, the average farm size in the U.S. is 418 acres and about 85% of U.S. farms operate less than 500 acres. Thus, we set the threshold of 500 acres to define small farms for the purposes of this survey.

Another goal of this study is to determine the relationship between people's preferences for participants in the food supply chain and people's willingness-to-pay (WTP) a premium for organic vs. non-organic bread. As such, the survey also elicited people's beliefs about the distribution of benefits across the supply channel resulting from the sale of organic and non-organic loaves of bread. People were asked to indicate how much they thought each of the following participants in the supply chain, as a whole, profited from the sale of a single loaf of organic and non-organic bread: small farmers, large farmers, agribusiness processors, and supermarkets. Participants responded by checking a box with competing dollar amounts between $0.01 to $0.05, $0.06 to $0.10, and $0.11 to $0.15.

To determine the extent to which the measured preferences for distributions of outcomes across the food supply chain can explain people's food preferences, the survey contained a question where people were asked to indicate the maximum premium they would be willing to pay for an organic loaf of bread over a conventional loaf of bread, assuming both were the same brand name. Finally, as will be explained in more detail in a subsequent sub-section, to determine the extent to which fairness concerns motivate people's willingness-to-pay a premium for organic bread, several additional questions were asked. In particular, respondents were asked to state: (i) how much they would normally expect to pay for a single loaf of conventional and organic bread (i.e., expected market prices) and (ii) the most they would be willing to pay for a conventional and organic loaf of bread if no other market participant (farmers, agribusinesses, or grocery store) made any profit and they were the only one that benefited from the purchase.

3.2. Fairness Models and Econometric Methods

To illustrate the general modeling approach, first ignore preferences for equality/efficiency and assume that individual i's utility from the purchase a loaf option j is:

$$U_{ij} = \alpha_j - \beta P_{ij} + \delta_1 \pi_{SF,ij} + \delta_2 \pi_{LF,ij} + \delta_3 \pi_{AB,ij} + \delta_4 \pi_{GS,ij}, \quad (1)$$

where P_{ij} is the price of the j^{th} loaf of bread, π_{SF}, π_{LF}, π_{AB}, π_{GS} are profits to

each of the members of the food supply chain (small farmers, large farmers, agri-business processors, and grocery stores, respectively), a_j is an alternative specific constant related to the utility of having and not-having a loaf of bread, β denotes the marginal utility of income, and δk describe people's preferences for the k^{th} type of business. One interpretation of the δ_k parameters are that they represent a measure of pure altruism; concern for the financial well-being of another than cannot be explained by concerns for inequality, efficiency, etc.

Because the fairness models introduced by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) involve a comparison of benefits to self with benefits to others, it is useful to re-write (1) to determine the net benefit a consumer receives from the purchase of bread option *j*. In particular, we calculate the "selfish" consumer surplus of an option by finding the price level, P, that makes a consumer indifferent to the buying when $\delta_1=\delta_2=\delta_3=\delta_4=0$, meaning all other parties in the supply chain receive no benefit. Without loss of generality, assume people are indifferent to buying and not buying when $U_{ij}=0$. Thus, willingness-to-pay for the loaf of bread is determined as $WTP=a_j/\beta$ or $a_j=\beta WTP$. Substituting this expression into equation (1) and rearranging yields:

$$U_{ij} = \beta(WTP - P_{ij}) + \delta_1\pi_{SF,ij} + \delta_2\pi_{LF,ij} + \delta_3\pi_{AB,ij} + \delta_4\pi_{GS,ij}, \quad (2)$$

where $(WTP-P_{ij})$ represents the consumer's benefit or consumer surplus from the purchase of option *j*. It is important to note that, following the Train and Weeks (2005), *WTP* is a coefficient directly estimatable from the respondent's choices, and as such, the consumer benefit (or consumer surplus) from purchasing an option can be determined as the difference between estimated willingness-to-pay and price. Given this framework, we let πC,ij denote the ith consumer's benefit or "profit" from bread option *j*, $\pi_{C,ij}=(WTP-P_{ij})$.

To determine the extent to which the fairness considerations explain food purchase behavior, we modify equation (2) as follows:

$$U_{ij} = \beta\pi_{C,ij} + \delta_1\pi_{SF,ij} + \delta_2\pi_{LF,ij} + \delta_3\pi_{AB,ij} + \delta_4\pi_{GS,ij} + \lambda Fair_{ij}, \quad (3)$$

where the variable $Fair_{ij}$ corresponds to a particular notion of fairness or inequality aversion proposed in the literature and described below.

First, we consider the model of inequality aversion advocated by Fehr and Schmidt (1999, henceforth FS). One portion of their inequality aversion motive can be captured as follows:

$$FSa_{ij} = -\frac{1}{4}[\max(\pi_{SF,ij} - \pi_{C,ij}, 0) + \max(\pi_{LF,ij} - \pi_{C,ij}, 0) + \max(\pi_{AB,ij} - \pi_{C,ij}, 0) + \max(\pi_{GS,ij} - \pi_{C,ij}, 0)] \quad (4)$$

If FSa_{ij} in equation (4) is substituted into the variable $Fair_{ij}$ in equation (3), then λ provides a measure of people's aversion to disadvantageous inequity. The FS model also proposes that people are averse to being in an advantageous position as well. Thus, the other inequality aversion motive can be captured as follows:

$$FSb_{ij} = -\frac{1}{4}[\max(\pi_{C,ij} - \pi_{SF,ij}, 0) + \max(\pi_{C,ij} - \pi_{SF,ij}, 0) + \max(\pi_{C,ij} - \pi_{AB,ij}, 0) + \max(\pi_{C,ij} - \pi_{GS,ij}, 0)] \quad (5)$$

If FSb_{ij} in equation (5) is into substituted the variable $Fair_{ij}$ into equation (3), then λ provides a measure of people's aversion to advantageous inequity. If one in-cludes a linear term for the payoff to each participant in the supply chain, as in equation (3), then the advantageous and disadvantageous motives are not sepa-rately identified, i.e., there is a linear identity between the linear payoff terms and the sum of the advantageous and disadvantageous measures in equations (4) and (5). Thus, following Engelmann and Strobel (2004), we investigate the ability of the FS model to explain behavior by substituting the following expression into equation (3)

$$Fair_{ij} = FSa_{ij} + FSb_{ij}. \quad (6)$$

Secondly, we consider the model of equity, reciprocity, and competition (ERC) proposed by Bolton and Ockenfels (2000, henceforth ERC) which assumes people dislike a difference between their own payoff and the average payoff. The ERC motivation can be expressed as

$$Fair_{ij} = ERC_{ij} = -100 \times \left| \frac{1}{5} - \frac{\pi_{C,ij}}{EFF_{ij}} \right|, \quad (7)$$

where EFF_{ij}=, which represents the sum of profits to all five participants in the

supply chain, including the consumer. Equation (7) shows that people prefer the average profit to be as close as possible to their own profit. The key contrast between the FS and ERC models is that in the ERC model, people are happy if they receive the average payoff regardless of whether some people earn more or less, but the same is not necessarily true in the FS model. In the FS model, individual i can be made better off by re-distributing payoffs.

The third model we consider is motivated by Charness and Rabin (2002), who posit that people are motivated, in part, by total surplus or efficiency. Thus, to implement the Charness and Rabin (2002) model, we simply substitute the following expression into equation (3):

$$Fair_{ij} = EFF_{ij}. \tag{8}$$

Finally, we explore an intuitive notion of inequality aversion, and investigate whether people are concerned about the standard deviation of profits across all supply chain participants,

$$Fair_{ij} = SD_{ij} = -\sqrt{\frac{1}{5}\sum_{k}(\pi_{kj} - EFF_{ij}/5)^2} \tag{9}$$

Because the standard deviation provides a depiction of profit variability, it can be interpreted as an alternative motivation for inequality aversion.

The question we now consider is how to empirically estimate the parameters of the model in equation (3). First, we adopt the random utility framework popularized by McFadden, and assume that the indirect utility function given in (3), while known by the respondent, is only observable to the analyst with error. Second, given that the question format requested people to respond on a cardinal scale (where the rating directly corresponds to the chance of making a purchase - i.e., 0=0% chance of making a purchase, 1=10% chance of making a purchase, etc.), we assume that an ordinary least squares model is appropriate, where the person's rating is a proxy for the utility derived from an option, i.e., $U_{ij}=Rating_{ij}+\varepsilon_{ij}$. Third, because each respondent was asked to answer 12 questions regarding how likely they were to buy a loaf of bread, the random errors are unlikely to be independent within-subject. Stated differently, unobservable hetero-

geneity is likely to be present across subjects. Indeed, Erlei (2008) recently pointed out the importance of heterogeneity of preferences and showed that it plays an important role in understanding laboratory behavior. Taken together, these considerations imply the following empirical model:

$$Rating_{ij} - 5 = \beta(WTP - P_{ij}) + \delta_1 \pi_{SF,ij} + \delta_2 \pi_{LF,ij} + \delta_3 \pi_{AB,ij} + \delta_4 \pi_{GS,ij} + \lambda Fair_{ij} + u_i + \epsilon_{ij} \quad (10)$$

where $Fair_{ij}$ is one of the measures in equations (6), (7), (8), or (9), u_i ~N(0,σu²) is a subject-specific random effect, and ε_{ij} ~N(0,σ_ε^2) is an over all random error term. Note that in (10), the constant 5 has been subtracted from the dependent variable. This step was taken so that the estimated value, *WTP*, corresponds to the dollar amount that makes people indifferent to buying a loaf of bread when all other parties receive profit equal to zero (by construction, a rating of 5 implies indifference toward buying, i.e., 50% chance of buying).

To test the relative performance of the four fairness models outlined above, i.e., equations (6), (7), (8), or (9), we consider how well the models performed in predicting out of sample by using cross-validation. In particular, the sample was randomly split in half, and each of the four fairness models was estimated using one half of the data. Then we investigated how well the estimated models predicted the "hold out" sample, which was the other half of the data not used in the estimation. This process was then repeated by switching the estimation and hold-out samples. To judge out-of-sample prediction performance, two model selection criteria were used: mean squared error (MSE) and the out-of-sample log likelihood function (OSLLF) approach. The MSE is simply the average of the squared difference between the estimated rating and the actual rating of the desirability of each bread option. A model with a lower MSE is preferred. The OSLLF method ranks models by likelihood function values observed at out of sample observations. The OSLLF selects the model with the highest out-of-sample log likelihood function value and is calculated as:

$$OSLLF = \sum_{i=1}^{N/2} \sum_{j=1}^{J} \log \Phi(Z_{ij}), \quad (11)$$

where $\Phi(Z_{ij}) = \frac{1}{\sqrt{2\pi\phi}} exp - \frac{(Rating_{ij}^{H} - Rating_{ij}^{E})^2}{2\phi}$ is the probability density function for an out-of-sample observation, is the actual rating of option j in the "hold-out" or out-of-sample dataset, and is the predicted rating of option j using parameters obtained by fitting a model to the "estimation" dataset.

3.3. Estimating the Fairness-Induced Willingness-to-Pay Premium for Organic Food

In addition to investigating consumer concerns for distribution of benefits across the food supply chain, it is also of interest to determine the extent to which such considerations explain people's preferences for organic food over conventional food. Once the coefficients for a given fairness model have been estimated, they can be combined with people's stated WTP, people's perceived prices for organic and non-organic bread, and people's beliefs about the profit levels for each participant in the supply chain to calculate the predicted utility for organic and non-organic bread:

$$\widehat{U_i^O} = \hat{\beta}(WTP_i^O - P_i^O) + \hat{\delta_1}\pi_{SF,i}^O + \hat{\delta_2}\pi_{LF,i}^O + \hat{\delta_3}\pi_{AB,i}^O + \hat{\delta_4}\pi_{GS,i}^O + \hat{\lambda}Fair_i^O \quad (12)$$

and

$$\widehat{U_i^C} = \hat{\beta}(WTP_i^C - P_i^C) + \hat{\delta_1}\pi_{SF,i}^C + \hat{\delta_2}\pi_{LF,i}^C + \hat{\delta_3}\pi_{AB,i}^C + \hat{\delta_4}\pi_{GS,i}^C + \hat{\lambda}Fair_i^C, \quad (13)$$

where the O and C superscripts denote organic and conventional bread, respectively, WTP_i^k is individual i's stated willingness-to-pay for the k^{th} type of bread when no other party in the food supply chain profits, P^{ik} is i's stated belief about what they would pay for the k^{th} type of bread in the grocery store, and where $\pi_{t,i}^{k}$ is the i's stated belief about the profits accruing to the t^{th} business type for the k^{th} type of bread. With equations (12) and (13), the premium people are willing to pay for organic bread can be determined by finding the price difference between organic and conventional bread, (P^O-P^C), that makes a person in-

different to purchasing organic or conventional non-organic bread:

$$(P_i^O - P_i^C) = (WTP_i^O - WTP_i^C) - \frac{1}{\hat{\beta}}[\hat{\delta_1}(\pi_{SF,i}^C - \pi_{SF,i}^O) + \hat{\delta_2}(\pi_{LF,i}^C - \pi_{LF,i}^O) + \hat{\delta_3}(\pi_{AB,i}^C - \pi_{AB,i}^O) + \hat{\delta_4}(\pi_{GS,i}^C - \pi_{GS,i}^O) + \hat{\lambda}(Fair_i^C - Fair_i^O)] \quad (14)$$

This price difference can be interpreted as the estimated premium consumers are predicted to be willing to pay for organic bread over conventional non-organic bread. This measure can be decomposed into two parts. The first term in the right hand side of the equality (14) represents the organic premium motivated by concerns unrelated to the distribution of payouts accruing to parities in the food supply chain such as concerns about the environment, health, quality, and etc., and the second term represents the organic premium explained solely by concerns for payoffs accruing to other participants in the supply chain. Thus, we can investigate the extent to which preferences for the distribution of outcomes drive WTP for organic vs. non-organic bread. The portion of the willingness-to-pay premium for organic bread explained by payoff-induced motives is:

$$\frac{-\frac{1}{\hat{\beta}}[\hat{\delta_1}(\pi_{SF,i}^C - \pi_{SF,i}^O) + \hat{\delta_2}(\pi_{LF,i}^C - \pi_{LF,i}^O) + \hat{\delta_3}(\pi_{AB,i}^C - \pi_{AB,i}^O) + \hat{\delta_4}(\pi_{GS,i}^C - \pi_{GS,i}^O) + \hat{\lambda}(Fair_i^C - Fair_i^O)]}{(P_i^O - P_i^C)}$$

Finally, to investigate external validity, we simply calculate the correlation coefficient between the predicted willingness-to-pay premium for organic bread given in equation (14) and the person's actual stated willingness-to-pay premium for organic bread obtained in the survey.

4. Results

Overall 207 completed surveys were returned. After accounting for undeliverable addresses, a response rate of 11.5% is implied. Although the response rate is somewhat low, each individual answered twelve rate questions, producing 2,484 observations for our analysis. This implies a sampling error of less than

3%. Despite the fact that only 207 individuals responded, we emphasize that we were able to obtain data from a much more diverse subject pool than what would have been the case had a convenience sample of students been used. For example, 60% of respondents were female (which is likely a result of the fact that we asked the primary food shopper in the household to complete the survey), the mean age of the respondents was about 56 years old (with a standard deviation of 15 years), 55% of the sample had earned a bachelor's degree, 17% had children under the age of 12 in the household, and only a small fraction of the sample (15%) said they or someone in their immediate family farmed or ranched for a living. Importantly, we do not claim that our estimates of preferences for relative payoffs are representative of the U.S. population per se but rather ask, for this sample of people, whether they exhibit preferences for the distribution of payoffs across the food supply chain and test whether these measured preferences related the willingness-to-pay a premium for organic food. Nevertheless, low response rate may be indicative of non-response bias. To address this issue, summary statistics of our sample were compared with data from the Current Population Survey of the U.S. Census Bureau (see table 1). In general, the characteristics of our sample correspond reasonably well with the population in terms of gender, annual income, and location while it differs by age and education.[23)]

23) One way to account for differences in the sample and the population is to create sample weights to force the sample to mimic the population in terms of selected demographic characteristics. To investigate whether our results are unduly influenced by non-representativeness, we have constructed sample weights for our data using iterative proportional fitting techniques. We have estimated all the models reported in this paper with these post-stratification weights applied to the data, but the estimation results are very similar to the unweighted results in every cases. As such, we simply report the results from the unweighted models. Results from the weighted models can be obtained from the authors on request.

Table 1. Characteristics of Survey Respondents as Compared to U.S. Population

Category		Population Proportions[a]	Sample Proportions
Gender	Female	51.6%	59.9%
Age	18−34 years	30.7%	8.0%
	35−44 years	19.2%	15.5%
	45−54 years	19.5%	26.7%
	55−64 years	14.5%	21.4%
	65 + years	16.2%	28.3%
Education	No Bachelor's degree	73.8%	45.4%
	Bachelor's degree or higher	26.2%	54.5%
Income[b]	Less than $20,000	19.0%	6.4%
	$20,000 to $99,999	60.8%	69.0%
	$100,000 or more	20.2%	24.6%
Region[c]	Northeast	18.1%	16.6%
	Midwest	22.0%	31.6%
	South	36.6%	32.1%
	West	23.2%	19.8%

[a] Percent of population from the U.S. Census.

[b] Annual household income.

[c] U.S. Census regions.

Table 2 reports estimates for each of four fairness models. For each model, except for the FS model, the coefficient β is positive, meaning the marginal utility of income is positive and consumers care about their own benefit or "profit." All model specifications indicate people are willing to pay about $1.45 for a loaf of bread assuming no other participant in the supply chain benefits from the sale. The "altruism" coefficients for payoffs to small farmers, δ_1, are positive and statistically significant in each model, meaning people primarily care about the benefits to small farmers.

In addition to these "altruism" parameters, the key results relate to the estimate of the parameter λ, which correspond to the various fairness concerns. The

only fairness motivation that was statistically significant was in the ERC model, but here the coefficient was of the opposite sign than expected - i.e., people preferred receiving payouts the diverged from the average payout. These findings indicate that, after holding constant the payoffs to each participant in the supply chain, choices are either unaffected by concerns for inequity or efficiency or are affected in non-intuitive ways.

Nevertheless, because the models suggest people care about the payouts to small farmers (i.e., altruism), beliefs about payoff differences might explain preferences for organic bread if people believe small farmers benefit from selling organic bread. The bottom portion of table 2 shows the premium for organic bread over conventional bread which results from payouts accruing to different parties in the food supply chain. The portion of premium explained by differences in relative payouts ranges from 39.7% to 48.8%. One might question why these values are so large when none of the fairness parameters are statistically significant. The answer is because people care about small farmers (and the coefficients for this particular participant are large), and because people perceive small farms to derive a large benefit from organic foods. To illustrate this, table 3 shows people's beliefs about the payoffs to different participants in the supply chain for organic and non-organic bread resulting from the sale of a single loaf of bread. As can be seen in table 3, people believe all participants in the food supply chain benefit more from selling organic bread than conventional bread, but small farmers are believed to benefit more than others. The bottom portion of table 2 also shows that for the SD and EFF models, that the predicted willingness-to-pay premium for organic bread is positively and significantly correlated with people's actual stated willingness-to-pay.

In general, table 2 suggests that aside from altruism toward small farmers, the fairness models proposed in the literature have very little explanatory power in the food choice context. One possible reason is that unlike simple distributional games, a person's own payoff is much less transparent when making a purchase as it is a result of consumer surplus - the difference between WTP and price. This lack of transparency in determining one's own benefit may cause people to be less sensitive to comparisons of self vs. others than is assumed in the models proposed

in the literature. As such, we consider whether modifications to the models might improve fit. In particular, we revised the fairness motivations expressed in equations (5), (6), (7), and (8), by excluding profits to self, $\pi_{C,ij}$, and assume that people in the FS and ERC models are concerned about inequity as it relates to small farmers, the group that table 1 indicates respondents were most concerned about. For example, the original *FS* model in equation (5) is modified to assuming that people dislike differences between small farmers' profits and profits earned by others in the food supply chain, as shown below
where

$$FSa_{ij}^{M} = -\frac{1}{3}[\max(\pi_{LF,ij} - \pi_{SF,ij}, 0) + \max(\pi_{AB,ij} - \pi_{SF,ij}, 0), + \max(\pi_{GS,ij} - \pi_{SF,ij}, 0)]$$

and

$$FSb_{ij}^{M} = -\frac{1}{3}[\max(\pi_{SF,ij} - \pi_{LF,ij}, 0) + \max(\pi_{SF,ij} - \pi_{AB,ij}, 0), + \max(\pi_{SF,ij} - \pi_{GS,ij}, 0)]$$

where FSa_{ij}^{M} and FSb_{ij}^{M} represent dislike from disadvantageous and advantageous inequity based on the deviation from small farmers' profit, not consumer surplus. In the same manner, ERC motive in equation (6) can be rewritten as

$$ERC_{ij}^{M} = -100 \times \left| \frac{1}{4} - \frac{\pi_{SF,ij}}{EFF_{ij}^{M}} \right|,$$

where $EFF_{ij}^{M} = \pi_{SF,ij} + \pi_{LF,ij} + \pi_{AB,ij} + \pi_{GS,ij}$.

Table 2. Estimates of Competing Fairness Models

Parameters	Models			
	SD	ERC	FS	EFF
β	1.596	1.280**	−2.404	2.261**
	(1.888)[a]	(0.055)	(3.343)	(0.841)
WTP	1.475**	1.422**	1.474**	1.474**
	(0.145)	(0.019)	(0.145)	(0.145)
δ_1	14.298**	13.981**	15.301**	15.301**
	(0.956)	(0.824)	(1.180)	(1.180)
δ_2	−0.993	−1.344	−	−
	(0.936)	(0.835)		
δ_3	−0.554	−0.465	0.538	0.538
	(1.226)	(0.491)	(0.959)	(0.959)
δ_4	−1.499	−1.768**	−0.491	−0.491
	(0.975)	(0.817)	(1.170)	(1.170)
λ[b]	−0.604	−0.001*	3.732	−0.933
	(4.253)	(0.000)	(3.345)	(0.836)
	2.729**	2.734**	2.729**	2.729**
	(0.312)	(0.312)	(0.312)	(0.312)
Portion[c]	0.410	0.488	0.397	0.432
Correlation[d]	0.324**	−0.026	0.104	0.251**
	(0.000)	(0.727)	(0.160)	(0.001)
MSE[e]	9.609	9.596	9.610	9.609
OSLLF[f]	−5340.913	−5339.478	−5341.041	−5341.046
No. of Obs.	2,484	2,484	2,484	2,484
No. of Respondents	207	207	207	207

Note: * and ** represents statistical significance at the 10% and 5% levels, respectively.

[a] Numbers in parentheses are asymptotic standard errors.

[b] SD = −standard deviation(self, small farmer, large farmer, agribusiness, grocery store),
EFF = self + small farmer + large farmer + agribusiness + grocery store,
ERC = $-100\times|(1/5)-(self/EFF)|$,
FS = FSa + FSb = −1/4[max(small farmer−self, 0) + max(large farmer−self, 0) + max(agribusiness−self, 0) + max(grocery store−self, 0)] −1/4[max(self−small farmer, 0) + max(self−large farmer, 0) + max(self−agribusiness, 0) + max(self−grocery store, 0)].

[c] Numbers are the trimmed mean of portion of estimated people's premiums on organic over conventional that result solely from fairness concerns (part B) versus other factors, such as safety, health, or environmental concerns (part A) by discarding the five lowest and highest values.

[d] Correlation between calculated people's premium for organic versus conventional and stated people's willingness−to−pay for organic.

[e] MSE is mean squared error between predicted and stated rate.

[f] OSLLF is the estimated likelihood function value observed at stated rate values.

Table 3. Beliefs about the Distribution of Profits across the Food Supply Chain Resulting from the Sale of a Single Organic and Non-Organic Loaf of Bread

Supply Chain Participants	Conventional Non-Organic	Organic	Difference in Organic and Non-Organic	Percent Increase from Non-Organic to Organic
Small farmers	$0.059	$0.073	$0.014	23.73%
Large farmers	$0.079	$0.089	$0.010	12.66%
Agribusiness	$0.089	$0.094	$0.005	5.62%
Grocery store	$0.100	$0.108	$0.008	8.00%

The results of these modified models are reported in table 4. According to the MSE and OSLLF criteria, all models in table 4 exhibit better out-of-sample predictive performance than their respective counterparts in table 2. Furthermore, the estimate parameters related to fairness concerns, λ, are now all statistically significant and of expected sign. For SD model, the coefficient on the standard deviation of profits across supply chain excluding consumers' profits is 5.655, meaning consumer prefer an equal distribution of profits among the agents in marketing channel. That λ in the ERC model is positive indicates consumers dis-like small farms receiving payoffs that differ from the average payoff to the four other business types, and a positive λ in FS model indicates that participants dis-like the payoffs to small farms diverging from payoffs to large farms, agribusi-nesses, and grocery stores. Finally, the positive coefficient for λ in the EFF model variable implies that, all else equal, people prefer higher total profits to the four businesses. In each of the four models shown in table4, 38.8% to 42.3% of total willingness-to-pay premium for organic foods can be explained by consumers' concerns for the distribution of profits. Importantly, the correlation between the estimated premium and people's stated willingness to pay premium is positive and statistically significant for all four models. Comparing the relative performance of the models indicates that, overall, the modified FS model provides the best fit to the data.

Table 4. Estimates of Modified Fairness Models

Parameters	Models			
	SD	ERC	FS	EFF
β	1.340**	1.326**	1.333**	1.328**
	(0.058)[a]	(0.057)	(0.057)	(0.058)
WTP	1.587**	1.756**	1.644**	1.474**
	(0.146)	(0.162)	(0.147)	(0.145)
δ_1	14.565**	14.280**	14.686**	13.815**
	(0.819)	(0.817)	(0.820)	(0.709)
δ_2	−0.854	−1.717**	−0.896	−1.486**
	(0.835)	(0.864)	(0.833)	(0.713)
δ_3	1.350*	−0.620	0.789	−0.948*
	(0.736)	(0.500)	(0.574)	(0.492)
δ_4	−1.113	−2.040**	−1.136	−1.977**
	(0.827)	(0.840)	(0.823)	(0.706)
λ[b]	5.655**	0.016**	5.727**	0.553*
	(1.764)	(0.005)	(1.406)	(0.301)
	2.741**	2.731**	2.738**	2.729**
	(0.313)	(0.312)	(0.312)	(0.312)
Portion[c]	0.405	0.423	0.415	0.388
Correlation[d]	0.300**	0.306**	0.286**	0.315**
	(0.000)	(0.000)	(0.000)	(0.000)
MSE[e]	9.557	9.592	9.537	9.609
OSLLF[f]	−5334.278	−5338.985	−5331.771	−5341.041
No. of Obs.	2,484	2,484	2,484	2,484
No. of Respondents	207	207	207	207

Note: * and ** represents statistical significance at the 10% and 5% levels, respectively.

[a] Numbers in parentheses are asymptotic standard errors.

[b] SD = −standard deviation(small farmer, large farmer, agribusiness, grocery store),
EFF = small farmer + large farmer + agribusiness + grocery store,
ERC = $-100\times|(1/4)-(\text{small farmer/EFF})|$,
FS = FSa + FSb = $-1/3$[max(large farmer−small farmer, 0) + max(agribusiness−small farmer, 0) + max(grocery store−small farmer, 0)] $-1/3$[max(small farmer−large farmer, 0) + max(small farmer−agribusiness, 0) + max(small farmer−grocery store, 0)].

[c] Numbers are the trimmed mean of portion of estimated people's premiums on organic over conventional that result solely from fairness concerns (part B) versus other factors, such as safety, health, or environmental concerns (part A) by discarding the five lowest and highest values.

[d] Correlation between calculated people's premium for organic versus conventional and stated people's willingness−to−pay for organic.

[e] MSE is mean squared error between predicted and stated rate.

[f] OSLLF is the estimated likelihood function value observed at stated rate values.

5. Conclusions

This study investigated the extent to which several "pro-social" models can explain people's food choices. We also sought to determine whether estimated concerns for others' payoffs can explain people's willingness to pay a premium for organic food. Using data from a mail survey administered to U.S. households, we found that although people exhibit altruistic preferences toward small farmers, the existing models of inequality aversion and efficiency proposed in the literature to not exhibit much explanatory power. The models exhibited much better explanatory power, when we modified the models to account for the fact that a person's own surplus is less transparent in a food purchasing context than in simple distributional games. The modified Fehr and Schmidt (1999) model, in which people are assumed to receive disutility from inequity relative to small farmers, provided the best out-of-sample prediction performance.

Our results also indicate that other-regarding preferences explain a non-trivial portion of people's willingness-to-pay a premium for organic food, and that the estimated models of other-regarding behavior exhibit reasonably high external validity as they are significantly related to people's stated willingness-to-pay a premium for organic food. Our findings indicate that concerns for inequality can be observed in a field context, though not in the same manner as in simple laboratory distributional experiments.

These results have implications for marketers and policy makers alike. For example, although fairness is often listed as one of the core principles of the organic movement (Padel et al., 2009) government labeling regulations in the U.S., for example, to not make any provision for fairness concerns. That is, a product can receive an organic label or certification regardless of the extent to which principles such as "fairness," "equity," or "justice" are violated. This may be one reason that many "food activists" have turned their back on the organic movement in the U.S. and are beginning to channel their efforts toward local foods and community supported agriculture. Indeed, if one wants to maximize benefits to small farmers, this can easily be accomplished by purchasing from farmers markets or

other direct marketing channels.

Of course, just because the U.S. standards for organic certification do not mandate particular distributional outcomes, this does not prevent some consumers from believing that organic foods yield distributions of outcomes they deem desirable. Indeed, our results show that, on average, people perceive that as compared to conventional foods, organic foods generate a larger benefit (i.e., bigger pie) to all participants in the food supply chain and that small farmers receive a larger share of the benefits (i.e., a bigger piece of the pie). For many food products, such beliefs are likely inaccurate and some food manufacturers and retailers may utilize or exploit such beliefs to their own benefit. This would suggest that consumers suffer from a cost of ignorance - i.e., they would make alternative choices if they knew the characteristics they were actually buying - and might suggest there is value in information provision or altering organic certification standards to bring standards in line with people's beliefs.

VI Conclusions

The self-interest hypothesis assumes that all people are exclusively motivated by their material self-interest. This hypothesis is a convenient simplification and there are, no doubt, situations in which almost all people behave as if they were strictly self-interested. In particular, for comparative static predictions of aggregate behaviour self-interest models may make empirically correct predictions because models with more complex motivational assumptions predict the same comparative static responses. However, the evidence presented in this paper also shows that fundamental questions of social life cannot be understood on the basis of the self-interest model. The evidence indicates that other-regarding preferences are important for bilateral negotiations, for the enforcement of social norms, for understanding the functioning of markets and economic incentives. They are also important determinants of cooperation and collective action and the very existence of cooperative institutions that enforce rules and norms may be due to the existence of other-regarding preferences. The examples that we have given in Section 5 of this chapter do of course not exhaust the potential impact of such preferences on economic and social processes. We did not mention the impact of other-regarding preferences on voting behaviour and the demand for redistribution (Tyran 2004, Ackert et al. 2004, Hahn 2004 a,b, Fong 2005), on contract choices (Güth, Klose, Königstein and Schwalbach 1998, Anderhub, Gächter and Königstein 2002, Cabrales and Charness 2004, Fehr, Klein and Schmidt 2004), on the hold up problem (Hacket 1994, Ellingsen and Johannesson 2004), on the optimal allocation of ownership rights (Fehr, Kremhelmer and Schmidt 2004) on trust (Bohnet and Zeckhauser 2004) and how they may undermine explicit incentives (Bohnet,

Frey and Huck 2001, Gneezy 2003, Fehr and Rockenbach 2003, Fehr and List 2004, Falk and Kosfeld 2004, Irlenbusch and Sliwka 2003). This long list of examples suggests that other-regarding preferences affect social and economic life in many domains. If they are neglected social scientists run the risk of providing incomplete explanations of the phenomena under study or - in the worst case - their explanations may be wrong.

However, although in view of the prevailing modelling practices in economics it is natural to emphasize the existence of a substantial share of subjects with other-regarding preferences, one should not forget the fact that many subjects often show completely selfish behaviours. Moreover, many of the examples we have discussed in Section 5 show that the interaction between self-interested actors and actors with other-regarding preferences may play a key role for the understanding of the outcomes of many experiments. Depending on the strategic environment selfish actors may induce actors with other-regarding preferences to behave as if completely selfish but the converse is also often true: actors with other-regarding preferences induce selfish actors to change their behaviour in fundamental ways. In order to fully understand the interaction between selfish and non-selfish actors, social scientists need rigorous formal models of other-regarding preferences. In Section 3 we have documented the current state of the art in this domain. While the current models clearly present progress relative to the self-interest approach the evidence reported in Section 4 also makes it clear that further theoretical progress is warranted. There is still ample opportunity for improving our understanding of other-regarding behaviour.

REFERENCES

Alrøe, H. F., Byrne, J., Glover, L., 2006. Organic Agriculture and Ecological Justice: Ethics and Practice. In: Halberg, N., Alrøe, H. F., Knudsen, M. T., Kristensen, E.S. (Eds.). Global Development of Organic Agriculture: Challenges and Prospects. CABI Publishing, Wallingford, UK.

Andreoni, J., Miller, J., 2002. Giving According to GARP: An Experimental Test of the Rationality of Altruism. Econometrica 70 (2), 737-753.

Andreoni, J., Vesterlund, L., 2001. Which is the Fair Sex? Gender Differences in Altruism. Quarterly Journal of Economics. 116 (1), 293-312.

Andreoni, James 1989. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." Journal of Political Economy 97, 1447-1458.

Andreoni, James and Miller, John, 1993. "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence", Economic Journal 103, 570-585.

Arrow, Kenneth J., 1981. "Optimal and Voluntary Income Redistribution." In: Rosenfield, Steven (ed), Economic Welfare and the Economics of Soviet Socialism: Essays in Honor of Abram Bergson, Cambridge: Cambridge University Press.

Becker, Gary S., 1974. "A Theory of Social Interactions." Journal of Political Economy 82, 1063-1093.

Bellemare, Christian and Sabine Kröger, 2003. "On Representative Trust," Working Paper, Tilburg University.

Ben-Shakhar, Gershon, Gary Bornstein, Astrid Hopfensitz and Frans van Winden, 2004. "Reciprocity and Emotions: Arousal, Self-Reports and Expectations,", Discussion Paper, University of Amsterdam.

Bereby-Meyer, Y., Niederle, M., 2005. Fairness in Bargaining. Journal of Economic Behavior & Organization 56 (2), 173-186.

Berg, Joyce, John Dickhaut, and Kevin McCabe, 1995. "Trust, Reciprocity and Social History," Games and Economic Behavior X, 122-142.

Binmore, Ken, 1998. Game Theory and the Social Contract: Just Playing, MIT Press, Cambridge, Massachusetts.

Binmore, Kenneth, John Gale and Larry Samuelson, 1995. "Learning to be Imperfect: The Ultimatum Game", Games and Economic Behavior 8, 56-90.

Blount, Sally, 1995. "When Social Outcomes aren't Fair: The Effect of Causal Attributions

on Preferences," Organizational Behavior and Human Decision Processes LXIII, 131-144.

Bolle, Friedel and Alexander Kritikos, 1998. "Self-Centered Inequality Aversion versus Reciprocity and Altruism", mimeo, Europa-Universität Viadrina.

Bolton, G. E., Ockenfels, A., 2000. ERC: A Theory of Equity, Reciprocity, and Competition. American Economic Review 90(1), 166-193.

Bolton, G.E., Ockenfels, A., 2006. Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. American Economic Review96,1906-1911.

Bolton, Gary and Rami Zwick, 1995. "Anonymity versus Punishment in Ultimatum Bargaining", Games and Economic Behavior 10, 95-121.

Bolton, Gary E. and Ockenfels, Axel, 2000. A theory of equity, reciprocity and competition. American Economic Review 100, 166-193.

Bolton, Gary E., 1991. "A Comparative Model of Bargaining: Theory and Evidence." American Economic Review 81, 1096-1136.

Bolton, Gary E., Jordi Brandts, and Axel Ockenfels, 1998. "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game", Experimental Economics 3, 207-221.

Bosman, Ronald and Frans van Winden, 2002. "Emotional Hazard in a Power-to-Take-Experiment," Economic Journal 112, 147-169.

Bosman, Ronald, Matthias Sutter and Frans van Winden, 2005. The Impact of Real Effort and Emotions in the Power-To-Take Game," Journal of Economic Psychology 26, 407-429.

Brandts, Jordi and Gary Charness, 2004. "Gift-Exchange with Excess Supply and Excess Demand", mimeo, Pompeu Fabra, Barcelona.

Camerer, C. F., 2003. Behavioral Game Theory, Experiments in Strategic Interaction. Princeton University Press, Princeton.

Camerer, Colin F. and Thaler, Richard H., 1995. Ultimatums, Dictators and Manners. Journal of Economic Perspectives 9, 209-19.

Cameron, Lisa A., 1999. " Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia." Economic-Inquiry 37(1), 47-59.

Carpenter, Jeffrey P., 2000. "Punishing Free-Riders: The Role of Monitoring-Group Size, Second-Order Free-Riding and Coordination", mimeo, Middlebury College.

Chang, J.B. and J.L. Lusk. 2009. "Fairness and Food Choice." Food Policy 34: 483-491.

Charness, G., Rabin, M. 2002. Understanding Social Preferences with Simple Tests. Quarterly Journal of Economics 117 (3), 817-869.

Charness, Gary, 1996. "Attribution and Reciprocity in a Labor Market: An Experimental Investigation," mimeo, University of California at Berkeley.

Charness, Gary, 2000. "Responsibility and Effort in an Experimental Labor Market", Journal of Economic Behavior and Organization 42, 375-384.

Cohen, Dov and Richard Nisbett, 1994. "Self-Protection and the Culture of Honor - Explaining Southern Violence," Personality and Social Psychology Bulletin 20, 551-567.

Cooper, David J., and Carol Kraker Stockman, 1999. "Fairness, Learning, and Constructive Preferences: An Experimental Investigation", mimeo, Case Western Reserve University.

Costa-Font, M., Gil, J.M., Trail, W.B., 2008. Consumer Acceptance, Valuation of and Attitudes towards Genetically Modified Food: Review and Implications for Food. Food Policy 33, 99-111.

Costa-Gomes, Miguel, and Klaus G. Zauner, 1999. "Learning, Non-equilibrium Beliefs, and Non-Pecuniary Payoff Uncertainty in an Experimental Game", mimeo, Harvard Business School.

Cox, James C., 2000. "Trust and Reciprocity: Implications of Game Triads and Social Contexts", mimeo, University of Arizona at Tucson.

Cox, James C., Friedman, Daniel, and Gjerstad, Steven, 2004. "A Tractable Model of Reciprocity and Fairness," mimeo, University of Arizona.

Cox, James C., Sadiraj, Klarita, and Sadiraj, Vjollca, 2001. "Trust, Fear, Reciprocity and Altruism," mimeo, University of Arizona.

Danish Research Center for Organic Farming, 2000. Principles of Organic Farming. Discussion Document for the DARCOF Users Committee. Tjele.

Daughety, Andrew, 1994. "Socially-Influenced Choice: Equity Considerations in Models of Consumer Choice and in Games", mimeo, University of Iowa.

Davis, Douglas, and Charles Holt, 1993. Experimental Economics, Princeton: Princeton University Press.

Dawes, Robyn M., and Richard Thaler, 1988. "Cooperation," Journal of Economic Perspectives II, 187-197.

Dillman, D. A., 2000. Mail and Internet Surveys. John Wiley & Sons, Inc, New York.

Dufwenberg, Martin and Kirchsteiger, Georg, 2004. "A Theory of Sequential Reciprocity," Games and Economic Behavior 47, 268-298.

Durham, C. A., Andrade, D. 2005. Health vs. Environmental Motivation in Organic Preferences and Purchases. Selected Paper for Presentation at the American Agricultural Economics Association annual meeting, July 24-27, Rhode Island.

Eichenberger, Rainer and Felix Oberholzer-Gee, 1998. "Focus Effects in Dictator Game Experiments", mimeo, University of Pennsylvania.

Ellingsen, Tore and Magnus Johannesson, 2000. "Is There a Hold-up Problem?", Stockholm School of Economics, Working Paper No. 357.

Engelmann, D., Strobel, M., 2004. Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. American Economic Review 94 (4), 857-869.

Erlei, M., 2008. Heterogeneous Social Preferences. Journal of Economic Behavior & Organization 65 (3), 436-457.

Fahr, Renè and Bernd Irlenbusch, 2000. "Fairness as a Constraint on Trust in Reciprocity: Earned Property Rights in a Reciprocal Exchange Experiment", Economics Letters 66, 275-282.

Falk, Armin and Fischbacher, Urs, 2005, "A Theory of Reciprocity," Games and Economic Behavior, forthcoming.

Falk, Armin, Fehr, Ernst, and Fischbacher, Urs, 2000a. "Informal Sanctions", Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 59.

Falk, Armin, Fehr, Ernst, and Fischbacher, Urs, 2000b. "Testing Theories of Fairness - Intentions Matter", Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 63.

Falk, Armin, Fehr, Ernst, and Fischbacher, Urs, 2000c. "Appropriating the Commons", Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 55.

Falk, Armin, Fehr, Ernst, and Fischbacher, Urs, 2003. "On the Nature of Fair Behavior," Economic Inquiry 41, 20-26.

Falk, Armin, Simon Gächter, and Judith Kovács, 1999. "Intrinsic Motivation and Extrinsic Incentives in a Repeated Game with Incomplete Contracts", Journal of Economic Psychology.

Fehr, E., Kirchsteiger, G., Riedl, A., 1993. Does Fairness Prevent Market Clearing? An Experimental Investigation. Quarterly Journal of Economics 108 (2), 437-459.

Fehr, E., Naef, M., Schmidt, K. M., 2006. Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments: Comment. American Economic Review 96 (5), 1912-1917.

Fehr, E., Schmidt, K. M., 1999. A Theory of Fairness, Competition, and Cooperation. Quarterly Journal of Economics 114 (3), 817-868.

Fehr, Ernst and Armin Falk, 1999. "Wage Rigidity in a Competitive Incomplete Contract Market", Journal of Political Economy 107, 106-134.

Fehr, Ernst and Bettina Rockenbach, 2003. "Detrimental Effects of Sanctions on Human Altruism", Nature 422, 137-140.

Fehr, Ernst and John List, 2004. "The Hidden Costs and Returns of Incentives - Trust and Trustworthiness among CEOs," Journal of the European Economic Association 3, sss-sss.

Fehr, Ernst and Schmidt, Klaus M., 2004. "Fairness and Incentives in a Multi-task Principal-agent Model," Scandinavian Journal of Economics 106, 453-474.

Fehr, Ernst and Tougareva, Elena, 1995: "Do High Monetary Stakes Remove Reciprocal Fairness? Experimental Evidence from Russia." Mimeo. Institute for Empirical Economic Research, University of Zurich.

Fehr, Ernst and Urs Fischbacher, 2004. "Third Party Punishment and Social Norms", Evolution and Human Behavior 25, 63-87.

Fehr, Ernst, and Gächter, Simon, 2000. "Cooperation and Punishment in Public Goods Experiments", American Economic Review 90, 980-994.

Fehr, Ernst, and Gächter, Simon, 2002. "Do Incentive Contracts Undermine Voluntary Cooperation", Working Paper No. 34, Institute for Empirical Research in Economics, University of Zurich.

Fehr, Ernst, Gächter, Simon and Kirchsteiger, Georg, 1997. "Reciprocity as a Contract Enforcement Device", Econometrica 65, 833-860.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl, 1993. "Does Fairness prevent Market Clearing? An Experimental Investigation," Quarterly Journal of Economics CVIII, 437-460.

Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl, 1998. "Gift Exchange and Reciprocity in Competitive Experimental Markets", European Economic Review 42, 1-34.

Fehr, Ernst, Klein, Alexander and Schmidt, Klaus M., 2004. "Contracts, Fairness, and Incentives." CESifo Working Paper No. 1215, Munich.

Fehr, Ernst, Kremhelmer, Susanne and Schmidt, Klaus M., 2004. "Fairness and the Optimal Allocation of Property Rights." Mimeo, University of Munich.

Fehr, Ernst, Urs Fischbacher, Bernhard Rosenbladt, Jürgen Schupp and Gert Wagner, 2002. A Nation-wide Laboratory - Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys," Schmollers Jahrbuch 122, 519-543.

Fischbacher, Urs, Simon Gächter and Ernst Fehr, 1999. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment", Working Paper No. 16, Institute for Empirical Research in Economics, University of Zurich,.

Forsythe, R. L., Horowitz, J., Savin, N. E., Sefton, M., 1994. Fairness in Simple Bargaining

Games. Games and Economic Behavior 6, 347-369.
Frey, Bruno and Hannelore Weck-Hannemann, 1984. "The Hidden Economy as an 'Unobserved' Variable", European Economic Review 26, 33-53.
Gächter, Simon and Armin Falk (1999): "Reputation or Reciprocity?", Working Paper No. 19, Institute for Empirical Research in Economics, University of Zürich.
Geanakoplos, John, Pearce, David, and Stacchetti, Ennio, 1989. "Psychological Games and Sequential Rationality." Games and Economic Behavior 1, 60-79.
Gintis, Herbert, 2000. "Strong Reciprocity and Human Sociality", Journal of Theoretical Biology 206, 169-179.
Goeree, Jacob, and Holt, Charles, 2001. "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions," American Economic Review 91, 1402-1422.
Gracia, A., de Magistris, T., 2008. The Demand for Organic Foods in South of Italy: A Discrete Choice Model. Food Policy 33, 386-396.
Greenberg, Jerald, 1990. "Employee Theft as a Reaction to Underpayment Inequity: The Hidden cost of Pay Cuts", Journal of Applied Psychology 75, 56 -568.
Grossman, Sanford and Hart, Oliver, 1983. "An Analysis of the Principal-Agent Problem, Econometrica 51, 7-45.
Gul, Faruk and Pesendorfer, Wolfgang, 2005. "TheCanonical Type Space for Interdependent Preferences", mimeo, Princeton University.
Güth, W., Schmittberger, R., Schwarze, B., 1982. An Experimental Analysis of Ultimatum Bargaining. Journal of Economic Behavior and Organization 3 (4), 367-388.
Güth, Werner and Eric van Damme, 1998. "Information, Strategic Behavior and Fairness in Ultimatum Bargaining: an Experimental Study", Journal of Mathematical Psychology 42, 227-247.
Güth, Werner, Hartmut Kliemt and Axel Ockenfels, 2000. "Fairness versus Efficiency - An Experimental Study of Mutual Gift-Giving", mimeo, Humboldt University of Berlin.
Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, 1982. "An Experimental Analysis of Ultimatium Bargaining," Journal of Economic Behavior and Organization III, 367-88.
Hannan, Lynn, John Kagel, and Donald Moser, 1999. "Partial Gift Exchange in Experimental Labor Markets: Impact of Subject Population Differences, Productivity Differences and Effort Requests on Behavior", mimeo, University of Pittsburgh.
Harper, G., Makatouni, A., 2002. Consumer Perception of Organic Food Production and Farm Animal Welfare. British Food Journal 104, 287-299.
Harsanyi, John, 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", Journal of Political Economy 63, 309-321.

Hart, Oliver and Moore, John, 1990. "Property Rights and the Nature of the Firm", Journal of Political Economy 98, 1119-58.

Hart, Oliver and Moore, John, 1999. "Foundations of Incomplete Contracts." Review of Economic Studies 66, 115-138.

Hoffman, Elisabeth, Kevin McCabe, and Vernon Smith, 1996. "On Expectations and Monetary Stakes in Ultimatum Games," International Journal of Game Theory 25, 289-301.

Hoffman, Elisabeth, Kevin McCabe, and Vernon Smith, 1996. "Social Distance and Other-regarding Behavior," American Economic Review 86, 653-660.

Hoffman, Elisabeth, Kevin McCabe, Keith Shachat, and Vernon Smith, 1994. "Preferences, Property Right, and Anonymity in Bargaining Games", Games and Economic Behavior 7, 346-380.

Holmström, Bengt and Milgrom, Paul, 1991. "Multi-task Principal-Agent Analyses." Journal of Law, Economics, and Organization 7 (Sp.), 24-52.

Huang, C. L., 1996. Consumer's Preferences and Attitudes towards Organically Grown Produce. European Review of Agricultural Economics 23(3), 331-342.

Huck, Steffen, Müller, Wieland, and Normann, Hans-Theo, 2001. "Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets," Economic Journal 111, 749-766.

International Federation of Organic Agriculture. 2008. Available (last up date 14 April 2008): http://www.ifoam.org/about_ifoam/principles/index.html.

Irlenbusch, Bernd and Dirk Sliwka, 2005. "Transparency and Reciprocity and Employment Relations," Journal of Economic Behavior and Organization 56, 383-403.

Isaac, Mark R., James M. Walker, Arlington W. Williams, 1994. " Group Size and the voluntary Provision of Public Goods", Journal of Public Economics 54, 1-36.

Kagel, John H, Chung Kim and Donald Moser, 1996. "Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs", Games and Economic Behavior 13, 100-110.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler, 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," American Economic Review LXXVI, 728-41.

Kirchsteiger, Georg, 1994. "The Role of Envy in Ultimatum Games", Journal of Economic Behavior and Organization 25, 373-389.

Kolm, Serge-Christophe, 1995. "The Economics of Social Sentiments: The Case of Envy", Japanese Economic Review 46, 63-87.

Kolodinsky, J., 2008. Affect of Information? Labeling Policy and Consumer Valuation of

rBST free and Organic Characteristics of Milk. Food Policy 22, 616-623.
Konow, J., 2000. Fair Shares: Accountability and Dissonance in Allocation Decisions. American Economic Review 90 (4), 1072-1091.
Kuhar, A., Juvancic, L., 2005. Modeling Consumer's Preferences towards Organic and Integrated Fruits and Vegetables in Slovenia. Selected Paper prepared for Presentation at the 97rd EAAE Seminar "The Economics and Policy of Diet and Health", University of Reading, April 21-22.
Laffont, Jean-Jacques and Tirole, Jean, 1993. A Theory of Regulation and Procurement. Cambridge (Mass.): MIT-Press.
Ledyard, John, 1995. "Public Goods: A Survey of Experimental Research", Chap. 2 in: Alvin Roth and John Kagel (eds.), Handbook of Experimental Economics. Princeton: Princeton University Press.
Levine, David, 1998. "Modeling Altruism and Spitefulness in Experiments", Review of Economic Dynamics 1, 593-622.
Lind, Allan and Tom Tyler, 1988. The Social Psychology of Procedural Justice. New York and London: Plenum Press.
List, J. A., 2006. The Behavioralist Meets the Market: Measuring Social Preferences and
List, John and Todd Cherry, 2000. "Examining the Role of Fairness in Bargaining Games", mimeo, University of Arizona at Tucson.
Lusk, J. L., Briggeman, B. C., 2009. Food Values. American Journal of Agricultural Economics 91(1), 184-196.
Makatouni, A., 2002. What Motivates Consumers to Buy Organic Food in the UK? British Food Journal 104, 345-352.
McCabe, Kevin, Mary Rigdon and Vernon Smith, 2000. "Positive Reciprocity and Intentions in Trust Games", mimeo, University of Arizona at Tucson.
Miller, Sven (1997): "Strategienuntersuchung zum Investitionsspiel von Berg, Dickhaut, McCabe", Diploma thesis, University of Bonn.
Milock, K., Wier, M., Andersen, L. M., 2004. Consumer's Demand for Organic Foods-Attitudes, Value and Purchasing Behaviour. Selected Paper for Presentation at the XIII Annual Conference of European Association of Environmental and Resource Economics, June 25-28, Budapest, Hungary.
Neilson, William, 2005. "Axiomatic Reference Dependence in Behavior toward Others and toward Risk", mimeo, Department of Economics, Texas A&M University.
Nöldeke, G., Schmidt, K.M., 1995. Option Contracts and Renegotiation: A Solution to the Hold-Up Problem. Rand Journal of Economics 26, 163-179.
Offerman, Theo, 1999. "Hurting hurts more than helping helps: The Role of the

self-serving Bias", mimeo, University of Amsterdam.
Ostrom, Elinor, 1990. "Governing the Commons - The Evolution of Institutions for Collective Action", New York: Cambridge University Press
Ostrom, Elinor, 2000. "Collective Action and the Evolution of Social Norms", Journal of Economic Perspectives 14, 137-158.
Padel, S., Foster, C., 2005. Exploring the Gap between Attitudes and Behaviour. British Food Journal 107 (8), 606-625.
Padel, S., Röcklinsberg, H., Schmid, O., 2009. The Implementation of Organic Principles and Values in the European Regulation for Organic Food. Food Policy (2009), doi:10.1016/j.foodpol.2009.03.008
Rabin, Matthew, 1993. "Incorporating Fairness into Game Theory and Economics." American Economic Review, 83(5), 1281-1302.
Reputation Effects in Actual Transaction. Journal of Political Economy 114 (1), 1-37.
Rotemberg, Julio, 2004. "Minimally Acceptable Altruism and the Ultimatum Game," mim-eo, Harvard Business School.
Roth, Alvin E., 1995. "Bargaining Experiments," in: J. Kagel and A. Roth (eds.): Handbook of Experimental Economics, Princeton, Princeton University Press.
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., Michael W. K. Malouf, and J. Keith Murningham, 1981. "Sociological versus strategic Factors in Bargaining", Journal of Economic Behavior and Organization 2, 153-177.
Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara, and Shmuel Zamir, 1991. "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," American Economic Review 81, 1068-95.
Samuelson, Paul A., 1993. "Altruism as a Problem Involving Group versus Individual Selection in Economics and Biology." American Economic Review 83, 143-148.
Sandbu, Martin E., 2002. "A Theory of Set-Dependent Fairness Preferences", mimeo, Harvard University.
Scott, Robert, 2003. "A Theory of Self-enforcing, Indefinite Agreements", Columbia Law Review 108, 1641-1699.
Segal, Ilya, 1999. "Complexity and Renegotiation: A Foundation for Incomplete Contracts." Review of Economic Studies 66(1), 57-82.
Segal, Uzi and Sobel, Joel, 2004. "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings." Mimeo, University of California at San Diego.

Seidl, Christian and Stefan Traub, 1999. "Taxpayers' Attitudes, Behavior, and Perceptions of Fairness in Taxation", mimeo, Institut für Finanzwissenschaft und Sozialpolitik, University of Kiel.

Selten, Reinhard and Axel Ockenfels, 1998. "An Experimental Solidarity Game", Journal of Economic Behavior and Organization, 34, 517-539.

Sen, Amartya, 1995. "Moral Codes and Economic Success", C. S. Britten and A. Hamlin (eds.), Market Capitalism and Moral Values, Edward Eldar, Aldershot.

Sethi, Rajiv and E. Somananthan, 2000. Understanding Reciprocity, mimeo, Columbia University.

Sethi, Rajiv and E. Somananthan, forthcoming. Preference Evolution and Reciprocity, Journal of Economic Theory.

Slonim, Robert, and Alvin E. Roth, 1997. "Financial Incentives and Learning in Ultimatum and Market Games: An Experiment in the Slovak Republic," Econometrica 65, 569-596.

Smith, Adam, 1759, reprinted 1982. The Theory of Moral Sentiments. Indianapolis: Liberty Fund.

Smith, Vernon L., 1962. "An Experimental Study of Competitive Market Behavior," Journal of Political Economy 70, 111-137.

Sonnemans, Joep, Arthur Schram and Theo Offerman, 1999. "Strategic Behavior in Public Good Games - When Partners drift apart", Economics Letters 62, 35-41.

Suleiman, Ramzi, 1996. "Expectations and Fairness in a modified Ultimatum Game", Journal of Economic Psychology 17, 531-554.

Sunding, D. L., 2003. The Role for Government in Differentiated Product Markets: Looking to Economic Theory. American Journal of Agricultural Economics 85(3), 720-724.

Torjusen, H., Lieblein, G., Wandel, M., Francis, C. A., 2001 Food System Orientation and Quality Perception among Consumers' and Producers of Organic Food in Hedmark County, Norway. Food Quality and Preference 12, 207-216.

Train, K., Weeks, M., 2005. Discrete Choice Models in Preference Space and Willingness-to-Pay Space. In: Scarpa, R., Alberini, A. (Eds.). Applications of Simulation Methods in Environmental and Resource Economics. Springer, Netherlands.

Umberger, W. J., McFadden, D. D. T., Smith, A. R., 2009. Does Altruism Play a Role in Determining U.S. Consumer Preferences and Willingness to Pay for Natural and Regionally Produced Beef? Agribusiness 25(2), 268-285.

Veblen, Thorsten, 1922. The Theory of the Leisure Class - An Economic Study of

Institutions, George Allen Unwin, London (first published 1899).
Verhoef, P. C., 2005. Explaining Purchase of Organic Meat by Dutch Consumers. European Review of Agricultural Economics 32 (2), 245-267.
Zajac, Edward, 1995. "Political Economy of Fairness", Cambridge, Massachusetts: MIT Press.
Zizzo, Daniel and Andrew Oswald, 2000. "Are People Willing to Pay to Reduce Others' Income", mimeo, Oxford University.