OTHER-REGARDING PREFERENCES

- This section of the course is concerned with understanding *other-regarding preferences*, which often play an important role in transactions in which there are no clear arms-length terms of trade, and hence the existing surplus must be created and divided in alternative ways.

- These issues are very important in economics since:

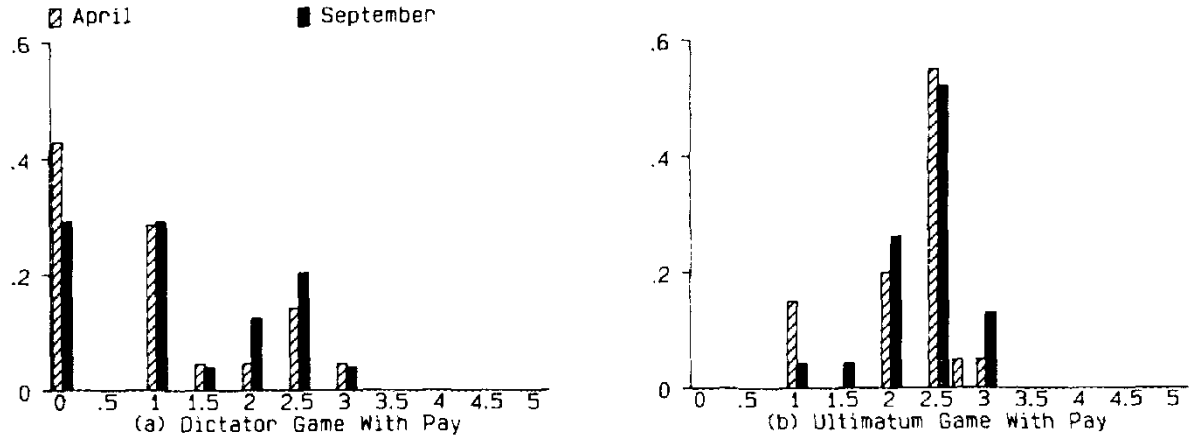
    1. The standard theory of purely self-regarding rational economic actors predicts that unless there is a scope for building a reputation, people will not trust one another and they will have no concerns for other people.
    2. Yet from casual observation it is evident that other-regarding preferences and trust are pervasive and conducive to successful economic transactions; moreover, their lack may seriously inhibit an economy's potential to prosper.

- We will first discuss some well-known evidence that illustrates departures from self-interest model of individuals. We will then come to a more systematic evidence on other-regarding preferences and also to some existing theories of such preferences.

Classic Evidence for Other-Regarding Preferences

Dictator Game

- This is the most elementary game, or, rather, an individual decision situation that has a potential to shed light on other-regarding preferences.
- In this "game", there are two players: proposer and receiver. The proposer decides how to split a pie (amount of money) of a fixed size between himself and the receiver. The decision is then implemented. The role of the receiver is completely passive.
- Typical lab implementation: the pie of \$10 or EUR 10 is provided by the experimenter (as manna from heaven). Subjects are randomly split into proposers and receivers. There is anonymity about the decisions of individual proposers or pairing of the proposers and receivers.
- This "game" was first introduced into the literature by **Forsythe et al. (1994)**.
- Here are their results:



- It is apparent that many subjects do not behave completely selfishly, contrary to the conventional theory based on self-regarding preferences.
- This result has been replicated many times ever since. In a typical experiment, usually more than 60% of subjects pass a positive amount of money, with the mean transfer being roughly 20% of the endowment. The exact amount of giving is sensitive to procedural details, though. See the next subsection for more discussion.
- We will talk about theoretical foundations for this result later on.
- This received wisdom has more recently been challenged, though. **Cherry et al. (2002)** report that when the endowment is earned rather than received from the experimenter and if proposers are granted full anonymity (even from the experimenter), passing positive amounts of money virtually disappears.
- **List (2007)** shows that if the choice set of the proposer is enlarged to include taking money away from the receiver, transferring positive amounts largely disappears, and the modal answer becomes taking away as much as possible from the receiver. If, on top of that, endowment is earned, about 70% of proposers do not make any (positive or negative) transfers, whereas most of the remaining ones take as much as possible from the receiver.

TABLE 1
AGGREGATE BEHAVIOR

Treatment (*N*)	Rate of Positive Offers	Median Offer	Mean Offer	Average Positive Offer*
Baseline (24)	.71	$1.00	$1.33	.38
Take ($1) (46)	.35	$0.00	$0.33	.31
Take ($5) (50)	.10	−$4.50	−$2.48	.42
Earnings (47)	.06	$0.00	−$1.00	.40

* Reported as a percentage of the total amount available in the allocation decision (average positive offer ignores zero and negative offers).

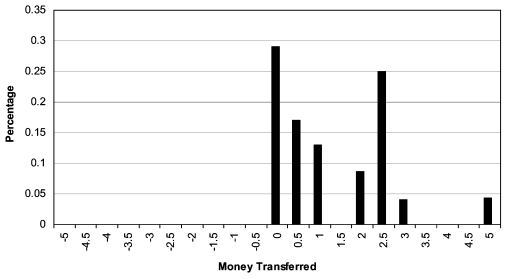


FIG. 1.—Baseline treatment (data online table B1)

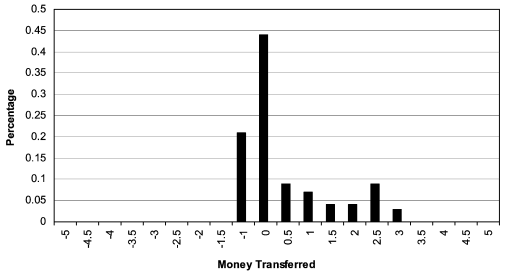


FIG. 2.—Treatment Take ($1) (data online table B2)

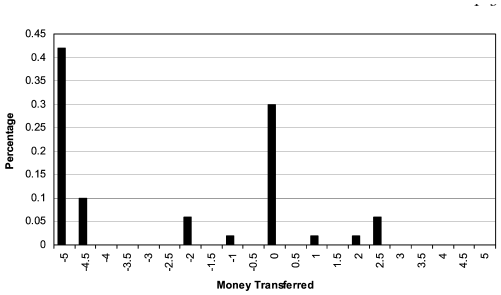


FIG. 3.—Treatment Take ($5) (data online table B3)

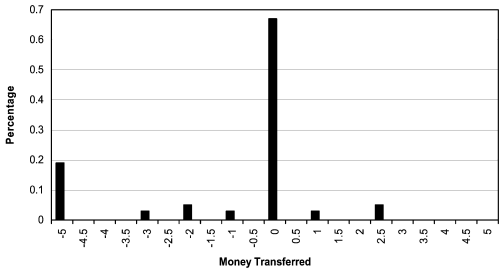


FIG. 4.—Treatment earnings (data online table B4)

Ultimatum Bargaining

- This is a true game in that it appends the Dictator Game with a second stage, in which the receiver can either accept the split proposed by the proposer, in which case the split is implemented, or can reject it, in which case nobody gets anything.
- This game was introduced into the literature by **Guth et al. (1982)**.
- Compared to the Dictator Game, it potentially gives the receiver some strategic power in affecting the outcome of the split. However, assuming self-regarding preferences only and applying the concept of *subgame-perfect equilibrium* to this game, the prediction is that the proposer will offer the receiver next to nothing, and the receiver will accept it, because something is better than nothing.
- However, this is not what is observed in the lab, where low positive offers are often rejected. This may be a rational strategy since many small offers are indeed rejected by receivers.
- Here are the results of Guth et al. (1982):

- Also this result has been replicated many times since. In typical lab results from developed countries, the modal offer is 40% or 50% of the pie; there are virtually no offers above 50%; some, but very few offers are as low as 20% or lower, and these get rejected about half of the time.
- Clearly, the empirical results differ significantly from the prediction based on subgame-perfect equilibrium with self-regarding preferences.
- This result, though, is sensitive to a number of procedural details:

    1. framing as pie-splitting vs. framing as a sale transaction (**Hoffman et al., 1994**): in sale framing, with the proposer being the seller, the median share of the "pie" offered to the buyer (receiver) goes down from 50% to 40%
    2. if the roles of proposer and receiver are allocated based on some objective performance ranking (**Hoffman et al., 1994**), with the better performers being allocated into the role of proposers, then the median offer falls further to 30%
    3. if groups of subjects decide collectively in the two roles (**Bornstein and Yaniv, 1998**), offers are lower (35%) than offers made by individuals (44%) under otherwise identical circumstances (including available pie per person in 3- and 7-person groups)
    4. the higher the stakes/size of the pie (**Slonim and Roth, 1998**), the lower (percentage-wise) the offers are and low (percentage-wise) offers are more likely to be accepted among experienced players; hence the receivers may think in terms of absolute rather than relative payoffs; alternatively, as the cost of rejection goes up, rejection becomes less likely

Table 4
Naive decision behavior in easy games.

Game	c account to be distributed (DM)	Demand of player 1 (DM)	Decision of player 2
A	10	6.00	1
B	9	8.00	1
C	8	4.00	1
D	4	2.00	1
E	5	3.50	1
F	6	3.00	1
G	7	3.50	1
H	10	5.00	1
I	10	5.00	1
J	9	5.00	1
K	9	5.55	1
L	8	4.35	1
M	8	5.00	1
N	7	5.00	1
O	7	5.85	1
P	6	4.00	1
Q	6	4.80	0
R	5	2.50	1
S	5	3.00	1
T	4	4.00	0
U	4	4.00	1

Table 5
Experienced decision behavior in easy games.

Game	c = amount to be distributed (DM)	Demand of player 1 (DM)	Decision of player 2
A	10	7.00	1
B	10	7.50	1
C	9	4.50	1
D	9	6.00	1
E	8	5.00	1
F	8	7.00	1
G	7	4.00	1
H	7	5.00	1
I	4	3.00	0
J	4	3.00	0
K	5	4.99	0
L	5	3.00	1
M	6	5.00	0
N	6	3.80	1
O	10	6.00	1
P	9	4.50	1
Q	8	6.50	1
R	7	4.00	0
S	6	3.00	1
T	5	4.00	0
U	4	3.00	1

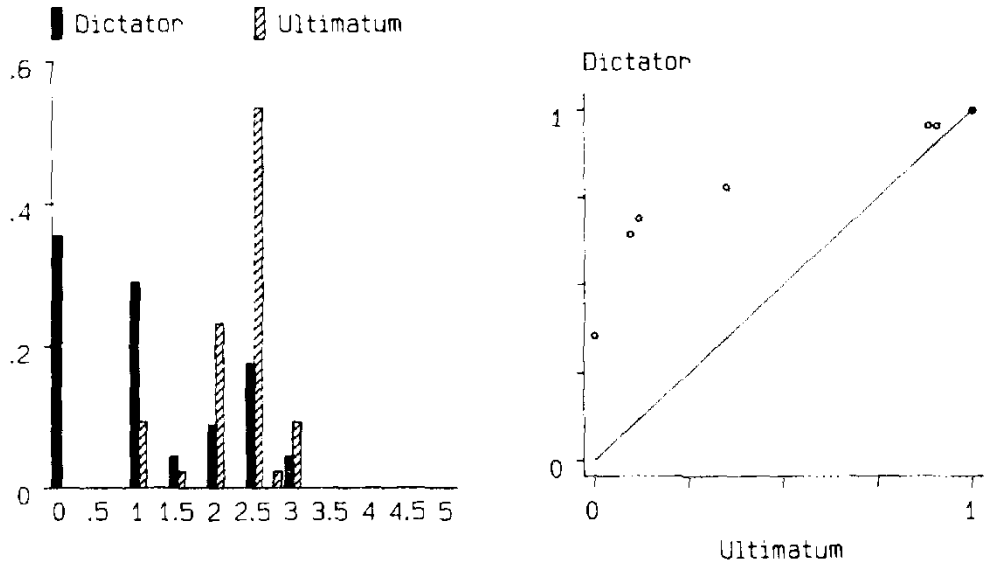


FIG. 3. Dictator with pay (pooled) vs ultimatum with pay (pooled).

Two-Stage Bargaining

- In real life, bargaining usually has more than one offer stage. Offers are usually reacted to by counter-offers, which are reacted to by counter-counter-offers, etc. One can envision an infinite-horizon version of this process (see **Rubinstein, 1982**), but we will focus our attention to the simple two-stage bargaining process here. Hence the proposer makes an offer; if accepted, the split is implemented; if rejected, the receiver makes a counter-offer; if accepted, the split is implemented; if rejected, both parties get zero. One can also introduce some costs of delayed agreement (discounting) by having the size of the pie shrink somewhat going into the counter-offer stage.

- Suppose the size of the pie is X in the first stage and Y in the second stage, with $X > Y$. Assuming self-regarding preferences only and applying the concept of *subgame-perfect equilibrium* to this game, the prediction is that in the second stage, the receiver will offer the proposed next to nothing, and the proposer will accept it. Knowing this, the receiver knows in the first stage that he can get at least $Y - \varepsilon$ from the game, and will hence reject any lower offer from the proposer. As a result, the best the proposer can do is to offer Y to the receiver in the first stage, which will be accepted, and pocketing $X - Y$ himself.

- In one of the later implementations, **Goeree and Holt (2000)** run an experiment of this sort. Their finding:

- Clearly, the empirical results differ significantly from the prediction based on subgame-perfect equilibrium with self-regarding preferences.

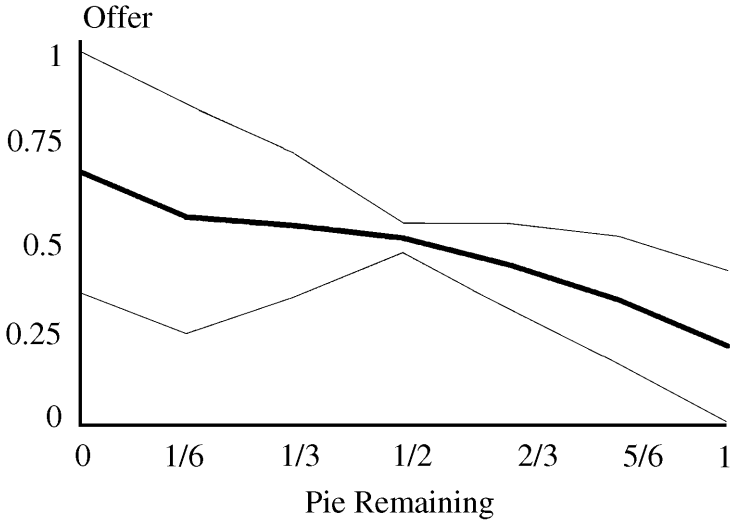


Fig. 1. Average first-stage offers (dark line) and standard deviations (thin lines).

Categorization of Other-Regarding Preferences

- Consider a reference group of n players indexed by $i \in \{1, .., n\}$ and let $\pi = (\pi_1, ..., \pi_n)$ denote the vector of their monetary payoffs that results from some economic interaction. The classic evidence discussed above suggests that individuals care about well-being and actions of other individuals. In particular, it suggests that the utility of individual i does not depend only on his or her own the monetary payoff π_i, but also on monetary payoffs of other individuals $j \neq i$ that are members of some relevant reference group, and also on own and other individuals' actions a_j. That is, i's utility function is given by $u_i(\pi_i, \pi_{-i}; a_i, a_{-i})$. Dependence on the actions would reflect elements of reciprocity (a_{-i}) and self-perception (a_i).

- It is reasonable to assume in most circumstances that $u_i(\cdot)$ is at least weakly increasing in the own payoff π_i. If $u_i(\pi_i, \pi_{-i}; a_i, a_{-i}) = \pi_i$, we talk about **purely self-regarding**, or **purely selfish** individual. This is a standard assumption in traditional economic theory.

Regard for Others' Payoffs, but not Their Actions

- If an individual is not purely selfish, there are several types of other-regarding preferences discussed in the literature. Actions will be omitted from the list of arguments of the utility function in this subsection.

  - **Altruism**: individual subjective well-being is increasing in the material well-being of the others. That is, $u_i(\cdot)$ is strictly increasing in the elements of π_{-i}. (This definition corresponds to *non-paternalistic altruism*. In further disaggregation, if i cares about a particular consumption bundle that j obtains, we talk about *paternalistic altruism*. For example, you may not care if other people spend their wealth increase on consumption goods, but may be happier if they spend it on education.) **Pure altruism** corresponds to the case when $u_i(\cdot)$ depends only on π_{-i} and is increasing in all of its elements.

– **Spite (envy):** this is the opposite of altruism in that individual subjective well-being is decreasing in the material well-being of the others. That is, $u_i(\cdot)$ is strictly increasing in the elements of π_{-i}. **Pure spite** corresponds to the case when $u_i(\cdot)$ depends only on π_{-i} and is decreasing in all of its elements.

– **Efficiency maximization:** individual cares about the sum of payoffs of all individuals:

$$u_i(\pi) = \sum_j \pi_j.$$

– **Maxmin (Rawlsian) preferences:** individual cares about the minimum of payoffs among all individuals.

$$u_i(\pi) = \min_j(\pi_j).$$

Note that in terms of the utility function proposed by Fehr and Schmidt, in case of $n = 2$, this corresponds to the case when $\alpha_i = 0$ and $\beta_i = 1$.

– **Competitive preferences:** individual cares about either the absolute or the relative difference in payoffs versus the other individuals. For **absolutely competitive preferences**,

$$u_i(\pi) = v_i(\pi_i - \pi_1, .., \pi_i - \pi_{i-1,}\pi_i - \pi_{i+1,}.., \pi_i - \pi_n),$$

where $v_i(\cdot)$ is increasing in all of its arguments. For relatively competitive preferences,

$$u_i(\pi) = w_i(\pi_i/\pi_1, .., \pi_i/\pi_{i-1,}\pi_i/\pi_{i+1,}.., \pi_i/\pi_n),$$

where $v_i(\cdot)$ is increasing in all of its arguments.

– **Inequality aversion (fairness):** given one's payoff, an individual's subjective well-being is maximized if other individuals have the same payoffs, and it is lower the further away payoffs of the other individuals are (in either direction in some metric, such as absolute or relative) from one's own payoff. The two best-known theories of inequality aversion are due to **Fehr and Schmidt (1999)** and **Bolton and Ockenfels (2000)**. We will come back to them in more detail. Fehr and Schmidt propose the following utility function:

$$\begin{aligned} u_i(\pi) = \pi_i &- \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{\pi_j - \pi_i, 0\} \\ &- \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{\pi_i - \pi_j, 0\} \end{aligned} \tag{1}$$

with $\beta_i \leq \alpha_i$ and $0 \leq \beta_i < 1$. This is an example of *self-centered fairness* (i.e., in comparison to oneself). In this setup, α_i measures the strength of *disadvantageous inequality* aversion of player i, whereas β_i measures the strength of *advantageous inequality* aversion; it is assumed that the latter is never stronger than the former, and that the advantageous inequality aversion is never so strong as to make the utility decline in own payoff, given the payoff of the other players ($\beta_i < 1$). On the other hand, the assumption that $0 \leq \beta_i$ is less convincing, since one could easily envision people who like advantageous inequality. The standard selfish preferences correspond to the special case $\alpha_i = \beta_i = 0$.

- It turns out that most cases can be parameterized by a simple utility function of the form

$$u_i(\pi) = (1-\rho_i)\pi_i + \sigma_i \sum_{j \neq i} \pi_j + (\rho_i - \sigma_i)\min(\pi_1, .., \pi_n). \tag{2}$$

  In this parametrization, ρ_i and σ_i capture the extent of non-selfish preferences. In particular, ρ_i parameterizes the extent to which the individual cares about the well-being of the worst-off individual relative to own well-being and σ_i parameterizes the extent of altruism or spite or competitiveness relative to the extent of caring for the worst-off individual. We then have that:

  - Pure selfishness: $\rho_i = 0, \sigma_i = 0$
  - Pure altruism: $\rho_i = 1, \sigma_i = 1$
  - Pure spite: cannot be modeled using these preferences
  - Efficiency maximization: $\rho_i = 0.5, \sigma_i = 0.5$
  - Maxmin: $\rho_i = 1, \sigma_i = 0$
  - Absolute competitiveness for $n = 2$: $\rho_i = \sigma_i \to -\infty$
  - Inequality aversion a'la Fehr and Schmidt for $n = 2$: $\rho_i = \beta_i, \sigma_i = -\alpha_i$

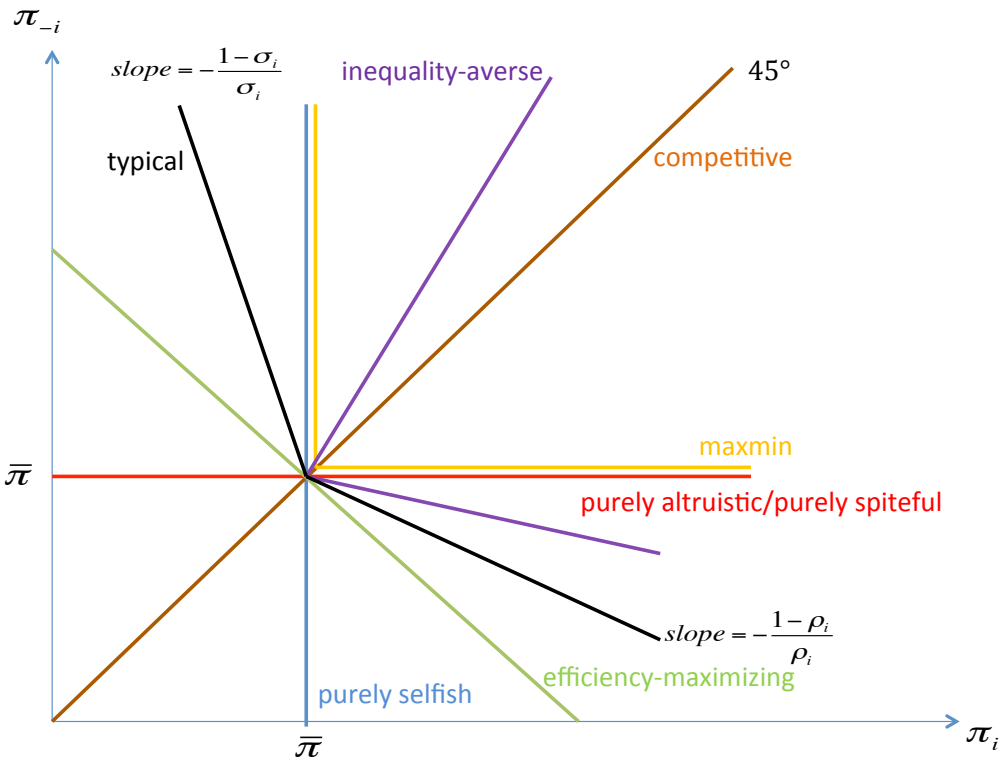
- Note that if $n = 2$, this parametrization simplifies to

$$u_i(\pi_i, \pi_{-i}) = (1-\rho_i)\pi_i + \sigma_i\pi_{-i} + (\rho_i - \sigma_i)\min(\pi_i, \pi_{-i}),$$

  implying that

$$u_i(\pi_i, \pi_{-i}) = \begin{cases} (1-\rho_i)\pi_i + \rho_i\pi_{-i} & \text{if } \pi_i \geq \pi_{-i} \\ (1-\sigma_i)\pi_i + \sigma_i\pi_{-i} & \text{if } \pi_i < \pi_{-i} \end{cases}.$$

  Hence the indifference curves in the (π_i, π_{-i}) space in the region $\pi_i \geq \pi_{-i}$ have the slope $-(1-\rho_i)/\rho_i$, whereas indifference curves in the region $\pi_i < \pi_{-i}$ have the slope $-(1-\sigma_i)/\sigma_i$. The following figure illustrates such indifference curves through the point $(\overline{\pi}, \overline{\pi})$ for different kinds of preferences:



Regard for Others' Payoffs, Mediated by Their Actions

- In this case, an individual may display other-regarding preferences, but whether this does happen and what type these preferences have depends on actions of other individuals that are observed before the individual decides on his or her own action. The most commonly discussed one is reciprocity, which is in some settings also referred to as conditional cooperation. Reciprocity consists of two elements, which may also be present in preferences individually:

  - **Positive reciprocity:** an individual i is altruistic toward other individual(s) who have displayed altruism toward i in their previous actions. That is, if $a_j > 0$ can be understood as an altruistic action by the others, then $u_i(\pi_i, \pi_{-i}; a_i, a_{-i})$ is increasing in π_j if $a_j > 0$.
  - **Negative reciprocity:** an individual i is spiteful toward other individual(s) who have displayed spite toward i in their previous actions. That is, if $a_j < 0$ can be understood as a spiteful action by the others, then $u_i(\pi_i, \pi_{-i}; a_i, a_{-i})$ is decreasing in π_j if $a_j < 0$.

- Up to date, there is no universally accepted theory of reciprocity. For some well-known attempts at modeling it, see **Charness and Rabin (2002)** and **Falk and Fischbacher (2006)**.

How do Various Theories of Other-Regarding Preferences Help Us to Explain Behavior Observed in Experiments?

Inequality aversion (fairness): Fehr and Schmidt (QJE, 1999)

- Fehr and Schmidt are motivated by the question of how it is possible that fairness seems to matter a lot for payoff outcomes in some settings (dictator and ultimatum games being the prime example), yet it does not seem to matter in others (e.g., market experiments). Likewise, these is a lot of evidence of cooperation in some settings (such as in Prisoners' Dilemma or Public Goods games, we will talk about these later), even though self-interested behavior would predict quite the opposite, but there is also a lot of evidence of lack of cooperation in other settings. The authors try to come up with a simple model of preferences that could organize all these observations.

- They propose a model of preferences that can achieve this objective. Also, importantly, they show how the heterogeneity of preferences interacts with the economic environment to determine whether a more or a less equitable distribution of gains from trade prevails. In particular, they show that this environment determines which type of preferences becomes decisive for the ultimate outcome.

- Now consider the implication of this assumption for the **Ultimatum Game**. Normalize the size of the pie to 1 and let s denote the offer made by the proposer (player 1) to the receiver/responder (player 2).

- **Proposition 1:** Under the preferences given by (1), it is a dominant strategy for the responder to accept any offer $s \geq 0.5$, to reject s if

$$s < S(\alpha_2) \equiv \alpha_2/(1+2\alpha_2) < 0.5,$$

and to accept $s > s'(\alpha_2)$. If the proposer knows the preferences of the responder, he will offer

$$s^* \begin{cases} = 0.5 & \text{if } \beta_1 > 0.5 \\ \in [S(\alpha_2), 0.5] & \text{if } \beta_1 = 0.5 \\ S(\alpha_2) & \text{if } \beta_1 < 0.5 \end{cases}$$

in subgame-perfect equilibrium. (For the rest of the proposition, see the paper.)

- Hence this proposition explains why there are no offers above 50%, the offers of 50% are always accepted, and that low offers are likely to be rejected (depends on α_2). For example, $\alpha_2 = 1/3$ implies an acceptance threshold of 0.2.

- Now consider a **Market Game** where a number of suppliers, each holding one unit of the good, are trying to sell it to a single buyer, who only demands one unit of the good. It has been robustly demonstrated in the past experimentally that all gains from trade go to the buyer in this case, and this very unequal distribution of gains from trade is due to seller competition. You can view this as an extended ultimatum game in which there are multiple proposers and a single respondent can accept or reject the highest offer. Due to competition, the unique subgame-perfect equilibrium is for at least two offers to be equal to 1, in line with existing evidence. But does this result still survive with inequality-averse preferences?

- **Proposition 2:** Under the preferences given by (1), for any admissible parameters (α_i, β_i), $i \in \{1, .., n\}$, there is a unique subgame-perfect equilibrium outcome in which at least two proposers offer $s = 1$, one of which is in turn accepted by the responder.

- The key observation here is that even though every proposer views the ultimate outcome as very inequitable, due to competition forces, no single player can enforce an equitable outcome. Given that there will be inequality anyway, each proposer has a strong incentive to outbid his competitors in order to turn part of the inequality to his advantage and to increase his own monetary payoff.

- So the overall implication of the theory is that fairness will have strong implications in one-on-one or small-group interactions, which may include bilateral negotiations over surplus in markets with search friction, such as the labor market, but not in homogeneous good markets and large group interactions, where forces of competition will dominate.

- However, now consider the implications of this theory for the **Dictator Game**. Let player 1 be the proposer and player 2 be the receiver, and let s be the share allocated by the proposer to the receiver. The proposer determines s by maximizing

$$U_1(s) = 1 - s - \alpha_1 \max\{2s - 1, 0\} - \beta_1 \max\{1 - 2s, 0\}.$$

First note that $s > 0.5$ is never optimal. Hence it is enough to consider the case $s \leq 0.5$. Then

$$U_1(s) = 1 - \beta_1 - s(1 - 2\beta_1).$$

As a result, with the exception of the knife-edge case of $\beta_1 = 0.5$, it is optimal to choose either $s = 0$ if $\beta_1 < 0.5$ or $s = 0.5$ if $\beta_1 > 0.5$. However, we saw from the results of Forsythe et al. (1998) that even though these indeed are the two modal decisions, there is also a non-negligible fraction of proposers who choose $s \in (0, 0.5)$. To account for this empirical fact, the theory may sometimes need to be slightly modified by dropping linearity for convexity of the cost of inequality in inequality. For example, under the modified preferences

$$U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} [\max\{x_j - x_i, 0\}]^{\gamma_i} - \beta_i \frac{1}{n-1} \sum_{j \neq i} [\max\{x_i - x_j, 0\}]^{\delta_i} \qquad (3)$$

with $\gamma_i, \delta_i > 1$, the analogous analysis implies that the optimal s for the proposer is given by

$$s^* \equiv \frac{1}{2} \left[1 - \frac{1}{(2\beta_1 \delta)^{\frac{1}{\delta - 1}}} \right]$$

if $\beta_1 \geq 1/(2\delta)$ and it is given by $s^* = 0$ otherwise.

- Next, let's turn our attention to a simple game of cooperation (to be discussed later, we will come back to this discussion then). In particular, consider a **Public Goods Game** with the multiplication factor of $a \in (1/n, 1)$. In this game, which we will refer to as the one-stage game, it is a dominant strategy to contribute nothing. Now consider a two-stage version of the game in which players can mete out punishments by reducing other players' payoffs at a marginal cost of $c \in (0, 1)$.In particular, each player starts the game with an endowment of $y > 0$ and the ultimate payoff of player i is given by

$$x_i(g_1, ..., g_n, p_1, ..., p_n) = y - g_i + a \sum_{j=1}^{n} g_j - \sum_{j=1}^{n} p_{ji} - c \sum_{j=1}^{n} p_{ij},$$

where g_j is the contribution of player j, p_{ij} is the punishment imposed by player i on player j and c is the marginal cost of punishment. Under standard preferences,

nobody will punish in the second stage ($p_{ij} = 0$ for all i and j), because punishments are privately costly. Hence the unique subgame-perfect equilibrium is to contribute nothing ($g_i = 0$ for all i) and not punish at all for all the players. **Fehr and Gachter (1996)** document, however, that although the one-stage game prediction fits experimental data very well, the two-stage game prediction does a poor job of predicting behavior. In the latter game, cooperators punish defectors and lower contribution levels are associated with higher received punishments. Thus defectors do not gain from free-riding because they are being punished.

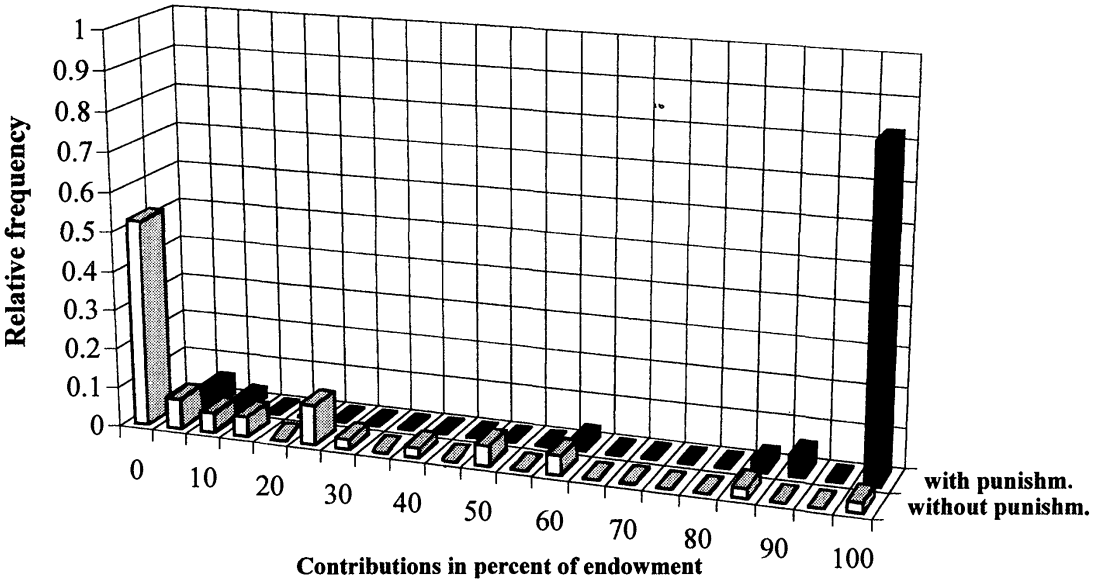


FIGURE II
Distribution of Contributions in the Final Period of the Public Good Game with Punishment (*Source:* Fehr and Gächter [1996])

- Can this inequality-aversion theory account for these disparate findings?
- **Proposition 4:** Consider the one-stage game and suppose players have preferences as given by (1). Then:

  (a) If $a + \beta_i < 1$, then it is a dominant strategy for player i to choose $g_i = 0$.

  (b) Let k denote the number of players with $a + \beta_i < 1$. If $k/(n-1) > a/2$, then there is a unique equilibrium with $g_i = 0$ for all $i \in \{1, ..., n\}$.

  (c) If $k/(n-1) < (a + \beta_j - 1)/(\alpha_j + \beta_j)(< a/2)$ for all players j with $a + \beta_j > 1$, then other equilibria with positive contribution levels exist. In these equilibria, all k players with $a + \beta_i < 1$ choose $g_i = 0$, while all other players contribute $g_j = g \in [0, y]$.

- That is, if the benefit to the player, direct + derived from reducing advantageous inequality, is less than the cost of contributing, the player does not contribute (a). Even if there are players for whom the reverse is the case, if there are too few of them, they will not contribute either since the would suffer too much from the disadvantageous inequality imposed on themselves by their contributions (b). Finally, if there are sufficiently many players who could sustain cooperation, they can do so even if other do not contribute, but only if the sucker feeling is not too strong (c).

- Fehr and Schmidt then go on arguing that one can also identify a set of parameter values for α_i's and β_i's that consistently predicts behavior across different games.

- **Proposition 5:** Consider the two-stage game and suppose players have preferences as given by (1). Suppose that there is a group of $n_C \geq 1$ of "conditionally cooperative enforcers" with preferences that obey $a + \beta_i \geq 1$ and

$$c < \frac{\alpha_i}{(n-1)(1+\alpha_i) - (n_C - 1)(\alpha_i + \beta_i)},$$

  whereas all other players do not care about inequality, i.e., $\alpha_i = \beta_i = 0$. Then there is the following subgame-perfect equilibrium:

  (a) In the first stage, each player contributes $g_i = g \in [0, y]$.

  (b) If each player does so, then there are no punishments in the second stage. If one of the non-cooperators deviates and chooses $g_i < g$, then each cooperative enforcer chooses $p_{ji} = (g - g_i)/(n_C - c)$, while all other players do not punish. If one of the cooperative enforcers chooses $g_i < g$ or if any player chooses $g_i > g$ or if more than one player deviates from g, then a Nash equilibrium of the punishment game is being played.

- That is, if there is a sufficiently large group of sufficiently motivated conditional cooperative enforcers, then cooperation can be sustained. Sometimes, it may be sufficient to have a single cooperative enforcer if his $\alpha_i/(1 + \alpha_i) > c(n - 1)$.

- One quibble about Proposition 5 is that there is a continuum of equilibria. But one can use an intuitive refinement to argue that $g_i = y$ for all i is the most "reasonable one": it generates the unique efficient and symmetric payoff vector.

Inequality aversion (fairness): Bolton and Ockenfels (AER, 2000)

- Similar motivation as Fehr and Schmidt, similar approach, working in parallel.
- They call their model ERC: *equity, reciprocity and competition.*
- Like Fehr and Schmidt, Bolton and Ockenfels illustrate that ERC is capable of explaining why inequality aversion affects outcomes in bilateral bargaining or small groups, whereas in markets and large groups it is overridden my forces of market competition.
- Formally, consider a set of n players indexed by $i \in \{1, .., n\}$ and let $x = (x_1, ..., x_n) \geq 0$ denote the vector of monetary payoffs. The utility of player $i \in \{1, .., n\}$ is given by

$$U_i(x) = u_i(x_i, \sigma_i), \tag{4}$$

  where

$$\sigma_i(x) = \begin{cases} x_i / \left(\sum_{j=1}^n x_j\right) & \text{if } \sum_{j=1}^n x_j > 0 \\ 1/n & \text{if } \sum_{j=1}^n x_j = 0 \end{cases} \tag{5}$$

  is the share of player i's payoff in the total pecuniary payoff. The usual monotonicity and differentiability assumptions are made on u_i: it is continuous and twice differentiable on its domain, $u_{i1} \geq 0$, $u_{i11} \leq 0$ (preference for more money, but with diminishing marginal utility), $u_{i2}(x_i, 1/n) = 0$ and $u_{i22}(x_i, \sigma_i) < 0$ (aversion to inequality). Hence equal division is the *social reference point.*
- Plotting the utility function for the case of two players, we see that the underlying idea is the same as in Fehr and Schmidt. The only technical difference is that FS use a piecewise linear utility function, whereas BO use a smooth and concave utility function; but as we discussed, the FS model can easily be extended to build concavity into the utility function, although the kink at the equality will remain.
- As a result, BO obtain similar results as FS in that their model can explain equal outcomes in ultimatum bargaining and dictator games, but very unequal outcomes in market games.

Systematic Evidence on Other-Regarding Preferences

Andreoni and Miller (ECA, 2002)

- The authors ask a very basic question: is there a rational preference relation over payoffs to oneself and to another individual?
- They judge on this by testing whether subject choices in continuous dictator games with different relative prices of giving and different budgets satisfy the **Generalized Axiom of Revealed Preference**.
- **Definition:** A bundle (of payoffs to oneself and to another individual) A is *(strictly) directly revealed preferred* to B if B was (strictly) in the budget set when A was chosen.
- **Definition:** If A is directly revealed preferred to B, B is directly revealed preferred to C,... to Y, and Y is directly revealed preferred to Z, then A is *indirectly revealed preferred* to Z.
- **Generalized Axiom of Revealed Preference (GARP):** If A is indirectly revealed preferred to B, then B is not strictly directly revealed preferred to A.
- Experimental design:
- Number of subjects: 176. Number of GARP violators: 18. Serious violators: 3.
- Identification of preference types:
- There is also evidence that about one quarter of the subjects are willing to sacrifice some of their own payoff if such act shrinks the payoff of the other subject as well. In 84 percent of cases, this happens when the reduction in the other subject's payoff is larger than the reduction in own payoff.

TABLE I

ALLOCATION CHOICES

Budget	Token Endowment	Hold Value	Pass Value	Relative Price of Giving	Average Tokens Passed
1	40	3	1	3	8.0
2	40	1	3	0.33	12.8
3	60	2	1	2	12.7
4	60	1	2	0.5	19.4
5	75	2	1	2	15.5
6	75	1	2	0.5	22.7
7	60	1	1	1	14.6
8	100	1	1	1	23.0
9[a]	80	1	1	1	13.5
10[a]	40	4	1	4	3.4
11[a]	40	1	4	0.25	14.8

[a]Were only used in session 5, others used in all sessions.

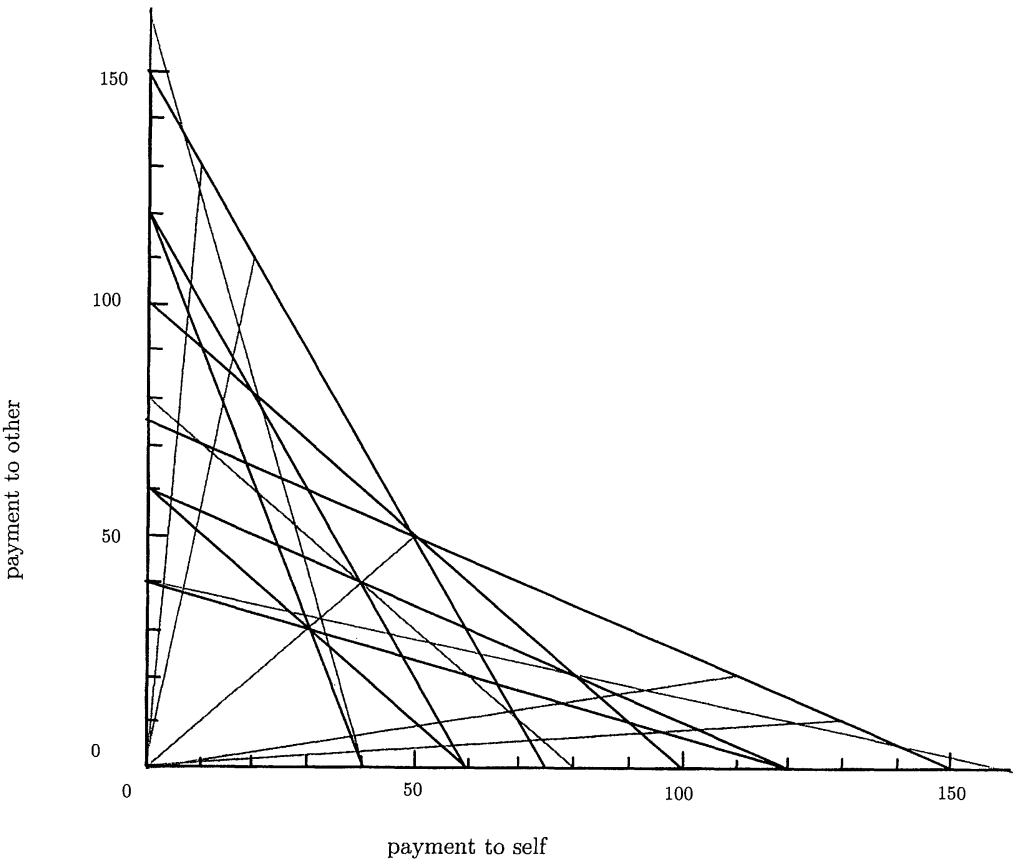


FIGURE 1.—Budget constraints offered subjects.

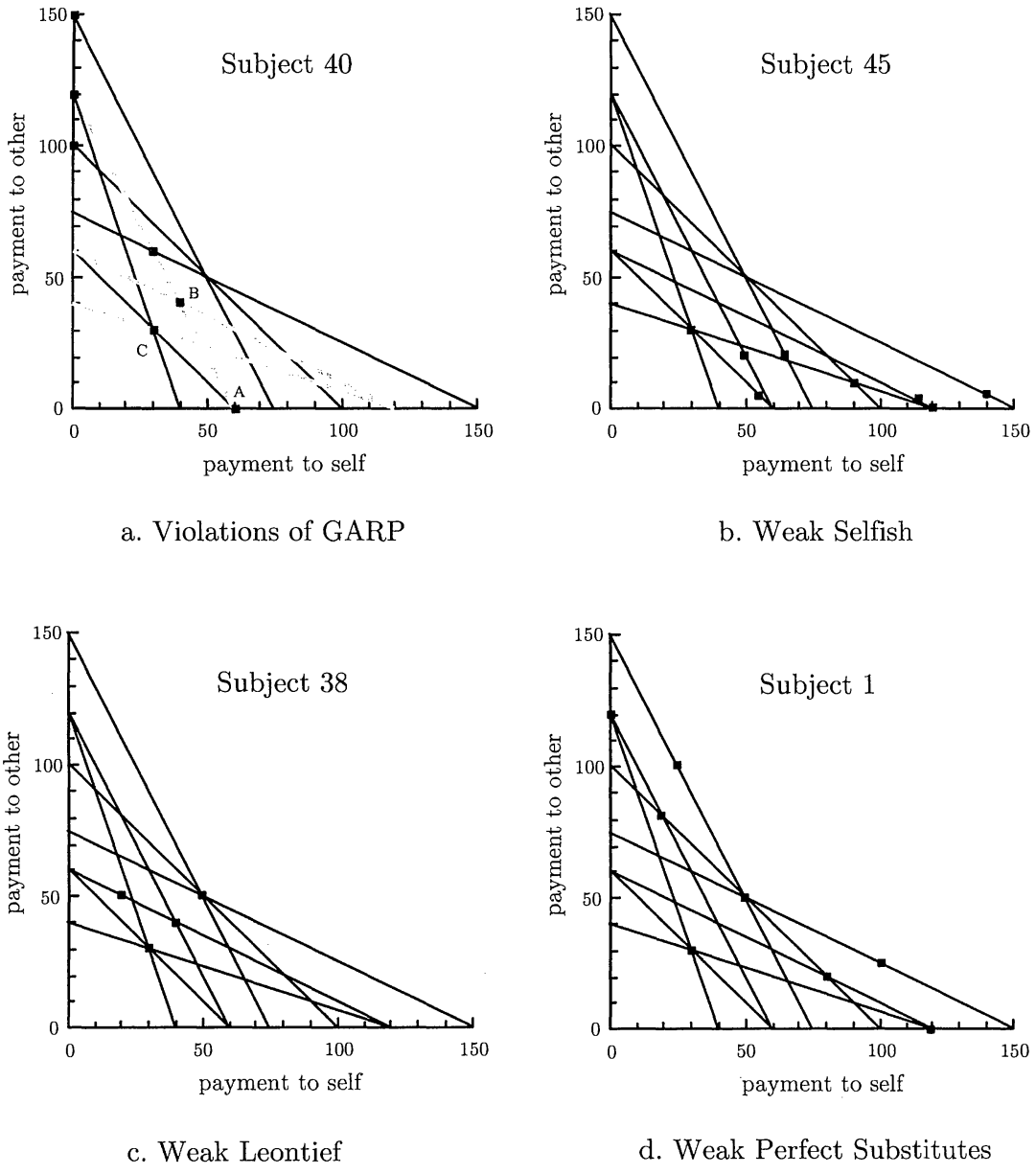


FIGURE 2.—Analyzing individual preferences.

TABLE III

SUBJECT CLASSIFICATION BY PROTOTYPICAL UTILITY FUNCTION

	Fit		
Utility Function	Strong	Weak	Total
Selfish	40	43	83 (47.2%)
Leontief	25	28.5[a]	53.5 (30.4%)
Perfect Substitutes	11	28.5[a]	39.5 (22.4%)

[a]One subject was equidistant from strong Leontief and Substitutes.

Charness and Rabin (QJE, 2002)

- Charness and Rabin pay attention to both payoff-motivated other-regarding preferences and to reciprocity. We will discuss only the former here, the latter is left to your own reading.
- They run 7 different dictator games:

TABLE I
GAME-BY-GAME RESULTS

Two-person dictator games		Left	Right
Berk29 (26)	B chooses (400,400) vs. (750,400)	.31	.69
Barc2 (48)	B chooses (400,400) vs. (750,375)	.52	.48
Berk17 (32)	B chooses (400,400) vs. (750,375)	.50	.50
Berk23 (36)	B chooses (800,200) vs. (0,0)	1.00	.00
Barc8 (36)	B chooses (300,600) vs. (700,500)	.67	.33
Berk15 (22)	B chooses (200,700) vs. (600,600)	.27	.73
Berk26 (32)	B chooses (0,800) vs. (400,400)	.78	.22

- In terms of the utility function

$$u_i(\pi_i, \pi_{-i}) = \begin{cases} (1-\rho_i)\pi_i + \rho_i\pi_{-i} & \text{if } \pi_i \geq \pi_{-i} \\ (1-\sigma_i)\pi_i + \sigma_i\pi_{-i} & \text{if } \pi_i < \pi_{-i} \end{cases}.$$

  that we discussed before, they define four different preference types:

  1. **competitive preferences**: $\sigma_i \leq \rho_i \leq 0$; hence $u_i(\pi_i, \pi_{-i})$ is increasing in π_i and decreasing in π_{-i}, the more so if $\pi_i < \pi_{-i}$; hence the indifference curves are upward-sloping everywhere
  2. **inequality-aversion preferences**: $\sigma_i < 0 < \rho_i < 1$; hence $u_i(\pi_i, \pi_{-i})$ is increasing in π_i and π_{-i} if $\pi_i \geq \pi_{-i}$ and increasing in π_i and decreasing in π_{-i} if $\pi_i < \pi_{-i}$, as in Fehr and Schmidt (1999); hence the indifference curves are downward sloping if $\pi_i > \pi_{-i}$ and upward-sloping if $\pi_i < \pi_{-i}$
  3. **social welfare preferences**: $0 < \sigma_i \leq \rho_i \leq 1$; hence $u_i(\pi_i, \pi_{-i})$ is increasing in both π_i and π_{-i}, with the relative weight on π_i being larger if $\pi_i < \pi_{-i}$; hence the indifference curves are downward sloping everywhere

- They find the following distribution of preference types implied by choices in these dictator games:

TABLE III
CONSISTENCY OF BEHAVIOR WITH DISTRIBUTIONAL MODELS
WHEN THE PREDICTION IS UNIQUE
(Entries are chances taken over total chances.)

Class of games	Narrow self-interest	Competitive	Difference aversion	Social welfare
B's behavior in the dictator games	132/206 (64%)	104/196 (53%)	49/106 (46%)	54/62 (87%)