A Theory of Fairness, Competition, and Cooperation

Ernst Fehr and Klaus M. Shmidt

published in Quarterly Journal of Economics, 1999

November 14, 2011
Almost all economic models assume that all people are exclusively pursuing material self-interest and do not care about social goals *per se*.

There is substantial evidence that fairness matters and affects the behavior of many people.

There is also evidence that seems to suggest that fairness considerations are rather unimportant in competitive environments.
Introduction

- Similarly conflicting evidence exists with regard to cooperation.
- Reality provides many examples indicating that people are more cooperative than is assumed in the standard model.
- There are also conditions under which a vast majority of subjects completely defect as predicted by the self-interest model.
Introduction

- There is a bewildering variety of evidence about when people care about fairness, and act cooperatively (or selfishly).
- The paper asks whether there is a single principle which can account for those observations.
- They argue that the answer is affirmative: the key is self-centered inequality aversion.
- No other deviation from the traditional approach is required.
Present-biased preferences

- Inequality aversion means that people resist inequitable outcomes and are willing to give up some material payoff to move in the direction of more equitable outcomes.
  - The notion of fairness considered here is thus *outcome-based*.
  - This is to be distinguished from *intention-based* fairness proposed by Rabin (1995).

- Inequality aversion is self-centered if people do not care per se about inequality that exists among other people but are only interested in the fairness of their own material payoff relative to the payoff of others.
Introduction

- Rabin (1995) proposes an intention-based model of fairness: people want to be nice to those who treat them fairly and want to punish those who want to hurt them.
- The notional of intention-based seems to capture the fundamental nature of our fairness concerns.
- The drawback is that it requires a psychological game, where the payoff depends not only on the strategies but also on the beliefs.
- The analysis of such a game can be quite complicated: for instance, Rabin’s model is restricted to two-person normal form games.
- Outcome-based models might be regarded as a reduced form of intention-based models.
An individual is inequality averse if he dislikes outcomes that are perceived as inequitable.

This question raises the difficult question of how individuals measure or perceive the fairness of outcomes.

The reference outcome that is used to evaluate a given situation is itself the product of complicated social comparison processes.

One key insight in social psychology is that relative material payoffs matter.
A simple model of inequality aversion

Two assumptions are made:

- There are subjects who dislike inequitable outcomes in addition to those who are purely selfish.
- Subjects suffer more from inequality that is to their material disadvantage than from inequality that is to their material advantage.

The utility function is given by

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

where \( \alpha_i \geq \beta_i \) and \( 1 > \beta_i \geq 0 \).
A simple model of inequality aversion

Some remarks:

- The utility function is linear in inequality aversion as well as in $x_i$ – while this may not be fully realistic, even this simple specification can explain surprisingly many observations.
- $\alpha_i \geq \beta_i$ means that a subject is loss averse in social comparisons.
- $\beta_i \geq 0$ rules out the existence of subjects who like to be ahead of others – in the context of the experiments here, this has no substantial impact.
- $1 > \beta_i$ is ruled out because it results in implausible.
The model is applied to two games: the ultimatum game and a simple market game. It is well known that in the ultimatum game, the gains from trade are shared relatively equally well. In market games, very unequal distributions are frequently observed. The challenge to any alternative theory is to explain both “fair” outcomes in the ultimatum game and competitive and rather “unfair” outcomes in market games.
In the ultimatum game, a proposer and a responder bargain about the distribution of a surplus of fixed size, normalized to one.

Let $s \in [0, 1]$ denote the responder’s share.

The proposer offers a share $s$ which may be accepted or rejected by the responder.

If the offer is rejected, both obtains a payoff of zero.

The unique subgame perfect equilibrium is that the proposer offers $s = 0$ and the responder accepts.
Ultimatum game

- This sharp theoretical prediction is in stark contrast to experimental evidence.
  - There are virtually no offers above 0.5.
  - The vast majority of offers in almost any study is in \([0.4, 0.5]\).
  - There are almost no offers below 0.2.
  - Low offers are frequently rejected, and the probability decreases with \(s\).
Proposition

*It is a dominant strategy for the responder to accept any offer \( s \geq 0.5 \), to reject \( s \) if \( s < s'(\alpha_2) := \alpha_2/1 + 2\alpha_2 \) < 0.5.*

- It is easy to see that the responder accepts any \( s \geq 0.5 \).
- If the responder accepts \( s < 0.5 \), the payoff is \( s + \alpha_2(1 - 2s) \).
- The responder accepts the offer if \( s - \alpha_2(1 - 2s) > 0 \iff s > \alpha_2/(1 + \alpha_2) \).
Proposition

If the proposer knows the preferences of the responder, he will offer

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

- The proposer would never offer \( s > 0.5 \), because \( s = 0.5 \) is surely accepted.
- If \( \beta_1 > 0.5 \), his utility is strictly increasing in \( s \) for all \( s \leq 0.5 \).
- If \( \beta_1 < 0.5 \), he simply offers the minimum amount that is accepted by the responder.
Proposition

If the proposer does not know the preferences of the responder but know that $\alpha_2$ is distributed according to $F(\alpha_2 \text{ with support } [\alpha, \bar{\alpha}])$, he will offer

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

- Without knowing the true value of $\alpha_1$, the offer may now be rejected.
The probability that an offer $s < 0.5$ is accepted is

$$p = \begin{cases} 
1 & \text{if } s \geq s'(\alpha), \\
F(s/(1-2s)) & \text{if } s'(\alpha) > s > s'(\overline{\alpha}), \\
0 & \text{if } s'(\overline{\alpha}) \geq s.
\end{cases}$$

- If $\beta_1 > 0.5$, the same argument applies.
- If $\beta_1 < 0.5$, there must exist an optimal offer $s \in (s'(\alpha), s'(\overline{\alpha})]$. 
It is well established that in a broad class of market games, prices converge to the competitive equilibrium.

The convergence can be observed even if the equilibrium is very “unfair”.

One example to capture this empirical feature is a simple market game in which many sellers want to sell one unit of a good to a single buyer who demands only one unit.
Market game with proposer competition

- There are $n - 1$ proposers who simultaneously propose a share $s_i \in [0, 1]$ to the responder.
- The responder has the opportunity to accept or reject the highest offer $\bar{s} := \max_i \{s_i\}$.
- If the responder rejects $\bar{s}$, no trade takes place and all players receive zero.
- A subgame perfect equilibrium is that at least two proposers make an offer of one, and the responder reaps all gains from trade.
- The prediction is confirmed in experiments.
Proposition

For any \((\alpha_i, \beta_i)\), there is a unique subgame perfect equilibrium in which at least two proposers offer \(s = 1\) which is accepted by the responder.

- Note first that as in the ultimatum game, the responder accepts any offer \(\bar{s} \geq 0.5\).
- That the responder rejects an offer \(\bar{s} < 0.5\) cannot be equilibrium either since proposer \(i\) can offer \(s_i = 0.5\) and improves his payoff.
- Hence, on the equilibrium path, \(\bar{s}\) must be accepted.
Market game with proposer competition

- Consider a candidate equilibrium with $\bar{s} < 1$.
- Any player can increase his payoff by offering $s_i$ slightly higher than $\bar{s}$.
- The only candidate that survives is then $\bar{s} = 1$. 
Market game with responder competition

- There is one proposer and $n - 1$ responders.
- The proposer offers $s$.
- Each responder simultaneously decides whether to accept or reject $s$.
- If more than one responder accept, the winner is determined by a random draw.
- If no one accepts, all receive zero.
Market game with responder competition

- The subgame perfect equilibrium is that the proposer offers $s \geq 0$ which is accepted by at least one responder.
- Not much evidence is obtained for this game.
- In one experimental study, though, offers are rather low compared to those typically observed in the ultimatum game.
Market game with responder competition

Proposition

Suppose that $\beta_1 < (n - 1)/n$. Then there exists a subgame perfect equilibrium in which all responders accept any $s \geq 0$, and the proposer offers $s = 0$. The highest offer that can be sustained in equilibrium is

$$\bar{s} = \min_{i \in \{2, \ldots, n\}} \frac{\alpha_i}{(1 - \beta_i)(n - 1) + 2\alpha_i + \beta_i} < 0.5.$$
Competition and fairness

- The model shows the limited role of fairness considerations in competitive environments.
- The crucial observation is that 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, which turns part of the inequality to his advantage.
- Competition renders fairness irrelevant if and only if none of the competing players can punish the monopolist by destroying some of the surplus and enforcing a more equitable outcome.
Since its publication, the paper has had enormous impacts on the study of fairness.

Its enormous acceptance and success depend partly on its tractability.

Despite its success, as always, it has had a fair share of criticisms and objections.

The recent debate: Binmore and Shaked (2010), Fehr and Schmidt (2010).