Reasons for Conflict: Lessons from Bargaining Experiments

by

ARMIN FALK, ERNST FEHR, AND URS FISCHBACHER *

In this paper we experimentally study the effects of fairness, spite, and reputation formation on conflict. We show that fairness preferences are a potential source of conflict and that intentions play an important role in the perception of fairness. Further, we show that feelings of spite may affect the occurrence of conflict. Finally, we study reputation formation as a possible source of conflict. We show that people invest in a reputation of being a tough bargainer. This does not automatically increase conflict, however. The reason is that through reputation, information about one's opponent is much better than in anonymous bargaining. (JEL: A 13, D 63, D 23, C 92, K 42)

1 Introduction

In this paper we experimentally study three possible sources of conflict: fairness, spite, and reputation formation. *Fairness* is a possible source of conflict because the feeling of being treated unfairly may lead to destructive, efficiency-reducing actions. A prime example is in labor relations, where unfair treatment by the firm (e.g., low wages or poor working conditions) may provoke workers to initiate strikes or to shirk. In several questionnaire studies (e.g., AGELL AND LUNDBORG [1995], BEWLEY [1999]) personnel managers indicate that even in recessions firms are unwilling to cut wages because they fear that pay cuts will be perceived as unfair by the workers and will hence invite shirking or destructive actions. Although by now a large literature has emerged that shows the relevance of fairness considerations in general, the precise determinants of what is considered as fair and unfair are still largely unknown. One important question concerns the role of intentions. Is perceived fairness solely determined by the consequences of an action, or (also)

* Financial support by the Swiss National Science Foundation (Project No. 1214-05100.97) and by the MacArthur Foundation (Network on Economic Environments and the Evolution of Individual Preferences and Social Norms) is gratefully acknowledged. This paper is also part of the EU-TMR Research Network ENDEAR (FMRX-CTP98-0238). We thank Christoph Engel, Gerd Gigerenzer, Jörg Oechssler, and the participants in the 20th International Seminar on New Institutional Economics in Wörlitz for helpful comments.

Journal of Institutional and Theoretical Economics
JITE 159 (2003), 171–187 © 2003 Mohr Siebeck – ISSN 0932-4569

by the intention behind that action? This question is of central importance for the understanding of conflicts, because if perceived fairness is not solely driven by consequences, there is scope for procedural rules that may help to mitigate conflicts.

Fairness (reciprocity) is one possible type of nonstandard preference that may cause conflict; *spite* is another one. Whereas a reciprocally motivated person engages in conflict only to retaliate for unfair behavior, a spiteful one has an interest in *reducing another person's payoff irrespective of whether that person was fair or unfair.* Naturally this may be a source of inefficiency and conflict: In principal–agent relations, e.g., spiteful agents shirk under a much broader set of circumstances than selfish agents. The reason is that spiteful agents aim at a payoff advantage relative to the principal, and by shirking they can redistribute income from the principal to the agent. Moreover, spiteful agents will also shirk in circumstances where fair agents do not shirk, because they want to redistribute income from the principal irrespective of whether they have been treated fairly or not. Fair agents, in contrast, will not shirk if they have been treated fairly. Thus, for spiteful agents the incentive compatibility condition will be very different from that for selfish or fair agents.

Considering *reputation formation* as a potential source of conflict is straightforward as well. In order to reach profitable outcomes in the future, bargaining parties have an incentive to build up a reputation of being tough bargainers. This in turn requires costly investments, e.g., strikes. In other words: If reputation formation is possible, conflicts can be used *strategically*.

The influence of fairness, spite, and reputation on conflict are not easy to study with field data. First, many economic relations are ongoing, and it is therefore impossible to disentangle fairness from repeated-game effects. Second, studying the effect of reputation requires a *ceteris paribus* variation of an economic relationship with and without the possibility of building up a reputation. This is impossible with field data. Third, unless the researcher knows precisely the rules of the game and the associated payoff functions, it is impossible to distinguish between motives of fairness and spite and to detect potential efficiency losses due to conflict.

In this paper we therefore study reasons and consequences of conflicts with the help of a very simple two-person bargaining game that was implemented in a controlled laboratory environment. The game under study is the so-called ultimatum game, which was first studied by GÜTH, SCHMITTBERGER, AND SCHWARZE [1982]. In this game a first mover, called the *proposer,* allocates a given pie between himself and a second mover, called the *responder*. In case the responder accepts, the proposed allocation is implemented. In case he rejects, both receive nothing. This game is well suited to study reasons for conflict. First, it captures in a stylized way a bargaining situation with opposing interests, i.e., with the potential for conflict. Second, the game is very simple in structure, which means that conflict cannot be attributed to the complexity and possible misunderstanding of the game. The simplicity of the game also implies that the motives that drive behavior in this game can be elicited straightforwardly. Third, the game makes consequences of conflict very obvious: If agreement fails, both parties are left with the inefficient outcome of zero payoffs.

Our main results are the following. In Section 2 we present evidence that fairness motives can in fact lead to conflict. We show, however, that perceived fairness is strongly influenced by an opponent's intention. The same unfavorable outcome is perceived as less unfair and accordingly associated with less conflict if one cannot infer unfair intentions from the action that led to the particular outcome. In Section 3 we discuss how the possibility for responders to build up a reputation shapes bargaining behavior and thus the occurrence of conflict. In particular we find that the desire to signal one's bargaining strength leads to tougher bargaining. However, even though reputation changes the outcome of the bargaining process, it does not lead to more frequent conflicts. The reason for this is that reputation reduces the proposers' uncertainty about the responders' acceptance thresholds and therefore allows a more fine-tuned offer policy.

Each section starts with a short description of the phenomenon and the description of the experimental design. This is followed by a discussion of the main results.

2 *Fairness: Consequences or Intentions?*

There is by now considerable evidence that fairness considerations affect economic behavior in many important areas. In bilateral bargaining situations agents frequently agree on rather egalitarian outcomes although the standard model with purely selfish preferences predicts rather unequal outcomes (ROTH [1995], CAMERER AND THALER [1995]). In competitive experimental labor markets with incomplete contracts, fairness considerations give rise to efficiency wage effects that generate stable deviations from the perfectly competitive outcome (FEHR AND FALK [1999]; BROWN, FALK, AND FEHR [2002]).[1]

The ultimatum game provides a prototypical experiment for studying the importance of fairness motives for the occurrence of conflicts. Standard economic theory (assuming common knowledge of rational and selfish actors) predicts offers of zero (or the smallest amount possible) and the acceptance of any positive offer. Contrary to this prediction, in many studies, the following three regularities have been observed: (i) there are practically no offers that exceed 0.5, (ii) the modal offers lie in a range between 0.4 and 0.5, and (iii) offers below 0.2 are extremely rare. Table 1 gives an overview of ten ultimatum-game experiments, which all confirm findings (i) to (iii). Moreover, the table reveals that results (i) to (iii) are robust with respect to changes in the cake size as well as with respect to different countries and cultures.

[1] Note that in this paper, we take fairness preferences as given. The focus is how fairness affects economic outcomes. Another very interesting and related question is concerned with the evolutionary roots of fairness preferences. For this question see for instance HUCK AND OECHSSLER [1999] or SETHI AND SOMANATHAN [2001]. Evolutionary models dealing with conflict are ARCE M. AND SANDLER [2003] and BRENNAN, GÜTH, AND KLIEMT [2003].

Table 1
Percentage of Offers below 0.2 and between 0.4 and 0.5 of the Total Pie in the Ultimatum Game

Study	No. of observations	Stake size (country)	Percentage of offers with $c < 0.2$	Percentage of offers $0.4 \leq c \leq 0.5$
CAMERON [1999]	35	Rp. 40.000 (Indonesia)	0	66
CAMERON [1999]	37	Rp. 200.000 (Indonesia)	5	57
FORSYTHE et al. [1994]	67	$5 and $10 (U.S.)	0	82
GÜTH, SCHMITTBERGER, AND SCHWARZE [1982]	79	DM 4–10 (Germany)	8	61
HOFFMAN, MCCABE, AND SMITH [1996]	24	$10 (U.S.)	0	83
HOFFMAN, MCCABE, AND SMITH [1996]	27	$100 (U.S.)	4	74
KAHNEMAN, KNETSCH, AND THALER [1986]	115	$10 (U.S.)	?	75[a]
ROTH et al. [1991]	116[b]	~ $ 10 (U.S., Slovakia, Israel, Japan)	3	70
SLONIM AND ROTH [1997]	240[c]	SK 60 (Slovakia)	0.4[d]	75
SLONIM AND ROTH [1997]	250[c]	SK 150 (Slovakia)	8	69
Aggregate result of all studies[e]	875		3.8	71

Note: [a] Percentage of equal splits. [b] Only observations of the final period. [c] Observations of all 10 periods. [d] Percentage of offers below 0.25. [e] Without KAHNEMAN, KNETSCH, AND THALER [1986].

Source: FEHR AND SCHMIDT [1999].

The most important reason why subjects make relatively high offers in the ultimatum game is the rejection risk of a low offer. Because low offers are generally rejected, the offer that maximizes the proposers' payoff is significantly higher than zero and typically not far below 50 percent. (In the experiment presented in Section 3, the

payoff-maximizing offer is for instance 40 percent.) The puzzle in the ultimatum game therefore is not the proposers' behavior but the rejection behavior of the responders. Rejecting positive offers implies that responders are willing to take costly actions in order to punish proposers for making unfairly low offers. This concern for fairness can best be understood as a form of *negative reciprocity*. In general it means that people are willing to punish unfair behavior even if doing so is costly to them. This is the type of conflict behavior we study in this section.[2] The crucial question to answer is, what shapes fairness perceptions? When do people feel treated fairly or unfairly? In particular we are interested in the question whether fair-minded people respond to fair or unfair *intentions* or whether they respond solely to fair or unfair *outcomes.*

Recently, different theoretical approaches have been developed that provide rigorous, albeit different, answers to this question. One class of fairness models – the inequity aversion models of FEHR AND SCHMIDT [1999] and BOLTON AND OCKENFELS [2000] – is based on the assumption that fairness intentions are behaviorally irrelevant. Another class of models (e.g., RABIN [1993], FALK AND FISCHBACHER [1999], DUFWENBERG AND KIRCHSTEIGER [1999]) assigns fairness intentions a major behavioral role.

To isolate the role of fairness intentions one ideally needs a treatment in which first movers can signal their fairness intentions and a treatment in which such signals are ruled out. The signaling of fairness intentions rests on two premises: (i) The first mover's choice set actually allows the choice between a fair and an unfair action, and (ii) the first mover's choice is under the *full* control of the first mover. In this paper we analyze the first premise by manipulating the choice set of the first mover in a series of reduced ultimatum games.

2.1 Experimental Design

We study four different versions of a simple reduced ultimatum game (compare FALK, FEHR, AND FISCHBACHER [2003]). In all four games the pie that can be split between the proposer and the responder is 10 points. The proposer has to decide between two actions, X and Y (see Table 2). Action X is the same in all games, namely the "unfair" offer 8/2, i.e., for the proposer to keep 8 and to offer 2 to the responder. Action Y varies across games. In the 5/5 game the alternative to keeping 8 is to offer an equal split, i.e., to offer 5/5. In the 2/8 game the alternative offer to 8/2 is 2/8. In the 8/2 game the proposer has in fact no alternative at all, i.e., he is forced to propose 8/2. Finally, in the 10/0 game the alternative offer is 10/0. In all games the proposed allocation is implemented if the responder accepts the offer. Payoffs are zero for both parties if the responder rejects the offer. In order to get data on the responder's behavior independent of the proposer's decisions, we employed the strategy method, i.e., responders had to specify complete strategies. Thus, every

[2] For an overview on reciprocity see FEHR AND GÄCHTER [2000].

responder had to indicate his action at both decision nodes, i.e., for the case of an *X* and for the case of a *Y* offer, before knowing the proposer's actual decision.

Table 2
Four Mini Ultimatum Games
(payoff proposer/payoff responder)

Game	*X* offer	*Y* offer
5/5	8/2	5/5
2/8	8/2	2/8
8/2	8/2	8/2
10/0	8/2	10/0

At the beginning of the experiment subjects were randomly assigned to the proposer or the responder role, and they kept that role in all four games. Subjects faced the games in varying order, and in each game they played against a different anonymous opponent. They were informed about the outcomes of all four games, i.e., about their opponents' choices, only *after* they had made their decision in all games. This procedure avoids income effects and rules out the possibility that subjects' behavior is influenced by previous decisions of their opponents. After the end of the fourth game, subjects received a show-up fee of CHF 10.00 plus their earnings from the experiment. For each point earned they received CHF 0.80, so that in all four games together CHF 32.00 (about $23 at the time) was at stake. The experiment took approximately 40 min. All experiments discussed in this paper were computerized and programmed with the experimental software z-Tree (FISCHBACHER [1999]).

2.2 Behavioral Predictions

In analyzing the four games we concentrate on the rejection of the *X* offer. Note that *X* is the same in all games. From a consequentialistic point of view the variation of the *Y* alternatives should therefore make no difference for the question whether *X* is rejected or accepted. If intentions matter, however, rejections may differ, because the same offer *X* may signal different things across games.

The standard economic model with selfish preferences predicts that in all games the allocation 8/2 is *never* rejected. After all, having 2 is better than having 0, i.e., if only material payoffs matter, rejecting is strictly dominated by accepting the *X* offer.

The inequity aversion models by FEHR AND SCHMIDT [1999] and BOLTON AND OCKENFELS [2000] predict that if *X* is rejected, it should be rejected with the same rate across all games. Since these models capture people's dislike for inequality, they are consistent with positive rejection rates. However, since they disregard the fact that identical outcomes may be perceived as more fair or less fair, depending

on the intentions of the first mover, they are not consistent with different rejection rates of the *X* offer across the four games.

Models of reciprocity that are based on the idea that unfair *intentions* trigger rejections are consistent with different rejection rates across the four games (RABIN [1993], FALK AND FISCHBACHER [1999], and DUFWENBERG AND KIRCHSTEIGER [1999]). The reason for this is that, depending on the alternative to the unfair 8/2 offer, the choice of the 8/2 offer signals different intentions of the proposer. Intuitively, one would expect, e.g., that in the 5/5 game a proposal of 8/2 clearly signals an unfair intention, because the proposer could have proposed the egalitarian offer 5/5. In the 2/8 game, offering 8/2 may still be perceived as an unfair intention, but probably less so than in the 5/5 game, because the only alternative available to 8/2 gives the proposer much less than the responder. In a certain sense, therefore, the proposer has an excuse for not choosing 2/8, because one cannot unambiguously infer from his unwillingness to propose an unfair offer *to himself* that he wanted to be unfair to the responder. Thus, we would expect the rejection rate of the 8/2 offer in the 5/5 game to be higher than in the 2/8 game. In the 8/2 game the proposer has no choice at all, so that the proposer's behavior cannot be judged in terms of fairness intentions. Responders can only judge the fairness of the *outcome* 8/2, and if they exhibit sufficient aversion against inequity, they will reject. Since any attribution of unfairness to the proposer's behavior is ruled out here, we expect an even lower rejection rate than in the 2/8 game. Finally, offering 8/2 in the 10/0 game may even be perceived as a fair (or less unfair) action, so that the rejection rate of 8/2 is likely to be the lowest in this game.

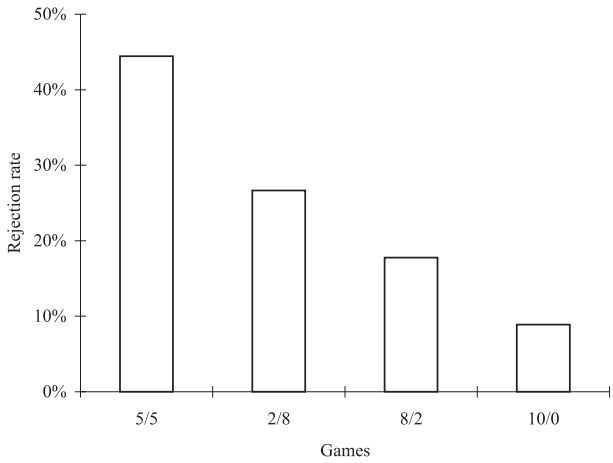
2.3 Experimental Results

Figure 1 shows the main results. The bars represent the percentage of responders that reject the 8/2 offer in the different games. The rejection rate in the 5/5 game is highest: 44.4 percent rejected the 8/2 offer. 26.7 percent rejected the 8/2 offer in the 2/8 game, 18 percent in the 8/2 game, and 8.9 percent in the 10/0 game.

These results indicate that perceived fairness is substantially shaped by the intentions of subjects' opponents. The fact that the rejection rate for the same offer varies between 44.4 percent and 8.9 percent is unambiguous evidence against a purely consequentialistic notion of fairness. In their fairness judgment people carefully distinguish between motives and consequences. If someone has no choice to behave fairly (as in the 10/0 game) or if he is left without any alternative (as in the 8/2 game), the perceived unfairness is clearly different from that in situations where a proposer has the option to reach a fair outcome (as in the 5/5 game).[3]

[3] Similar results are found, e.g., in FALK, FEHR, AND FISCHBACHER [2002], MCCABE, RIGDON, AND SMITH [2003], CHARNESS [2003], BRANDTS AND SOLA [2001], and BLOUNT [1995]. Authors who report evidence questioning the importance of intentions are BOLTON, BRANDTS, AND OCKENFELS [1998], COX [2000], and OFFERMAN [2002]. See FALK, FEHR, AND FISCHBACHER [2001] for a detailed discussion of possible explanations and differences between these studies.

Figure 1
Rejection Rates across the Four Mini Ultimatum Games



These results are of great practical and theoretical interest. At the theoretical level the question obviously concerns the correct modeling of fairness preferences, but also standard utility theory in general. Standard utility theory assumes that the utility of an action is determined solely by its consequences and not by the intention behind it. Given the behavioral importance of intentions, this *consequentialism* inherent in standard utility models is strongly questioned. At the practical level the issue is important because the nature of conflicts is likely to be affected by attribution of intentions. Fairness attributions are likely to influence decision-making in firms and other organizations as well as in markets and the political arena. The decision to cut wages or fringe benefits clearly decreases the material payoffs of workers. If the response of the negatively affected workers or unions also takes into account the firm's intentions, it will be much easier to prevent opposition, strike, or conflict when the firm can credibly claim that it is somehow forced to take the action – by law, by international competition, by the threat of bankruptcy, or by some other external force. It is, therefore, no coincidence that the rhetoric of business leaders often appeals to the phrase "there is no alternative". If there is indeed no alternative, it is not possible to attribute unfair intentions to the action, because the decision-maker cannot be held responsible for the action. If, in contrast, there are obvious alternative actions available, it is much easier for the affected parties to attribute unfair intentions to the action, and as a consequence, opposition and conflict will be much stronger.[4]

[4] The attribution of intentions is also important in law (HUANG [2000]). Intentions often distinguish between whether the same action is a tort or a crime and whether a tort should involve punitive damages. Other distinctions made in criminal law con-

The fact that fairness perceptions are not purely consequentialistic points to the potential of procedural rules to circumvent conflicts: Our argument is that if an unfavorable outcome is not intentionally caused by the favored person, conflict is less likely to occur. Institutions that guarantee procedural fairness may achieve this. To illustrate, assume, e.g., that a fair *random* device determines about the split of a pie, i.e., no unfair intentions can be inferred from any outcome. This resembles the situation in the 8/2 game where the proposers had no alternative and where the rejection rate was very low – even though the responders received only a small share of the pie. The scope for fair procedures to mitigate distributional conflict is further shown, e.g., in LIND AND TYLER [1988] or BOLTON, BRANDTS, AND OCKENFELS [2002]. These findings suggest that communities, groups, or societies that ensure procedural justice are better equipped to deal with conflicting interests and structural changes and are more likely to reach socially desirable outcomes.

3 *Spite as a Source of Conflict*

In the previous section we have shown that the perception of being treated in an unfair manner may trigger negative reciprocity, i.e., rejection and conflict. A negative reciprocal action aims at punishing an unfair action. This is why in the 5/5 game above subjects reject the 8/2 offer. Spiteful preferences imply that someone values the payoff of others negatively, *irrespective* of whether they behaved fairly or unfairly. Whereas a reciprocal responder rejects an offer in the ultimatum game only if it was considered as unfair, a spiteful player rejects an offer if doing so increases the payoff difference to his or her advantage. Interestingly, the spite motive has so far been neglected by most fairness theories.[5]

3.1 Experimental Design

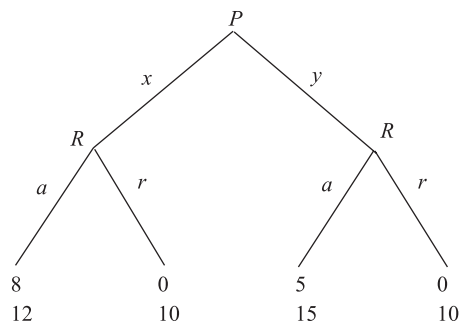
In the 5/5 game we cannot distinguish between spite and negative reciprocity. Both motives are compatible with rejection of the 8/2 offer. To study the potential of spite for conflict we therefore conducted another ultimatum game. The game in this section is basically the 5/5 game with the only difference that the payoff of the responder was increased by 10 tokens at *each* final node (compare Figure 2). Thus, after the proposer chooses the move X, the responder faces the choice between 8/12 and 0/10. As in the games discussed above, we employed the strategy method, i.e., responders had to specify complete strategies.

---

cern whether an action is taken purposely, knowingly, recklessly, or negligently (see Model Penal Code §2.02(1)–(2)). Thus, the penal code (which represents a codified broad sense of justice) distinguishes quite carefully between the consequences of an action and the intentions underlying the action

[5] See, however, the model by LEVINE [1998], who models spiteful types and altruistic types in the same framework.

Figure 2
Game Tree of the Experiment Testing for Spite



Note that the rejection of the 8/12 offer is incompatible with preferences for fairness as conceptualized in most fairness models.[6] The reason is that most conceptions of fairness rely on an equitable payoff distribution to determine fair and unfair outcomes. Thus, if the responder receives a higher payoff than the proposer, the responder is not treated unfairly. Therefore there is no reason to punish the proposer by rejecting the 8/12 offer. However, if responders have spiteful preferences, rejections may occur. Note that if the responder prefers 0/10 over 8/12, he is willing to give up 2 tokens to increase the payoff difference from 4 to 10 units. The rejections of the 8/12 offer in this game thus provide a measure of the quantitative importance of a preference for spite. Finally note that a rejection of the 5/15 offer cannot be explained with a preference for spite, because a rejection of the 5/15 offer leaves the payoff difference intact.

3.2 Results

The results from this game in fact reveal the existence of spiteful behavior. 24 percent of the responders rejected 8/12 in favor of 0/10, and only 4 percent rejected 5/15 in favor of 0/10. This suggests that a nonnegligible fraction of the subjects are willing to spend money to increase the payoff difference between themselves and the reference agent – who is in this case the proposer.

This result is confirmed by a public-goods experiment with a subsequent punishment stage (FALK, FEHR, AND FISCHBACHER [2001]), where the punishment behavior of about 20 percent of the subjects cannot be explained with preferences for fairness but is compatible with spite. Compared to fairness, the quantitative

[6] An exception is DUFWENBERG AND KIRCHSTEIGER [1999]. According to their model, a rejection of the 8/12 offer is possible if the responder thinks that it is unfair that the proposer did not make the 5/15 offer, which is the most profitable offer to him.

importance of spite is certainly lower. Nevertheless spite seems to be a robust and important source of efficiency-reducing conflict.

4 Does Reputation Increase Conflict?

In the previous two sections we have concentrated on one-shot games where reputation concerns cannot play a role. In reality, however, many relations involve ongoing repeated interactions, where it is possible to condition behavior on each other's history. Labor relations, e.g., usually last for several years. But also the relation between unions and industry interest groups are long-term relations. In this type of long-lasting relationships conflicts may be used strategically. Players in a repeated game have a material incentive to make a costly signal that they are tough bargainers. In the context of an ultimatum game this may induce responders to reject offers that they would accept in a one-shot environment. This investment in reputation is – in addition to fairness motives – a potential source of conflict, in particular if proposers refuse to raise their offers. On the other hand, reputation implies that people know much about their opponent. Given that responders are quite heterogeneous in their rejection behavior, knowing a responder's rejection record reduces uncertainty about the responder's type and allows a better fine tuning on the side of the proposer. This second effect of reputation is *ceteris paribus* likely to reduce conflict.

4.1 Experimental Design

To study the effect of reputation on conflict we contrast bargaining behavior in two treatments (compare FEHR AND FISCHBACHER [2002]). In the baseline treatment (B-treatment), reputation was ruled out by design, while in the reputation treatment (R-treatment) proposers were informed about how their responders had acted in the previous periods. In both treatments subjects played the ultimatum game for 10 periods. The total pie that had to be distributed was 10 points, and proposers were free to make any of the offers 0, 1, ..., 5.[7] Responders had to indicate an acceptance threshold. If the actual offer was larger than or equal to that threshold, the offer was automatically accepted; otherwise it was rejected. In both treatments, proposers and responders were never matched with the same subject more than once. In the B-treatment the interaction between proposers and responders was completely anonymous, and nothing was known about the history of one's opponent. In the R-treatment, on the other hand, a proposer was informed about the rejection behavior of the responder. This information was presented in a table, where each proposer was informed about which offers his or her current responder had received and how many of these offers he or she had accepted. The longer the

[7] The action space of the proposers was restricted to 5 in order to keep the reputation record clear and transparent (see below). It is known from other ultimatum-game experiments that offers above 50 percent are extremely rare.

game had been going on, the better was the information a proposer received about his counterpart.

4.2 Results

The main results are shown with the help of Figures 3 to 6. Figure 3 displays the evolution of the average thresholds and offers in the two treatments. Let us first consider the thresholds. In the B-treatment the average threshold is about 3, which is well above the level predicted by standard economic theory and supports the relevance of fairness motives. Figure 3 also shows that – as hypothesized – the thresholds are substantially higher in the R- than in the B-treatment. On average thresholds are slightly below 4. Responders obviously understand that it may pay to signal a tough reputation, i.e., to reject offers they would have accepted in a one-shot setting. The increase in thresholds is supported by Figure 4, which shows the average thresholds of each individual, sorted from low to high thresholds.[8] It is clear from Figure 4 that thresholds are significantly higher in the R- than in the B-treatment. For example, while in the B-treatment more than 60 percent of the responders have an average threshold of at least 3, the corresponding number in the R-treatment is 80 percent. In the R-treatment 50 percent of the subjects' average thresholds are above or equal to 4.

Figure 4 also reveals that the heterogeneity between subjects is quite substantial. In the B-treatment, e.g., four subjects have average thresholds between 0 and 1, eight subjects between 1 and 2, and five subjects above 4. This heterogeneity of fairness preferences is important in that it implies a large amount of uncertainty on the side of the proposers about a particular responder's type. As a consequence, it makes a fine-tuned offer policy extremely difficult, particularly in the B-treatment.

How is the offer policy of proposers affected by the different treatments? Figure 3 shows that not only do thresholds increase, but also offers are higher in the R- than in the B-treatment. While the average offer is 3.2 in the B-treatment, it is 3.9 in the R-treatment. Obviously proposers adapt their offers to the higher thresholds.

How does the increase in thresholds and offers influence the rejection behavior? Figure 5 provides the answer. The figure shows the evolution of the frequency of rejections in the B- and the R-treatment. Two things are worth pointing out. First, in both treatments there is a general tendency of declining conflict. In particular, in the first three periods there is a sharp drop in the rate of rejections. This drop is more pronounced in the R- than in the B-treatment, which is compatible with the interpretation that the learning of the opponents' thresholds is faster in the R- than in the B-treatment. Second, if anything, rejections are *lower* in the R- than in the B-treatment. The average rejection rate across all periods is 25.3 percent in the

[8] The graph therefore shows the transposed cumulative distribution of the average thresholds.

Figure 3
Evolution of Thresholds and Offers

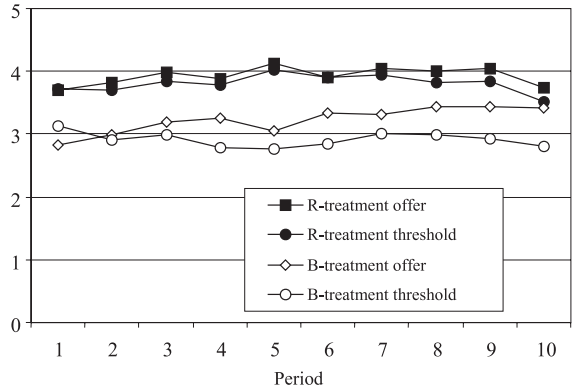
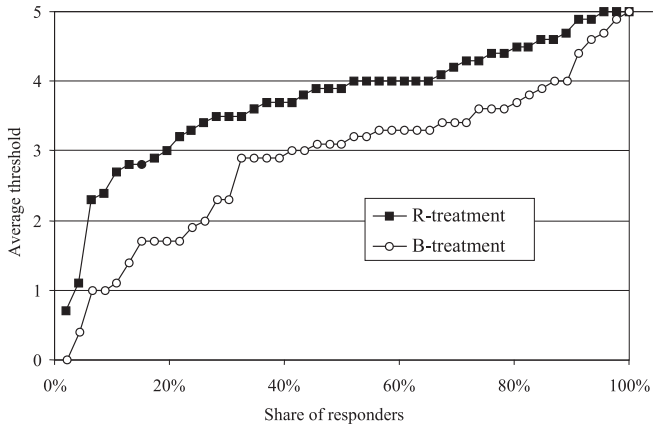


Figure 4
Average Thresholds of Subjects (sorted)



B- and 24.1 percent in the R-treatment. This difference, however, is not statistically significant.[9]

To sum up our observations so far: Responders invest heavily in their reputation, with the result that thresholds are significantly higher in the R- than in the

[9] To test for the difference in rejection rates we used a Mann–Whitney test to compare session averages ($p = 0.309$). As a second test, we ran regressions with robust standard errors (sessions as clusters). Again, no significant difference was found.

Figure 5
Evolution of Rejections in the B- and the R-Treatment

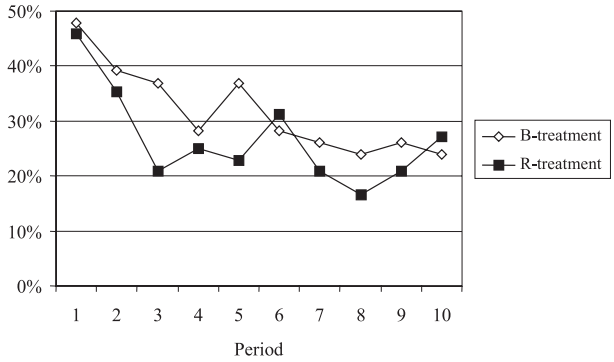
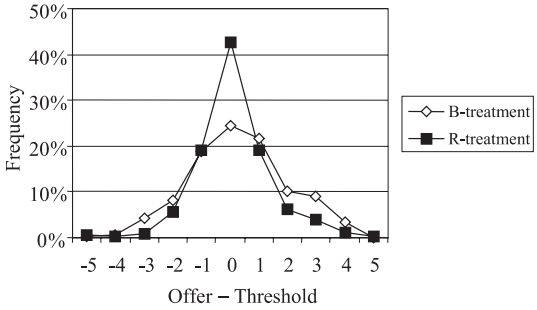


Figure 6
Histogram of Difference between Offer and Threshold



B-treatment. However, conflict and rejections are actually slightly lower in the R- than in the B-treatment, indicating that proposers adapt their offer policy to the different strategic environments. A closer look at the offer policy reveals that the difference between average offers and thresholds is larger in the B- than in the R-treatment (see also Figure 3). Figure 6 shows a histogram with the differences between offers and threshold in both treatments. Notice that in the B-treatment most observations lie to the right of the middle, indicating that most of the time proposers' offers exceeded responders' thresholds. We interpret this positive difference between offer and threshold as a "risk premium": Given the heterogeneity of fairness preferences of responders (compare Figure 4), proposers are very uncertain about their opponents' thresholds. They are therefore forced to pay such a premium in order to avoid too frequent rejections.

Figure 6 also shows that the risk premium is systematically lower in the R- than in the B-treatment. In particular, in more than 40 percent of the cases proposers make an offer that exactly corresponds to the respective threshold. In the B-treatment this number is less than 25 percent. This means that proposers are rather successful in fine-tuning their offers in the R-treatment. This fine tuning is possible because proposers have detailed information about the individual threshold levels in the R- but not in the B-treatment. This holds the more, the longer the game is played (see the sharp drop of rejections in the first three periods in Figure 5).

Taken together, the possibility of forming a reputation does not increase conflict. Instead it leads to higher thresholds, higher offers, and thus a more even share of the pie than in the B-treatment. Rejections are slightly lower in the R- than in the B-treatment, which is equivalent to saying that overall efficiency is increased as well. Reputation is not only beneficial to the responders but also has at least a relative advantage for proposers: Given that the responder's reputation implies better knowledge of the opponent, proposers can fine-tune their offers. As a result, they reach a similar acceptance rate in the R- to that in the B-treatment with a lower "risk premium". In fact, average payoffs for proposers are only slightly lower in the R- than in the B-treatment.

A final remark is in place. In our setup only one side of the traders could create a reputation. In many real-world settings reputation is two-sided. It is an interesting question whether allowing both sides to invest in a tough reputation increases or decreases the likelihood of conflict in bargaining. This merits further investigation.

References

AGELL, J., AND P. LUNDBORG [1995], "Theories of Pay and Unemployment: Survey Evidence from Swedish Manufacturing Firms," *Scandinavian Journal of Economics*, 97, 295–307.

ARCE M., D., AND T. SANDLER [2003], "An Evolutionary Game Approach to Fundamentalism and Conflict," *Journal of Institutional and Theoretical Economics,* 159, 132–154.

BEWLEY, T. F. [1999], *Why Wages Don't Fall during a Recession*, Harvard University Press: Cambridge, MA.

BLOUNT, S. [1995], "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Process,* 63, 131–144.

BOLTON, G. E., J. BRANDTS, AND A. OCKENFELS [1998], "Measuring Motivations for the Reciprocal Responses Observed in a Simple Dilemma Game," *Experimental Economics,* 1, 207–220.

– –, – –, AND – – [2002], "Fair Procedures: Evidence from Games Involving Lotteries," Working Paper, University of Barcelona.

BOLTON, G., AND A. OCKENFELS [2000], "ERC – A Theory of Equity, Reciprocity and Competition," *American Economic Review,* 90, 166–193.

BRENNAN, G., W. GÜTH, AND H. KLIEMT [2003], "Trust in the Shadow of the Courts," *Journal of Institutional and Theoretical Economics,* 159, 16–36.

BRANDTS, J., AND C. SOLA [2001], ''Reference Points and Negative Reciprocity in Simple Sequential Games," *Games and Economic Behavior*, 36, 138–157.

BROWN, M., A. FALK, AND E. FEHR [2002], "Contractual Incompleteness and the Nature of Market Interactions," Working Paper No. 38, IEW, University of Zurich.

CAMERER, C., AND R. THALER [1995], "Ultimatums, Dictators, and Manners," *Journal of Economic Perspectives,* 9, 209–219.

CAMERON, L. [1999], "Raising the Stakes in the Ultimatum Game: Experimental Evidence from Indonesia," *Economic Inquiry*, 37, 47–59.

CHARNESS, G. [2003], "Attribution and Reciprocity in a Simulated Labor Market: An Experimental Investigation," *Journal of Labor Economics*, forthcoming.

COX, J. C. [2000], "Trust and Reciprocity: Implications of Game Triads and Social Contexts," Working Paper, Department of Economics, University of Arizona, Tucson.

DUFWENBERG, M., AND G. KIRCHSTEIGER [1999], "A Theory of Sequential Reciprocity," Working Paper, CentER for Economic Research, Tilburg.

FALK, A., E. FEHR, AND U. FISCHBACHER [2001], "Informal Sanction," Working Paper No. 59, IEW, University of Zurich.

– –, – –, AND – – [2002], "Testing Theories of Fairness," Working Paper No. 63, IEW, University of Zurich.

– –, – –, AND – – [2003], "On the Nature of Fair Behavior," *Economic Inquiry*, 41, 20–26.

– – AND U. FISCHBACHER [1999], "A Theory of Reciprocity," Working Paper No. 6, IEW, University of Zurich.

FEHR, E., AND A. FALK [1999], ''Wage Rigidities in a Competitive Incomplete Contract Market*", Journal of Political Economy,* 107, 106–134.

– – AND U. FISCHBACHER [2002], "Retaliation and Reputation," Mimeo, University of Zurich.

– – AND K. SCHMIDT [1999], "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

– – AND S. GÄCHTER [2000], "Fairness and Retaliation – The Economics of Reciprocity," *Journal of Economic Perspectives*, 14, 159–181.

FISCHBACHER, U. [1999], "z–Tree: Zurich Toolbox for Readymade Economic Experiments," Working Paper No. 21, IEW, University of Zurich.

FORSYTHE, R., J. HOROWITZ, N. SAVIN, AND M. SEFTON [1994], "Fairness in Simple Bargaining Experiments," *Games and Economic Behaviour*, 6, 347–369.

GÜTH, W., R. SCHMITTBERGER, AND B. SCHWARZE [1982], "An Experimental Analysis of Ultimatium Bargaining," *Journal of Economic Behavior and Organization,* 3, 367–388.

HOFFMAN, E., K. MCCABE, AND V. SMITH [1996], "On Expectations and Monetary Stakes in Ultimatum Games," *International Journal of Game Theory*, 25, 289–301.

HUANG, P. H. [2000], "Reasons within Passions: Emotions and Intentions in Property Rights Bargaining," *Oregon Law Review,* 79, 435–478.

HUCK, S., AND J. OECHSSLER [1999], "The Indirect Evolutionary Approach to Explaining Fair Allocations," *Games and Economic Behavior,* 28, 13–24.

KAHNEMAN, D., J. L. KNETSCH, AND R. THALER [1986], "Fairness as a Constraint on Profit Seeking: Entitlements in the Market," *American Economic Review*, 76, 728–741.

LEVINE, D. [1998], "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics,* 1, 593–622.

LIND, E. A., AND T. R. TYLER [1988], *The Social Psychology of Procedural Justice,* Plenum Press: New York.

MCCABE, K. A., M. L. RIGDON, AND V. L. SMITH [2003], "Positive Reciprocity and Intentions in Trust Games," *Journal of Economic Behavior and Organization,* forthcoming.

OFFERMAN, T. [2002], "Hurting Hurts More than Helping Helps: The Role of the Self-Serving Bias," *European Economic Review,* 46, 1423–1437.

RABIN, M. [1993], "Incorporating Fairness into Game Theory and Economics," *American Economic Review,* 83, 1281–1302.

ROTH, A. V. [1995], "Bargaining Experiments," pp. 253–348 in: J. E. Kagel and A. E. Roth (eds.), *Handbook of Experimental Economics*, Princeton University Press, Princeton, NJ.

ROTH, A., V. PRASNIKAR, M. OKUNO-FUJIWARA, AND S. ZAMIR, [1991], "Bargaining and Market Behaviour in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review*, 81, 1068–1095.

SETHI, R., AND E. SOMANATHAN [2001], "Preference Evolution and Reciprocity," *Journal of Economic Theory,* 97, 273–297.

SLONIM, R., AND A. ROTH [1997], "Financial Incentives and Learning in Ultimatum and Market: An Experiment in the Slovak Republic," *Econometrica*, 66, 569–596.

Armin Falk
Ernst Fehr
Urs Fischbacher
Institute for Empirical Research in Economics
University of Zürich
Blümlisalpstrasse 10
8006 Zürich
Switzerland
E-mail:
falk@iew.unizh.ch
efehr@iew.unizh.ch
fiba@iew.unizh.ch

Social Intelligence in Games

Comment

by

GERD GIGERENZER AND RICHARD MCELREATH

1 Introduction

In his essay "An anthropologist on Mars," the neurologist Oliver Sacks describes his encounter with Temple Grandin, a highly remarkable autistic person who holds a Ph.D. and teaches at Colorado State University. Like other autists, Temple Grandin seems to be largely devoid of social intelligence. When Temple watches the social games normal people play, the social rituals they follow, and their powerful emotions that can change from love today to hate tomorrow, she feels, as she put it, like an anthropologist on Mars. As SACKS [1995, p. 270] describes her, "Lacking it [social intelligence], she has instead to 'compute' others' intentions and states of mind, to try to make algorithmic, explicit, what for the rest of us is second nature."

Oliver Sacks studies people who have neurological disorders. Economists and cognitive psychologists usually do not; they study normal people. Nevertheless, there is a similarity between the autistic personality and the rational models that many economists and some psychologists embrace. If by "rational" we mean conforming to the classical expected-utility model, or backward induction, then we have an "autistic" conception of human rationality. Just like Temple Grandin, *homo economicus* – defined in that way – lacks social intelligence and is puzzled by the strange behavior of normal people.

Experimental economists rely on simple games to elicit behavior that is blatantly at odds with the assumption that players are attempting to maximize their own expected utility. Consensus is growing that the rational-choice model is descriptively wrong. This message has been hammered home by two groups of researchers: (i) experimental economists studying social games such as the ultimatum game, and (ii) psychologists demonstrating that individual people's judgments and decisions violate the axioms or consequences of the expected-utility calculus, the laws of probability, or logic. Real humans are capable of acting and punishing altruistically and of valuing fairness, and are guided by emotions, obligations, feelings of guilt, and other moral sentiments. What to do with a model of rational man who is socially unintelligent, that is, a model that is descriptively wrong?

One reaction is just to go on, close one's eyes, and ignore this research. A second reaction is to engage in a "repair program": One sticks with the expected-utility

Journal of Institutional and Theoretical Economics
JITE 159 (2003), 188–194 © 2003 Mohr Siebeck – ISSN 0932-4569

framework, but adds one variable that has been experimentally shown to influence choice, such as regret. Or one tinkers with the utility and probability functions, and changes the shape of the curves to accommodate other experimental results, as in prospect theory. There is a third way, a radical departure from the repair program: to dispense with the expected-utility framework entirely, and start afresh on the basis of empirically rooted knowledge about the human mind and its capabilities. This, in my view, is the program underlying the work summarized in FALK, FEHR, AND FISCHBACHER [2003].[1] For instance, in the book *Bounded Rationality: The Adaptive Toolbox*, Reinhard Selten and one of us formulated as our goal "to promote bounded rationality as the key to understanding how actual people make decisions without utilities and probabilities" (GIGERENZER AND SELTEN [2001, p. i]). Quantitative probabilities, utilities, and optimization do not play a large role in the actual processes of the human mind, whereas fast and frugal heuristic processes, such as emotions, name recognition, aspiration levels, imitation learning, limited search, stopping rules, and one-reason decision-making, do (GIGERENZER, TODD, AND THE ABC RESEARCH GROUP [1999]).

2 *Do Intentions or Only Consequences Matter?*

In three experiments, FALK, FEHR, AND FISCHBACHER [2003] use the ultimatum game as a window on the actual behaviors, motivations, and cognitions of people in bargaining situations. Specifically, they study three possible reasons why people reject or do not reject a given offer: the roles of intention, spite, and reputation. We applaud this experimental approach, because it brings social intelligence into a previously nonsocial theory of rationality in games.

Standard utility theory, so the authors write, "assumes that the utility of an action is determined solely by the consequences of an action and not by the intention behind it" (FALK, FEHR, AND FISCHBACHER [2003, p. 178]). By restricting the options of the proposer to only two possible offers, so that intentions and behavior are no longer coupled, the first experiment ingeniously demonstrates that intentions and not only outcomes count. We are quite fond of experimental economics, but as you can infer from the analogy with autism, we are impatient with the slow move towards understanding the Martians, that is, us humans.

It is both interesting and frustrating that economists still debate whether or not inferred intentions matter to people. A glance into the empirical literature on moral development in children or on notions of justice in adults would have revealed dozens or hundreds of studies that have experimentally demonstrated that even very young

[1] It was surprising to hear Armin Falk saying in the discussion that the work on fairness, intention, spite, and other forms of social intelligence is a straightforward extension of expected-utility maximization. To us, that seems to be a strategic move to reduce conflict with hard-nosed utility maximizers; Herbert Simon and Allan Newell used the same move many years ago when they claimed that their new conception of the mind as computer was in the tradition of the earlier *Gestalt* psychology (GIGERENZER [2000, ch. 2]).

children already distinguish between bad outcomes that a person caused intentionally and those that were caused unintentionally or because there was no choice. Children punish differently in these situations, as do courts in countries where the legal rules distinguish between intended and unintended harm. In Kohlberg's theory of moral development, the most influential theory on this topic, the distinction between consequences and intentions of an action arises in the human mind at Stage 2, which occurs around the ages of five to six (KEASY [1978]). Note that Stage 2 is not the highest level of moral development in his theory; there are six stages in total. Conducted in a wide variety of situations, these experimental studies in psychology have already provided the empirical evidence that Falk et al.'s first experiment found in a specific social situation, the ultimatum game. But this experimental literature is not discussed by Falk et al. This omission is part of a game played by most social scientists, and not only economists: Read and cite only what your peers in your journals have written, and ignore the rest. It is an "identification" game that is played all too frequently, and we are all guilty of participating in it.

Imagine playing the ultimatum game with a computer, and assume that you do not attribute intentions to it. The same humans who reject low offers from other humans would rarely reject low offers from computers. BLOUNT [1995] performed an ultimatum-game experiment in which responders played both against a computer making random offers and against human subjects. She found that people rejected low offers from other subjects, as is the usual result, but very rarely rejected offers from the computer. The issue is not low offers alone; it is the intentions behind them that make you feel exploited, angry, and ready to retaliate.

There are theories, such as FEHR AND SCHMIDT [1999] and BOLTON AND OCKENFELS [2000], that assume that intentions of fairness are behaviorally irrelevant. Falk et al.'s experiment using the ultimatum game provides convincing experimental evidence against that position (see also MCCABE, SMITH, AND LEPORE [2000]). This is a definite step towards helping the anthropologists understand the Martians. However, in this case the anthropologists – i.e., the economists – are themselves Martians. In addition, there is already a large amount of literature available showing that inferred intentions matter. One therefore wonders why this particular experiment needed to be conducted: Could it really be that economists did not know the answer all along?

3 On the Attribution of Spiteful Behavior

Falk et al. present inferred intentions as an explanation for why responders reject or do not reject low offers, and they add spite and reputation to this list. Spite is, interestingly, another social behavior that autists can hardly understand, just like deception and lying (GRANDIN [1995]). However, spitefulness is rarely a personality trait, but rather a behavior that emerges in a specific class of social interactions, such as competitive interactions in which coming out first, rather than maximizing one's

individual gain, matters. Moreover, it is not clear whether the experiment succeeds in demonstrating the operation of spite. The authors may infer the existence of malevolence far too quickly, or else the short description of this experiment leaves out relevant information. In a strict sense, they seem to consider spite to be any situation in which the ego lowers its own payoff to reduce the payoff of the other individual. This is the same as the definition used in evolutionary biology.

Here are two alternative interpretations. First, from the point of view of the responder, it is of advantage if the proposer offers the choice *Y* rather than *X* (see Figure 2 in FALK, FEHR, AND FISCHBACHER [2003]), because then the responder can get 15 points rather than 12 points. If the responder becomes angry when the proposer nevertheless offers choice *X*, the responder may think of punishing the proposer by responding with the option [0;10], which is interpreted as indicating spite. However, the behavior of the responder can equally be interpreted as retaliation: You did not provide me with my best outcome, so I will do the same to you. Consistent with this alternative interpretation is the observed result that when the proposer offers the choice *Y*, this kind of retaliation is rarely observed. The experiment, however, does not seem to be able to distinguish between the two interpretations. The rejoinder to this alternative explanation is that the experiment was a one-shot game and thus retaliation in order to build up reputation or change the other's behavior makes no sense. We address this rejoinder in the next section.

There is a second alternative interpretation that also does not need to assume a spiteful personality. If the responder did not consider the final 10 units of payoff to be part of the fairness calculation (and why should they, since the responder gets this portion of the payoff regardless of what the proposer does?), then rejection does not have to indicate spite. Instead, it can indicate the same motivations that led to rejections in the first experiment, namely that the proposer could have offered a fair [5;5] split, but instead offered the [8;2] distribution. Just as in Experiment 1, some responders rejected the [8;2] offer. The possible mistake here is in taking for granted that the subjects will sum up the payoffs in the same way as the experimenters, i.e. that there is only one way for a subject to understand the payoffs. The nature of payoffs can be ambiguous to the subjects in such experiments. Experimenters have a normative theory about what aspects of the payoffs matter, but the subjects often do not have such a theory, or they construct a different one. This makes the results of the second experiment much harder to interpret than the authors suggest.

4 Are One-Shot Games Always One-Shot?

The distinction between the experimenter's and the participant's theory of what matters in a given experiment can be of major importance. Georg Elwert once tried to convince a group of Ayizo peasants in Benin, West Africa, to play the ultimatum game and failed.[2] The Ayizo were concerned about features of the game: "Who is the other?" "What will people say if I behave this way?" and "I don't believe that this

[2] Personal communication, June 2002.

will be a one-shot game." For these peasants, the importance of the status and the power of the other player, as well as concern with the social acceptability of one's actions made them refuse to play a game that aims at implementing anonymity. This distinction between the experimenter's and the participant's views of what matters in an experiment is also relevant for the distinction between one-shot and repeated games with Western participants. We will make two points. The first has been made before; the second, we believe, is new.

4.1 Heuristics Adapted to the Evolutionary Past

The first point is that people have a tendency to understand a new situation by analogy with a prototype entrenched in their mind. Such a prototype can be a situation that was typical in the history of human evolution, that is, in our Pleistocene past. The outcomes of economics experiments often suggest that people behave as if they lived in a world in which the games were repeated:[3] They reject unfair offers, punish altruistically, seek reputation, and otherwise act in ways consistent with strategies that evolve in repeated games. Some evolutionary psychologists' preferred explanation is a version of the disequilibrium argument: People cannot tell the difference between finite and repeated games (at some level of cognition at least), since there were very few finite games in our Pleistocene past. A similar point is made when subjects behave as if the encounter were not really anonymous. PINKER [1997, p. 42] writes that "our brains are not wired to cope with anonymous crowds ... and other newcomers to the human experience." There was little anonymity in the environment of evolutionary adaptiveness, so the story goes, and thus people may not take advantage of anonymity.

In principle, these sorts of arguments are cogent. But it cannot be that simple. People *are* able to tell at least some situations of anonymity from others: For instance, people in restrooms, especially men, wash their hands more when other people are present (PEDERSEN, KEITHLY, AND BRADY [1986]). And, as Falk et al.'s third experiment shows, people often behave differently in repeated games than they do in one-shot games. There may have been a sufficient number of finite and anonymous interactions in the Pleistocene past, as well. What is likely happening in these experiments is that a specific set of cues are responsible for alerting subjects to the duration of the game setting. In some repeat experiments, these cues are weak or absent. In others, they are strong. The default assumption may well be a long time horizon, or perhaps the opposite. The default assumption may even vary as a function of culture, occupation, or experience. More attention to the specific cues that subjects use to select among strategies would help to demystify this issue.

[3] The crucial distinction, of course, is not between one-shot and repeated play, but between low and high probabilities of continued interaction. We will use "finite" and "repeated" to refer to variation along this dimension.

4.2 Negotiating the Rules of the Game

But there is a second phenomenon that may account for some one-shot situations in which people behave as if there were more trials to come. The argument here is not that people in the experimental situation carry over behavioral strategies that are adapted to a past world where bargaining was repeated and nonanonymous, and where reputation could be formed. The argument is that in the real world, agents can do more than follow the rules of a game; they can change the rules. That is, a one-shot game can actually be changed into a repeated game. Here is an anecdote to illustrate this point (GIGERENZER [1996, p. 324]):

> "A small town in Wales has a town fool. He once was offered the choice between a pound and a shilling, and he took the shilling. The proposer was quite amused. Someone else tried it again, and the fool again took the shilling. People came from everywhere to witness this phenomenon. Over and over, they offered him a choice between a pound and a shilling. He always took the shilling."

The town fool's choice in the original one-shot game looks irrational. But here we have a situation in which a particular choice increases the probability of getting to choose again. The fool's unusual choice changed the one-shot game into a repeated game. More generally, in a world that offers greater uncertainty than a well-defined game, the rules are not always completely fixed, and – if one is smart or lucky enough – they can even be changed. The crucial point is the existence of uncertainty and the possibility of change. As Benjamin Franklin reminded us in 1789, "in this world nothing is certain, except death and taxes." Part of social intelligence is the ability to change the rules, not only to perform well enough within the rules.

Why are attempts to change rules a point worth considering? Unlike the ultimatum game, typical real-world bargaining – with the exception of activities such as auctions – does not have fixed rules. Buyers and sellers, employers and unions often negotiate the rules in the process, along with the outcome. In the language of economic models, aspects of the rules are endogenous. The introduction of social intelligence into economic analysis has methodological consequences. If people are socially intelligent, they may behave in ways that are inappropriate to one-shot games, and even try to change the nature of the game.

4.3 Beyond Homo Autisticus

To conclude, we return to the analogy between rational choice and autism. Autism is a mental disease. People with autism have difficulties in judging the intentions of others, and they do not reckon with spiteful behavior, deception, or lying. FALK, FEHR, AND FISCHBACHER [2003] conclude that what they call "nonstandard preferences"– fairness intentions, spite, and reputation – influence people's behavior in the ultimatum game. We fully agree that these forms of social intelligence matter. But we also feel that economic theory deserves a more realistic and radical point of

departure than that of a surprised autist who studies social behavior. More courage may be needed.

References

BLOUNT, S. [1995], "When Social Outcomes Aren't Fair: The Effect of Causal Attributions on Preferences," *Organizational Behavior and Human Decision Processes*, 63, 131–144.

BOLTON, G. E., AND A. OCKENFELS [2000], "ERC – A Theory of Equity, Reciprocity and Competition," *American Economic Review*, 90, 166–193.

FALK, A., E. FEHR, AND U. FISCHBACHER [2003], "Reasons for Conflict: Lessons from Bargaining Experiments," *Journal of Institutional and Theoretical Economics,* 159, 171–187.

FEHR, E., AND K. SCHMIDT [1999], "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

GIGERENZER, G. [1996], "Rationality: Why Social Context Matters," pp. 319–346 in: P. B. Baltes, and U. M. Staudinger (eds.), *Interactive Minds: Life-Span Perspectives on the Social Foundation of Cognition*, Cambridge University Press: Cambridge.

– – [2000], *Adaptive Thinking: Rationality in the Real World*, Oxford University Press: New York.

– – AND R. SELTEN (eds.) [2001], *Bounded Rationality: The Adaptive Toolbox*, MIT Press: Cambridge, MA.

– –, P. TODD, AND THE ABC RESEARCH GROUP [1999], *Simple Heuristics that Make us Smart*, Oxford University Press: New York.

GRANDIN, T. [1995], *Thinking in Pictures*, Vintage Books: New York.

KEASEY, C. B. [1978], "Children's Developing Awareness and Usage of Intentionality and Motives," pp. 219–260 in: C. B. Keasey (ed.), *Nebraska Symposium on Motivation*, Vol. 25, University of Nebraska Press: Lincoln, NE.

MCCABE, K. A., V. L. SMITH, AND M. LEPORE [2000], "Intentionality Detection and 'Mind-reading': Why does Game Form Matter?" *Proceedings of the National Academy of Sciences of the United States of America*, 97, 4404–4409.

PEDERSEN, D. M., S. KEITHLY, AND K. BRADY [1986], "Effects of an Observer on Conformity to Handwashing Norm," *Perception and Motor Skills*, 62, 169–170.

PINKER, S. [1997], *How the Mind Works*, W. W. Norton: New York.

SACKS, O. [1995], *An Anthropologist from Mars*, Vintage Books: New York.

Gerd Gigerenzer
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94
14195 Berlin
Germany
E-mail:
gigerenzer@mpib-berlin.mpg.de

Richard McElreath
Department of Anthropology
University of California, Davis
One Shields Avenue
Davis, CA 95616-8522
USA
E-mail:
mcelreath@ucdavis.edu

Intentions Matter: Lessons from Bargaining Experiments

Comment

by

JÖRG OECHSSLER

Many economists are uncomfortable with concepts like altruism, spite, reciprocity, or fairness. Even those who acknowledge their potential importance believe that such concepts are too vague to apply in theoretical work and too imprecise to be tested for with empirical or experimental data. The accomplishment of the paper by FALK, FEHR, AND FISCHBACHER [2003] (henceforth Falk et al.) is to show that this belief is unjustified and that it is indeed possible to differentiate among those concepts experimentally once they are properly defined.

A sensible research program in this area should work as follows: In the first stage, one tests existing theories. If they are found to be lacking – and in fact, a vast experimental literature shows that many commonly observed phenomena cannot be explained as the behavior of a selfish, own-payoff-maximizing *homo oeconomicus* – new theories are to be developed that can explain the data. Recently, several new theories have been developed that assume diverse forms of other-regarding preferences. But in a third stage, those new theories must be scrutinized with new data. Falk et al.'s paper belongs to the third stage.

The new theories that are being compared in their paper are:

(1) Inequity aversion preferences (see, e.g., FEHR AND SCHMIDT [1999] and BOLTON AND OCKENFELS [2000]). Inequity aversion is a particular form of fairness that assumes that individuals are interested in maximizing a utility function in which both the individual's absolute payoff and a measure of equality enter. The important feature is that utility is increasing with respect to the second argument: the more equal is the distribution of payoffs among all individuals, the greater is the utility. In particular, one even dislikes inequality that is in one's own favour.

(2) Spiteful preferences, as suggested, e.g., by LEVINE [1998]. Spiteful preferences are characterized by the fact that the individual will incur a cost in order to lower others' payoffs even if this increases inequality (whereas an inequity-averse individual will do so only if it lowers inequality).

(3) Reciprocity-based preferences as introduced by RABIN [1993], DUFWENBERG AND KIRCHSTEIGER [1999], and FALK AND FISCHBACHER [1999]. Reciprocity is

Journal of Institutional and Theoretical Economics
JITE 159 (2003), 195–198 © 2003 Mohr Siebeck – ISSN 0932-4569

defined as the tendency to respond kindly to acts that are perceived as kindly or fair, and unkindly otherwise.

The important feature that distinguishes reciprocal preferences from both inequity aversion and spite is that reciprocity necessarily must be based on the perception of the degree of kindness or unkindness, respectively, that is judged to motivate the acts of first movers. For inequity aversion and spite only the realized outcomes matter.

Falk et al. present a simple and very neat experiment to distinguish among these different forms of preference. Table 1 summarizes the evidence from selected treatments (a "✓" indicates that the evidence of a particular treatment is supportive of the theory). Consider first the 8/2 treatment, which I consider as a benchmark. Since the proposer was devoid of any meaningful choice, the responder could not deduce anything about the proposer's intention. Thus, reciprocity cannot play a role in this treatment. The fact that 18% of subjects rejected the 8/2 allocation is compatible with both inequity aversion and spite. It is compatible with inequity aversion because the subjects are willing to give up 2 units in order to reduce inequality (0/0 is more equal than 8/2). But it could just as well be spite, because each individual gave up 2 units in order to reduce the other's payoff by 8.

Table 1
Theories and Experimental Evidence

Theory	References	Main implications	Treatments		
			8/2	5/5 10/0	5/15
Inequity aversion	FEHR AND SCHMIDT [1999] BOLTON AND OCKENFELS [2000]	only *outcomes* matter; trade-off between own payoff and equity	✓		
Reciprocity	RABIN [1993] DUFWENBERG AND KIRCHSTEIGER [1999] FALK AND FISCHBACHER [1999]	*intentions* matter; tit for tat; psychological game theory		✓	
Spite	LEVINE [1998]	outcomes matter; *relative* payoff maximization	✓		✓

Next, consider the 5/5 treatment. When the responder considers an 8/2 offer, there is no difference in terms of *outcomes* from the corresponding subgame of the 8/2

treatment. Thus, theories based on inequity aversion and spite predict no change in behavior. The fact that there is a drastic increase in rejection rates to 44.4% is a clear sign that intentions matter. The 10/0 treatment shows the same for positive reciprocity.

Given that inequity aversion and spite can both explain the behavior of the 18% of subjects in treatment 8/2, treatment 5/15 in Figure 2 of Falk et al. provides a clean way of distinguishing the two motivations. Namely, no inequity-averse individual should reject the 8/12 offer, since that would decrease the own payoff and increase inequality. Yet 24% did so, which leaves spite as the only explanation. The fact that 4% did even reject the 5/15 offer may also be explained by spite if spite is defined as relative payoff maximization, that is, maximization of the ratio of payoffs.[1]

Interestingly, 24% of subjects were spiteful in the 5/15 treatment, whereas 18% were either spiteful or inequity-averse in the 8/2 treatment. If the percentage of spiteful people in all subject groups is roughly the same, one may conclude from the evidence in Falk et al.'s paper that inequity aversion is no relevant explanation at all, since all behavior patterns can be explained by spite or reciprocity. Proponents of inequity aversion will, no doubt, point to the dictator game as proof of the applicability of their theory. However, I am not convinced. The fact that subjects in double-blind dictator games allocate substantially smaller amounts to receivers than in less anonymous settings points more toward norm-driven behavior than toward utility maximization based on inequity aversion.[2]

From a theoretical perspective it is also of interest whether nonstandard preferences of the type discussed here might arise in an evolutionary process. It is often argued that preferences that deviate from the selfish own-payoff-maximizing paradigm would not withstand evolutionary forces, since – almost by definition – they induce the individuals endowed with such preferences to maximize the "wrong" objective function. This argument, however, overlooks possible strategic effects of nonstandard preferences. In fact, it has been shown that spite or envy and reciprocity can survive in an evolutionary process. However, either the interaction must take place in small groups (HUCK AND OECHSSLER [1999]) or preferences of opponents must be observable (e.g., SETHI AND SOMANATHAN [2001]), which may be a strong assumption.

As far as I know, no similar results are available with respect to the survival of preferences with inequity aversion. It is also difficult to imagine how inequity aversion could ever confer an evolutionary advantage in dictator games, because, in contrast to the ultimatum game, in the dictator game a fair-minded proposer only loses resources without gaining a strategic advantage.

If I had to sum up the experiments of Falk et al. in two words, it would be the following: *intentions matter.* Of course, this insight can only be surprising to us economists.

---

[1] There is also evidence for relative-payoff maximization (or spite) in experiments on oligopoly behavior (see, e.g., HUCK, NORMANN, AND OECHSSLER [1999]).

[2] For example, when the effect of norms is weak, as in the double-blind dictator games of HOFFMAN et al. [1994], the median allocation is 0.

References

BOLTON, G., AND A. OCKENFELS [2000], "ERC – A Theory of Equity, Reciprocity and Competition," *American Economic Review,* 90, 166–193.

DUFWENBERG, M., AND G. KIRCHSTEIGER [1999], "A Theory of Sequential Reciprocity," Mimeo, CentER for Economic Research, Tilburg.

FALK, A., AND U. FISCHBACHER [1999], "A Theory of Reciprocity," *Working Paper No. 6,* University of Zürich.

– –, E. FEHR, AND U. FISCHBACHER [2003], "Reasons for Conflict: Lessons from Bargaining Experiments," *Journal of Institutional and Theoretical Economics,* 159, 171–187.

FEHR, E., AND K. SCHMIDT [1999], "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics,* 114, 817–868.

HOFFMAN, E., K. MCCABE, K. SHACHAT, AND V. SMITH [1994], "Preferences, Property Rights, and Anonymity in Bargaining Games," *Games and Economic Behavior,* 7, 346–380.

HUCK, S., H.-T. NORMANN, AND J. OECHSSLER, [1999], "Learning in Cournot Oligopoly: An Experiment," *Economic Journal*, 109, C80–C95.

– – AND J. OECHSSLER [1999], "The Indirect Evolutionary Approach to Explaining Fair Allocations," *Games and Economic Behavior,* 28, 13–24.

LEVINE D. [1998], "Modelling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics,* 1, 593–622.

RABIN, M. [1993], "Incorporating Fairness into Game Theory and Economics," *American Economic Review,* 83, 1281–1302.

SETHI, R., AND E. SOMANATHAN [2001], "Preference Evolution and Reciprocity," *Journal of Economic Theory,* 97, 273–297.

Jörg Oechssler
Department of Economics
University of Bonn
Adenauerallee 24
53113 Bonn
Germany
E-mail:
oechssler@uni-bonn.de