Chapter 3 Trust, Fairness and Reciprocity

3.1 Trust

There are numerous situations in everyday life in which we make decisions in social interactions under uncertainty. In many of these situation we have to form expectations about the behavior of another person, who might harm or promote us. These are situations of trust and are a feature of our social life. Deutsch (1973, pp. 146-148) highlights several meanings of trust, such as trust as social conformity, as virtue, as risk-taking, or as confidence. I understand trust as confidence in the benevolent behavior of another person, following recent work (e.g. Boon & Holmes, 1991; Good, 1988). This interpretation can be easily mapped to the investment game or to other trust games. In a trust game one player chooses between a safe option with a relatively low payoff or a risky option that offers a potential higher payoff depending on the other player's subsequent decision (see Figure 2).

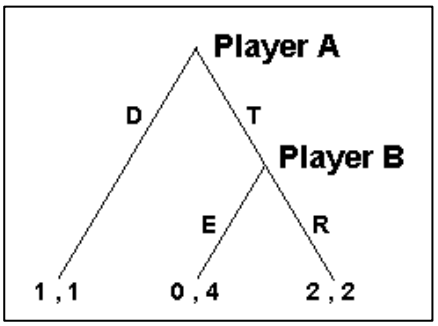

Figure 2. Trust game.

The first value of the pairs of values represents player A's payoff, whereas the second value represents player B's payoff. The game has one subgame-perfect equilibrium consisting of D and E.[1]

The choice of the risky option can be interpreted as *trust*, as it expresses the player's confidence about the benevolent behavior of the other person. The choice of the safe option can be interpreted as *distrust*. Subsequently player B decides whether to exploit or to reciprocate the trusting decision of player A. The game is a simplified version of the

[1] To derive the Nash equilibria for the Trust game one has to specify a complete strategy for player B that specifies how player B reacts to player A's possible decisions. A strategy for player B of, for instance, "E-R" implies that player B will choose E if player A chooses D and player B will choose R if player A chooses T. The Trust game has two Nash equilibria: The first consists of D and EE, the second of D and RE. However, the last one is less convincing, because it says that player B would reciprocate in the hypothetical case that player A chose T. In contrast, the game has only one single subgame-perfect equilibrium.

investment game as both players have only two possible choices.

It is characteristic of a trust game that a player who distrusts the other player cannot perceive whether his distrusting behavior was justified or whether the other player would have reciprocated trust. This is different from the prisoner's dilemma, which has often been used to study mutual trust, in which players play the same roles (see Figure 3).

		Player B	
		C	D
Player A	C	3, 3	1, 4*
	D	4*, 1	2*, 2*

Figure 3. Prisoner's dilemma.

The asterisks represent the best reply strategy for one player given the other player's strategy. The cell with two asterisks present the Nash equilibrium.

The two players in the prisoner's dilemma have two pure strategies: cooperate (C) or defect (D). The players' payoff is greater when they both cooperate than when they both defect. However, if the opponent cooperates, defecting leads to higher individual payoff and similar if the opponent defects, this also leads to a higher individual payoff. Hence, defecting always leads to a higher payoff than cooperating, and thus defecting is the dominant strategy for both players. Therefore the prisoner's dilemma also establishes a social dilemma in which individuals' self-interest leads to an inefficient, socially least desired outcome.

Many studies using the prisoner's dilemma were conducted between 1950 and 1970. They could demonstrate that the magnitude of cooperation is positively associated with, among other factors, the possibility of communication, punishment options, reversibility of decisions, long-term interests or anticipated continued interaction, low ambiguity, easy contact between participants, and low incentives (for overviews see Deutsch, 1973; Gallo & McClintock, 1965; Pruitt & Kimmel, 1977). Unfortunately most of these early studies used only very small or imaginary incentives, making the validity of the results questionable. Gumpert, Deutsch, and Epstein (1969), for instance, showed that the magnitude of cooperation was substantially smaller when participants played with high, real payoff compared to imaginary payoffs, supporting the apprehension that individuals might not take a game seriously when playing with imaginary payoffs (on the impact of incentives see also Hertwig & Ortmann, 2001; Smith & Walker, 1993). Although, the prisoner's dilemma is appropriate for studying mutual trust in symmetrical relationships it is unjustified to generalize the experimental results to situations of asymmetrical

relationships. In contrast, an asymmetrical relationship, with different roles of individuals and diverging bargaining powers, is better described with the investment game.

Berg et al. (1995) conducted the investment game without repetition in an experiment. They found that participants in the role of player A sent on average $5.20 of their endowment of $10 to their counterpart, whereas participants in the role of player B returned on average $4.70 of the trebled endowment. Only 2 of 32 participants sent nothing to their counterpart, consistent with the game-theoretical prediction, whereas 5 participants sent their entire endowment. Investments of $10 led to a return of on average $10.20, while investments ranging from $5 to $6 led to average returns of $6.10. Interestingly these average return rates show, besides their deviation from the game-theoretical prediction, that some participants in the role of player B did not reciprocate high investments with high returns, making investments not very profitable.

Van Huyck et al. (1995) investigated the investment game, which they called the "peasant-dictator game," with some modifications. The main differences from Berg et al.'s study are that only participants in the role of player A obtained an endowment. In addition, participants played the game repeatedly, but after each period, participants were re-matched with different opponents, and players' positions were reassigned for each game. Finally, the amounts of endowments and the rates of return were varied. It resulted in investment rates being rather low compared to the results of Berg et al. (1995). Only under a treatment with high rates of return of 400% were substantial investments of on average 34% made, whereas under treatments with lower rates of returns (100%, 67%, and 25%) the average investments were below 10%. I presume that the low investment rates in Van Huyck et al.'s study can be attributed to two reasons. First, because of the low rates of return and lack of repetition of the game the chance of being exploited was probably judged very high by participants in the role of player A, leading to low investments. Additionally, due to players' roles being reassigned after each period, participants could easily justify low investments that led to small payoffs to player B, as the opponents would also have the opportunity to keep their endowments as player A in subsequent games.

Besides having an asymmetrical payoff structure, the investment game differs from the prisoner's dilemma by the several options available to the players, which enables a subtle measurement of the magnitude of trust. Another particularity of the investment game consist in the role of the second player. Her decision expresses reciprocity and fairness. If player A trusts player B by making a high investment, player B can reciprocate trust with a fair return, but what is a "fair" return?

3.2 Fairness and Reciprocity

A large body of research demonstrating that people are not only motivated by self-interest but are also motivated to reach *fair* outcomes has been done by using the "ultimatum game." The ultimatum game is a two-person sequential bargaining game in which player A is provided with an endowment. Player A makes a proposal of how to split the endowment between the two players. After player A's proposal player B decides whether to accept or reject the proposal. If player B accepts, both earn the proposed amounts; if player B rejects, both earn nothing. The game-theoretical analysis predicts that player B will accept any positive amount, so player A should offer only a minimum amount and demand almost all of the endowment for him- or herself.[2]

However, experimental results deviate from this prediction. In the ultimatum game first studied by Güth, Schmittberger, and Schwarz (1982), participants in the role of player A made on average an offer of 33% of the endowment to player B (mode=50%, *SD*=15%), rejecting the equilibrium prediction. In addition, contrary to the game-theoretical prediction around 17% of all proposals were rejected, even though they offered positive amounts of on average 21% of the endowment to player B. Similar results have been reported by Forsythe, Horowitz, Savin, and Sefton (1994), Hoffman, McCabe, and Smith (1996), Kahneman, Knetsch, and Thaler (1986), Ochs and Roth (1989), and Roth, Prasnikar, Okuno-Fujiwara, and Zamir (1991).

If an individual in an interaction behaves cooperatively and trusts another person, then the other person behaves *reciprocally* if she also cooperates and is beneficial to the first individual. A broader definition also implies negative reciprocity, so that harmful behavior is answered with punishing behavior. In the investment game, reciprocity implies that high investments are reciprocated with high returns. In the investment game studied by Berg et al. (1995), participants in the role of player B returned on average 30% of the trebled investments when an investment was made. This return rate is rather low given that only a return rate of at least 34% provides player A with a payoff above his endowment. A closer inspection of the data shows that there are two predominate return rates. Most frequently nothing of the trebled investment is returned, consistent with the game-theoretical prediction. Second-most frequently 70% of the trebled investment is returned to player A, yielding equal final payoffs for both players. This result shows that some

[2] There is another subgame-perfect equilibrium: If player A offers nothing to player B, player B should be indifferent about rejecting or accepting; if player B accepts, player A should offer nothing to player B, which is the second subgame-perfect equilibrium.

individuals actually act in their own self-interest and exploit the other player that made an investment, whereas other individuals made substantial returns. I also reanalyzed Berg et al.'s (1995) data by determining the correlation between the investment rate and the return rate, which turned out to be marginal r=.02. In sum, the occurrence of substantial investments in the experiment expresses trust, but the magnitude of reciprocity was rather small.

Similar games to the investment game have recently been receiving increasing attention (Bolle, 1998; Fehr, Gächter, & Kirchsteiger, 1997; Fehr, Kirchsteiger, & Riedl, 1993; Gneezy, Güth, & Verboven, 1998; Güth, Ockenfels, & Wendel, 1997; Jacobsen & Sadrieh, 1996; Ortmann, Fitzgerald, & Boeing, 2000). Güth et al. (1997), for instance, studied a trust game in which the players have only two pure strategies (similar to the one in Figure 2). Player A can either "trust" the other player, thereby attaining the opportunity of a higher payoff, or "distrust" the opponent, yielding low payoffs for both players. If player A trusts player B, player B can "reciprocate" trust by choosing an option that leads to equal payoffs for both players or "exploit" player A by choosing an option that leads only to a very high payoff for player B. Under various treatments the majority of participants trust their opponents (21 out of 28). Similar to Berg et al.'s (1995) results the magnitude of reciprocity was rather low. Only a small number of participants reciprocated trust by a decision leading to equal payoffs.

In contrast to the studies of Berg et al. (1995) and Güth et al. (1997) Kirchler, Fehr, and Evans (1996) could show substantial reciprocal behavior. Participants who took on the role of employee reciprocated high wages, chosen by other participants who acted as employers, with high effort. However in Kirchler et al.'s study the costs for higher effort levels were relatively low, making reciprocal behavior presumably more likely. In sum, the experimental results demonstrate that individuals often do not conform to the game-theoretical prediction and their behavior expresses substantial trust. However, the results concerning the magnitude of reciprocity are mixed.

Most of these recent experiments that demonstrate how concerns for fairness affect behavior have been conducted by experimental economists. However, the insight that individuals are motivated to reach fair outcomes has a long tradition and has been claimed by many social psychologists for years. Pruitt (1967), for instance, argues that individuals gain two forms of utility in experimental games – one that emerges from the monetary payoffs and one that comes from the social comparison. Sermat (1964) claims that not only the monetary payoffs but also the social acceptance of one's behavior influences the

decisions in a game. This early research has influenced theoretical concepts of fairness and distributive justice. Beginning with the work of Adams (1963), equity theory became the dominant standard of investigating the justice of allocation decisions. According to the equity theory (Adams, 1965; Deutsch, 1975; Walster, Berscheid, & Walster, 1973; Walster, Walster, & Berscheid, 1978), people are motivated to reach fair outcomes. According to the equity principle, a distribution is fair if the profit of an interaction relative to the contribution is equal across all interaction partners. Deutsch (1975) highlights two other principles: equality (everyone receives the equivalent regardless of the contribution) and need (everyone receives relatively what they require). Interestingly and contrary to experimental economists, most psychologists in this research field base their theory on the assumption of self-interested individuals: Due to an immanent conflict between the self-interested individuals in a social system, principles emerge on how resources should be distributed. Violations of these principles lead to punishment and compliance to rewards, hence, the principles are learned and internalized (Mikula, 1980). For such a reinforcement mechanism to function anonymity must not exist, and ongoing interactions are a necessity; therefore, the underlying game-theoretical model—although often not stated as such—is an indefinitely repeated game.

Several authors incorporate fairness within a social utility approach (Bolton, 1991, 1997; Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999; Loewenstein, Thompson, & Bazerman, 1989; Messick & Sentis, 1979; Messick & Sentis, 1985). They have postulated that not only one's own monetary payoff contributes to the utility of a decision's consequence but also the other individuals' payoff. MacCrimmon and Messick (1976) argued that six motives affect decisions in social situations: self-interest, self-sacrifice, altruism, aggression, cooperation, and competition. Messick and Sentis (1985), on the other hand, proposed an additive utility model with only two motives, a self-interested and a social component, which measure the difference between one's own and the other actor's payoff. Loewenstein et al. (1989) incorporated three aspects in their utility function: one's own payoff, a potential positive and a potential negative difference between one's own and the other's payoff. Bolton (1991) assumed that the absolute payoff and the difference of one's own and the other's payoff contribute to the overall utility. However, in his model, only if one's own payoff is less than the other's payoff, this can produce a negative utility, whereas an unequal split to one's own advantage has no negative effect. These studies show that individuals compare their own payoffs with those of others and dislike inequality if this cannot be justified (e.g. by different individuals' contributions). However, if

inequality occurs individuals' absolute displeasure connected to a split that disfavors oneself is greater compared to an unequal split that favors oneself. More recently special attention has been devoted to the role of individuals' intentions. It appears as crucial who is responsible for an unfair unequal distribution (Falk & Fischbacher, 2000; Rabin, 1993). Individuals appear to accept unequal distributions of payoffs if the outcome is not caused by their presumably kindly interaction partner—in other words, if the intentions of their interaction partner were good (Camerer & Thaler, 1995).

I argue that most of the social utility approaches encounter two major problems. First, most of these models make the assumption that peoples' preferences concerning allocation decisions are static. I think it is obvious that the preferences depend on the dynamic process of the allocation problem. The large body of research concerning "procedural justice" has shown that the representation of a fair allocation often depends on the way this allocation was attained (Barrett-Howard & Tyler, 1986; Folger, 1977; Tyler, 1994; Tyler & Lind, 1992). Therefore constructing a static model that neglects, for instance, the behavior of the interaction partner will presumably only have limited validity. Second, the authors of the proposed utility models usually do not claim that individuals actually calculate utilities. In contrast, it is argued that individuals behave "as if" they attempt to maximize expected utility. From a psychological perspective this "as if" argument is dissatisfying since it neglects the cognitive process underlying any decision. But even for an economist this view has its disadvantages, because by neglecting the decision process, I claim, it will be difficult to predict when people will actually behave in a manner consistent with the utility approach.