Summary

Trust, reciprocity, and fairness are central aspects of our social life. In many social interactions, we form expectations about the behavior of another person who might harm or benefit us. If we trust someone, this person can either reciprocate or exploit our trusting behavior. In a situation of trust and reciprocity, fairness often plays an important role in defining an appropriate behavior that reciprocates trust. Past research has illustrated how trust and fairness are influenced by various factors. However, what is still missing are models that describe the underlying decision process involved in social interactions of trust and fairness. The present work, therefore, focuses on the development and evaluation of models that can describe the decision process. For this purpose, simple decision strategies are proposed; these are psychologically plausible because they take human cognitive limitations into account.

The present work has several goals: First, people's motivation to reach fair outcomes should be investigated. Second, the performance of simple strategies should be explored by letting strategies compete with one another. Finally, the fit of simple strategies in predicting individuals' decisions in a social interaction should be examined.

Trust and fairness are studied by using a two-person sequential bargaining game. Both players of this so-called "investment game" receive an endowment. Player A decides how much of his endowment he wants to invest. The invested amount is augmented and delivered to player B. Player B then decides how much of the augmented amount she wishes to return to player A. If player A *trusts* that player B will *reciprocate* the expressed trust with a *fair* return, then player A makes an investment.

However, the game-theoretical prediction for this game is straightforward: To maximize the monetary payoff, player B will return nothing to player A. This can be anticipated by player A; hence, no amount is sent to player B, which leads to an inefficient outcome. The situation changes dramatically if the game is repeated indefinitely with a certain continuation probability. In this situation, player B will hesitate to exploit player A, fearing an end to investments by player A in subsequent periods. Such a situation also makes an investment for player A reasonable. Given that social interactions usually have an ongoing character, it is argued that the indefinitely repeated game is a more appropriate model for social interaction involving trust and fairness.

The first study demonstrates that individuals in the investment game without repetition deviate from the game-theoretical prediction by making substantial investments

and returns. In contrast, reciprocal behavior could not be shown as the correlation between returns and investments was rather small. In an additional experimental condition the payoffs were doubled, leading to lower average investments and returns, supporting the prediction that high incentives move behavior in the direction of the game-theoretical prediction.

In the second study, attention was drawn to the indefinitely repeated investment game. To explore how motivation for fairness might influence individuals' behavior the endowment for player B was varied; in one condition no endowment was given to player B, whereas in the other condition both players received an endowment in every period. Equity theory predicts that people are motivated to reach fair outcomes. An outcome is fair if the profit of an interaction relative to the contribution is equal across all interaction partners. Accordingly, it is predicted that in the condition with an endowment for both players, the return rates will be higher compared to the condition with no endowment for player B. Equity theory also predicts higher investments if player B receives no endowment, as then only a substantial investment enables both players to end up with equal final payoffs. Consistently, the return rates were much higher in the condition with an endowment for both players. On the contrary, the investment rates were not much higher in the condition with no endowment for player B. Additionally, the repeated game promotes reciprocity, as a substantial positive correlation between the investment and return rates was obtained. Although the influence of a motivation for fair outcomes could be illustrated, it also became clear that individuals' behavior varied substantially, supporting the claim for process models that describe the dynamics of the social interaction.

Before process models were investigated experimentally, the performance of various strategies was first evaluated. It turned out that many strategies form Nash equilibria; that is, the strategies are mutually best replies to each other. The psychologically plausible assumption of small, unsystematic errors in people's selection or execution of strategies promoted the use of the limited evolutionary stability concept for the evaluation of strategies. For a selected set of strategies, a few strategies could be distinguished as limited evolutionarily stable. To generalize these results to a large set of strategies evolutionary processes were simulated. Two strategies frequently evolved that led to efficient payoffs. One strategy for player A invests the entire endowment and repeats the investment unless the return is below player A's endowment, in which case no investment is made for all subsequent periods. The frequently evolved strategy for player B always makes a return

that gives player A a payoff just above player A's endowment. These two simple strategies were also frequently obtained under a condition in which strategies could make use of a higher complexity. On the other hand, if small execution errors occurred strategies for player A frequently evolved that incorporate a mechanism that tolerates low returns.

In the final study strategies were developed to describe individuals' decision processes in the repeated investment game. It turned out that the simple strategies outperformed a learning model and a baseline model in predicting participants' decisions. The strategy that best predicts participants' decisions for player A is one that invests the entire endowment in the first period and repeats the investment as long as substantial returns are made. After repeated low returns the strategy makes no investment for all following periods. The strategy that best predicts participants' decision for player B is one that makes a return that leads to equal final payoffs for both players if substantial investments are made and makes no return if player A only makes very small investments.

The present work was able to illustrate the extent of trust, reciprocity, and the motivation for fair outcomes under varying conditions. A comparison of the strategies developed to describe peoples' decisions and those found in the evolutionary simulations shows some striking similarities. For instance, all strategies for player A make an initial high investment, which is repeated if substantial returns are made. However, the strategies developed for predicting participants' decisions have some extra features. The strategies for player A incorporate a mechanism for tolerating low returns, which were found in the evolutionary simulations only if small execution errors were assumed. The strategies for player B incorporate a mechanism for punishing low investment.

The evolutionary simulations illustrate that simple strategies are sufficient for obtaining efficient outcomes and are not outperformed by more complex strategies. Simple strategies were also able to predict a substantial proportion of individuals' decisions. They could explicate the underlying dynamic process that is involved in a situation of trust, reciprocity, and fairness.