BACKGROUND

Chapter 2 A Model of Trust and Reciprocity

2.1 Descriptions of Asymmetrical Social Interactions

In an asymmetrical social interaction individuals' decisions and the decisions' consequences are not interchangeable. The social interaction between an employee and an employer is an example of an asymmetrical interaction. What are distinguishing characteristics of the interactions? The interaction consists of two individuals; it is possibly *ongoing*; and decisions are often made *sequentially*. First, the employee, for instance, decides whether to spend extra time at work, which would produce an extra surplus for the firm, and subsequently the employer decides whether to return some of the extra surplus to the employee in the form of a wage increase. The relationship is asymmetrical because the roles of the two individuals and their decisions are not interchangeable. The employee decides whether to spend extra effort. The employer decides on how a potential surplus should be allocated. If the employee decides not to spend high effort then the employer may not have the possibility to increase the wage, due to low profit. The relationship is ongoing since after each wage negotiation work carries on and the employee may adjust the effort level; subsequently the employer may alter the wage and it is often unclear when the relationship will finally conclude.

I am interested in the cognitive decision process involved in a social interaction like the one described above. Interactive decision making is the field of game theory. I argue that to understand the cognitive decision process in a social interaction it is important to understand individuals' strategic advantages or disadvantages that can be articulated with a game-theoretical analysis. In contrast to individual decision making, interactive decision making implies that the ones who are involved in an interaction are affected by their own decisions *and* by the decision of others. A game-theoretical analysis assumes that the individuals are aware of the interdependence, that individuals believe that the others are also aware of the interdependence, and that all involved in the interaction take these interdependencies of their decisions into account (Aumann, 1989; Fudenberg & Tirole, 1991; Samuelson, 1997). A social interaction can usually be described by a game that is defined by a set of players, the possible options of each player (their strategies), and the payoffs that are the consequences of the combinations of decisions. Depending on the consequences the game can cause certain problems of conflict or coordination. Main characteristics of a game are given by the payoff structure, which can be symmetrical or asymmetrical; the information the players have about the decision of the other players; and possible repetitions of the interactions.

First, if the roles in a social interaction are interchangeable then this results in an identical set of strategies and a symmetrical payoff structure. However, if the roles are not interchangeable, which is often the case in real-life situations, the set of strategies differ and asymmetrical payoff structures result. This often has strong implications for the bargaining powers of the players involved in strategic interactions. If the consequences of a "disagreement" between two individuals is much worse for one of the individuals then this individual has less bargaining power than the other, who might not have much to lose. Second, if in a social interaction the players make their decisions simultaneously or have no information about the other player's behavior the game can be described in a "normal form" payoff matrix. In contrast, if one player has information about the other player's decision, that is, when the players make decisions sequentially, then this "extensive form" game can be presented in a sequential or extensive tree structure. Whereas games in the normal form are usually analyzed by the Nash equilibrium concept, games in the extensive form are analyzed by the subgame-perfect equilibrium concept (see Fudenberg & Tirole, 1991 pp. 83-100). A Nash equilibrium is a combination of strategies such that each player's strategy is an optimal response ("best reply strategy") to the other players' strategies. An extensive game can be represented as a tree, so that a subgame consists of a node of the tree together with all of the informationally closed sub-tree that follows. A Nash equilibrium is called a subgame-perfect equilibrium if all players' strategies induce a Nash equilibrium in every subgame (in the following, Nash equilibria are occasionally only called equilibria). Third, an interaction can be either ongoing, so that it is described by a repeated game, or a single event, in which case it is modeled with a game without repetition. If a game is repeated this often has dramatic consequences for the game-theoretical analysis, because the players can take past decisions of their opponents into account when making their present decisions, and they have to consider that their present decisions may affect the future decision of others.

How can an asymmetric relationship, like the one of an employee and an employer, be described in a game-theoretical notion? Due to its symmetrical payoff structure and the simultaneous decision process, the famous prisoner's dilemma (Luce & Raiffa, 1957) would be an improper description. Instead, such a dependency relationship between two individuals resembles the situation of an investment game or “peasant-dictator game” (Berg, Dickhaut, & McCabe, 1995; Van Huyck, Battalio, & Walters, 1995).

2.2 The Investment Game and its Game-Theoretical Solution

The investment game is a two-person sequential bargaining game (see Figure 1). Both players receive an endowment of, for instance, DM 10. Player A can invest any amount of the endowment, which is then augmented, producing some surplus before it is delivered to player B. Player B decides how much of the then, for example, tripled amount she wishes to return to player A.

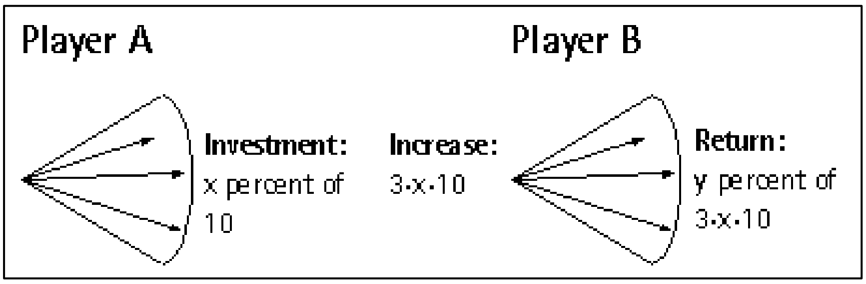

Figure 1. Investment game.

The game provides the opportunity of a surplus for both players. If player A *trusts* player B and believes that she will return a substantial amount, then player A will invest his entire endowment (i.e. producing an efficient outcome that maximizes the payoff for both players). If player B *reciprocates* the expressed trust with a *fair* return then both player will ultimately be better off than if player A *distrusts* and if player B *exploits* player A. However, the subgame-perfect equilibrium for the investment game without repetition 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 in which no surplus is produced. The social interaction that is represented by the investment game can be called a social dilemma. Because both players follow their self-interest in maximizing their payoff, they will finally end up with the minimum mutual payoff, which is socially least desired. The socially most desired outcome, that is, an efficient outcome with the maximum mutual payoff, is not obtainable if both players follow their self-interest.

Most attempts to resolve a social dilemma affect the payoff structure of the game. For instance, one could think of a contract that binds player B to make a certain return if player A makes a particular investment. The employer, for instance, could make an agreement with the employee that if the employee's effort leads to a particular surplus for the company, the employer will pay a particular bonus. If the agreement is broken by the employer the employee could enforce his rights, thereby inducing costs for the employer, which in fact makes it in the interest of the employer to keep the contract. However, the possibility of a binding contract changes the entire game by modifying the players' payoffs. In any case, if a situation is encountered that consists of a social dilemma, then a contract could change the situation very effectively and resolve the dilemma. There are several reason why a binding contract might not be established. First, the possibility of a contract might simply not exist. Second, the costs for arranging a contract might exceed the possible risk of being exploited, so that a contract is not worthwhile. Finally, even if a contract could be arranged the contract leaves some eventualities unspecified, which would again allow player B to take advantage of her opponent.

Another possibility to resolve the dilemma lies in future interactions. The relationship between an employer and employee is usually *ongoing*. If the employer does not reciprocate the employee's high effort by a wage increase the employee will presumably reduce her effort in the future. Therefore it is in the employer's self-interest to reinforce high effort with wage increases to motivate his employee. Even if a relationship between an employer and employee ends, the employer might care about his good reputation so that he will not exploit the employee in his or her final interactions. Therefore I argue that the investment game is not a proper model of the *ongoing* relationship between an employer and employee, because it incorporates no repetition. A better model for the relationship is the indefinitely repeated investment game.

First, two kinds of repetitions have to be distinguished: finite repetition and indefinite repetition. If the investment game is repeated finitely the game-theoretical prediction does not change substantially, because in the final period of the game the same prediction will be made as for the "one-shot" game without repetition. By backward induction this prediction carries "backward" to the first period. In an indefinitely repeated game, the stage-game (i.e. one single period of the investment game) may be repeated with a certain continuation probability. This does not mean that the game will last forever. For instance, with a continuation probability of 0.5 (the first period happens with a probability of 1) the probability that the game will at most last for three periods is 0.875 and the expected number of periods is only two. The uncertainty about the end of the game changes the game-theoretical prediction dramatically, given a sufficiently large continuation probability. Given that the last period is not defined, backward induction cannot be applied. In an indefinitely repeated investment game player B will hesitate to exploit player A, because there is always the chance of another period in which player A will not repeat the investment if he was exploited by player B in the previous period. This shows that player A has the threat of withdrawing future investments, thereby punishing player B's potential exploitive behavior. By moving from the one-shot or finitely repeated game to the indefinitely repeated investment game the social dilemma is resolved.

The argument I made for the investment game can be generalized: I argue that most of our social interactions are ongoing and therefore an indefinitely repeated game should be selected as a model for the social interaction. Furthermore the decision strategies and motives we have learned or that have been socially evolved are adapted to a world with ongoing relationships in which we know precisely who our interaction partners are. A one-shot game (i.e. an anonymous interaction without any repetition) is a "pathological situation" that is suitable to demonstrate our moral behavior as it will "stand out from the background" (Binmore, 1994, p. 130), but it is inappropriate to reflect the majority of our social interactions to which we are adapted. If we wish to understand why, for instance, our fairness motives survive, "one must look at those situations in which it is easy to overlook that a rule [e.g. fairness rule] is being used at all because it works so smoothly" (Binmore, 1994, p. 130).