Institute for Empirical Research in Economics
University of Zurich

Working Paper Series
ISSN 1424-0459

Working Paper No. 203

The Effect of Sunk Costs on the Outcome of Alternating-Offers Bargaining between Inequity-Averse Agents

Christian Ewerhart

September 2004

The Effect of Sunk Costs on the Outcome of Alternating-Offers Bargaining between Inequity-Averse Agents[1]

Christian Ewerhart[2]

University of Zurich

September 2004

Abstract

The paper analyzes the infinite-horizon alternating-offers bargaining game between agents with inequity-averse preferences. Without prior investments, the model predicts a shift of the outcome towards equal division. Asymmetric investments affect the ex-post bargaining outcome, giving an advantage to the party that contributed more. Under suitable circumstances, this effect may significantly mitigate the hold-up problem. In fact, in a symmetric set-up, if production is sufficiently profitable, and parties are sufficiently patient, then the first-best investment levels can be approximated without a contract.

Keywords: Alternating-offers bargaining, inequity aversion, hold-up problem

JEL classification: C72, C91

[1] Acknowledgements to be added.

[2] Postal address: Institute for Empirical Research in Economics, Winterthurerstrasse 30, CH–8006 Zurich, Switzerland. E-mail: christian.ewerhart@iew.unizh.ch.

1. Introduction

The alternating-offers bargaining model developed by Ståhl (1972) and Rubinstein (1982) has found many useful applications in the economic literature. The model considers the bargaining process between two parties in a discrete-time framework. Specifically, in the initial period, Party 1 is called up to make an offer on how to divide a given pie of a perfectly divisible good. Party 2 may then accept or reject the offer. After an rejection, negotiations are delayed to the next period. At the begining of that period, the value of the good depreciates for both parties, and Party 2 is called up to make an offer. Party 1 can then either accept or reject this offer. In the infinite-horizon version of the bargining game, the right to make an offer goes back and forth between the two parties until one offer is accepted, and payoffs can be made.

While the model is analytically very convenient, experimentalists such as Roth and Ochs (1989) found that the assumed utility specification may not fully explain the behavior that is typically observed in the laboratory. For example, subjects might consider the equal division of the pie as a focal point which influences the bargaining outcome. To reconcile the theoretical prediction with the experimental results, Bolton (1991) proposed to use a refined utility specification: In his model, an agent is motivated not only by his or her absolute monetary payoff, but also by the relative size of this payoff compared to the other agent's payoff. Bolton's approach explains a number of experimental findings on alternating-offers bargaining games, such as the

posing of disadvantageous counteroffers after a rejection, and the occurrence of rejections. However, and in spite of the explanatory power of the chosen utility specification, Bolton expresses a feeling of uneasiness with it. At the end of the article, he writes:

> *"Why are people, at least in some situations, willing to pay for fair treatment? It is a key question, as of yet without an answer."*
>
> G. Bolton (1991, p. 1129)

In this paper, we address the above issue from a theoretical perspective. We consider simultaneous investment decisions of two agents that subsequently divide the jointly produced output by negotiations. Contractual arrangements are excluded. For this specific situation, we show that fairness preferences may lead to more efficient investment decisions when compared to standard preferences. In fact, if production is sufficiently profitable, then the first best investment levels can be approximated for sufficiently patient agents. Thus, a community consisting of individuals with fairness preferences would be economically more successful than another community consisting of individuals with standard preferences.

To avoid potential misunderstandings, we wish to stress that fairness (or inequity aversion) is not outright altruism. An altruistic agent is someone who may prefer giving away some of his payoff to another agent. This is

not necessarily the case for an agent with fairness preferences. An agent with fairness preferences will in general prefer a higher payoff to a lower payoff, even if the payoff difference would be given to some other agent. For example, in a bargaining situation, agents with fairness preferences will normally exhibit strictly conflicting interests. In difference to a standard utility specification, however, the loss of a marginal share below the "fair" division creates a larger disutility for an agent with fairness preferences than an equal-sized loss of a marginal share above this value.

This general definition of fairness has been introduced independently in two papers. Fehr and Schmidt (1999) develop a theory of inequity aversion vis-à-vis an individual opponent, and apply it to explain the experimental data drawn from the ultimatum games, the various models of competition and cooperation, as well as the dictator and gift-exchange games. Bolton and Ockenfels (2000) consider a set-up where an agent compares his or her payoff to the average payoff of the agents in a given reference group. Their model explains experimental results from the ultimatum and dictator games, the prisoner's dilemma, the gift exchange game, and Bertrand markets. Both papers derive the economic consequences of the assumption that the agents' preferences incorporate fairness considerations. However, neither of these papers sets out to provide a formal answer to the above question.[3]

[3]Bolton and Ockenfels (2000) conjecture that an answer may be drawn from evolutionary models of cooperation, as proposed by Güth (1995) and others.

In a recent contribution, Tröger (2002) provides a foundation for the economic rationale of fairness in a set-up that is similar to ours.[4] His paper considers a two-stage game of joint production between two agents. In the first stage, one agent makes a unilateral investment. In the second stage, there is a Nash demand game that determines the distribution of the gains from investment. While the Nash demand game has a continuum of equilibrium outcomes, Tröger shows that evolutionary forces may select one equilibrium in which the division of the gains from investment critically depends on prior investment. Tröger argues that this result stands in contrast to the established paradigm of backwards induction, and to the irrelevance of sunk costs suggested by standard economic theory. Tröger's analysis offers valuable insights into the origin of fair bargaining *outcomes*. However, his analysis cannot rationalize fairness *preferences* as put forward in the studies cited above. Indeed, in his model evolution does not discriminate between agents with different utility specifications. Instead, as mentioned above, evolution serves as a device selecting one of many a priori equivalent equilibria. All agents in Tröger's model have standard preferences. His model therefore appears less suitable as a foundation for fairness preferences.

The present paper rationalizes fairness preferences in a hold-up situation based on the idea that such preferences may create better incentives for a joint production than the traditional utility specification. The model as-

[4]See also Ellingson and Robles (2002), and Goree and Holt (2002).

sumes a disutility from inequality, so that unfair outcomes come as a cost to the individual members of a society. Still, as mentioned before, the division of the gains of production is a non-cooperative game with strictly opposing preferences over the efficient frontier. The beneficial effect of fairness preferences comes about because, in the strategic interactions with the other agent, disutilities from inequality may lead to a more accurate stock-taking of past contributions, which would be "sunk" in a society of individuals with standard utility specification.

In the formal analysis, we consider first the infinite-horizon alternating-offers bargaining game for the case of fairness preferences. We derive an explicit solution to this bargaining game for a given reference point. Thereafter, it is assumed that both parties have the ex-ante possibility to invest in a way that increases joint output. As we can show, under suitable assumptions, the modification of the utility specification dramatically changes the results of the bilateral investment game: While traditional utility specifications lead to severe underinvestment due to the hold-up problem, the integration of inequity aversion may promote investments very close to the efficient level.

The rest of the paper is structured as follows. Section 2 analyzes the infinite-horizon bargaining model between inequity-averse agents. In Section 3, we introduce an ex-ante investment stage and derive our main result Theorem 2. Section 4 concludes. The Appendix contains technical proofs.

2. Bargaining between parties with a reference point

The game structure is as in Rubinstein (1982). There are two parties $i = 1, 2$, who must agree to share a pie of size one. An agreement is a pair $(s_1, s_2) \in \Re^2$ satisfying $s_1 + s_2 = 1$. We do not restrict shares to be nonnegative. There are infinitely many periods. In all even periods 0, 2, 4 etc., party 1 proposes an agreement (s_1, s_2) that party 2 can accept or reject. If 2 accepts any offer, the game ends. If 2 rejects 1's offer in period $2k$, then in period $2k+1$, player 2 can in turn propose an agreement (s_1, s_2) that 1 can accept or reject. If 1 accepts, the game ends. If 1 rejects, then he can make an offer in the subsequent period, and so on.

Utility for party i is assumed to be strictly increasing in the share s_i allocated to party i. In addition, we assume that the parties have a common reference agreement $(s_1^{\text{ref}}, s_2^{\text{ref}})$ satisfying $s_1^{\text{ref}} + s_2^{\text{ref}} = 1$. Following the existing approaches to modeling fairness preferences, the reference point enters the utility function so that the disutility of a marginal reduction of the share is larger for shares $s_i < s_i^{\text{ref}}$ than for values $s_i > s_i^{\text{ref}}$. Specifically, we assume that the utility function of party i is given by

$$u_i(s_i, s_i^{\text{ref}}, t) := \delta_i^t(s_i - \alpha_i(s_i - s_i^{\text{ref}})^+ - \beta_i(s_i^{\text{ref}} - s_i)^+). \tag{1}$$

Here, the parameter α_i measures the marginal disutility from inequality when receiving more than the reference level, and β_i measures the marginal disutility from inequality when receiving less than the reference level, where

$0 \leq \alpha_i \leq \beta_i < 1$. The parameter $\delta_i \in (0;1)$ is the discount factor of party i, and $(s)^+ := \max\{0; s\}$. If no agreement is reached in finite time, then utility is zero for both parties.[5]

It follows from (1) that the reference agreement, if obtained in the initial period, will generate a reference utility pair $(u_1^{\text{ref}}, u_2^{\text{ref}}) := (s_1^{\text{ref}}, s_2^{\text{ref}})$. Moreover, as the utility function of party i is strictly increasing in its share s_i, for $i = 1, 2$, it should be clear that a party can achieve a utility level above the reference value only if it obtains more than in the reference agreement. In the linear analysis without social utility components (i.e., $\alpha_1, \alpha_2, \beta_1, \beta_2 = 0$), the reference agreement will not affect the outcome of the game. However, as will become clear later, when players care about the relative sizes of their payoffs, this parameter will be of relevance.

We start the analysis of the game by determining the bargaining set that results from possible agreements in the initial period. In the linear case the bargaining set is just a straight line. With inequity-averse preferences, however, the bargaining set possesses a characteristic kink at the reference agreement, as depicted in Figure 1. The following proposition describes the

- place Figure 1 here -

[5]The reader will note that the form of preferences is similar to the specification in Fehr and Schmidt (1999). In fact, for $s_i^{\text{ref}} = 1/2$, and for a common discount factor δ, the utility specification amounts to

$$u_i(s_i, \frac{1}{2}, t) = \delta^t s_i - \frac{\alpha_i}{2}(\delta^t s_i - \delta^t s_j)^+ - \frac{\beta_i}{2}(\delta^t s_j - \delta^t s_i)^+,$$

where $s_j = 1 - s_i$ is the share allocated to party j. As we will see in Section 3, the purpose of the somewhat generalized set-up is to allow for the possibility that the parties consider a division of the pie as fair that gives shares of different size to the individual parties.

bargaining set formally. It can be derived without much difficulty from the utility specification (1). Now, and throughout the paper, we will denote by j the party that is not i.

Proposition 1. *Denote by* $(u_1^{\text{ref}}, u_2^{\text{ref}})$ *the reference point of the bargaining parties. Assume that an agreement* (s_1, s_2) *is reached in period* $t = 0$*, yielding a utility pair* (u_1, u_2)*. Then* $u_j = g_j(u_i, u_i^{\text{ref}})$*, where*

$$g_j(u_i, u_i^{\text{ref}}) := 1 - u_i - \sigma_i(u_i - u_i^{\text{ref}})^+ - \frac{\sigma_j}{1+\sigma_j}(u_i^{\text{ref}} - u_i)^+,$$

and

$$\sigma_i := \frac{\alpha_i + \beta_j}{1 - \alpha_i} \geq 0.$$

Proof. Assume first $u_i \leq u_i^{\text{ref}}$. Then,

$$u_i = (1 + \beta_i)s_i - \beta_i s_i^{\text{ref}}. \tag{2}$$

On the other hand, $u_j \geq u_j^{\text{ref}}$ and therefore

$$u_j = (1 - \alpha_j)s_j + \alpha_j s_j^{\text{ref}}. \tag{3}$$

Replacing s_j by $1 - s_i$ and s_j^{ref} by $1 - s_i^{\text{ref}}$ in (3), and subsequently eliminating s_i using (2) yields the assertion for $u_i \leq u_i^{\text{ref}}$. The proof for the case $u_i > u_i^{\text{ref}}$ is analogous and therefore omitted. □

As we show in the Appendix, a straightforward adaptation of an argument given by Shaked and Sutton (1984) reveals that the bargaining game possesses a unique subgame-perfect equilibrium, yielding a utility of u_i^* for the

party i that makes the initial offer in the respective subgame. The pair (u_i^*, u_j^*) is characterized by the two equations

$$g_1(u_2^*) = \delta_1 u_1^* \text{ and } g_2(u_1^*) = \delta_2 u_2^*, \tag{4}$$

where we ignore the second argument of the functions $g_i(.,.)$ for simplicity. These equations allow a graphical interpretation, as shown in Figure 1. We denote by

$$\gamma_i := \frac{1-\delta_j}{1-\delta_i\delta_j} \tag{5}$$

the equilibrium share for party i in the linear model, when i makes the first proposal. Recall that $1-\gamma_i = \delta_j\gamma_j$ for $i = 1, 2$. The solution of the bargaining game can then be summarized as follows.

Theorem 1. *Assume that the parties have fairness preferences (1) based on the common reference agreement* $(s_i^{\text{ref}}, s_j^{\text{ref}})$. *Assume also that bargaining is efficient, i.e.,* $g_j(0, s_i^{\text{ref}}) > 0$. *Consider a subgame in which party* i *makes the initial offer. Then the unique subgame-perfect equilibrium in the subgame gives party* i *a utility of*

$$u_i^*(s_i^{\text{ref}}, s_j^{\text{ref}}) = \begin{cases} \gamma_i(1 - \dfrac{\sigma_i}{1+\sigma_i} s_j^{\text{ref}}\} & \text{if } s_i^{\text{ref}} \leq \gamma_i^-, \\ \dfrac{(1-\delta_j)(1+\sigma_j) + \{\sigma_i(1+\sigma_j) + \delta_j\sigma_j\} s_i^{\text{ref}}}{(1+\sigma_i)(1+\sigma_j) - \delta_i\delta_j} & \text{if } \gamma_i^- < s_i^{\text{ref}} < \gamma_i^+, \\ \gamma_i(1 + \sigma_j s_j^{\text{ref}}) & \text{if } s_i^{\text{ref}} \geq \gamma_i^+, \end{cases}$$

where

$$\gamma_i^- \quad := \quad \frac{\delta_i\gamma_i}{1+\sigma_i\gamma_j} \in (0; \gamma_i),$$

$$\gamma_i^+ \quad := \frac{\gamma_i(1+\sigma_j)}{1+\sigma_j\gamma_i} \in [\gamma_i; 1). \tag{6}$$

The utility of the other party in the subgame is equal to $\delta_j u_j^*(s_j^{\text{ref}}, s_i^{\text{ref}})$. *The outcome is achieved in the initial period. Moreover, the functions* $u_i^*(s_i^{\text{ref}}, s_j^{\text{ref}})$ *are continuous for* $i = 1, 2$.

Proof. See the Appendix. □

The first case in the statement of Theorem 1 corresponds to a situation where the reference share for party i is comparably low. In such a situation, the bargaining set, when compared to the one in Figure 1, would be shifted up and to the left along the dotted diagonal. The subgame-perfect equilibrium would then predict a utility profile that is on the lower right segment of this bargaining set. As a consequence, party i receives more than his or her reference share, while party j receives less than in the reference agreement. The second case in the statement of the Theorem is depicted in Figure 1. Here, party i, if the first to make an offer, receives again more than the reference share, and party j less. Finally, in the third case, the bargaining set would be shifted down and to the right, when compared to the one shown in Figure 1. Now the outcome would lie on the left upper segment of the bargaining set, giving party i a share below the reference point, and party j a share above it.

In a way summarizing the above mechanics, Figure 2 shows the equilibrium utility of the first-moving party as a function of the reference level. The - place Figure 2 here -

critical property of this function is that, in contrast to the model with linear utility specification, the equilibrium utility u_i^* is strictly increasing in the reference share s_i^{ref} for values below γ_i^+. As we will see in Section 3, this feature of the equilibrium utility implies that incentives to invest are much stronger for inequity-averse parties than for parties with a linear utility. Maybe surprisingly at first sight, the equilibrium utility is strictly decreasing in the reference share s_i^{ref} for values $s_i^{\text{ref}} > \gamma_i^+$. This effect comes about because for high reference values, the equilibrium outcome of the bargaining game will be on the upper left segment of the bargaining set (see Figure 1). A marginal increase in s_i^{ref} shifts the bargaining set downwards and to the right, decreasing the equilibrium utility for party i.

Another feature of the equilibrium is that for reference levels in the interval $(\gamma_i^-; \gamma_i^+)$, the bargaining procedure will result in a share for party i that is close to s_i^{ref}. In fact, as can be seen from Figure 1, the more patient the parties, the closer will be the equilibrium utility to the reference share in this area. The consequence of this effect is that sufficiently patient parties receive almost precisely the reference share, provided that the reference agreement is close enough to an equal division of the pie. This feature will be of central relevance in our discussion of efficient investment in the subsequent section.

3. Sunk costs

We will now extend the model by assuming that the size of the pie to be divided between the two parties depends on specific investments made in some ex-ante stage. The objective will be to compare incentives for investment for individuals with a standard utility function to the corresponding incentives of individuals with fairness preferences.

To simplify matters, we will consider a symmetric set-up from now onwards. Let the common parameters of inequality aversion be denoted by $\alpha := \alpha_1 = \alpha_2$ and $\beta := \beta_1 = \beta_2$. Let $I_1 \geq 0$ and $I_2 \geq 0$ denote the investment levels of parties 1 and 2, respectively. The costs of investments I_i for party i are denoted by $c(I_i)$. The function $c(.)$ satisfies $c(0) = 0$, and is assumed to be strictly increasing, convex, and twice differentiable. Investments affect the ex-post surplus, i.e., the total value of an agreement $y = y(I_1, I_2) \geq 0$. The production function $y(.,.)$ is assumed to be symmetric, i.e., $y(I_i, I_j) = y(I_j, I_i)$. Moreover, we require $y(.;.)$ to be strictly increasing in both I_1 and I_2, and to be strictly concave as well as differentiable.

The efficient investment levels (I_1^*, I_2^*) are characterized by the first-order conditions

$$\frac{\partial y}{\partial I_i}(I_i, I_j) = c'(I_i),\ i = 1, 2.$$

Holmström (1982) showed that an efficient investment cannot be obtained in this set-up with a sharing rule that depends only on y. E.g., if parties

with linear utility functions divide the surplus according to the sharing rule suggested by alternating offers bargaining then investment levels $(I_1^{\text{lin}}, I_2^{\text{lin}})$ will be characterized by the first-order conditions

$$\gamma_1 \frac{\partial y}{\partial I_1}(I_1, I_2) = c'(I_1) \tag{7}$$

$$(1-\gamma_1) \frac{\partial y}{\partial I_2}(I_1, I_2) = c'(I_2). \tag{8}$$

As a consequence of the so-called double marginalization of rents, investment is inefficient for $\gamma_1 \in (0;1)$. In the context of specific investments made into a long-term relationship, this incentive problem causes the well-known hold-up problem (see, e.g., Williamson, 1975, Grossman and Hart, 1986, Che and Hausch, 1999, and Maskin and Tirole, 1999).

We will now investigate the impact of inequity aversion on this outcome. The probably most natural candidate for a utility specification would start from net present values for the individual parties given by

$$m_i(I_i, I_j, s_i, t) = \delta^t s_i y(I_i, I_j) - c(I_i),$$

and define inequity-averse utility for party i as

$$\widehat{U}_i(m_i, m_j) := m_i - \widehat{\alpha}(m_i - m_j)^+ - \widehat{\beta}(m_j - m_i)^+$$

for parameters $\widehat{\alpha}$ and $\widehat{\beta}$. Unfortunately, this specification leads to a non-stationary bargaining problem when the parties choose heterogeneous investment levels in the ex-ante stage. Indeed, as the size of the pie shrinks

over time, while investments costs do not decline, the fair division of the cake would attribute a share that increases in t to the party that invested more ex ante, and a share that is decreasing in t to the party that invested less.

To avoid this complication, we will assume in the sequel that the parties have a stationary reference level in the form of a share s_i^{ref} of the overall pie that is created by the joint contribution. The utility component representing inequity aversion will be assumed to vary proportionally with the size of the pie. E.g., the utility loss of receiving a payoff of 100 currency units when the other party receives 120 is, measured in monetary terms, just twice the utility loss of receiving 50, when the other party ends up with 60. Under these assumptions, utility is given by

$$U_i(I_i, I_j) = \delta^t(s_i - \alpha(s_i - s_i^{\text{ref}})^+ - \beta(s_i^{\text{ref}} - s_i)^+)y(I_i, I_j) - c(I_i), \qquad (9)$$

where $0 < \delta < 1$ is the common discount factor. This specific form in not fully unplausible in our view, and turns out to be analytically very convenient. Also, while we solve the model only for this specific utility specification, we conjecture that the results do not depend on it in a critical way.

Our next assumption concerns the choice of the reference share s_i^{ref}. We believe that there are many plausible candidates for such a reference point. Moreover, the reference point could depend in principle on many characteristics of the contributing parties and the situation. In this study, we will focus on one specific case, and assume that individuals are concerned with

an ex-post equal economic profit. To derive this reference point, note that the agreement $(s_1^{\text{ref}}, s_2^{\text{ref}})$ that puts parties 1 and 2 into an equivalent economic post-bargaining position satisfies

$$s_1^{\text{ref}} y(I_1, I_2) - c(I_1) = s_2^{\text{ref}} y(I_1, I_2) - c(I_2). \tag{10}$$

A simple calculation using $s_1^{\text{ref}} + s_2^{\text{ref}} = 1$ shows that the reference share for party i is then given by

$$s_i^{\text{ref}}(I_i, I_j) = \frac{1}{2} + \frac{c(I_i) - c(I_j)}{2y(I_i, I_j)}, \tag{11}$$

provided that $y(I_i, I_j) > 0$. In the sequel, we will assume that the parties engage ex post in alternating-offers bargaining using this reference level.[6]

Proposition 2. *For any given investment level I_j of firm j, the reference share $s_i^{\text{ref}}(I_i, I_j)$ is strictly increasing in I_i.*

Proof. See the Appendix. □

We will now state conditions under which efficient pair of investments $(I_1, I_2) = (I^*, I^*)$ can be approximated as an equilibrium outcome between inequity-averse parties. One central condition concerns the profitability of production. With an unprofitable production, the reference share may react very strongly to small changes in investments. However, as can be seen from Figure 2, if

[6]The reference share described by (11) may appear intuitively too high off the equilibrium path, e.g., if party i chooses an inefficiently high investment level. However, it will be noted that a lower reference share off the equilibrium path would ceteris paribus lead to lower profit for the deviating party, and thus enforce our argument.

investments are lowered, then the equilibrium utility u_i^* does not decline at the same rate as the reference value. This could make it attractive for one party i to lower the investment somewhat so that the reference share would drop below γ_i^-.[7]

It turns out that the appropriate degree of profitability depends on the degree of inequity aversion, and is given by

$$\frac{c(I^*)}{y(0,I^*)} < \frac{\alpha+\beta}{2-\alpha+\beta}. \tag{12}$$

As the right-hand side of the inequality is always smaller than 1, the condition presupposes that $c(I^*) < y(0,I^*)$ in the first place. Note that, for linear utilities, the right-hand side of (12) will disappear so that the condition cannot be satisfied in the traditional set-up. As we show in the proof of the subsequent result, condition (12) makes sure that the reference share of a deviating player cannot fall below γ^-. With these conditions we can prove our main result:

Theorem 2. *Consider the symmetric two-stage model of joint production introduced in the text above. Assume that the parties' utility functions exhibit inequity-aversion as specified in (9) with $\beta > 0$. Assume also that reference shares are formed on the basis of equal economic performance, as specified*

[7]Indeed, one can show that the equilibrium share function for party i has always an upwards kink at γ^-, as suggested by Figure 2. This feature of the equilibrium implies that a party's objective function is non-concave for low values of the reference share, i.e., in a neighborhood of γ^-. Condition (12), specified below, ensures that this non-concavity is not reached by $s_i^{\text{ref}}(I_i, I_j)$ if the other party j chooses an investment level I_j close to I^*.

in (10). Assume finally that unilateral production is sufficiently profitable as captured by condition (12). Then for any $\varepsilon > 0$, there is a common discount factor $\widetilde{\delta}$ such that for any $\delta \in (\widetilde{\delta}; 1)$, the equilibrium investment level of each party lies in the interval $(I^* - \varepsilon; I^* + \varepsilon)$.

Proof. See the Appendix. □

The idea of the proof is as follows. From our earlier results, it follows that for sufficiently patient parties, and for a reference share s_i^{ref} close to 1/2, the outcome of the bargaining game will close to the reference share. The reference share, however, is a function of the relative size of investments made in the first period. This provides nearly efficient incentives for ex-ante investments in an intermediary range of investment levels. The profitability assumption implies now that the reference value is not too responsive to investments. E.g., as can be seen from (11), if production become more efficient ceteris paribus, then s_i^{ref} moves to 1/2. The profitability assumption thereby ensures that the reference level never falls below γ_i^-, even for a one-sided deviation to zero investments. As can be seen from Figure 2, this guarantees that the equilibrium utility u_i^* is a concave function of the reference share, and therefore of the investment in the first-stage. It is possible that for a very high investment, the reference share will exceed γ_i^+. However, the kink at γ_i^+ reduces the marginal return from investment and makes overinvestment less attractive. As a consequence, a deviation is not attractive. For $\delta \to 1$, this implies efficient incentives for investment.

4. Conclusion

This paper has studied the infinite-horizon alternating-offers bargaining game for the case of inequity-averse preferences. It has been shown that under inequity aversion, there is a shift of the bargaining outcome towards fair division. When the parties may engage in ex-ante investments to enlarge the surplus that is to be divided ex post, then the outcome of the non-cooperative bargaining game will attribute a larger share to the party that invested more. This effect was shown to reduce incentive problems in team production and hold-up problems. We conclude that communities consisting of inequity-averse individuals might be able to offer a better protection of relationship-specific investments than economies consisting of individuals with standard utility specification. In particular, this argument would provide an economic rationale for the apparent prevalence of fairness preferences in the laboratory.

Appendix

This appendix contains proofs of Theorem 1, Proposition 2, and Theorem 2.

Proof of Theorem 1. The proof has two parts. In the first part, we show that the bargaining game between parties with a reference point has a unique subgame-perfect equilibrium. We are explicit here because of the possibility of negative utility from unfair outcomes, which is usually excluded in standard approaches. Cf., e.g., Rubinstein (1982, Assumption A-2). In the second part of the proof, we derive explicitly the parties' equilibrium utilities as a function of the reference level and of the fairness parameters.

Part 1. The second step of the subsequent argument is adapted from Fudenberg and Tirole (1991, Subsection 4.4.2), who use the following notation. Let $\underline{v}_i$ and $\overline{v}_i$, respectively, denote the infimum and the supremum of the set of utilities that i may obtain in some subgame-perfect equilibrium of the bargaining game, in which i makes the initial offer. Similarly, let $\underline{w}_i$ and $\overline{w}_i$, respectively, denote the infimum and the supremum of the set of utilities that i may obtain in some subgame-perfect equilibrium of the bargaining game, in which j makes the initial offer.

Step 1. We claim that in any Nash equilibrium of the bargaining game, party i obtains a non-negative utility. To see why, note that otherwise, party i could deviate profitably to the following strategy. If $s_i^{\text{ref}} \geq 0$, then i would always propose a share of $s_i = s_i^{\text{ref}}$, and reject all offers made by party j. If

$s_i^{\text{ref}} < 0$, then i would always propose a share of

$$s_i = \frac{\alpha_i}{1-\alpha_i}|s_i^{\text{ref}}|,$$

and reject all offers made by party j. It is easy to check that this strategy guarantees a utility of at least zero. This proves the claim. By symmetry, the claim is likewise true for party j, so that party i cannot get more than $g_i(0)$ in any Nash equilibrium. We therefore get

$$0 \leq \underline{v}_i \leq \overline{v}_i \leq g_i(0). \tag{13}$$

Step 2. Consider now an offer by party i. It is clear that party j accepts any offer above $\delta_j \overline{v}_j$. Hence

$$\underline{v}_i \geq g_i(\delta_j \overline{v}_j). \tag{14}$$

Also, party i will not offer more than $\delta_j \overline{v}_j$. Thus,

$$\overline{w}_j \leq \delta_j \overline{v}_j. \tag{15}$$

Party i's offer will be either accepted or rejected. If accepted, it can give i at most $g_i(\delta_j \underline{v}_j)$. If rejected, party i obtains at most $\delta_i \overline{w}_i$. Therefore

$$\begin{aligned} \overline{v}_i &\leq \max\{g_i(\delta_j \underline{v}_j), \delta_i \overline{w}_i\} \\ &\leq \max\{g_i(\delta_j \underline{v}_j), \delta_i^2 \overline{v}_i\}, \end{aligned} \tag{16}$$

where we have used (15) in the second inequality. We claim that

$$\max\{g_i(\delta_j \underline{v}_j), \delta_i^2 \overline{v}_i\} = g_i(\delta_j \underline{v}_j). \tag{17}$$

Indeed, if (17) is violated, then $\overline{v}_i \leq \delta_i^2 \overline{v}_i$ from (16). But then $\underline{v}_i = \overline{v}_i = 0$ by (13) and $\delta < 1$. However, this would imply $\underline{v}_j \geq g_j(0)$ by (14), which implies $\overline{v}_j = \underline{v}_j = g_j(0)$ using (13) again. But then, because the efficient frontier is attractive, we have $0 < \delta_j \overline{v}_j < g_j(0)$. Using (14) again, this yields

$$\underline{v}_i > g_i(\delta_j \overline{v}_j) > 0,$$

and contradicts $\underline{v}_i = 0$. This proves our claim (17). Combining this with (16), we have therefore shown

$$\overline{v}_i \leq g_i(\delta_j \underline{v}_j). \tag{18}$$

Step 3. From inequalities (14), (18), and their symmetric counterparts, noting that $g_i(.)$ and $g_j(.)$ are strictly decreasing and mutually inverse functions, we find the two central inequalities

$$g_j(\underline{v}_i) \leq \delta_j g_j(\delta_i \underline{v}_i), \tag{19}$$

$$g_j(\overline{v}_i) \geq \delta_j g_j(\delta_i \overline{v}_i). \tag{20}$$

From (13) and from the fact that the function

$$u_i \longmapsto g_j(u_i) - \delta_j g_j(\delta_i u_i)$$

is strictly decreasing on the interval $[0; g_i(0)]$, it follows that (19) and (20) must be equalities. Thus, in any subgame-perfect equilibrium, the party i that makes the initial offer receives a utility $u_i^* := \underline{v}_i = \overline{v}_i$, characterized by

$$g_j(u_i^*) = \delta_j g_j(\delta_i u_i^*). \tag{21}$$

Consider now the party j that does not make the initial offer. A rejection to an offer by i secures party j a utility of $\delta_j u_j^*$. Thus, $\underline{w}_j \geq \delta_j u_j^*$. Using (15), we get that the equilibrium utility for party j amounts to

$$\delta_j u_j^* = \underline{w}_j = \overline{w}_j. \tag{22}$$

This shows the uniqueness of the subgame-perfect equilibrium utilities. The argument that even the subgame-perfect equilibrium profile is unique given that utilities are unique can be found in Fudenberg and Tirole (1991, p. 116) and is therefore omitted.

Part 2. Starting from a subgame-perfect equilibrium in the bargaining game characterized by a share u_i^* for the first-moving party i, we have precisely one of the following three cases.

Case A. $u_i^* \leq u_i^{\text{ref}}$. By Proposition 1, in this case,

$$g_j(u_i) = 1 - \frac{1}{1+\sigma_j} u_i - \frac{\sigma_j}{1+\sigma_j} s_i^{\text{ref}} \tag{23}$$

for $u_i = u_i^*$ and for $u_i = \delta_i u_i^*$. Equilibrium is characterized by

$$\delta_j g_j(\delta_i u_i^*) = g_j(u_i^*). \tag{24}$$

Combining the last equations and rearranging yields

$$u_i^* = \gamma_i(1 + \sigma_j s_j^{\text{ref}}). \tag{25}$$

Thus, case A occurs only if

$$\gamma_i(1 + \sigma_j s_j^{\text{ref}}) \leq s_i^{\text{ref}}. \tag{26}$$

Rewriting (26) yields $s_i^{\text{ref}} \geq \gamma_i^+$, where γ_i^+ is defined as in the statement of the Theorem. However, it is likewise clear that if $s_i^{\text{ref}} \geq \gamma_i^+$, then (25) constitutes an equilibrium outcome.

Case B. $\delta_i u_i^* \geq u_i^{\text{ref}}$. We find from Proposition 1 that

$$g_j(u_i) = 1 - (1 + \sigma_i)u_i + \sigma_i s_i^{\text{ref}}$$

for $u_i = u_i^*$ and for $u_i = \delta_i u_i^*$. Thus, in equilibrium we have

$$u_i^* = \gamma_i\{1 - \frac{\sigma_i}{1 + \sigma_i} s_j^{\text{ref}}\}. \tag{27}$$

Hence, in case B we have

$$\delta_i \gamma_i\{1 - \frac{\sigma_i}{1 + \sigma_i} s_j^{\text{ref}}\} \geq s_i^{\text{ref}}.$$

Rearranging yields $s_i^{\text{ref}} \leq \gamma_i^-$. It is not difficult to check that if this is true, we have an equilibrium given by (27).

Case C. $u_i^* > u_i^{\text{ref}}$ and $\delta_i u_i^* < u_i^{\text{ref}}$. Proposition 1 yields for this case

$$g_j(u_i^*) = 1 - (1 + \sigma_i)u_i^* + \sigma_i s_i^{\text{ref}} \tag{28}$$

$$g_j(\delta_i u_i^*) = 1 - \frac{1}{1 + \sigma_j}\delta_i u_i^* - \frac{\sigma_j}{1 + \sigma_j} s_i^{\text{ref}}. \tag{29}$$

Plugging equations (28) and (29) into (24) and rearranging yields the formula for u_i^* given in the statement of the Theorem for values $\gamma_i^- < s_i^{\text{ref}} < \gamma_i^+$. Because existence and uniqueness of the subgame-perfect equilibrium is guaranteed for all $u_i^{\text{ref}} \in [0; 1]$, case C will occur if and only if neither case A nor case B occurs.

The assertion concerning the utility of the second-moving party j follows from (22). A long, but straightforward calculation shows that $u_i^*(s_i^{\text{ref}}, s_j^{\text{ref}})$ is continuous at $s_i^{\text{ref}} = \gamma_i^-$ and at $s_i^{\text{ref}} = \gamma_i^+$. This proves the Theorem. □

Proof of Proposition 2. Fix $I_j \geq 0$, and write $c_j := c(I_j)$. Because the function $c(I_i)$ is strictly increasing, convex, and differentiable in I_i, there is an inverse function $\phi(c_i)$ that is strictly increasing, concave, and differentiable in $c_i = c(I_i)$. To prove the claim, it suffices to show that the function

$$\varphi(c_i) := \frac{c_i - c_j}{y(\phi(c_i), I_j)}$$

is strictly increasing in c_i. The first-order derivative reads

$$\varphi'(c_i) = \frac{y(\phi(c_i), I_j) - (c_i - c_j)\dfrac{\partial y}{\partial I_i}(\phi(c_i), I_j)\phi'(c_i)}{y(\phi(c_i), I_j)^2}.$$

Obviously, the function $\varphi(.)$ is strictly increasing in c_i as long as $c_i \leq c_j$. But for $c_i > c_j$, this is likewise the case because $y \geq 0$, and the function that maps c_i into $y(\phi(c_i), I_j)$ is strictly concave. □

Proof of Theorem 2. The strategy of the proof is to construct, for any given δ that is sufficiently close to 1, a subgame-perfect equilibrium, such that the corresponding levels of bilateral investments approximate (I^*, I^*). Fix for the moment some discount factor $\delta \in (0; 1)$. Note that in this symmetric set-up, the parameters γ_i, γ_i^-, γ_i^+, and σ_i are independent of i, so that we can drop the index i in the sequel. From Proposition 2, we know that for

any given investment level $I_j \geq 0$ for party j, there are investment levels $I_i^+ = I_i^+(I_j) \in \Re_0^+ \cup \{\infty\}$ and $I_i^- = I_i^-(I_j) \in \Re_0^+$ such that $s_i^{\text{ref}} \geq \gamma^+$ if and only if $I_i \geq I_i^+$, and such that $s_i^{\text{ref}} < \gamma^-$ if and only if $I_i < I_i^-$. Fix now an investment level I_j. Then

$$U_i(I_i, I_j) = \{\mu_i(I_i, I_j) + \lambda_i(I_i, I_j) s_i^{\text{ref}}(I_i, I_j)\} y(I_i, I_j) - c(I_i),$$

where the functions $\mu_i(I_i, I_j)$ and $\lambda_i(I_i, I_j)$ are given by Theorem 1. These functions are piecewise constant, with jumps at values $I_i^-(I_j)$ and $I_i^+(I_j)$. Rewriting yields

$$U_i(I_i, I_j) = \{\mu_i + \frac{\lambda_i}{2}\} y(I_i, I_j) - (1 - \frac{\lambda_i}{2}) c(I_i) - \frac{\lambda_i}{2} c(I_j), \tag{30}$$

where we ignore the arguments of μ_i and λ_i. The first-order conditions are given by

$$\{\mu_i + \frac{\lambda_i}{2}\} \frac{\partial y}{\partial I_i} - (1 - \frac{\lambda_i}{2}) \frac{\partial c}{\partial I_i} = 0, \tag{31}$$

for $i = 1, 2$. Denote by $(I_1^\#(\delta), I_2^\#(\delta))$ the solution to the first-order conditions (31) assuming that the parameters μ_i and λ_i are as in the intermediary case $I_i^- \leq I_i < I_i^+$. A straightforward calculation using (31) and Theorem 1 shows that for $i = 1, 2$, we have $I_i^\#(\delta) \to I^*$ for $\delta \to 1$. We will prove now that $(I_1^\#(\delta), I_2^\#(\delta))$ is an equilibrium in the investment stage. Clearly, for δ sufficiently close to one, $|c(I_i^\#(\delta)) - c(I_j^\#(\delta))|$ is arbitrarily small, so that $s_i^{\text{ref}}(I_1^\#(\delta); I_2^\#(\delta))$ is close to $1/2$. At the same time,

$$\gamma^- \to \frac{1}{2+\sigma} \text{ and } \gamma^+ \to 1 - \frac{1}{2+\sigma}$$

for $\delta \to 1$, so that for δ sufficiently close to 1, we get

$$I_i^-(I_j^\#(\delta)) < I_i^\#(\delta) < I_i^+(I_j^\#(\delta)).$$

Thus, the first-order condition combined with the concavity of (30) yields that $I_i^\#(\delta)$ is an optimal response to $I_j^\#(\delta)$ on the interval

$$[I_i^-(I_j^\#(\delta)); I_i^+(I_j^\#(\delta))].$$

To check that $I_i^\#(\delta)$ is optimal also globally, we consider first the deviation to some $I_i \geq I_i^+(I_j^\#(\delta))$. By the continuity of $U_i(I_i, I_j^\#(\delta))$ at $I_i = I_i^+(I_j^\#(\delta))$, it suffices to show that $\partial U_i(I_i, I_j^\#(\delta))/\partial I_i$ is jumping downwards at $I_i = I_i^+(I_j^\#(\delta))$. To see why this is true, note that

$$\frac{\partial U_i}{\partial I_i} = \frac{\partial u_i^*}{\partial s_i^{\text{ref}}} \frac{\partial s_i^{\text{ref}}}{\partial I_i} y + u_i^* \frac{\partial y}{\partial I_i} - \frac{\partial c}{\partial I_i},$$

and recall from Theorem 1 that $\partial u_i^*/\partial s_i^{\text{ref}}$ is jumping downwards at $I_i = I_i^+(I_j^\#(\delta))$, while $\partial s_i^{\text{ref}}/\partial I_i > 0$ by Proposition 2. This proves that an upwards deviation is not attractive. To ensure that also a downward deviation is not profitable, it suffices to guarantee that $I_i^-(I_j^\#(\delta)) \leq 0$ or, equivalently, that

$$\frac{1}{2} + \frac{c(0) - c(I_j^\#(\delta))}{2y(0, I_j^\#(\delta))} \geq \frac{\delta\gamma}{1+\sigma\gamma}. \tag{32}$$

This condition is satisfied if

$$\frac{c(I_j^\#(\delta))}{y(0, I_j^\#(\delta))} \leq 1 - \frac{2\delta\gamma}{1+\sigma\gamma}.$$

For δ close to one, this condition, in turn, follows from the profitability assumption (12). Thus, the investment level $I_i^\#$ is an optimal response to

$I_j^{\#}$. The analysis of player j's investment decision follows the same steps and is therefore omitted. This proves the assertion. $\square$

References

Bolton, G., and A. Ockenfels, 2000, ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review* **90**, 166-193.

Bolton, G., 1991, A Comparative Model of Bargaining: Theory and Evidence, *American Economic Review* **81**, 1096-1136.

Che, Y.-K-, and D. Hausch, 1999, Cooperative Investments and the Value of Contracting, *American Economic Review* **89**, 125-147.

Ellingson, T., and J. Robles, 2002, Does Evolution Solve the Hold-up Problem?, *Games and Economic Behavior* **39**, 28-53.

Fehr, E., and K. Schmidt, 1999, A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics* **114**, 817-868.

Fudenberg, D., and J. Tirole, 1991, *Game Theory*, MIT Press, Cambridge.

Goeree, J., and C. Holt, 2000, Asymmetric Inequality Aversion and Noisy Behavior in Alternating-Offer Bargaining Games, *European Economic Review* **44**, 1079-1089.

Grossman, S., and O. Hart, 1986, The Costs and Benefits of Ownership, A Theory of Lateral and Vertical Integration, *Journal of Political Economy* **94**, *691-719.*

Güth, W., 1995, An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives, *International Journal of Game Theory* **24**, *323-344.*

Holmström, B., 1982, Moral Hazard in Teams, *Bell Journal of Economics* **13**, 324-340.

Maskin, E., and J. Tirole, 1999, Unforeseen Contingencies and Incomplete Contracts, *Review of Economic Studies* **66**, 83-114.

Ochs, J., and A. Roth, 1989, An Experimental Study of Sequential Bargaining, *American Economic Review* **10**, 6-38.

Rubinstein, A., 1982, Perfect Equilibrium in a Bargaining Model, *Econometrica* **50**, 97-109.

Shaked, A., and J. Sutton, 1984, Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model, *Econometrica* **52**, 1351-1364.

Ståhl, I., 1972, *Bargaining Theory*, Stockholm: Economic Research Institute, Stockholm School of Economics.

Tröger, T., 2002, Why Sunk Costs Matter for Bargaining Outcomes: The Evolutionary Approach, *Journal of Economic Theory* **10**2, 375-402.

Williamson, O., 1975, *Markets and Hierarchies: Analysis and Antitrust Implication*s. New York: The Free Press.

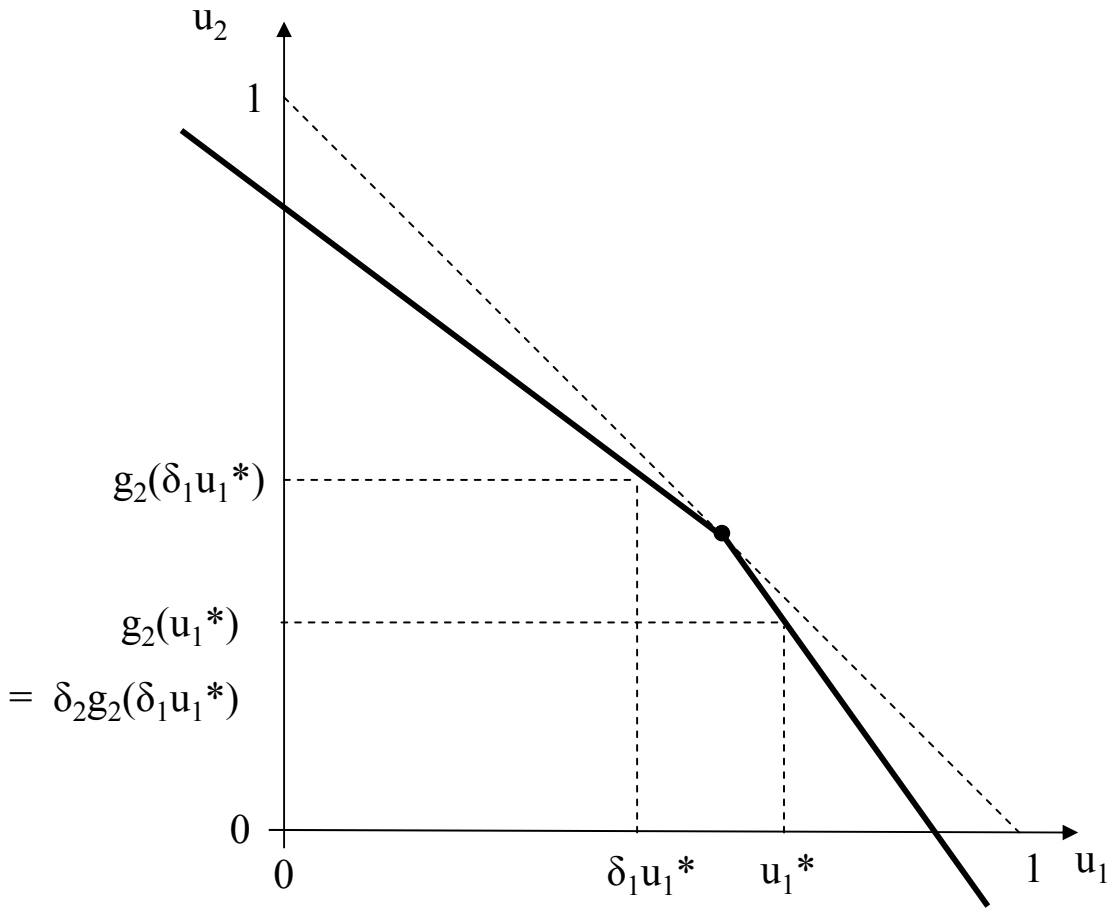

Figure 1

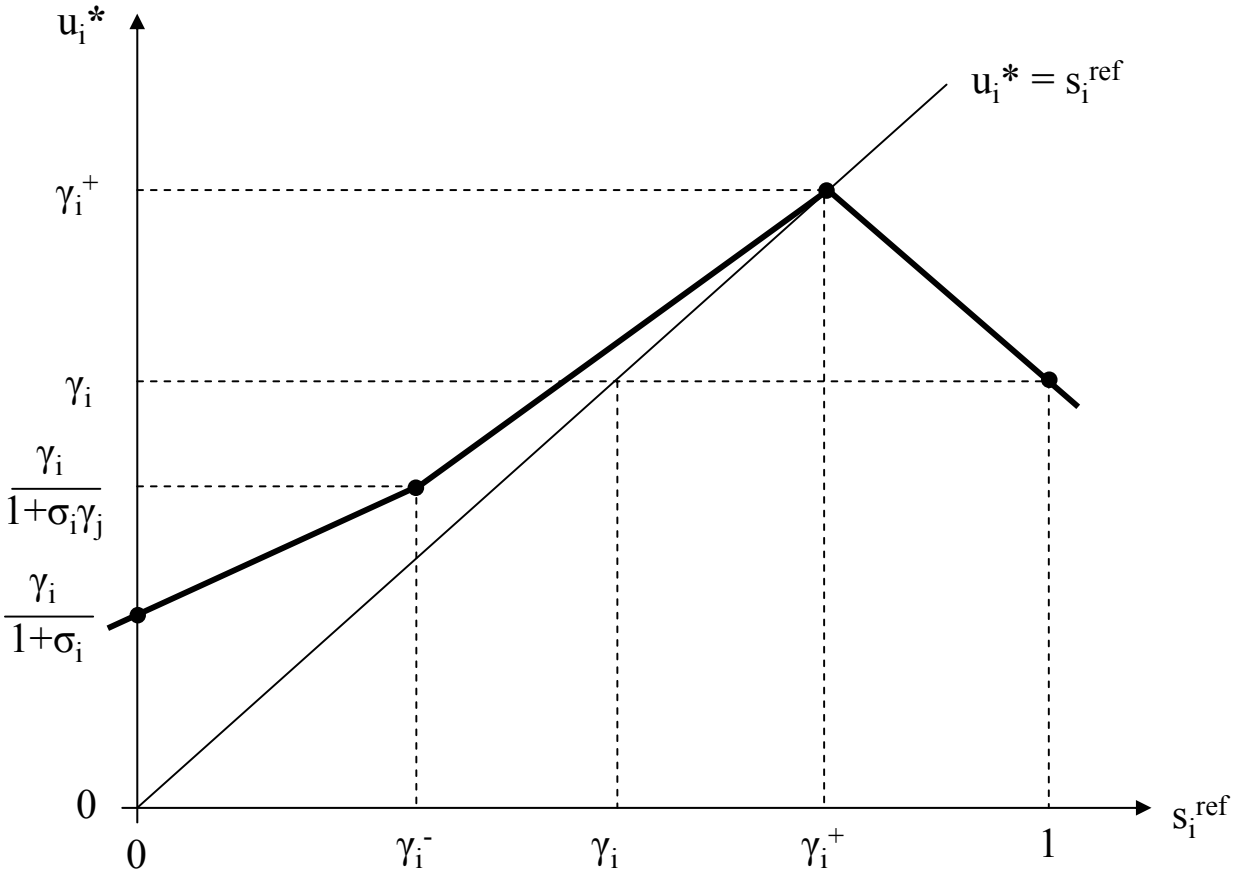
u_i^*
$u_i^* = s_i^{ref}$
γ_i^+
γ_i
$\frac{\gamma_i}{1+\sigma_i\gamma_j}$
$\frac{\gamma_i}{1+\sigma_i}$
0
s_i^{ref}
0
γ_i^-
γ_i
γ_i^+
1

Figure 2