Christian Ewerhart*

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

Abstract

When investments are specific to a relationship and contracting possibilities are incomplete, the efficiency of a joint venture may be severely impaired by ex-post opportunistic and hold-up type behavior. How is the logic of this argument affected by inequity aversion? In this paper I show that incentives to invest are stronger with inequity aversion because a higher investment by an individual agent increases not only the total surplus to be divided, but also, generally, the relative share of the surplus obtained by this agent in the ex-post negotiation. In fact, when production is sufficiently profitable and agents are sufficiently patient, then first-best investment levels may be approximated without any contract.

JEL-Classification: C78, L22.

Keywords: Alternating-Offers Bargaining; Inequity Aversion; Hold-Up Problem.

1 Introduction

It is by now a well-established proposition that the need to engage in relationship-specific investments may lead to efficiency losses when comprehensive contracting is either too costly or simply impossible. E.g., long-term relationships between a buyer and a seller of an intermediate good may be adversely affected by ex-post opportunistic behavior (Williamson (1975)). Even if ex-post negotiations are efficient, the anticipation of a potential appropriation of quasi rents may cause severe hold-up and underinvestment problems when contracting possibilities are incomplete (Klein, Crawford, and Alchian (1978); Grout (1984); Williamson (1985); Grossman and Hart (1986); Tirole (1986); Hart and Moore (1988); Hart (1995)) or when at least renegotiation cannot be ruled out (Che and Hausch (1999)).

* Institute for Empirical Research in Economics, Winterthurerstrasse 30, CH-8006 Zurich, Switzerland. E-mail: christian.ewerhart@iew.unizh.ch.

** I would like to thank Ernst Fehr for useful discussions and encouragement. For helpful suggestions, I am grateful to Peter-Jürgen Jost and Alfred Wagenhofer (the editors), two anonymous referees, Patrick Schmitz, participants of the GEABA conference in Bonn in late 2004, and seminar participants in Bern and Zurich.

For example, when the development and realization of an innovative business model such as an internet search engine requires the long-term collaboration of two or more highly skilled individuals with complementary backgrounds, then it may not always be feasible to write down at an early stage which contributions should be made by the individuals, and how the potential surplus should be shared in the end. As a consequence, when the time has come to bargain about an ownership structure of the emerging firm, there may be the imminent risk of opportunistic behavior by a party that has made only a comparably small contribution, but that nevertheless remains indispensable to fully realize the project's proceeds. Opportunism is feasible in such a situation because early unilateral investments that could not be protected by suitable contractual arrangements will be sunk, and should play no role in ex-post negotiations between rational agents with traditional utility specification. Thus, because the outside option in the negotiations does not reflect prior contributions, ex-post opportunistic behavior is feasible, which may cause the well-known hold-up inefficiency.

This inefficiency proposition of disintegrated production is intuitively challenged by a number of contributions showing that bargaining behavior of experimental subjects appears to be influenced not only by considerations of bargaining power, but also by "nonstrategic" factors such as prior contributions to a joint project. In fact, Selten (1978) has reported on *reward allocation experiments* in which first two subjects perform a common task in separate rooms and later one of the subjects is asked to distribute a sum of money between herself and the other subject. It turned out that subjects with the inferior contribution exhibited a tendency to distribute rewards in proportion to the announced contributions, despite the full discretion given to the distributing subject. More recently, experimental examinations of bargaining over the division of a jointly produced surplus have been performed in particular by Hackett (1993), Königstein (2000), and Gantner, Güth, and Königstein (2001). A common finding in these contributions is that experimental bargaining behavior is different when the surplus is created by the subjects themselves rather than provided for free. Moreover, the share of realized surplus is increasing in the relative investment, which suggests a mitigation of the hold-up problem. Overall, the predictions of the traditional utility model have been surprisingly poor in this context[1].

Recent work has sought to reconcile these observations with the theoretical approach[2] by examining the proposition that the ex-post bargaining game may allow several equilibria, so that evolutionary forces or other criteria of equilibrium selection may deter-

1 Further factors affecting experimental bargaining behavior include fair and justified results (Güth, Schmittberger and Schwarze (1982)), social conventions (Roth, Malouf, and Murnighan (1981)), relative payoff considerations (Ochs and Roth (1989); Bolton (1991), perceived intentions (Rabin (1993); Dufwenberg and Kirchsteiger, (1998)), verbal coordination and commitment (Ellingsen and Johannesson (2004a; 2004b)), and even noise (Goeree and Holt, (2000)). See Güth (1994) for empirical evidence of distributive justice.

2 See, e.g., Binmore, Shaked, and Sutton (1985).

mine outcomes of the bargaining stage that vary with initial investments[3]. In this paper, I show that innovative approaches to modeling inequity aversion suggest an alternative resolution of this tension between the traditional and the social psychological approach to bargaining with prior production, even when the bargaining parties face no explicit deadline and negotiations allow only one equilibrium. In the formal analysis, I equip economic agents with social preferences entailing inequity aversion, as suggested by Bolton (1991), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000). I then consider a bilateral investment game in which two rational agents jointly produce a potential surplus which is divided ex post according to the infinite-horizon alternating-offers procedure suggested by Rubinstein (1982). The results are as follows.

With inequity aversion, both offers and acceptance levels in the bargaining game depend on the relative contributions in the investment stage. As a rule, a relatively larger (smaller) investment by one agent leads to offers that assign a larger (smaller) share to that agent. Similarly, a relatively larger (smaller) investment by one agent generally increases (decreases) the acceptance level of that agent and decreases (increases) the acceptance level of the other agent. The consequences of having the outcome of the bargaining game biased towards a "fair" division are striking. For any given level of inequity aversion, I show that when the situation is ex-ante symmetric, and when joint production is sufficiently profitable, then the *investment levels of patient agents approximate the first-best level without any contract*. Thus, when I extrapolate the evidence on fairness and inequity aversion from experimental situations to real-world decision making, then the theoretical findings of my paper imply that the severity of the hold-up problem may not be as omnipresent as suggested by traditional models.

The intuition behind this result is illustrated by the following outline of the model for "very patient" agents. Assume that two agents, 1 and 2, may ex ante invest I_1 and I_2 at costs $C(I_1)$ and $C(I_2)$, respectively, to allow a mutually beneficial ex-post agreement of value $Y(I_1,I_2)$. The efficient investment level I^{FB} for each agent can then be determined by maximizing total welfare $Y(I_1,I_2) - C(I_1) - C(I_2)$. But when the surplus Y is divided equally (via negotiations between very patient agents) and ex-ante investment levels are determined in a noncooperative way, then the equilibrium solution predicts the second-best investment level $I^{SB} \neq I^{FB}$. An inefficiency comes about here because agent 1, for instance, maximizes $Y(I_1,I_2)/2 - C(I_1)$, so that her marginal return is just half as large as it should be to achieve efficient investments.

However, with inequity averse agents, the division of the surplus is only equal (for very patient agents) under the condition that both agents choose the same investment level. Thus, higher levels of contributions can form a noncooperative equilibrium in the invest-

3 See in particular Tröger (2002), Ellingsen and Robles (2002), and Carmichael and MacLeod (2003). For instance, Tröger (2002) 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 a unique equilibrium in which the division of the gains from investment critically depends on prior investment. This result stands in sharp contrast to the established paradigms of both backwards induction and the irrelevance of sunk costs.

ment stage because deviations from the efficient level have consequences not only on the total surplus, but also on the relative share of the surplus that the agent obtains in the bargaining outcome. Assume that both agents invest approximately I^{FB}, and that agent 1 contemplates a deviation to a somewhat lower investment level. Such a deviation may appear attractive for agent 1 because the marginal cost of investment strictly exceeds her share in the marginal return from a higher surplus. However, because agent 2 is inequity averse, the outcome of the negotiations will also be affected, so that agent 1 would expect to obtain only a reduced share of the anyhow reduced surplus. This consideration makes a deviation for agent 1 unattractive. Indeed, it turns out that in a neighborhood of the equilibrium outcome, the alternating-offers bargaining game allocates to agents 1 and 2 each (provided they are very patient) an approximate net profit of $\{Y(I_1, I_2) - C(I_1) - C(I_2)\}/2$. This division of the surplus, however, ensures approximately efficient incentives to invest.

The paper is structured as follows. Section 2 describes the set-up, and derives the equilibrium for the traditional utility specification. In Sections 3 and 4, I analyze the infinite-horizon bargaining model between inequity averse agents. Section 5 contains a discussion of incentives for ex-ante investments, and a statement of the main result Theorem 2. Section 6 concludes. All proofs appear in the Appendix.

2 Bilateral investments

There are two agents $i = 1, 2$, who have the opportunity to engage in a profitable joint venture. By notational convention, agent j is the business partner of agent i, i.e., if $i = 1$ then $j = 2$, and if $i = 2$ then $j = 1$. The model has two stages. In the first stage (ex ante), the two agents make contributions to the joint venture in the form of relationship-specific investments. In the second stage (ex post), the investments lead to a potential surplus, which is contingent on the approval of both agents and therefore divided by multiperiod negotiations.

Throughout the paper, I assume that potential rents at the second stage cannot be protected effectively by a contract written and signed at the investment stage. This may be an intuitively plausible assumption, e.g., in a bilateral relationship with two-sided investments, when objective measures for costs and benefits from the joint venture are not available, and agents are unable to commit not to renegotiate. As an illustration, I note that in the introductory example of the internet search engine, both the contribution made by the respective individual and the intangible value of the created business concept would probably be more than difficult to quantify in precise terms.

Formally, let the size of the potential surplus $Y = Y(I_1, I_2) \geq 0$ depend on investments $I_1 \geq 0$ and $I_2 \geq 0$ made by agents 1 and 2, respectively, in the first stage. The production function $Y(.,.)$ is required to be strictly increasing in both arguments, and to be twice continuously differentiable with a negative definite Hessian. To achieve a symmetric set-up, I assume that $Y(I_1, I_2) = Y(I_2, I_1)$ for all I_1, I_2. In particular, this assumption allows me to write $Y(I_i, I_j)$ without causing a potential misinterpretation. The simplifying notation $(\partial Y/\partial I_i)(I_i, I_j)$ stands for $(\partial Y/\partial I_1)(I_1, I_2)$ if $i = 1$, and for $(\partial Y/\partial I_2)(I_1, I_2)$ if $i = 2$.

Costs of investments for agent i are denoted by $C(I_i)$. The function $C(.)$ satisfies $C(0) = 0$, and is assumed to be strictly increasing, convex, and twice continuously differentiable. Agent i's monetary profit is then given by

$$\Pi_i = s_i Y - C_i, \tag{1}$$

where s_i denotes the share of the surplus allocated to agent i in the second stage, and $C_i = C(I_i)$. To ensure an interior efficiency solution, I assume throughout the paper that $(\partial Y/\partial I_i)(0,0) > (\partial C/\partial I_i)(0)$ for $i = 1, 2$. The first-best investment level $I^{FB} = I_1^{FB} = I_2^{FB}$ is then characterized by the two first-order conditions

$$\frac{\partial Y}{\partial I_i}(I_i, I_j) = \frac{\partial C}{\partial I_i}(I_i), \tag{2}$$

for agents $i = 1, 2$.

Noncooperative negotiations in the second stage follow the alternating-offers procedure, as in Rubinstein (1982). The model allows infinitely many periods of negotiation. In all even periods 0, 2, 4, etc., agent 1 proposes an *agreement* (s_1, s_2) that agent 2 may either accept or reject. Here, and throughout the paper, an agreement is a pair (s_1, s_2) satisfying $s_1 + s_2 = 1$. If agent 2 accepts any offer, the game ends. If 2 rejects 1's offer in period $2k$, then in period $2k + 1$, agent 2 can in turn propose an agreement (s_1, s_2) that 1 can either accept or reject. If 1 accepts, the game ends. If 1 rejects, then she can make an offer in the subsequent period, and so on. Agents are assumed to discount their respective share s_i with a common discount factor $\delta \in (0; 1)$. In the hypothetical case that negotiations go on forever, both agents obtain a share of zero.

The bargaining model is a game of perfect information and is solved by deriving the subgame-perfect equilibrium (Selten (1975)). This concept requires that agents' strategies induce a Nash equilibrium in any subgame of the extensive form. Rubinstein (1982) notes that the infinite-horizon bargaining game has a stationary structure, which is due to the fact that when negotiations are delayed by just one period, then the roles of the two agents are exchanged and the discounting merely rescales the payoffs. This insight suggests the existence of a stationary equilibrium profile in these games. Indeed, it is well-known that there is a unique subgame-perfect equilibrium in the alternating-offers bargaining game with a traditional utility specification. In this equilibrium, the agent who makes an offer requests a share of $\gamma = 1/(1 + \delta)$, and the agent who receives an offer accepts all offers and only those that give a share of at least $1 - \gamma = \delta/(1 + \delta)$. Obviously, with these strategies, agents reach agreement $(\gamma, 1 - \gamma)$ without delay in the initial period.

Consider now negotiations about the surplus from a joint venture. In the traditional case with linear utility functions, investments are sunk, and the agents divide the surplus according to the sharing rule suggested by alternating-offers bargaining, i.e., independently of I_1 and I_2. In this case, second-best investment levels I_1^{SB} andI_2^{SB}, provided they are interior, are characterized by the first-order conditions

$$\gamma \frac{\partial Y}{\partial I_1}(I_1, I_2) = \frac{\partial C}{\partial I_1}(I_1) \tag{3}$$

$$(1-\gamma) \frac{\partial Y}{\partial I_2}(I_1, I_2) = \frac{\partial C}{\partial I_2}(I_2). \tag{4}$$

With agent 1's share $\gamma \in (0.5;\ 1)$, it is clear that $I_1 = I_2 = I^{FB} > 0$ satisfies neither (3) nor (4). On the other hand, if the second best is a boundary solution, then $I_1^{SB} = 0$ or $I_2^{SB} = 0$, so that again, investments are inefficient. Thus, the first-best investment levels cannot be implemented when returns from investments must be divided ex-post, not even in the limit for $\delta \to 1$.

3 Inequity aversion

Here, I modify the agents' utility specification to incorporate an aversion against unequal outcomes. Consider the utility function

$$U_i = \Pi_i - \frac{\alpha}{2}(\Pi_j - \Pi_i)^+ - \frac{\beta}{2}(\Pi_i - \Pi_j)^+, \tag{5}$$

where Π_i denotes agent i's monetary profit, as introduced in Section 2, and $(x)^+ = \max\{0; x\}$. The parameter α measures agent i's marginal disutility from inequality when agent i receives less than agent j, and β measures agent i's marginal disutility from inequality when agent i receives more than agent j, where $0 \leq \beta < 1$ and $\beta \leq \alpha$. This form of preferences is essentially the same as the specification in Fehr and Schmidt (1999). In fact, to rule out outright altruistic preferences in the bargaining game, the parameter β has been bounded from above by one. Of course, for $\alpha = \beta = 0$, equation (5) captures the traditional utility specification.

In the context of the investment game, the above utility specification allows a useful reformulation, provided that $Y(I_i, I_j) > 0$. Plugging (1) into (5) and rearranging yields

$$U_i = \{s_i - \alpha(s_i^{\text{ref}} - s_i)^+ - \beta(s_i - s_i^{\text{ref}})^+\}\, Y(I_i, I_j) - C(I_i), \tag{6}$$

where

$$s_i^{\text{ref}} = s_i^{\text{ref}}(I_i, I_j) = \frac{1}{2} + \frac{C(I_i) - C(I_j)}{2Y(I_i, I_j)} \tag{7}$$

is agent i's *reference share*. E.g., when $C(I_1) = 20$ and $C(I_2) = 10$, while $Y(I_1, I_2) = 50$, then agent 1's reference share would be $s_1^{\text{ref}} = 0.6$, and agent 2's reference share would be $s_2^{\text{ref}} = 0.4$, reflecting the relatively larger contribution of agent 1 at the investment stage .

I assume now that the course of the negotiations in itself will have no impact on the reference share: Agent $i = 1{,}2$ evaluates an outcome realized in the initial bargaining period as described by (6) for given parameters α and β. Moreover, the surplus Y, if realized only in some period $t > 0$ of the bargaining game, is discounted by the factor δ^t. If no agreement is reached within finite time, then agent i's utility is assumed to be $-C(I_i)$, thus reflecting the specificity of investments.

The reference share can be interpreted in broad terms as an "expectation" that the agent brings to the negotiation table. In this interpretation, an agent with a larger (lower) relative contribution would expect to receive more (less) than half of the surplus, because this division would be "only fair". The following proposition collects intuitive properties of the reference share.

Proposition 1. *Assume $(I_1, I_2) \neq (0{,}0)$. If both agents 1 and 2 were to receive their reference share, both would end up with the same monetary profit, i.e.*,

$$s_1^{\mathrm{ref}}\ Y - C_1 = s_2^{\mathrm{ref}}\ Y - C_2.$$

Moreover, the function $s_i^{\mathrm{ref}}(I_i, I_j)$ is strictly increasing in I_i, and strictly decreasing in I_j.

Proposition 1 says that implementing the reference share implies an identical payoff for both agents. Perhaps more interesting is that Proposition 1 also states that a higher unilateral contribution by some agent i leads ceteris paribus to a higher reference share for agent i, and to a lower reference share for agent j. This fact is a consequence of strictly decreasing marginal productivities and of nondecreasing marginal costs: While an increase in I_i causes a larger output, it also implies a more than proportional increase in the costs for agent i.

It could be argued that the reference share described by (7) may sometimes appear intuitively too high or too low. E.g., when one agent chooses an inefficiently high investment level, then the other agent could argue plausibly that this overinvestment has not been a useful contribution to the joint venture. However, lowering the reference share to reflect an inefficient overinvestment would tend to lead to lower profits for the deviating agent, and thus strengthen the conclusions of this paper. Moreover, although reference shares below zero or above one allow a natural interpretation in terms of compensating transfers between the agents, such reference shares are feasible under the assumptions of Theorem 2 below only when both agents deviate simultaneously at the investment stage[4].

4 The profit-sharing rule implied by (7) has been considered before in the literature. In a model of bilateral investments followed by a Nash demand game, Carmichael and MacLeod (2003) show that their "fair share" rule is generically part of the unique efficient equilibrium. In contrast to Carmichael and MacLeod, my analysis does not assume either efficiency or the behavioral rule. Rather, I show that the behavior in the unique subgame-perfect equilibrium between inequity-averse agents is approximately in line with the fair share rule. This feature of the equilibrium in turn implies close-to-efficient investments.

4 Negotiations

In the bargaining stage, agents make alternative proposals concerning the division of the surplus Y, which is realized only when the agents reach an agreement. However, after investments have been made, Y and C_i are constants. Therefore, I can measure each agent's utility $u_i = \sigma_i Y - C_i$ without loss of generality in terms of σ_i. With view on this fact, call

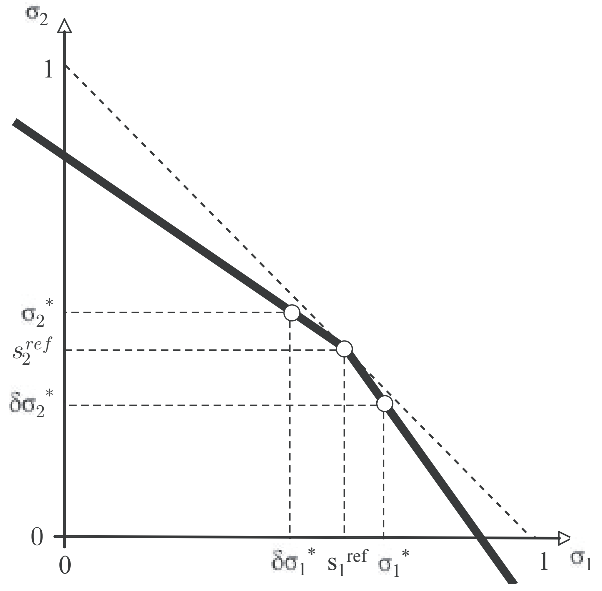
$$\sigma_i = \sigma_i(s_i,\ s_i^{ref}) = s_i - \alpha(\ s_i^{ref} - s_i)^+ - \beta(s_i - \ s_i^{ref})^+$$

agent *i's subjective share* of the total surplus. Then the set of feasible pairs (σ_1, σ_2) of subjective shares resulting from agreements (s_1, s_2) forms a bargaining set (the "utility possibility frontier"), as depicted in *Figure 1*. In the Appendix, I show that $\sigma_j = g_j\,(\sigma_i,\ s_i^{ref})$, where

$$g_j(\sigma_i,\ s_i^{ref}) = 1 - \sigma_i - \frac{\hat{\alpha}}{1+\hat{\alpha}}(\ s_i^{ref} - \sigma_i)^+ - \ \hat{\alpha}\ (\sigma_i - \ s_i^{ref})^+ \qquad (8)$$

with $\hat{\alpha} = (\alpha+\beta)/(1-\beta)$. Outcomes that differ from the reference agreement $(\ s_1^{ref}, s_2^{ref})$ lead to utility losses for at least one agent. Because these utility losses increase with the extent of inequality, the possibility frontier has a characteristic kink at $(\ s_1^{ref}, s_2^{ref})$. As an immediate consequence of Proposition 1, this kink will move along the dotted cross-diagonal from the top left down to the bottom right for increasing values of I_1, and similarly for decreasing values of I_2.

Figure 1: The bargaining equilibrium for inequity-averse agents



Also with inequity aversion, I can describe the subgame-perfect equilibrium in the Rubinstein (1982) bargaining game by equilibrium offers and acceptance levels for the two agents. Following the intuitive exposition of Muthoo (1999) let $(s_1^*, 1 - s_1^*)$ denote the agreement offered by agent 1 in equilibrium. When this offer is accepted by agent 2, then the subjective share for agent 1 will be $\sigma_1^* = \sigma_1(s_1^*, s_1^{\text{ref}})$, and the subjective share for agent 2 will be[5]

$$g_2(\sigma_1^*, s_1^{\text{ref}}) = \sigma_2(1 - s_1^*, s_2^{\text{ref}}) = \delta\sigma_2^*,$$

where $\sigma_2^* = \sigma_2(s_2^*, s_2^{\text{ref}})$, and s_2^* is agent 2's offer. That is, agent 2's subjective share from accepting agent 1's offer in period t must be equal to agent 2's subjective share from making an offer $(1 - s_2^*, s_2^*)$ herself in period $t + 1$, that would be accepted by agent 1. Similarly, when it is agent 2's turn, the subjective share for agent 1 would be

$$g_1(\sigma_2^*, s_2^{\text{ref}}) = \sigma_1(1 - s_2^*, s_1^{\text{ref}}) = \delta\sigma_1^*.$$

Therefore, the pair (σ_1^*, σ_2^*) of subjective shares retained by the respective proposers of an offer is characterized by the two equations

$$g_1(\sigma_2^*) = \delta\sigma_1^* \text{ and } g_2(\sigma_1^*) = \delta\sigma_2^*, \tag{9}$$

where I have suppressed the second argument of the functions $g_i(.,.)$ to simplify notation. These equations allow a useful graphical interpretation, as shown in *Figure 1*: Starting from σ_1^*, and heading along the dotted line first north, then west, I find $g_2(\sigma_1^*) = \delta\sigma_2^*$. Similarly, starting from σ_2^*, and heading first east, then south, the dotted line leads to $g_1(\sigma_2^*) = \delta\sigma_1^*$.

Solving the equations in (9) yields the subjective shares (σ_1^*, σ_2^*) that arise in the subgame-perfect equilibrium of the bargaining game. However, unlike the traditional utility specification, inequity aversion leads the agents into one of three qualitatively different scenarios, depending on their initial investments (cf. *Figure 2*). E.g., when agent 2's investment is much larger than agent 1's investment, then the reference share s_1^{ref} will be significantly smaller than one half, as illustrated in the top left diagram of *Figure 2*. In this situation, the bargaining outcome yields more than the reference share for agent 1 and less than the reference share for agent 2. I will describe the other two scenarios following the statement of Theorem 1.

There is second complication caused by the introduction of inequity aversion into the model. Although the agents can typically conclude a mutually beneficial agreement at the ex-post stage, a Pareto improvement may not always be feasible when investments differ too much. Indeed, entering into negotiations is individually rational for both agents when

5 For full details, in particular concerning uniqueness, see the proof of Theorem 1 in the Appendix.

$$g_i(0) \geq 0 \text{ for } i = 1,2,$$

or equivalently, when the reference share for agent 1 satisfies

$$s_1^{\text{ref}}(I_1, I_2) \in [-\tfrac{1}{\hat{\alpha}};\ 1 + \tfrac{1}{\hat{\alpha}}\ ]. \tag{10}$$

For the sake of expositional simplicity, I state the solution of the bargaining game only under condition (10). When this condition fails to hold, then the disutility from inequity is so large for at least one of the agents that there is no agreement that can be acceptable to both agents. Thus, agents individually prefer to choose their outside options in the bargaining stage. However, without return, there cannot be an equilibrium at the investment stage unless $I_1 = I_2 = 0$. The following result describes subjective shares in the bargaining equilibrium as a function of the reference share, provided that bargaining is individually rational.

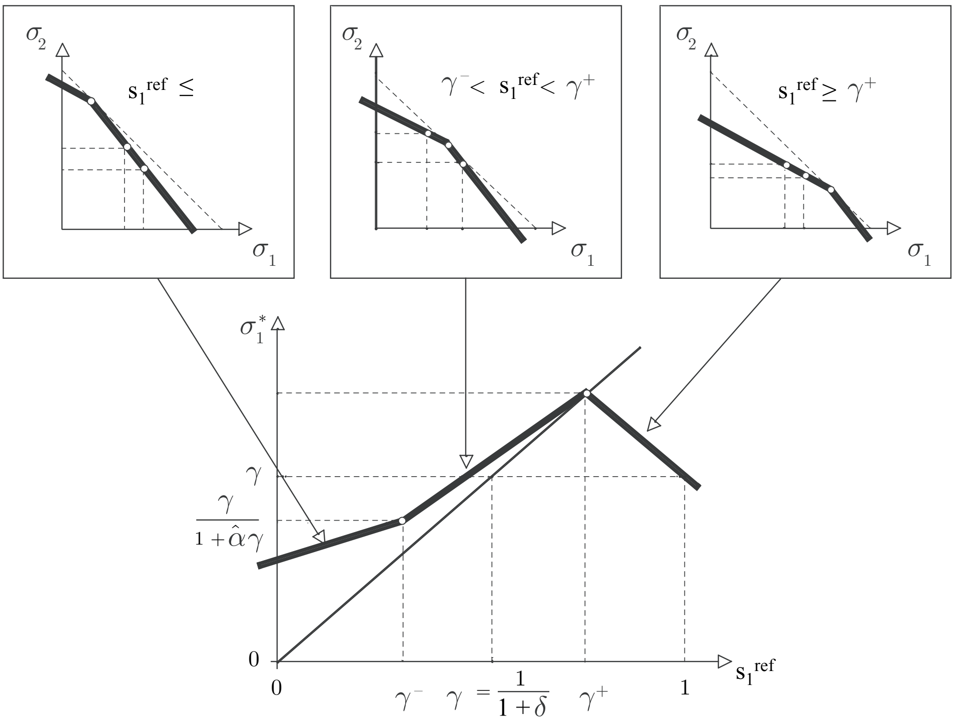
Theorem 1. *Let* $(I_1, I_2) \neq (0,0)$ *be the pair of investments made by the agents in the first stage, and let* $(s_1^{\text{ref}}, s_2^{\text{ref}})$ *denote the resulting reference agreement. Assume that entering into negotiations is individually rational for both agents. Then the unique subgame-perfect equilibrium of the bargaining game gives agent 1 a subjective share of* $\sigma_1(s_1^*, s_1^{\text{ref}})$ *and agent 2 a subjective share of* $\delta\sigma_2(s_2^*, s_2^{\text{ref}})$, *where*

$$\sigma_i(s_i^*, s_i^{\text{ref}}) = \begin{cases} \dfrac{1 + \hat{\alpha} s_i^{\text{ref}}}{(1 + \hat{\alpha})(1 + \delta)} & \text{if } s_i^{\text{ref}} \leq \gamma^-, \\[2ex] \dfrac{(1 - \delta)(1 + \hat{\alpha}) + \hat{\alpha}(1 + \hat{\alpha} + \delta) s_i^{\text{ref}}}{(1 + \hat{\alpha})^2 - \delta^2} & \text{if } \gamma^- < s_i^{\text{ref}} < \gamma^+, \\[2ex] \dfrac{1 + \hat{\alpha} - \hat{\alpha} s_i^{\text{ref}}}{1 + \delta} & \text{if } s_i^{\text{ref}} \geq \gamma^+, \end{cases}$$

with $\gamma^- = \delta/(1 + \hat{\alpha} + \delta)$ *and* $\gamma^+ = (1 + \hat{\alpha})/(1 + \hat{\alpha} + \delta)$. *The agreement is reached in the first bargaining period. Moreover, the function* $\sigma_i(s_i^*, s_i^{\text{ref}})$ *is continuous in* s_i^{ref} *for* $i = 1,2$.

The first case in the statement of Theorem 1 corresponds to a situation in which agent 1's contribution is significantly lower than that of agent 2 (but still large enough to make bargaining individually rational). In this situation, the bargaining set considered in *Figure 1* shifts up and to the left, as depicted in the top left diagram in *Figure 2*. The subgame-perfect equilibrium would then predict a profile of subjective shares that is located in the lower right segment of the frontier. Therefore, agent 1 receives more than her reference share, while agent 2 receives less. The second case in the statement of the Theorem is depicted in the diagram in the top center of *Figure 2*. Here, initial contributions I_1 and I_2 are approximately the same in size. Agent 1, due to her first-mover advantage, receives more than her reference share and agent 2 less. In the final case (top right diagram in *Figure 2* the bargaining set shifts down and to the right, when compared to the intermediate case. Now the outcome lies on the upper left segment of the bargaining set, giving agent 1 a share smaller than her reference share and agent 2 a share larger than her reference share.

Figure 2: Outcome of the bargaining game as a function of the reference share



The diagram at the bottom of *Figure 2* shows the subjective share of agent 1 as a function of the reference level (the diagram for agent 2 looks similar). In contrast to the model with linear utility specification, the subjective share σ_1^* obtained in equilibrium is strictly increasing in the reference share s_1^{ref} for values between zero and γ^+, which causes the stronger incentives to invest in the relevant range. Although it may be surprisingly at first sight, the subjective share is strictly decreasing in the reference share s_1^{ref} for values $s_1^{ref} > \gamma^+$. 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. A marginal increase in s_1^{ref} shifts the bargaining set downwards and to the right, adds to the inequity of the outcome, and decreases thereby the subjective share for agent 1.

A monotonicity property also holds for the equilibrium strategies, i.e., for the objective shares offered by the agents in the bargaining game.

Proposition 2. *Impose the same assumptions as in Theorem 1. Then the objective share s_i^* offered by agent i in equilibrium is strictly increasing in s_i^{ref} for $s_i^{\text{ref}} \leq \gamma^+$.*

Since minimum acceptance levels are identical to equilibrium offers, this finding provides a comparative statics of equilibrium agreements. The proof given in the Appendix is essentially a calculation that exploits the explicit expressions provided by Theorem 1. Proposition 2, in conjunction with Proposition 1, verifies an intuitive claim on equilibrium agreements in the bargaining stage. In the relevant range, a higher investment by an individual agent will not only increase the gross surplus created in the joint venture, but also improve the agent's position in the negotiations, so that the agent can obtain a higher percentage share of the surplus. As I discuss in the next section, this effect improves incentives in the investment stage in a significant way.

5 Investments

The crucial feature of the subgame-perfect equilibrium identified in Theorem 1 is that for reference levels in the interval $(\gamma^-; \gamma^+)$, negotiations will result in a subjective share for agent 1 that is close to s_1^{ref}, and to a subjective share for agent 2 that is close to δs_2^{ref}. Indeed, for $\delta \to 1$,

$$\frac{(1-\delta)(1+\hat{\alpha}) + \hat{\alpha}\{1+\hat{\alpha}+\delta\}\, s_i^{\text{ref}}}{(1+\hat{\alpha})^2 - \delta^2} \to s_i^{\text{ref}}. \tag{11}$$

The consequence of this effect is that sufficiently patient agents receive almost precisely the reference share in the relevant domain.

I now present conditions under which the efficient pair of investments is approximated by investment levels that result in a subgame-perfect equilibrium between inequity-averse agents. The central condition concerns the profitability of production. Indeed, it is intuitive that with a costly and unproductive technology, it may become an attractive option for an individual agent to reduce investments, even if marginal incentives are approximately in line with the efficiency criterion. As *Figure 2* indicates for the case $i = 1$, agent i's subjective share moves closely with the reference share only when investments are not too different. However, when the investment level I_1 is so much lower than I_2 that s_1^{ref} falls below γ^-, then the subjective share does not shrink as quickly as before, so that a reduction could indeed become attractive for agent 1. It turns out that the necessary level of profitability depends on the degree of inequity aversion[6]. This level is given by

$$\frac{C(I^{FB})}{Y(0, I^{FB})} < \frac{\beta+\alpha}{2-\beta+\alpha}. \tag{12}$$

6 The equilibrium share function for agent i has always an upwards kink at γ^-, as suggested by *Figure 2*. This feature of the equilibrium implies that agent i's indirect utility function $U_i(I_i, I_j)$ is non-concave as a function of I_i when $s_i^{\text{ref}}(I_i, I_j) = \gamma^-$. Condition (12) ensures that this non-concavity is not reached provided that I_j is close to I^{FB}.

In the traditional utility specification, the right-hand side of (12) disappears, so that in this case the condition cannot be satisfied.

Theorem 2. *Consider the symmetric two-stage model of joint production introduced in Section 2. Assume that the agents' utility functions exhibit inequity aversion with parameters* α *and* β *as specified in Section 3. Assume also that production is sufficiently profitable in view of* α *and* β *as captured by condition (12). Then for any* $\varepsilon > 0$ *there is a discount factor* $\widetilde{\delta}$ *such that for any* $\delta \in (\widetilde{\delta};1)$, *the equilibrium investment level of each agent lies in the interval* $(I^{FB} - \varepsilon; I^{FB} + \varepsilon)$.

The intuition for improved incentives is straightforward. Since the equilibrium shares of the two agents depend on their reference agreement and the reference agreement depends on their relative contributions, the bargaining solution gets closer to a proportional rule. Therefore, it gives agents a higher incentive to invest.

To see how approximate efficiency comes about in the formal framework, I again consider the case $\gamma^- < s_i^{\text{ref}} < \gamma^+$ in Theorem 1 and let agents become very patient ($\delta \to 1$). In this case $\sigma_i^* \to s_i^{\text{ref}}$, so that very patient agents will expect to receive approximately their reference share of the total surplus. Hence, for δ close to one, agent i chooses I_i ex ante to maximize a function that is "close" to

$$s_i^{\text{ref}}(I_i, I_j)\,Y(I_i, I_j) - C(I_i) = \frac{Y(I_i, I_j) - C(I_i) - C(I_j)}{2}. \tag{13}$$

Moreover, under the assumptions made on the production technology and cost functions, the investment I_i, which maximizes (13), will be close to the investment that maximizes the precise objective function. This consideration suggests that incentives are close to efficient here. Although the marginal returns are equally divided between the agents, the marginal costs will also be shared equally, which leads to socially desirable incentives at the investment stage.

6 Conclusion

Relationship-specific investments increase the potential gains or quasi rents from a cooperative joint venture with one specific business partner, but leave unaffected the gains from a joint venture with any other potential partner. When comprehensive contracting is difficult and the surplus from the joint venture must be divided at an ex-post stage by bilateral negotiations, then the returns on these investments may be subject to severe ex-post opportunism by the business partner. The anticipation of this hold-up problem typically leads to inefficient incentives to invest.

In this paper I argue that the hold-up inefficiency described above can be significantly mitigated for joint ventures between inequity-averse decision makers. In contrast to the case of a traditional utility specification, the firms will compare their net benefits from the joint venture, which entail in particular the respective contributions in the investment stage. Incentives to invest are stronger for inequity-averse agents than for traditional agents, because a higher contribution increases not only the total surplus to be divided, but also the expected share of the surplus. In fact, when the technology is sufficiently productive and the costs of a delay of negotiations go to zero, then the noncooperatively chosen investments converge to the efficient level.

In January 2005, Dieter Pfaff (personal communication) reported on an experiment that appears suitable for testing the predictions of my model. The experiment compares two set-ups. In the first set-up, there is a bilateral investment game, followed by negotiations about the division of the jointly produced surplus. In the second set-up, the surplus is shared equally without negotiations. Although the results are still preliminary, the data unambiguously suggests significantly stronger incentives, and even efficient investments, in the first set-up, which accords with my theoretical prediction.

While the discussion in this paper may provide some additional light on the potential role of psychological elements in the hold-up problem, the formal analysis relies on a number of simplifying assumptions, which further theoretical research may be able to relax. These assumptions primarily concern the symmetry of the set-up, the stationary specification of the reference share, the irrelevance of the course of the bargaining for agents' utilities, the perfect observability of individual investment levels, and the common knowledge of utility types. Recently, several researchers have obtained interesting results in two of these directions for the ultimatum bargaining procedure. In this case, asymmetric information on investment costs may reduce the individual contribution (Ellingsen and Johannesson (2005)), while asymmetric information about utility types may even improve efficiency (von Siemens (2005)).

Appendix

This Appendix contains the proofs of Propositions 1 and 2, Theorems 1 and 2, and equation (8).

Proof of Proposition 1. The first assertion is obvious. To prove the second assertion, 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 $C_i = C(I_i)$. It is sufficient to show now that the function $\varphi(C_i) = (C_i - C_j)/Y(\phi(C_i), I_j)$ is strictly increasing in C_i. Indeed,

$$\varphi'(C_i) = \frac{Y(\phi(C_i), I_j) - (C_i - C_j)\frac{\partial Y(\phi(C_i), I_j)}{\partial C_i}}{Y(\phi(C_i), I_j)^2} \tag{14}$$

is strictly positive for all $C_i > 0$. To see why, I note first that, as verified by explicit differentiation, the function that maps C_i into $Y(\phi(C_i), I_j)$ is strictly concave. But then, for $C_i > C_j$,

$$\frac{\partial Y(\phi(C_i), I_j)}{\partial C_i} < \frac{Y(\phi(C_i), I_j) - Y(\phi(C_j), I_j)}{C_i - C_j} \leq \frac{Y(\phi(C_i), I_j)}{C_i - C_j}.$$

It is also clear that $\varphi'(C_i)$ must be strictly positive if $C_i \leq C_j$. This argument proves that $s_i^{\mathrm{ref}}(I_j, I_i)$ is strictly increasing in I_i. But from $s_i^{\mathrm{ref}}(I_i, I_j) = 1 - s_j^{\mathrm{ref}}(I_j, I_i)$ it is obvious that $s_i^{\mathrm{ref}}(I_i, I_j)$ is also strictly decreasing in I_j. □

Proof of Equation (8). Clearly, because $\beta < 1$, the inequality $\sigma_i \leq s_i^{\mathrm{ref}}$ holds if and only if $s_i \leq s_i^{\mathrm{ref}}$. Consider the case $\sigma_i \leq s_i^{\mathrm{ref}}$. Then,

$$\sigma_i = (1 + \alpha)\, s_i - \alpha\, s_i^{\mathrm{ref}}\,. \tag{15}$$

On the other hand, $\sigma_j \geq s_j^{\mathrm{ref}}$. Therefore,

$$\sigma_j = (1 - \beta)\, s_j + \beta s_j^{\mathrm{ref}}\,. \tag{16}$$

Replacing s_j by $1-s_i$ and s_j^{ref} by $1- s_i^{\mathrm{ref}}$ in (16), and subsequently eliminating s_i using (15) yields the assertion for $\sigma_i \leq s_i^{\mathrm{ref}}$. The proof for the case $\sigma_i > s_i^{\mathrm{ref}}$ is analogous and therefore omitted. □

Proof of Theorem 1. The proof proceeds in three steps. The methods used in the second step are based on Shaked and Sutton (1984).

Step 1. In any equilibrium of the bargaining game, both agents obtain a nonnegative subjective share. To see why this is true, assume to the contrary that $\sigma_i^* < 0$ for some agent i. Then agent i could profitably deviate as follows. If $s_i^{\mathrm{ref}} \geq 0$, then agent i could always propose a share of $s_i = s_i^{\mathrm{ref}}$, and reject all offers made by agent j. This strategy clearly guarantees a subjective share of $\sigma_i \geq 0$ for agent i. On the other hand, if $s_i^{\mathrm{ref}} < 0$, then agent i could always propose a share of $s_i = |\, s_i^{\mathrm{ref}}\, |\beta/(1- \beta)$, and reject all offers made by agent j. This strategy would guarantee a subjective share of $\sigma_i = 0$ for agent i. I have therefore shown that $\sigma_i^* \geq 0$ in any equilibrium. Clearly, this intermediate result also implies that $\sigma_j^* \leq g_j\,(0,\, s_i^{\mathrm{ref}})$.

Step 2. In his Assumption A–2, Rubinstein (1982) excludes the possibility of negative subjective shares from unfair outcomes. To derive equations (9) under my somewhat more general conditions, I follow the exposition in Fudenberg and Tirole (1991, Subsection 4.4.2). Let $\underline{v}_i$ and $\overline{v}_i$, respectively, denote the infimum and the supremum of the set of subjective shares that agent i may obtain in some subgame-perfect equilibrium of the bargaining game, when agent i makes her initial offer. Similarly, let $\underline{w}_i$ and $\overline{w}_i$, respectively, denote the infimum and the supremum of the set of subjective shares that agent i can obtain in some subgame-perfect equilibrium of the bargaining game, when agent j makes the initial offer. Consider now an offer by agent i. It is clear that agent j accepts any offer above $\delta\overline{v}_j$. Hence

$$\underline{v}_i \geq g_i(\delta\overline{v}_j). \tag{17}$$

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

$$\overline{w}_j \leq \delta \overline{v}_j. \tag{18}$$

Agent j will either accept or reject agent i's offer. If accepted, the offer can give agent i at most $g_i(\delta \underline{v}_j)$. If rejected, agent i obtains at most $\delta \overline{w}_i$. Therefore

$$\overline{v}_i \leq \max\{g_i(\delta \underline{v}_j), \delta \overline{w}_i\} \leq \max\{g_i(\delta \underline{v}_j), \delta \overline{v}_i\}, \tag{19}$$

where (18) is used to derive the second inequality. It is claimed now that

$$\max\{g_i(\delta \underline{v}_j), \delta \overline{v}_i\} = g_i(\delta \underline{v}_j). \tag{20}$$

Indeed, if (20) is violated, then $\overline{v}_i \leq \delta \overline{v}_i$ from (19). But then $\underline{v}_i = \overline{v}_i = 0$ by Step 1 and $\delta < 1$. It is not difficult to derive $\underline{v}_j \geq g_j(0)$ from (17), and $\overline{v}_j \leq g_j(0)$ from Step 1, which together imply that $\overline{v}_j = \underline{v}_j = g_j(0)$. But then, because bargaining is individually rational, it must be the case that $0 < \delta \overline{v}_j < g_j(0)$. Using (17) again, this inequality implies $\underline{v}_i > g_i(\delta \overline{v}_j) > 0$, and contradicts $\underline{v}_i = 0$. This argument verifies claim (20). From (20) and (19), it follows that

$$\overline{v}_i \leq g_i(\delta \underline{v}_j). \tag{21}$$

The functions $g_i(.)$and $g_j(.)$ are strictly decreasing. Moreover, $g_i(.)$and $g_j(.)$ are mutually inverse functions, i.e., they satisfy $g_i(g_j(s_i, s_i^{ref}), s_j^{ref}) = s_i$ and a similar equality with exchanged roles for i and j. Therefore, using inequalities (17), (21), and their symmetric counterparts,

$$g_j(\underline{v}_i) \leq \delta g_j(\delta \underline{v}_i) \text{ and } g_j(\overline{v}_i) \geq \delta g_j(\delta \overline{v}_i). \tag{22}$$

From Step 1 and from the fact that the function $\sigma_i \mapsto g_j(\sigma_i) - \delta g_j(\delta \sigma_i)$is strictly decreasing on the interval $[0;\ g_i(0)]$, it follows that the inequalities (22) must hold with equality. Thus, in any subgame-perfect equilibrium, the agent i who makes the initial offer receives a subjective share $\sigma_i^* = \underline{v}_i = \overline{v}_i$, characterized by $g_j(\sigma_i^*) = \delta g_j(\delta \sigma_i^*)$. Consider now agent j, who receives the initial offer. A rejection of agent i's offer secures agent j a subjective share of $\delta \sigma_j^*$. Thus, $\underline{w}_j \geq \delta \sigma_j^*$. Using (18), it follows that the subjective share for agent j in equilibrium amounts to $\delta \sigma_j^* = \underline{w}_j = \overline{w}_j$. I have therefore proved uniqueness of subjective shares that result from any subgame-perfect equilibrium. The argument that even the subgame-perfect equilibrium profile is unique given that subjective shares ("utilities") are unique can be found in Fudenberg and Tirole (1991) and is therefore omitted.

Step 3. This part of the proof contains the explicit derivation of the agents' subjective shares. Consider a subgame-perfect equilibrium outcome (σ_1^*, σ_2^*) in the bargaining game. By Step 2,

$$g_j(\sigma_i^*) = \delta \sigma_j^* \text{ and } g_j(\delta \sigma_i^*) = \sigma_j^*. \tag{23}$$

Precisely one of the following three cases must obtain:

Case A. $\sigma_i^* \leq s_i^{\text{ref}}$. By equation (8), in this case,

$$g_j(\sigma_i) = 1 - \frac{1}{1+\hat{\alpha}}\sigma_i - \frac{\hat{\alpha}}{1+\hat{\alpha}} s_i^{\text{ref}} \tag{24}$$

for $\sigma_i = \sigma_i^*$ and for $\sigma_i = \delta\sigma_i^*$. Combining these two equations with (23) and rearranging yields

$$\sigma_i^* = \frac{1 + \hat{\alpha}\, s_j^{\text{ref}}}{1+\delta}. \tag{25}$$

Thus, case A occurs only if the right-hand side of (25) is $\leq s_i^{\text{ref}}$ or, equivalently, when $s_i^{\text{ref}} \geq \gamma^+$. Conversely, it is clear that if $s_i^{\text{ref}} \geq \gamma^+$, then (25) constitutes an equilibrium outcome.

Case B. $\delta\sigma_i^* \geq s_i^{\text{ref}}$. From equation (8),

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

follows for $\sigma_i = \sigma_i^*$ and for $\sigma_i = \delta\sigma_i^*$. Thus, in equilibrium,

$$\sigma_i^* = \frac{1 + \hat{\alpha} s_i^{\text{ref}}}{(1+\hat{\alpha})(1+\delta)}. \tag{26}$$

Hence, in case B,

$$\delta \frac{1 + \hat{\alpha} s_i^{\text{ref}}}{(1+\hat{\alpha})(1+\delta)} \geq s_i^{\text{ref}}.$$

Rearranging yields $s_i^{\text{ref}} \leq \gamma^-$. Conversely, it is straightforward to check that if $s_i^{\text{ref}} \leq \gamma^-$ then there is an equilibrium given by (26).

Case C. $\sigma_i^* > s_i^{\text{ref}}$ and $\delta\sigma_i^* > s_i^{\text{ref}}$. Equation (8) yields

$$g_j(\sigma_i^*) = 1 - (1+\hat{\alpha})\sigma_i^* + \hat{\alpha}\, s_i^{\text{ref}} \tag{27}$$

$$g_j(\delta\sigma_i^*) = 1 - \frac{1}{1+\hat{\alpha}}\delta\sigma_i^* - \frac{\hat{\alpha}}{1+\hat{\alpha}} s_i^{\text{ref}}. \tag{28}$$

Plugging equations (27) and (28) into (23) and rearranging yields the formula for σ_i^* given in the statement of Theorem 1 for values $\gamma^- < s_i^{\text{ref}} < \gamma^+$. Because the existence and uniqueness of the subgame-perfect equilibrium is guaranteed for all $s_i^{\text{ref}} \in [-1/\hat{\alpha};\ 1 + 1/\hat{\alpha}]$, case C will occur if and only if neither case A nor case B occurs.

The assertion concerning the subjective share of the second-moving agent j follows from (23). A straightforward calculation shows that $\sigma_i^*(s_i^*, s_i^{ref})$ is continuous at $s_i^{ref} = \gamma^-$ and at $s_i^{ref} = \gamma^+$. This proves Theorem 1. □

Proof of Proposition 2. Assume $s_i^{ref} \in [-1/\hat{\alpha}; \gamma^+]$. Then, by Theorem 1, $\sigma_i(s_i^*, s_i^{ref}) \geq s_i^{ref}$. Hence,

$$\sigma_i(s_i^*, s_i^{ref}) = s_i^*(1-\beta) + \beta s_i^{ref}.$$

Using the explicit expressions in Theorem 1 yields for $s_i^{ref} \leq \gamma^-$ that

$$s_i^* = \frac{1}{(1+\alpha)(1+\delta)} + \frac{\alpha(1+\alpha)(1+\delta) + \beta(\alpha + \delta(1+\alpha))}{(1+\alpha)(1-\beta)(1+\delta)} s_i^{ref},$$

and for $s_i^{ref} \in (\gamma^-, \gamma^+]$ that

$$s_i^* = \frac{(1+\alpha)(1-\delta)}{(1+\alpha)^2 - \delta^2(1-\beta)^2} + \frac{(1-\beta)(\alpha + \delta\beta)}{1 - \delta(1-\beta) + \alpha} s_i^{ref}.$$

The assertion follows. □

Proof of Theorem 2. Fix $\hat{\alpha}$ throughout. To generate an equilibrium candidate, consider a pair (I_1, I_2) with $I_1 \approx I_2$, so that $s_1^{ref}(I_1, I_2) \approx 1/2$. For $\delta \to 1$, limits are given by $\gamma^-(\delta) \to 1/(2+\hat{\alpha})$ and $\gamma^+(\delta) \to 1 - 1/(2+\hat{\alpha})$, so that $s_i^{ref}(I_i, I_j) \in (\gamma^-(\delta); \gamma^+(\delta))$ for δ sufficiently close to one. By Theorem 1, there are functions $\mu_i(\delta)$ and $\nu_i(\delta)$ such that $\sigma_i^*(s_i^{ref}) = \mu_i(\delta) + \nu_i(\delta) s_i^{ref}$. The functions $\mu_i(\delta)$ and $\nu_i(\delta)$ are well defined and continuously differentiable in a neighborhood of $\delta = 1$ by condition (12). With this notation, agent i's utility as a function of investments reads

$$U_i(I_i, I_j) = \{\mu_i + \nu_i s_i^{ref}(I_i, I_j)\} Y(I_i, I_j) - C(I_i)$$

$$= (\mu_i + \tfrac{\nu_i}{2}) Y(I_i, I_j) - (1 - \tfrac{\nu_i}{2}) C(I_i) - \tfrac{\nu_i}{2} C(I_j), \qquad (29)$$

where the arguments of μ_i and ν_i have been suppressed to ease notation. The corresponding first-order conditions are given by

$$(\mu_i + \tfrac{\nu_i}{2}) \frac{\partial Y}{\partial I_i} - (1 - \tfrac{\nu_i}{2}) \frac{\partial C}{\partial I_i} = 0, \qquad (30)$$

for agents $i = 1, 2$. The implicit function theorem ensures now that (30) admits a solution $(I_1^{\#}(\delta), I_2^{\#}(\delta))$ that depends continuously on δ in a neighborhood of $\delta = 1$. Indeed, given that the determinant of the Hessian of $Y(.,.)$ is strictly positive, it is straightforward to check that

$$\begin{pmatrix} \{\mu_1 + \frac{\nu_1}{2}\}\frac{\partial^2 Y}{\partial I_1^2} - (1 - \frac{\nu_1}{2})\frac{\partial^2 C}{\partial I_1^2} & \{\mu_1 + \frac{\nu_1}{2}\}\frac{\partial^2 Y}{\partial I_1 \partial I_2} \\ \{\mu_2 + \frac{\nu_2}{2}\}\frac{\partial^2 Y}{\partial I_1 \partial I_2} & \{\mu_2 + \frac{\nu_2}{2}\}\frac{\partial^2 Y}{\partial I_2^2} - (1 - \frac{\nu_2}{2})\frac{\partial^2 C}{\partial I_2^2} \end{pmatrix}$$

is non-singular for δ close to one. Hence, in the limit $I_i^{\#} \to I^{FB}$ for $\delta \to 1$. It remains to be shown that for δ close enough to one, the pair $(I_1^{\#}(\delta), I_2^{\#}(\delta))$ is indeed an equilibrium in the investment stage. Clearly, the approximate identity $I_i^{\#}(\delta) \approx I_j^{\#}(\delta)$ holds for δ close to one. Thus, the first-order condition combined with the concavity of (29) yields that $I_i = I_i^{\#}(\delta)$ is an optimal response to $I_j^{\#}(\delta)$ under the restriction $s_i^{\mathrm{ref}}(I_i, I_j^{\#}(\delta)) \in (\gamma^-(\delta); \gamma^+(\delta))$. To check that $I_i^{\#}(\delta)$ is optimal even globally, consider first the deviation to some I_i such that $s_i^{\mathrm{ref}}(I_i, I_j^{\#}(\delta)) \in [\gamma^+(\delta); 1 + 1/\hat{\alpha}]$. By the continuity of $U_i(I_i, I_j^{\#}(\delta))$, it suffices to show that $\partial U_i(I_i, I_j^{\#}(\delta))/\partial I_i$ jumps downwards at $I_i = I_i^+$, where $s_i^{\mathrm{ref}}(I_i^+, I_j^{\#}(\delta)) = \gamma^+(\delta)$. To verify this claim, consider the derivative

$$\frac{\partial U_i}{\partial I_i} = \frac{\partial \sigma_i^*}{\partial s_i^{\mathrm{ref}}} \frac{\partial s_i^{\mathrm{ref}}}{\partial I_i} Y + \sigma_i^* \frac{\partial Y}{\partial I_i} - \frac{\partial C}{\partial I_i},$$

and recall from Theorem 1 that $\partial \sigma_i^* / \partial s_i^{\mathrm{ref}}$ is jumping downwards at $I_i = I_i^+$, while $\partial s_i^{\mathrm{ref}} / \partial I_i > 0$ by Proposition 1. This argument proves that a deviation such that $s_i^{\mathrm{ref}}(I_i, I_j^{\#}(\delta)) \in [\gamma^+(\delta); 1 + 1/\hat{\alpha}]$ decreases agent i's utility. A deviation I_i such that $s_i^{\mathrm{ref}}(I_i, I_j^{\#}(\delta)) > 1 + 1/\hat{\alpha}$ yields no share in the surplus, and is therefore clearly suboptimal. To ensure that also a downward deviation is not profitable, it suffices to guarantee that $s_i^{\mathrm{ref}}(0, I_i^{\#}(\delta)) > \gamma^-(\delta)$. However, as a short calculation shows, for δ sufficiently close to one, this inequality follows from (12). Thus, the investment $I_i^{\#}(\delta)$ is an optimal response to $I_j^{\#}(\delta)$. That $I_i^{\#}(\delta) \to I^{FB}$ for $\delta \to 1$ has been shown above. □

References

Binmore, Kenneth, Avner Shaked, and John Sutton (1985), Testing Noncooperative Bargaining Theory: A Preliminary Study, *American Economic Review* 75, 1178-180.

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

Bolton, Gary E., and Axel Ockenfels (2000), ERC: A Theory of Equity, Reciprocity, and Competition, *American Economic Review* 90, 166-193.

Carmichael, Lorne, and W. Bentley MacLeod (2003), Caring about Sunk Costs: A Behavioral Solution to Holdup Problems with Small Stakes, *Journal of Law, Economics, and Organization* 19, 106-118.

Che, Yeon-Koo, and Donald B. Hausch (1999), Cooperative Investments and the Value of Contracting, *American Economic Review* 89, 125-147.

Dufwenberg, Martin, and Georg Kirchsteiger (2004), A Theory of Sequential Reciprocity, *Games and Economic Behavior* 47, 268-298.

Ellingsen, Tore, and Magnus Johannesson (2004a), Is there a Hold-up Problem?, *Scandinavian Journal of Economics* 106, 475-194.

Ellingsen, Tore, and Magnus Johannesson (2004b), Promises, Threats, and Fairness, *Economic Journal* 114, 397-420.

Ellingsen, Tore, and Magnus Johannesson (2005), Sunk Costs and Fairness in Incomplete Information Bargaining, *Games and Economic Behavior* 50, 155-177.

Ellingsen, Tore, and Jack Robles (2002), Does Evolution Solve the Hold-up Problem?, *Games and Economic Behavior* 39, 28-53.

Fehr, Ernst, and Klaus M. Schmidt (1999), A Theory of Fairness, Competition and Cooperation, *Quarterly Journal of Economics* 114, 817-868.

Fudenberg, Drew, and Jean Tirole (1991), *Game Theory*, MIT Press: Cambridge.

Gantner, Anita, Werner Güth, and Manfred Königstein (2001), Equitable Choices in Bargaining Games with Joint Production, *Journal of Economic Behavior and Organization* 46, 209-225.

Goeree, Jacob K., and Charles A. Holt (2000), Asymmetric Inequality Aversion and Noisy Behavior in Alternating-Offer Bargaining Games, *European Economic Review* 44, 1079-1089.

Grossman, Sanford, and Oliver D. Hart (1986), The Costs and Benefits of Ownership, A Theory of Lateral and Vertical Integration, *Journal of Political Economy* 94, 691-719.

Grout, Paul A. (1984), Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach, *Econometrica* 52, 449-460.

Güth, Werner (1994), Distributive Justice – a Behavioral Theory and Empirical Evidence, in Brandstätter, Hermann, and Werner Güth (eds.), *Essays on Economic Psychology*, Springer: Berlin, 153-176.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze (1982), An Experimental Analysis of Ultimatum Bargaining, *Journal of Economic Behavior and Organization* 3, 367-388.

Hackett, Steven C. (1993), Incomplete Contracting: A Laboratory Experimental Analysis, *Economic Inquiry* 31, 274-297.

Hart, Oliver (1995), *Firms, Contracts, and Financial Structure*, Oxford University Press: Oxford.

Hart, Oliver, and John Moore (1988), Incomplete Contracts and Renegotiation, *Econometrica* 56, 755-785.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian (1978), Vertical Integration, Appropriable Rents, and the Competitive Contracting Process, *Journal of Law and Economics* 21, 297-326.

Königstein, Manfred (2000), *Equity, Efficiency and Evolutionary Stability in Bargaining Games with Joint Production, Lecture Notes in Economics and Mathematical Systems*, Springer: Berlin.

Muthoo, Abhinay (1999), *Bargaining Theory with Applications*, Cambridge University Press: Cambridge.

Ochs, Jack, and Alvin E. Roth (1989), An Experimental Study of Sequential Bargaining, *American Economic Review* 10, 6-38.

Rabin, Matthew (1993), Incorporating Fairness into Game Theory and Economics, *American Economic Review* 83, 1281-1302.

Roth, Alvin E., Michael W.K. Malouf, and J. Keith Murnighan (1981), Sociological versus Strategic Factors in Bargaining, *Journal of Economic Behavior and Organization* 2, 153-177.

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

Selten, Reinhard (1975), Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games, *International Journal of Game Theory* 4, 25-55.

Selten, Reinhard (1978), The Equity Principle in Economic Behavior, in Gottinger H., and W. Leinfellner (eds.), *Decision Theory and Social Ethics*, Reidel: Dordrecht, 289-301

Shaked, Avner, and John Sutton (1984), Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model, *Econometrica* 52, 1351-1364.

Tirole, Jean (1986), Procurement and Renegotiation, *Journal of Political Economy* 94, 235-259.

Tröger, Thomas (2002), Why Sunk Costs Matter for Bargaining Outcomes: The Evolutionary Approach, *Journal of Economic Theory* 102, 375-402.

Williamson, Oliver E. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press: New York.

Williamson, Oliver E. (1985), *The Economic Institutions of Capitalism*, Free Press: New York.

Von Siemens, Ferdinand (2005), Bargaining under Incomplete Information, Fairness, and the Hold-up Problem, University of Munich.