

Transaction Costs and the Robustness of the Coase Theorem*

LUCA ANDERLINI
(*Southampton University and
Georgetown University*)

LEONARDO FELLI
(*London School of Economics*)

January 2001

ABSTRACT. This paper explores the extent to which the presence of ex-ante transaction costs may lead to failures of the Coase Theorem. In particular we identify and investigate the basic ‘hold-up problem’ which arises whenever the parties to a Coasian negotiation have to pay some *ex-ante costs* for the negotiation to take place. We then show that a ‘Coasian solution’ to this hold-up problem is not available. This is because a Coasian solution to the hold-up problem typically entails a negotiation about the payment of the costs associated with the future negotiation which in turn is associated with a fresh set of ex-ante costs, and hence with a new hold-up problem.

JEL CLASSIFICATION: C70, D23, D60, D80.

KEYWORDS: Transaction Costs, Hold-Up Problem, Coase Theorem, Coasian Negotiation.

ADDRESS FOR CORRESPONDENCE: Leonardo Felli, London School of Economics, Houghton Street, London WC2A 2AE, England. E-mail: lfelli@econ.lse.ac.uk.

*This paper is a revised version of part of the material contained in a working paper entitled “Costly Coasian Contracts” (Anderlini and Felli 1998). Both authors acknowledge financial support from the E.S.R.C. (Grant R000237825). This revision was completed while Leonardo Felli was visiting the Department of Economics at the University of Pennsylvania. Their generous hospitality is gratefully acknowledged. We are also grateful to seminar participants at the I.S.E.R. 2000 in Siena for stimulating comments.

1. Introduction

1.1. Motivation

The Coase theorem (Coase 1960) has had a pervasive influence on the way economists and legal scholars think about *inefficiencies*. It guarantees that provided that *property rights* are allocated, fully informed rational agent involved in an inefficient situation will ensure through negotiation that there are no unexploited gains from trade and hence an efficient outcome obtains.

In its strongest formulation, the Coase theorem is interpreted as guaranteeing an efficient outcome regardless of the “way in which property rights are assigned” (Nicholson (1989, p.725)) and whenever the potential mutual gains “exceed [the] necessary bargaining costs” (Nicholson (1989, p.726)).¹

The predictions entailed by the stronger version of the Coase theorem are startling. Whenever property rights are allocated, we should observe only outcomes which are *constrained efficient* in the sense that all potential gains from trade (net of transaction costs) are exploited. This clearly contradicts even the most casual observation of empirical facts. The economic world is full of inefficient situations in which obvious Pareto improving negotiation opportunities are available but they are left unexploited by the parties involved.

If we were to believe the predictions of the ‘strong’ Coase theorem, all these apparent inefficiencies would not be real inefficiencies at all. They should simply be viewed as the result of transaction costs which are ‘high’ relative to the potential gains from trade. We take the view that this is not a satisfactory explanation of these observed facts.

Our aim in this paper is to take issue with this strong version of the Coase Theorem and show that the impact of transaction costs can extend over and above their size relative to the potential gains from trade. This stems from the *strategic role* which transaction costs may play in a Coasian negotiation. It turns out that a key factor in determining the strategic role of transaction costs is whether they are payable *before*

¹This stronger version of the Coase theorem does *not* correspond to what is claimed in Coase (1960), but it is an interpretation of it which is sufficiently common to have found its way into basic microeconomic text-books such as the one quoted above.

or *after* the negotiation on the distribution of the unexploited gains from trade takes place. Our analysis below highlights the impact of *ex-ante* transaction costs.

The primary effect of *ex-ante* transaction costs is that they may generate a constrained inefficient outcome. In the most basic version of our model the agents may end up not exploiting any of the potential gains from trade, and hence the potentially beneficial negotiation may not even occur. When the choice of ex-ante costs is ‘gradual’, higher costs paid correspond to a more ‘detailed’ negotiation that allows the parties to exploit better the potential gains from trade and hence to achieve a higher surplus from the negotiation. In this case the agents will, in general, end up leaving part of the potential gains from trade unexploited. In other words, they will choose to negotiate an agreement which is less detailed than would be optimal, after the ex-ante costs are taken into account. In its simplest form, the strategic effect which drives our results below is not hard to outline.

Consider any ‘Coasian negotiation’ with the following features.² Two agents contemplate entering a negotiation which might yields a surplus of an arbitrary given size. Moreover, the two agents’ shares of the surplus generated by the negotiation are exogenously given, say because the extensive form which they must use in the negotiation is itself exogenously given.

Suppose now that there are ex-ante transaction costs associated with this negotiation. In particular, suppose that the agents must each pay a given cost before the negotiating phase begins. Then, if the *distribution* of ex-ante costs is such that one (or both) agents will not be able to recoup the ex-ante cost given the distribution of surplus, the negotiation will not take place. This is possible even when the total of ex-ante costs across the two agents is *less* than the surplus which the negotiation may generate so that it would be *socially efficient* for the agents to pay the ex-ante costs and negotiate the division of the surplus.

In this paper we model this situation taking the distribution of surplus as exogenously given, in Anderlini and Felli (1998) and Anderlini and Felli (2001a) we explore a variety of extensive forms in which the agents are allowed to bargain over

²By this we mean a situation in which the property rights of the agents are sufficiently well defined to allow them to enter into a negotiating phase, and that there are some un-exploited gains from trade. This obviously covers an extremely wide variety of possible situations, ranging from text-book like externalities to complex contingent contracts.

the distribution of surplus, provided of course that the ex-ante costs have been paid. We find that the problem we have described is ‘highly pervasive’ in the sense that in a whole variety of extensive forms, the agents will not negotiate an agreement even though it would be socially efficient to do so.

What we have just described is a version of a source of inefficiencies well known in contract theory as the ‘hold-up problem’ (Grout 1984, Grossman and Hart 1986, Hart and Moore 1988, among many others). The problem is particularly acute in our setting since it may be impossible for the negotiating parties to find a ‘Coasian solution’ to this hold-up problem for the following reasons.

Imagine that the two agents in the negotiation we have described attempt to resolve the inefficiency in the following (Coasian) way. Before the ex-ante costs are paid, they negotiate a transfer of money which will compensate the agent who is unable to recoup the ex-ante cost for his loss, of course contingent on his paying the ex-ante cost. Provided the sum of ex-ante costs does not exceed the surplus generated by the negotiation, such transfer can always be arranged so that both agents now benefit from paying the ex-ante costs and entering the negotiation. However, the problem which arises now is that the contingent compensating transfer can itself be viewed as a Coasian negotiation, which may involve a new set of ex-ante costs.

Suppose that the ‘second tier’ negotiation we have described does indeed involve a new set of strictly positive ex-ante costs. Suppose moreover that the second tier ex-ante costs must be paid for, and the second tier negotiation must occur, *before* the first order costs are paid for and the first order negotiation occurs.³

Then it is possible to show that the ex-ante costs associated with the second tier negotiation may not be paid. In particular there always exists an equilibrium such that the contingent compensating transfer is not negotiated and therefore does not take place. Hence, the ex-ante costs associated with the first tier negotiation will, in turn, not be paid, and the actual surplus will not materialize. This outcome is of particular interest since it is the only one which survives when we restrict attention to ‘renegotiation-proof’ equilibria. In other words if, in a Coasian spirit, we insist

³This is obviously an *assumption* as such. However, we believe it to be plausible in a wide variety of cases.

that the outcome of negotiation at every stage must yield a (constrained) efficient outcome, the overall outcome is not efficient and therefore not Coasian at all.

The hold-up problem described above is less pervasive in an environment in which each party to the negotiation has an incentive to bestow a higher share of the surplus to his partner for reasons that may have nothing to do with the partner's payment of the ex-ante transaction costs. In Section 6 below we consider one of these reasons. If the size of the surplus shared by the two parties depends on the distribution dictated by the negotiation, then each party may be willing to leave his counterpart more than his outside option in the attempt to increase the overall size of the surplus. This may result in a smaller range of parameter values for which the hold-up problem described above occurs.

1.2. *Related Literature*

It is clear that the original version of the Coase theorem (Coase 1960) explicitly assumes the absence of any transaction costs or other frictions in the bargaining process. Indeed, Coase (1992) describes the theorem as a provocative result that was meant to show how unrealistic is the world without transaction costs.⁴

Here, we go further by identifying the crucial strategic role played by ex-ante transaction costs (as opposed, for instance, to those transaction costs which are payable ex-post) which may lead to an outcome that is *constrained inefficient*.

We are certainly not the first to point out that the Coase theorem no longer holds when there are frictions in the negotiation process. There is a vast literature on bargaining models where the frictions take the form of incomplete and asymmetric information. With incomplete information, efficient agreements often cannot be reached and delays in bargaining may obtain.⁵ By contrast, the reduced form negotiation that we consider in our analysis is one of complete information. The source of inefficiencies in this paper can therefore be traced directly to the presence of transaction costs.

As in our analysis, Dixit and Olson (2000) are concerned with a classical Coasian public good problem in which they explicitly model the agents' ex-ante (possibly

⁴de Meza (1988) provides an extensive survey of the literature on the Coase theorem, including an outline of its history and possible interpretations.

⁵See Muthoo (1999) for an up-to-date coverage as well as extensive references on this strand of literature and other issues in bargaining theory.

costly) decisions of whether to participate or not in the bargaining process. In their setting they find both efficient and inefficient equilibria as opposed to the unique constrained inefficient equilibrium we derive in our setting. They also highlight the inefficiency of the symmetric (mixed-strategy) equilibria of their model.

The papers that are closest to the present one are Anderlini and Felli (2001a) and Anderlini and Felli (2001b). In particular, Anderlini and Felli (2001b) is also concerned with the hold-up problem generated by ex-ante contractual costs in a model of a Coasian negotiation and with the inefficiencies it generates. However, while the parties' bargaining power is taken as a primitive in the present analysis, in Anderlini and Felli (2001b) we take as primitive a general formulation of the 'game forms' that define the actual negotiation game played by the parties. We then show that the inefficiencies generated by the presence of ex-ante transaction costs are robust to this general formulation of the the protocol followed by the actual negotiation and any possible associated compensating transfers. In Anderlini and Felli (2001b) we also interpret our results in a different vein as the ones presented here. In particular, there we argue that the hold-up problem created by ex-ante transaction costs should be considered a major source of contractual incompleteness.

Anderlini and Felli (2001a) instead focuses on a specific extensive form of the parties' negotiation in the presence of ex-ante transaction costs: the alternating offers bargaining game with complete information with potentially infinite rounds of negotiation in which the players discount the future at a strictly positive rate (Rubinstein 1982). Ex-ante transaction costs are modeled as positive costs that both parties, at each round of negotiation, must pay to participate to that round of the bargaining game. In this environment, in Anderlini and Felli (2001a) we show that a whole plethora of inefficiencies may arise. First, for some values of these costs it is efficient to reach an agreement but the unique equilibrium outcome is one in which agreement is never reached. Secondly, even when there are equilibria in which an agreement is reached, we find that the model always has an equilibrium in which agreement is never reached, as well as equilibria in which agreement is delayed for an arbitrary length of time. Finally, the equilibrium in which agreement is never reached is pervasive in the sense that if the parties are given the opportunity to renegotiate out of these inefficient outcomes, the only equilibrium outcome that survives is the one in

which agreement is never reached, regardless of the value of the transaction costs.

1.3. Overview

We begin with a discussion of the possible interpretations of the ex-ante transaction costs in Section 2. In Section 3 we present the simplest possible model of the basic hold-up problem associated with a surplus-enhancing negotiation. This problem is analyzed in the case in which the ex-ante costs associated with the Coasian negotiation are discrete and are either complements or substitutes. In Section 4 we address the question of whether a Coasian solution to our basic hold-up problem is plausible. We do this by analyzing the possibility of a negotiated transfer from one party to the other before the payment of the transaction costs that are at the origin of the hold-up problem. We then turn (Section 5) to the analysis of a simple model in which the agents' choice of ex-ante costs is continuous. In Section 6 we present a model in which the size of the surplus available to the parties depends on its distribution. Section 7 offers some concluding remarks. To ease the exposition, we have relegated all proofs to the Appendix.

2. Ex-Ante Transaction Costs

We are concerned with Coasian negotiations in which the parties have to incur some ex-ante transaction costs, before they reach the stage in which the actual negotiation occurs.

The interpretation of these ex-ante transaction costs which we favor is that of time spent 'preparing' for the Coasian negotiation. Typically, a variety of tasks need to be carried out by the parties involved before the actual negotiation begins.

In those cases in which the negotiation of an agreement contingent on a state of nature is concerned, both parties need to conceive of, and agree upon, a suitable language to describe precisely the possible realizations of the state of nature. The parties also need to collect and analyze information about the 'legal environment' in which the agreement will be embedded. For instance, in different countries the same agreement will need to be drawn-up differently to make it legally binding and hence enforceable.

In virtually all settings in which a negotiation is required the parties need to spend time arranging a way to ‘meet’, and they need to ‘ earmark ’ some of their time schedules for the actual meeting.

In many cases, before a meaningful negotiation can start, the parties will need to collect and analyze background information which may be relevant to their understanding of the actual trading opportunities. These activities may range from collecting information about (for instance the credit-worthiness of) the other party, to actual ‘thinking’ or ‘complexity’ costs incurred to understand the negotiation problem. We view this type of ex-ante transaction costs as both relevant and important for the type of effects which we identify in our analysis below. However, it should be emphasized that our model does not *directly* apply to this type of costs. This is because in our model the size of the gains from trade is fixed and known to the parties. On the other hand, the lack of information and/or understanding of the negotiation setting which we have just described, would clearly make the size of the surplus uncertain for the parties involved. We have not considered the case of uncertain surplus for reasons of space and analytical convenience. However, we conjecture that the general ‘flavor’ of our results generalizes to this case.

We conclude this section with an observation. In many cases the parties to a negotiation will have the opportunity to delegate to outsiders many of the tasks which we have mentioned as sources of ex-ante transaction costs. The most common example of this is the hiring of lawyers. In these cases, the time costs which we have just discussed will be *monetized* at an appropriate rate. Abstracting from agency problems (between the negotiating party/principal and the lawyer/agent), which are likely to increase the ex-ante costs anyway, our analysis applies, unchanged, to the case in which the ex-ante transaction costs are payable in money.

3. Discrete Transaction Costs

3.1. Perfect Complements

Our model consists of two agents, called A and B , who face a ‘Coasian’ opportunity to realize some gains from trade. Without loss of generality we normalize the size of the surplus, which is realized if an agreement is reached, to be one, and we set the parties’ payoffs in the case disagreement to be equal to zero.

In the simple model we analyze in this section, once the negotiating phase is reached the division of surplus between the two agents is exogenously given and cannot be changed by the agents. This should be thought of as the result of the agents having exogenously given bargaining power in the negotiating phase (the *extensive form* of the bargaining game they play to divide the surplus is exogenously given).⁶

Let $\lambda \in [0, 1]$ be the share of the surplus which accrues to agent A if the parties engage in the negotiation and $1 - \lambda$ the share of the surplus which accrues to B .

For the negotiation to start, each agent has to pay a given *ex-ante transaction cost*. In other words, the agents reach the negotiating phase only if they both pay a certain amount before the negotiation begins.⁷ These costs should be thought of as representing a combination of the activities necessary for the gains from trade to materialize which we discussed in some detail in Section 2 above.

Let $c_A > 0$ and $c_B > 0$ be the two agents' ex-ante costs. Clearly, if $c_A + c_B > 1$ then the two agents will never reach the negotiation that yields the unit surplus, but then neither would a social planner since the total cost of the negotiation exceeds the surplus which it yields. We are interested in the case in which it would be socially efficient for the two agents to negotiate an agreement. Our first assumption guarantees that this is the case.

Assumption 1. *The surplus which the negotiation yields exceeds the total ex-ante costs which are payable for the negotiation to occur. In other words $c_A + c_B < 1$.*

Our two agents play a two-stage game. In period $t = 0$ they both simultaneously and independently decide whether to pay their ex-ante cost. Only if both agents pay their ex-ante cost at $t = 0$, do they have the possibility of negotiating an agreement yielding a surplus of size one at $t = 1$.⁸ We simplify the analysis of the game at $t = 1$ by taking it as a 'black box' yielding payoffs of λ to A and $1 - \lambda$ to B . If one or

⁶In Section 6 below, we consider the possibility that the *size* of the surplus may depend on its distribution across the agents.

⁷Notice therefore that we are implicitly assuming that the agents have some endowments of resources out of which the ex-ante costs can be paid.

⁸Notice that we are therefore assuming that the two agents' ex-ante costs are *perfect complements* in the 'technology' which determines whether the surplus-generating negotiation is feasible or not. We examine the cases of perfect substitutes, and of strategic complements in Subsections 3.2 and 3.3 below respectively.

	pay c_B	not pay c_B
pay c_A	$\lambda - c_A, 1 - \lambda - c_B$	$-c_A, 0$
not pay c_A	$0, -c_B$	$0, 0$

Figure 1: Normal form of the two-stage game with ex-ante costs.

both agents do not pay their ex-ante costs at $t = 0$, the game at $t = 1$ is trivial: the negotiation which yields the unit surplus is not feasible; the agents have no actions to take and they both receive a payoff of zero.

Throughout the paper, unless otherwise stated, by equilibrium we mean a *subgame perfect* equilibrium of the game at hand.

The normal form which corresponds to the two-stage game we have just described is depicted in Figure 1. From this it is immediate to derive our first two propositions, which therefore are stated without proof.

Proposition 1. *Let a pair of ex-ante costs $c_A > 0$ and $c_B > 0$ satisfying Assumption 1 be given. Then there exists a range of values — namely $\Lambda = [0, c_A) \cup (1 - c_B, 1]$ — for the distribution parameter λ such that the only equilibrium of the two-stage game represented in Figure 1 has neither agent paying the ex-ante cost, and therefore yields the no-agreement outcome.*

Proposition 2. *Let any value of the distribution parameter $\lambda \in [0, 1]$ be given. Then there exists a set $\mathcal{C} = \{c_A, c_B \mid \text{either } c_A > \lambda \text{ or } c_B > 1 - \lambda \text{ and } c_A + c_B < 1\}$ of pairs of ex-ante costs which satisfy Assumption 1, and such that the only equilibrium of the two-stage game represented in Figure 1 has neither agent paying the ex-ante cost, and therefore yields the no-agreement outcome.*

We view Propositions 1 and 2 together as implying that in the presence of ex-ante transaction costs, if the *distribution* of ex-ante costs across the parties is sufficiently ‘mis-matched’ with the given distribution of surplus, then the ex-ante costs will generate a version of the hold-up problem which will induce the agents not to negotiate an agreement even though it would be socially efficient to do so.

The intuition behind our results above is simple enough. If negotiating an agreement involves some costs which are payable ex-ante, the share of the surplus accruing to each party will not depend, in equilibrium, on whether the ex-ante costs are paid. Therefore, the parties will pay the costs only if the distribution of the surplus generated by the negotiation will allow them to recoup the cost ex-post. If the distribution of surplus and that of ex-ante costs are sufficiently ‘mis-matched’, then one of the agents will not be able to recoup the ex-ante cost. In this case, an agreement will not be reached, even though it would generate a total surplus large enough to cover the ex-ante costs of both agents.

Consider now, as a benchmark, the alternative setup in which both parties can pay the costs c_A and c_B after the negotiation has occurred and an agreement is reached. In other words the transaction costs can be paid ex-post rather than ex-ante.⁹ In this case the extensive form of the game is equivalent to a simple negotiation in which the size of the gains from trade is $1 - c_A - c_B$. The assumption we made on the black-box negotiation implies that in this case the two parties do indeed reach an agreement. Party A receives the share of surplus $\lambda (1 - c_A - c_B)$ while party B receives the share $(1 - \lambda) (1 - c_A - c_B)$. In other words when transaction costs can be paid ex-post the strong version of the Coase Theorem holds and a constrained efficient outcome is achieved.

We conclude this subsection with two observations. First of all, the simultaneity in the payment of the ex-ante costs is not essential to Propositions 1 and 2. Both results apply to the case in which the ex-ante costs are payable sequentially by the two agents before the actual negotiation begins.

Secondly, while the model has a unique equilibrium for the parameter configurations identified in Propositions 1 and 2, it has multiple equilibria whenever these two propositions do not apply. It is clear that, whenever both $\lambda > c_A$ and $(1 - \lambda) > c_B$, the model has two equilibria. One in which the ex-ante costs are paid and an agreement is reached, and another in which neither agent pays the ex-ante costs simply because he expects the other agent not to pay his cost either. The equilibrium in which the agreement is reached strictly Pareto-dominates the no-agreement equilibrium. Clearly, the

⁹These ex-post costs could, for example, be associated with registering the agreement with the relevant authorities.

multiplicity of equilibria disappears if the costs are payable sequentially. The latter observation will become relevant again in Section 4 below.

3.2. *Perfect Substitutes*

So far, we have assumed that the agents' ex-ante costs are 'perfect complements' in determining whether the negotiation of an agreement is feasible or not. The next proposition tells us that when the agents' ex-ante costs are *perfect substitutes* our constrained inefficiency results of Subsection 3.1 still hold, although the inefficiency may take a different form.

The intuition behind the next results is straightforward. In an environment in which the ex-ante costs may be paid by *either* agent the negotiation leads to a constrained efficient outcome if *at least one* of the two ex-ante costs is smaller than the size of the surplus. It is then easy to envisage a situation in which the share of the surplus accruing to each agent is strictly smaller than his ex-ante costs although there is enough surplus to cover the smallest of these costs. In this case, in equilibrium, the parties will not reach an agreement although it would be socially efficient to do so.

When the ex-ante costs are perfect substitutes, a new type of inefficiency can also arise in equilibrium. In particular, it is possible that the agents reach an agreement, but the equilibrium involves the highest of the two ex-ante costs being paid.

Formally, when the ex-ante costs are perfect substitutes Assumption 1 needs to be modified. Assumption 2 below identifies the range of ex-ante costs which guarantees that negotiating an agreement is socially efficient in this case.

Assumption 2. *The surplus which the agreement yields exceeds the minimum ex-ante cost payable for the negotiation to become feasible. In other words $\min\{c_A, c_B\} < 1$. Without loss of generality (up to a re-labeling of agents) let $c_A \leq c_B$. Hence $c_A < 1$.*

Consider now the model with ex-ante transaction costs that are perfect substitutes and let Assumption 2 above hold. The normal form of our new two-stage game is depicted in Figure 2.

We start with the case in which the value of the distribution parameter $\lambda \in [0, 1]$ is given. In such a case by varying the values of the transaction costs (c_A, c_B) it is always possible to generate two different types of inefficiencies.

	pay c_B	not pay c_B
pay c_A	$\lambda - c_A, 1 - \lambda - c_B$	$\lambda - c_A, 1 - \lambda$
not pay c_A	$\lambda, 1 - \lambda - c_B$	$0, 0$

Figure 2: Normal form when the ex-ante costs are perfect substitutes.

First of all, it is always possible that the ex-ante costs be such that the no-agreement outcome obtains. This is immediate from the payoffs in Figure 2 and it is stated without proof in our next proposition.

Proposition 3. *For any given $\lambda \in [0, 1]$ there exists a set $\hat{\mathcal{C}}_1 = \{c_A, c_B \mid 1 > c_A > \lambda \text{ and } c_B > (1 - \lambda)\}$ of pairs of ex-ante costs satisfying Assumption 2 such that the only equilibrium of the two-stage game represented in Figure 2 has neither agent paying the ex-ante cost, and therefore yields the no-agreement outcome.*

Secondly, given any $\lambda \in [0, 1]$, it is possible that the ex-ante costs are such that the second type of inefficiency we have mentioned above obtains in equilibrium. An agreement is negotiated, but it is the agent with the highest ex-ante cost who pays in equilibrium. Once again our claim follows immediately from the payoff matrix in Figure 2 and it is stated without proof in the following proposition.

Proposition 4. *For any given $\lambda \in [0, 1/2)$ there exists a set $\hat{\mathcal{C}}_2 = \{c_A, c_B \mid 1 > c_A > \lambda \text{ and } c_B < (1 - \lambda)\}$ of pairs of ex-ante costs satisfying Assumption 2 such that the only equilibrium of the two-stage game represented in Figure 2 has agent A not paying the ex-ante cost c_A , and B paying the ex-ante cost $c_B > c_A$.¹⁰*

Next, we consider the case in which we take as given the ex-ante costs (c_A, c_B) . In this case it is always possible that the value of the distribution parameter λ be such that one of the two types of inefficiencies we have identified obtains in equilibrium. Again, we state our next proposition without proof since it is an immediate consequence of the payoff matrix in Figure 2.

¹⁰Recall that we are assuming that $c_A \leq c_B$. To deal with the range $\lambda \in (1/2, 1]$ it is sufficient to modify Assumption 2 to read $c_B \leq c_A$ and $c_B < 1$. Proposition 4 does not hold when $\lambda = 1/2$.

Proposition 5. *Let a pair of ex-ante costs $c_A > 0$ and $c_B > 0$ satisfying Assumption 2 be given. Assume that $c_A \neq c_B$.*

If it is the case that $c_A + c_B \leq 1$, then for any λ in $[0, c_A)$ the unique equilibrium of the model is constrained inefficient in the sense that an agreement is negotiated, but it is B who pays the ex-ante cost $c_B > c_A$.

If it is the case that $c_A + c_B > 1$, then for any $\lambda \in [0, \min\{0, 1 - c_B\}]$ the (unique if $\lambda \neq 1 - c_B$) equilibrium of the model is, again, that an agreement is negotiated, but it is B who pays the ex-ante cost. Moreover, for any $\lambda \in (1 - c_B, c_A)$ the unique equilibrium of the model involves neither agent paying his ex-ante cost, and hence yields the no-agreement outcome.

3.3. Strategic Complements

Our goal in this subsection is to show that the analogues of Propositions 1 and 2 hold when the ex-ante costs are technologically perfect substitutes, but are ‘strategic complements’ in the game-theoretic sense.¹¹ We conjecture that this is true ‘in general’, but we limit our formal analysis to an example which is a modification of the model of the previous subsection.

The description of our next model is as follows. At $t = 0$ both agents decide simultaneously and independently whether to pay their ex-ante costs. If both agents decide not to pay the ex-ante costs, then negotiation is not feasible and both receive a payoff of zero. If either agent $i \in \{A, B\}$ pays the ex-ante cost c_i at $t = 0$ the negotiation of an agreement yielding one unit of surplus becomes feasible. If both pay their ex-ante costs at $t = 0$ the distribution parameter λ determines the agreement which is negotiated and the A ’s and B ’s payoffs are $\lambda - c_A$ and $1 - \lambda - c_B$ respectively.

However, if only one agent, say A , pays the ex-ante cost at $t = 0$, he is allowed to make a take-it-or-leave-it offer ℓ to B at $t = 1$. The value of ℓ is interpreted as an offer to make A ’s and B ’s payoffs equal ℓ and $1 - \ell$ respectively, minus any costs paid. This can be thought of as a crude way to say that if only one agent pays the ex-ante cost then the bargaining power shifts dramatically in his favor.

¹¹Intuitively, two decision variables are strategic complements if an increase in one induces an increase in the optimal choice (the ‘best response’ of the opposing player) of the other. See Fudenberg and Tirole (1996, Ch. 12).

Moreover, we assume that A , if he alone has paid the ex-ante cost, can, in principle, make some offers which would push agent B below his individual rationality constraint. In other words we assume that the take-it-or-leave-it offer ℓ must be in the interval $[-\epsilon, 1 + \xi]$ with ϵ and ξ some (possibly small) positive numbers.

At $t = 2$, B has two choices. He can either pay an ex-ante cost $c'_B > 0$ or pay nothing.¹² If he does not pay he does not observe A 's offer, but is still allowed to accept or reject it blind. If B decides to pay his ex-ante cost, he can then observe A 's offer and subsequently decide to accept or reject it.

The description of the extensive form which is played if it is B alone who pays the ex-ante cost at $t = 0$ is exactly symmetric to the case we have just described.

Notice that the strategic complementarity of the two agents' ex-ante costs is built into the extensive form game we have described precisely via the shift in bargaining power which obtains when one agent alone pays the ex-ante costs at $t = 0$.

Suppose now that the parameters λ , c_A and c_B are such that the agents would not negotiate an agreement in the model described in Subsection 3.1. Our next proposition then tells us that, in the model with strategic complementarities we have just described, they will not negotiate an agreement either.

Proposition 6. *Consider the model with ex-ante costs which are strategic complements described above in this subsection. Assume that either $\lambda < c_A$ or $1 - \lambda < c_B$. Then the unique equilibrium outcome of the model has neither agent paying the ex-ante cost at $t = 0$, and hence the no-agreement outcome obtains.*

4. The Impossibility of a Coasian Solution

In Section 3 we have argued that ex-ante transaction costs may give rise to a version of the hold-up problem which in turn generates an inefficient (no-agreement) outcome. The next natural question to ask is then whether a Coasian solution to the hold-up problem is generally available in the present set up. In other words: is it possible to add another stage of negotiation to our model (say $t = -1$), prior to the stage

¹²Notice that while Proposition 6 below restricts the values of c_A and c_B to be in an appropriate range, c'_B can take any positive (small) value.

in which the ex-ante costs are incurred, in which the agents can negotiate a ‘grand agreement’, which will resolve the hold-up problem and hence restore efficiency?

The answer to the above question is trivially ‘yes’, if at $t = -1$ a *truly grand* agreement can be negotiated *costlessly*, which specifies everything, including the payment of the ex-ante costs, *and* the division of the actual surplus at time $t = 1$. The answer, however, changes dramatically if the ‘grand agreement’ is itself costly.

We specify two crucial details of the grand agreement negotiation stage. First of all we assume, as seems plausible in the present context, that in order to be able to negotiate an agreement at $t = -1$ a fresh set of ex-ante costs must be incurred by the parties before $t = -1$, say at $t = -2$. Secondly, we restrict the agents to negotiate a *compensating transfer* at $t = -1$. In other words, we take a specific view on the agreements which the agents can enter at $t = -1$. Indeed, we restrict them to be transfers *contingent* on the payment of ex-ante costs at $t = 0$. This seems to be in the spirit of our model of Section 3, in that, in principle, it allows the agents to effectively transfer surplus between them, but it keeps the distribution of surplus in the last stage of the negotiating process, $t = 1$, exogenously fixed, as before.¹³

It is worth emphasizing at this point that we find that the presence of any *strictly positive* ‘second tier’ ex-ante costs might be sufficient to keep the addition of a grand agreement negotiation stage from resolving the hold-up problem of Section 3. We view this as a strength of the results we present in this section. Indeed, in many situations it would be sensible to assume that the second tier ex-ante costs are in fact at least as large as the ‘first tier’ ex-ante costs, on the grounds that the negotiation of a grand agreement, in an intuitive sense, is a more complex object than the negotiation of the agreement itself.

Formally, we modify the model of Section 3 as follows. There are now four time periods, $t \in \{-2, -1, 0, 1\}$. The sequence of decisions and events for the two agents (depicted schematically in Figure 3) is as follows. In period $t = -2$, the two agents decide simultaneously and independently whether to pay the second tier ex-ante costs (c_A^2, c_B^2) . If either or both agents decide not to pay these ex-ante costs, the period

¹³While the assumption that a fresh set of ex-ante costs arises at $t = -2$ is crucial for our result (Proposition 7), the restriction to the negotiation of *compensating transfers* is not (Anderlini and Felli 2001b).

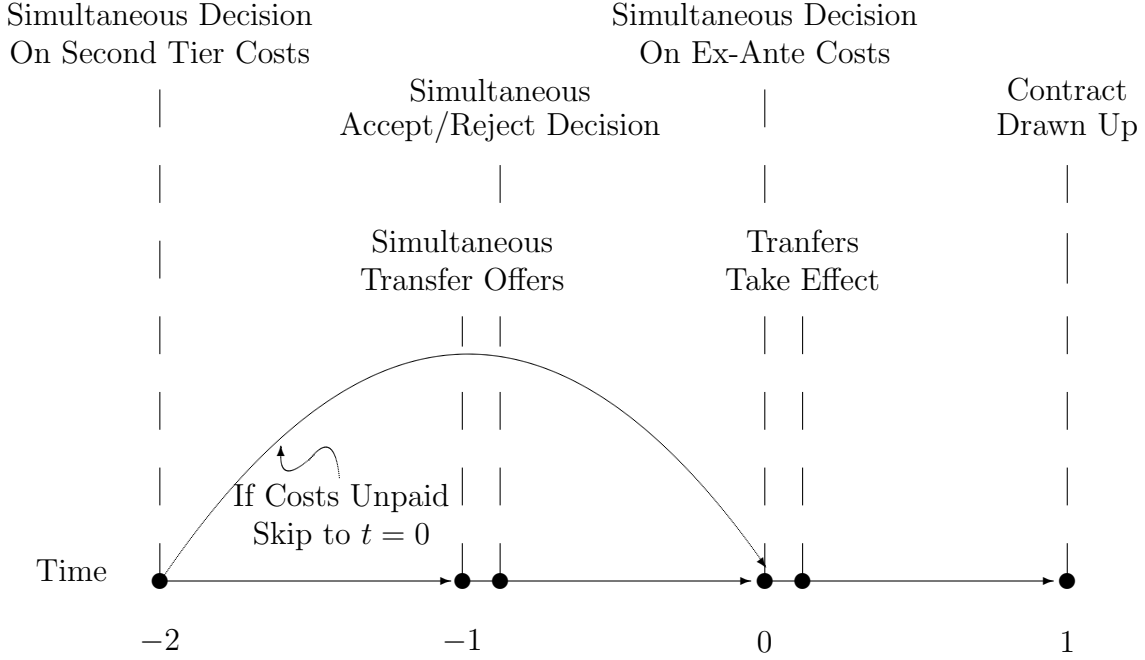


Figure 3: Timing in the two tier model.

$t = -1$ compensating transfers to be described shortly are automatically set equal to 0, and the agents effectively move directly to time $t = 0$. If, on the other hand both agents pay the second tier ex-ante costs, then period $t = -1$ compensating transfers can be negotiated.

For simplicity, we assume that (provided that both pay the second tier costs) at $t = -1$, both agents make simultaneous offers of contingent compensating transfers to each other. Formally, each agent $i \in \{A, B\}$ chooses a real number $\sigma_i \geq 0$, which is interpreted as a commitment to transfer the amount of wealth σ_i to the other agent, $j \neq i$, if and only if j pays the first tier ex-ante cost c_j^0 in period $t = 0$. Immediately after choosing σ_i , still in period $t = -1$, A and B simultaneously choose whether to accept or reject the other agent's offer of compensating transfer. Those offers which are accepted at this stage are binding in period $t = 0$.

The decisions and events in periods $t = 0$ and $t = 1$ are analogous to those described in Subsection 3.1. At $t = 0$, both agents choose simultaneously and independently whether to pay the first tier ex-ante costs (c_A^0, c_B^0) . Each agent $i \in \{A, B\}$ then incurs an ex-ante cost of c_i^0 at this time, and subsequently receives a compensating transfer of σ_j from agent $j \neq i$. Only if both agents have paid the first tier ex-ante

costs the $t = 1$ negotiation of the surplus-generating agreement becomes possible.

Provided both agents have paid their first tier ex-ante costs their payoffs are $\lambda - \gamma_A$ and $1 - \lambda - \gamma_B$ respectively, where γ_i denotes the total ex-ante costs paid by agent $i \in \{A, B\}$ during the entire game, minus any compensating transfer received from agent $j \neq i$, and plus any compensating transfers paid by i to j . If the surplus-generating agreement is not negotiated, then the two agents payoffs are simply $-\gamma_A$ and $-\gamma_B$ respectively.

The assumption that the total (for both tiers) of ex-ante costs must be low enough so that it is socially efficient for the parties to negotiate a grand agreement is easy to state for this version of our model.

Assumption 3. *Let $c_i = c_i^2 + c_i^0$ for $i \in \{A, B\}$. Then $c_A + c_B < 1$.*

It is apparent from the description of our model with compensating transfers above (cf. Figure 3) that this model, viewed from $t = 0$, is in fact identical to the simple model of Subsection 3.1, whenever both agents have chosen not to pay the second tier ex-ante costs. We can therefore ask whether the parameters of our model with compensating transfers are such that either Proposition 1 or Proposition 2 guarantee that, in the absence of compensating transfers, the no-agreement outcome is the unique equilibrium of the model. This motivates our next definition.

Definition 1. *Assume that either $c_A^0 > \lambda$ or $c_B^0 > 1 - \lambda$ so that, provided that neither agent has paid the second tier ex-ante cost then the only equilibrium outcome of the model is the no-agreement outcome (see Propositions 1 and 2 above). Then we say that the model with compensating transfers ‘yields the no-agreement outcome in the final stage’.*

We are now ready to state our next proposition. It tells us that, if the parameters of the model of Subsection 3.1 yield the no-agreement outcome, then adding a new stage to the model, with a second tier of positive ex-ante costs and compensating transfers may not solve the hold-up problem generated by the first tier ex-ante costs. In particular, the model with compensating transfers has multiple equilibria, and at least one equilibrium that yields the no-agreement outcome.

Proposition 7. *Consider the model with compensating transfers. Suppose that c_A^0 , c_B^0 and λ yield the no-agreement outcome in the final stage (cf. Definition 1), and assume that the second tier ex-ante costs are strictly positive for both agents ($c_i^2 > 0$ for $i \in \{A, B\}$). Then the model has multiple equilibria. In particular, there always exists an equilibrium in which neither agent pays either tier of ex-ante costs, and hence yields the no-agreement outcome. Moreover, there is also an equilibrium in which both agents pay both tiers of ex-ante costs and an agreement is negotiated.*

The reason why the model with compensating transfers always has one equilibrium in which none of the costs are paid is an obvious one. Recall that at each stage the two agents decide simultaneously and independently whether to pay their ex-ante costs. Moreover an agreement (or compensating transfers) is feasible only if both agents pay. It is then immediately clear that if one agent expects the other not to pay his ex-ante cost he should not pay either. The cost would be wasted since it has no effect on the remainder of the game. Therefore it is an equilibrium for both agents to pay none of the costs.

The intuition behind the existence of a subgame perfect equilibrium in which the parties do pay both tiers of ex-ante costs and negotiate a grand agreement is less straightforward.¹⁴

Imagine that some compensating transfers have been agreed. If the transfers are such that the first tier ex-ante costs are ‘covered’ for both agents (which is always possible in principle because of Assumption 3), then the terminal subgame of the model has two equilibria. One in which an agreement is negotiated, and another one in which neither agent pays the first tier ex-ante cost and the no-agreement outcome obtains. Note that these equilibria are strictly Pareto-ranked.

It is then possible to construct an equilibrium in which the agents switch (off-the-equilibrium-path) between equilibria of the terminal subgame, according to what transfers have been offered and agreed in the first stage of the game. The ‘switching point’ can always be constructed in such a way that it is in the interest of the agent whose share of the surplus exceeds his costs to compensate the other for the shortfall

¹⁴We are indebted with Stephen Matthews for suggesting to us the existence of this type of equilibrium in this model.

between his share of the surplus and both tiers of ex-ante costs. The ‘threat’ of switching to the inefficient equilibrium is ‘credible’ because the no-agreement outcome is always one of the possible equilibrium outcomes of the terminal subgame.

Two observations come to mind with respect to the equilibrium just described. First, even if the agreement is successfully negotiated this is done by paying two tiers of ex-ante costs rather than one. Therefore, even when an agreement is reached the equilibrium of the model is constrained inefficient.

Remark 1. *All equilibria of the model with compensating transfers are constrained inefficient. In particular, the inefficiency takes the form of the no-agreement outcome for some of the equilibria, while for others it takes the form of an agreement that is negotiated paying two tiers ex-ante costs rather than one.*

Secondly, and in our view more importantly, the equilibria of our model with compensating transfers in which an agreement is negotiated rely on the agents playing (off-the-equilibrium-path) an equilibrium in the terminal subgame which is strictly Pareto-dominated by another equilibrium of the same subgame. This runs against the intuition that the parties to a negotiation will be able to re-negotiate ex-post to an equilibrium which makes them both better off when one is available.

Imagine now that we impose the restriction that in the terminal subgame the agents must play the Pareto-efficient equilibrium when the subgame has two equilibria. Then, after the second tier ex-ante costs have been paid, they are *sunk* in a strategic sense. This means that the agent who has a ‘deficit’ in the last stage of the game, by subgame perfection, will accept any offer of compensating transfer which leaves him with a positive continuation payoff. Therefore in any equilibrium which obeys this new restriction, the compensating transfers will *not* take into account the second tier ex-ante costs. Therefore, one of the two agents will find it profitable not to pay the second tier ex-ante cost for which he would not possibly be compensated. This, in turn, means that compensating transfers will not be observed in equilibrium, and therefore yields the no-agreement outcome. This is the focus of Proposition 8 below.

The idea that some type of renegotiation-proofness is an appealing additional restriction to impose on the set of subgame perfect equilibria is not new, both in

contract theory (Grossman and Hart 1986, Hart and Moore 1988, Rubinstein and Wolinsky 1992, Aghion, Dewatripont, and Rey 1994, Nöldeke and Schmidt 1995, among others), and in game theory (Farrell and Maskin 1989, Abreu, Pearce, and Stacchetti 1993, Benoît and Krishna 1993, among many others).

Below, we give an informal definition of a *renegotiation-proof* equilibrium which applies to our model of this section.

Definition 2. *A subgame perfect equilibrium of the model with compensating transfers is renegotiation-proof if and only if the equilibria played in every proper subgame are not strictly Pareto-dominated by any other equilibrium of the same subgame.*¹⁵

Our next result says that if we restrict attention to renegotiation-proof equilibria, then the possibility of compensating transfers does not resolve the hold-up problem identified in Section 3. It is true that the model always has an equilibrium in which transfers take effect and an agreement is negotiated. But this equilibrium is not renegotiation-proof. Thus, although it may be tempting to select (in a Coasian ‘spirit’) the equilibrium with agreement among the two mentioned in Proposition 7 simply because it Pareto-dominates the no-agreement equilibrium, this type of selection is open to an objection which is, in our view, fatal.

Surely, if we are willing to select among Pareto-ranked equilibria in favor of the dominating one, we should also be willing to apply the same logic to every subgame. After all, once entered, every subgame is just like a game. However, if we apply this selection criterion to every subgame (in a recursively consistent way, of course), the only equilibrium of the entire game which survives is the constrained inefficient one, in which the no-agreement outcome obtains.

Proposition 8. *Consider the model with compensating transfers. Suppose that c_A^0 , c_B^0 and λ yield the no-agreement outcome in the final stage and that $c_A^2 > 0$ and $c_B^2 > 0$. Then every renegotiation-proof subgame perfect equilibrium of the model involves neither agent paying either tier of ex-ante costs and therefore yields the no-agreement outcome.*

¹⁵Notice that our informal definition is made particularly simple by the fact that our model with compensating transfers only has one ‘level’ of proper subgames.

We view Proposition 8 as saying that the possibility of compensating transfers does not resolve the hold-up problem identified in Section 3 in the following sense. Either, we are willing to accept the multiple equilibria identified in Proposition 7, and therefore to accept the no-agreement equilibrium as being just as plausible as the one in which an agreement is negotiated. Or, we attempt to select among Pareto-ranked equilibria in favor of the efficient ones. However in this case, we should apply this logic consistently to every subgame, and hence single out those equilibria which are renegotiation-proof. In this case only the no-agreement outcome survives.

Notice that the multiplicity of equilibria in the terminal subgame of our model is crucially dependent on the fact that the ex-ante costs are payable simultaneously by the agents. Therefore if the game is modified so that the costs are payable sequentially, all subgames have a unique equilibrium and the no-agreement outcome is certain to prevail.

In particular, consider the following modification of the extensive form depicted in Figure 3. At $t = 0$ A decides whether to pay the ex-ante cost c_A^0 . Then B observes A 's choice and decides whether to pay the ex-ante cost c_B^0 . The rest of the extensive form is identical to the one in Figure 3. The following proposition characterizes the equilibria of this modification of the model with compensating transfers.¹⁶

Proposition 9. *Consider the modified model with compensating transfers which we have just described. Suppose that c_A^0 , c_B^0 and λ yield the no-agreement outcome in the final stage and that $c_A^2 > 0$ and $c_B^2 > 0$. Then, the unique subgame perfect equilibrium of the model is for neither agent to pay either tier of ex-ante costs, and hence yields the no-agreement outcome.*

5. Continuous Transaction Costs

So far, we have assumed that the agents' decision regarding the ex-ante costs is 'lumpy'; the negotiation of an agreement is not feasible unless both agents sink a minimum, strictly positive, ex-ante cost. This is the reason why Propositions 1 and

¹⁶The alternative game in which B decides whether to pay the ex-ante cost before A is just a relabelling of the one we just described. Proposition 9 below obviously applies to this extensive form as well.

2 above refer to a *range* of the distribution parameter λ for any given ex-ante costs, and to a *range* of ex-ante costs given any value of the distribution parameter λ .

In this section we consider a model in which the agents have a continuous choice of ex-ante costs. Our model is still a ‘reduced form’ one, in that we do not model explicitly the effects of increased ex-ante costs paid by the agents.¹⁷ We simply postulate that the *size of the surplus* yielded by the agreement which the agents negotiate is an increasing function of the magnitude of the ex-ante costs paid by the two agents.

The interpretation of our reduced form increasing relationship between the ex-ante costs and the surplus generated by the negotiated agreement, we believe, is a natural one. We imagine a situation in which, as the agents pay larger amounts of ex-ante costs, more *detailed agreements* become feasible between them. The meaning of the word detail here can range from a more accurate description of the relevant variables, to a more detailed description of the possible states of nature (and therefore, in a dynamic model, to agreements with a longer time horizon), to an agreement which is better specified in legal terms, which as a consequence is more easily enforced, and therefore yields a higher level of surplus net of ‘enforcement costs’.

The results which we derive in this section are the analogues in our set-up of the general under-investment results stemming from a hold-up problem (Hart and Moore 1988). Formally, the model which we analyze is close to Holmström (1982), and can be described as follows.

The two agents, A and B , play a two stage game. At $t = 0$, both agents decide, simultaneously and independently, how much ex-ante transaction cost to pay. Agent $i \in \{A, B\}$ chooses a number $c_i \in [0, \bar{c}_i]$ with $\bar{c}_i \in (0, 1)$.¹⁸ At $t = 1$ the agents do not in fact have any choices to make; the pair of ex-ante costs (c_A, c_B) paid at $t = 0$, determines the size of the surplus that the negotiated agreement yields to the agents. This is then divided among them according to the exogenously given distribution parameter λ . We denote with $x(c_A, c_B)$ the surplus corresponding to the pair (c_A, c_B) ,

¹⁷In Anderlini and Felli (1996) we present a model (Section 8) in which the agents have a ‘gradual’ choice of ex-ante costs, and the effect of increased ex-ante costs paid is modelled explicitly as affording the agents a more detailed agreement which extends further into the future.

¹⁸Recall that we are assuming throughout that the agents pay their ex-ante costs out of their endowments which are available to them at $t = 0$.

with $x_A(c_A, c_B)$ and $x_B(c_A, c_B)$ its partial derivatives with respect to the first and second argument respectively, and with $x_{A,A}(c_A, c_B)$, $x_{B,B}(c_A, c_B)$ and $x_{A,B}(c_A, c_B)$ the second and cross partial derivatives with respect to the same arguments. We assume that x is a (twice-differentiable) strictly increasing and strictly concave function which satisfies the Inada conditions $\lim_{c_A \rightarrow 0} x_A(c_A, c_B) = \infty$, $\lim_{c_B \rightarrow 0} x_B(c_A, c_B) = \infty$, $x_A(c_A, c_B) = 0$ for all $c_A \geq \bar{c}_A$ and $x_B(c_A, c_B) = 0$ for all $c_B \geq \bar{c}_B$.¹⁹ We also assume that the ex-ante costs are complements in the sense that the cross partial derivative $x_{A,B}(c_A, c_B)$ is always positive.

Given a pair of ex-ante costs (c_A, c_B) , the payoffs accruing to A and B are given by $\lambda x(c_A, c_B) - c_A$ and $(1 - \lambda)x(c_A, c_B) - c_B$ respectively. We denote by c_A^* and c_B^* the (unique) equilibrium ex-ante costs which the agents pay in the game we have just described. Given that our assumptions on the function x guarantee an interior solution, the equilibrium is easy to characterize.

Remark 2. *The model with continuous ex-ante costs we have described above yields a unique equilibrium pair (c_A^*, c_B^*) , which can be characterized as follows by the corresponding first order conditions.*

$$x_A(c_A^*, c_B^*) = \frac{1}{\lambda} \quad \text{and} \quad x_B(c_A^*, c_B^*) = \frac{1}{1 - \lambda} \quad (1)$$

The efficiency benchmark with which to compare the equilibrium identified in Remark 2 is straightforward to define and to characterize.

Definition 3. *The socially efficient levels of ex-ante costs in the model with continuous costs are denoted by c_A^E and c_B^E . They are defined as the pair of ex-ante costs which maximize the difference between the surplus given by the negotiated agreement and the sum of ex-ante costs $x(c_A, c_B) - c_A - c_B$.*

There is a unique socially efficient pair of ex-ante costs (c_A^E, c_B^E) , which can be characterized as follows using the corresponding first order conditions.

$$x_A(c_A^E, c_B^E) = 1 \quad \text{and} \quad x_B(c_A^E, c_B^E) = 1 \quad (2)$$

¹⁹Therefore, we are assuming that there are decreasing returns to scale in the relationship between the ex-ante costs paid and the size of the surplus which the negotiation generates.

Using the concavity of x and the fact that the cross partial derivative of x is positive, it is easy to show that (1) together with (2) imply that $c_i^* < c_i^E$ for all $i \in \{A, B\}$. This is the content of our next proposition.

Proposition 10. *Let any value of the distribution parameter $\lambda \in (0, 1)$ for the model with continuous choice of ex-ante costs be given. Then, in equilibrium, both agents pay an inefficiently low level of ex-ante costs in the sense that $c_i^* < c_i^E$ for all $i \in \{A, B\}$. This obviously implies that $x(c_A^*, c_B^*) < x(c_A^E, c_B^E)$.²⁰*

Thus, the agents under-invest in the ex-ante costs that determine the degree of sophistication of the agreement they negotiate in equilibrium. Therefore, the agents negotiate an agreement which is constrained inefficient in the sense that it is less ‘detailed’ than would be socially efficient, even after transaction costs are taken into account.

The intuitive reason why the agents under-invest in their ex-ante costs according to Proposition 10 is simple to outline. Each party’s share of the surplus generated is fixed, the total surplus is an increasing concave function of both ex-ante costs and these costs are complements. Each agent invests in his ex-ante costs only up to the point at which *his own* marginal net return is zero. Such point is therefore below the point at which the marginal *social* (across both agents) net return on his investment in the ex-ante cost is equated to zero.

We conclude this section with an observation. Proposition 10 describes a stronger inefficiency result than Propositions 1 and 2 above since it yields an inefficient outcome *regardless* of the value of λ . Notice that our model of Subsection 3.1 can be viewed as a ‘special case’ (in which the assumption of concavity of x is violated) of our model with a continuous choice of ex-ante costs in which the size of the surplus yielded by the negotiated agreement is a *discontinuous* function of the ex-ante costs paid by the two agents. Intuitively, the ‘marginal’ conditions for efficiency are therefore easier to satisfy in the model of Subsection 3.1 than in our present set-up.

²⁰Notice that if we allow for λ to take the values 0 and 1 as well, we still obtain under-investment. For instance if $\lambda = 0$, A will choose $c_A = 0$, and B will set $0 < c_B < c_B^E$. If $\lambda = 1$, then $0 < c_A < c_A^E$ and $c_B = 0$.

6. The Size of the Surplus as a Function of Its Distribution

In our entire analysis so far we have assumed that the size of the potential surplus generated by the negotiated agreement is determined independently of its distribution across the two agents. Whether this is a good assumption or a bad one, depends on the details of the Coasian negotiation at hand.

It is easy to think of a Coasian negotiation in which the ‘expected quality’ of the object the parties negotiate on is determined (at least in part) by the action(s) of one (or both) agents, after the agreement has been successfully negotiated. This is, for instance, the case if the negotiation concerns a ‘principal’ and an ‘agent’. The action of the agent in this case affects the amount of surplus available, and is in turn affected by how much surplus the agent is able to appropriate. In Anderlini and Felli (1996), we analyze a simple principal-agent model in which the interplay of incentive compatibility and limited liability is the mechanism through which the size of the surplus depends on the share left to the agent.

For reasons of space, in our analysis below we use, again, a ‘reduced form’ model in which we simply assume that the size of the potential surplus depends on its distribution between the two agents.

Intuitively, when the size of the surplus depends on its distribution, the hold-up problem we have analyzed so far becomes less acute for the following reasons. When the distribution of surplus is negotiated between the two agents, one (or both) agent(s) may have an incentive to propose a ‘fairer’ distribution in order to increase surplus size. It is then possible that, as a result, the distribution of surplus which is agreed by the agents is ‘less mis-matched’ with the distribution of ex-ante costs than would be the case otherwise. As we know from Section 3 this may resolve the hold-up problem, and yield an equilibrium outcome in which the agreement is successfully negotiated.

We make our next point in a very simple setting. Assume that the extensive form of the negotiation between the two parties is such that agent A makes a take-it-or-leave-it offer of a share of the surplus to agent B . When the size of the surplus is independent of its distribution the unique subgame perfect equilibrium of this simple extensive form of the parties’ negotiation corresponds to the parameter value $\lambda = 1$ in the black-box negotiation assumed in Subsection 3.1 above. Proposition 2 then

implies that provided $c_B > 0$ the unique equilibrium outcome of the model is the one in which an agreement is never negotiated whatever the distribution of the ex-ante costs. Hence, this is a situation in which the hold-up problem described in Section 3 above is most acute. We then couple this extensive form with a simple functional form describing how the surplus size depends on its distribution, and we show that for *some* possible configurations of (positive) ex-ante costs, in equilibrium an agreement is in fact successfully negotiated. Thus the hold-up problem identified above becomes less acute when the size of the potential surplus generated by the agreement depends on its distribution across the two agents.

As before, x denotes the size of the surplus, while λ is the distribution parameter. The function $x(\lambda)$ embodies the dependence of surplus size on its distribution. We assume $x(\lambda)$ to be non-negative for every $\lambda \in [0, 1]$, and that for some $\lambda \in (0, 1]$, $x(\lambda)$ is strictly positive. The function $x(\cdot)$ is also assumed to be differentiable on its domain.

At $t = 0$, A and B decide simultaneously and independently whether to pay the given pair of ex-ante costs (c_A, c_B) which we assume to be affordable and strictly positive. If either agent does not pay the ex-ante costs the surplus is not available and the agents' payoffs are equal to minus any ex-ante costs paid. If, on the other hand, both pay the ex-ante costs, then A makes an offer λ to B , who may accept or reject it. If B accepts the offer, an agreement is successfully negotiated and A 's payoff is $\lambda x(\lambda) - c_A$, while B 's payoff is $(1 - \lambda)x(\lambda) - c_B$.

We start by identifying the socially efficient level of the distribution parameter λ .

Definition 4. *The socially efficient level of λ in the model we have just described is denoted λ^E , and is given by*

$$\lambda^E = \arg \max_{\lambda \in [0, 1]} x(\lambda) \quad (3)$$

Notice that, since we are assuming that the surplus is strictly positive for at least some λ , we also know that $x(\lambda^E) > 0$.

The analogue of Assumption 1 above which guarantees that it is socially efficient for the agents to negotiate an agreement is easy to state in our new model.

Assumption 4. Let $x(\lambda^E) > c_A + c_B$.

The value of the distribution parameter in any equilibrium in which an agreement is successfully negotiated is easy to characterize. Consider the subgame which begins after both agents have paid the ex-ante costs (c_A, c_B) . Then A will make an offer to B which maximizes his continuation payoff, subject to the constraint that it should be in B 's interest to accept the offer. In other words, A will offer to B a λ^* which solves

$$\max_{\lambda \in [0,1]} \lambda x(\lambda) \quad \text{s.t.} \quad (1 - \lambda)x(\lambda) \geq 0 \quad (4)$$

The solution to (4) is easy to find once we notice that the shadow price of the constraint in (4) is always zero because $x(\cdot)$ is non-negative. Moreover, since $x(\cdot)$ is strictly positive for some $\lambda \in (0, 1]$, it must be that $\lambda^*x(\lambda^*) > 0$. Using these facts and the first order conditions for (4) it is immediate that

Remark 3. If $x'(1) + x(1) > 0$ then $\lambda^* = 1$. Conversely, if $x'(1) + x(1) < 0$ then $\lambda^* \in (0, 1)$. Moreover, if $x'(1) + x(1) < 0$ then $x(\lambda^*) < x(\lambda^E)$.

We are now in a position to identify the range of ex-ante costs for which the fact that the size of the surplus depends on its distribution is sufficient to resolve the extreme hold-up problem described above.

We state Proposition 11 below without proof since it is an immediate consequence of Remark 3.

Proposition 11. Assume that $x'(1) + x(1) < 0$. Then there exists a range of ex-ante costs — namely $\mathcal{C}_\lambda = \{c_A, c_B \mid 0 < c_A < \lambda^* \text{ and } 0 < c_B < (1 - \lambda^*)\}$ — which satisfy Assumption 4 and such that the model has an equilibrium in which both agents pay the ex-ante costs and an agreement is successfully negotiated.

If A finds it profitable to ‘bribe’ B in order to increase the overall size of the surplus, it is possible that, in equilibrium, both agents will pay strictly positive ex-ante costs and negotiate an agreement. This is in contrast to the extreme form of

the hold-up problem described above that occurs when the size of the surplus is independent of its distribution and $\lambda = 1$.

A final observation is in order. The possible equilibria identified in Proposition 11 in which an agreement is successfully negotiated are still inefficient in the sense that $x(\lambda^*) < x(\lambda^E)$. Since A only appropriates a fraction of the surplus, it is not in his interest to offer B a λ which guarantees a global maximum of x .

7. Concluding Remarks

If the parties involved in a Coasian negotiation need to sink some ex-ante transaction costs before they can reach the negotiating phase of their interaction, the ex-ante costs may generate a version of the hold-up problem. If the distribution of ex-ante costs and the distribution of surplus generated by the negotiation are sufficiently ‘mismatched’, one of the two parties to the negotiation will not find it to his advantage to pay the ex-ante cost, even though the surplus generated by the agreement would be sufficient to cover the total ex-ante costs associated with it. Therefore, in equilibrium the agreement will not be negotiated. We have verified this claim in a variety of simple models. Moreover we showed that the hold-up problem we have identified is most acute when the agents have a continuous choice of ex-ante costs.

Unlike many other versions of this problem, under appropriate conditions, the hold-up problem generated by ex-ante transaction costs is unlikely to have a ‘Coasian solution’.²¹ This is because the Coasian negotiation that attempts to solve the hold-up problem is likely to generate a fresh set of ex-ante transaction costs and hence a new hold-up problem.

Lastly, we have explored a reduced form model in which the size of the surplus depends on its distribution. In this case it is apparent that the hold-up problem we identify is less pervasive than in the case in which the size of the surplus is not affected by its distribution.

²¹By this of course we mean a negotiated *contractual* solution. A large literature has examined the possible *institutional* solutions to the hold-up problem generated by contractual incompleteness (Grossman and Hart 1986, Hart and Moore 1990, Aghion and Tirole 1997, Rajan and Zingales 1998).

Appendix

Proof of Proposition 6: Since either $\lambda < c_A$ or $1 - \lambda < c_B$, it is clear that there is no equilibrium in which both agents pay the ex-ante cost at $t = 0$.

We only show that it is not possible that in any pure strategy equilibrium A alone pays the ex-ante cost at $t = 0$. Any equilibrium in which B alone pays the ex-ante cost at $t = 0$ can be ruled out in a symmetric way and we omit the details. Mixed strategy equilibria can be ruled out using standard arguments and we omit the details.

Suppose then that there is an equilibrium in which only A pays the ex-ante cost at $t = 0$. There are two cases to consider. Either B pays his cost to see A 's offer or he does not.

Suppose next that there is an equilibrium in which A only pays the ex-ante costs at $t = 0$ and subsequently B either accepts or rejects A 's offer without seeing it. Note that in this case B cannot condition his decision to accept or reject on the value of ℓ since he does not pay to see it. If B accepts in equilibrium, clearly A will set $\ell = -\epsilon$. But this would give an equilibrium payoff of $-\epsilon$ to B , and therefore yields a contradiction since B can always guarantee himself a payoff of zero by not paying any costs and rejecting any offer. If B rejects A 's offer blind in equilibrium, then A 's equilibrium payoff is $-c_A$ since no agreement is negotiated and A pays his ex-ante cost at $t = 0$. This is again a contradiction since A can guarantee himself a payoff of zero by not paying the ex-ante cost at $t = 0$ (and rejecting any offers made by B if he pays his ex-ante cost).

Lastly, consider the possibility of an equilibrium in which A alone pays the ex-ante costs at $t = 0$ and subsequently B pays his ex-ante cost c'_B to see the value of ℓ , and then accepts or rejects A 's offer. Notice that now B can condition his decision to accept or reject A 's offer on the actual value of ℓ . Using subgame perfection, it is immediate to see that, in equilibrium, it must be the case that B accepts all offers which guarantee that $1 - \ell > 0$ (his ex-ante cost is *sunk* when the accept/reject decision is made). Therefore, in equilibrium, A will offer precisely $\ell = 1$. It follows that in any equilibrium in which A alone pays the ex-ante cost at $t = 0$ and subsequently B pays to see A 's offer, B 's payoff is at most $-c'_B$. But this is a contradiction since B , as before, can guarantee himself a payoff of zero by not paying any costs and rejecting any offer. ■

Lemma A.1. *Consider the terminal subgame of the model with simple compensating transfers described in Subsection 4 which occurs after the pair of compensating transfers (σ_A, σ_B) has been agreed, as a function of the pair (σ_A, σ_B) . If the following inequalities are satisfied*

$$\lambda - \sigma_A + \sigma_B - c_A^0 \geq 0 \tag{A.1}$$

$$1 - \lambda + \sigma_A - \sigma_B - c_B^0 \geq 0 \tag{A.2}$$

the subgame has two equilibria. If both inequalities are strict one equilibrium strictly Pareto-dominates the other. The Pareto-superior equilibrium is such that both parties pay the ex-ante costs (c_A^0, c_B^0) and the agreement is successfully negotiated leaving the parties with continuation

payoffs $(\lambda - \sigma_A + \sigma_B - c_A^0)$ and $(1 - \lambda + \sigma_A - \sigma_B - c_B^0)$. The inferior equilibrium is such that both parties do not pay the ex-ante costs (c_A^0, c_B^0) and yields the no-agreement outcome. If either or both inequalities (A.1) and (A.2) are violated the terminal subgame has a unique equilibrium in which neither agent pays the ex-ante costs (c_A, c_B) , and hence yields the no-agreement outcome.

PROOF: The claim follows immediately from the fact that an agreement is feasible only if both A and B pay the ex-ante costs (c_A^0, c_B^0) , and from the observation that either agent i can guarantee a continuation payoff of zero by not paying his ex-ante cost c_i^0 . ■

Lemma A.2. *Consider the model with simple compensating transfers described in Section 4. If there exists an equilibrium of the model in which both $\sigma_A > 0$ and $\sigma_B > 0$, then there exists another, payoff equivalent, equilibrium of the model in which the transfers take the values $\tilde{\sigma}_A = \sigma_A - \sigma_B$ and $\tilde{\sigma}_B = 0$ if $\sigma_A \geq \sigma_B$, and $\tilde{\sigma}_A = 0$ and $\tilde{\sigma}_B = \sigma_B - \sigma_A$ if $\sigma_B \geq \sigma_A$.*

PROOF: We examine only the case in which $\sigma_A \geq \sigma_B$. The other case is a simple re-labelling of this one. To construct the new equilibrium, let the strategies of both agents be identical to the strategies in the original equilibrium, except for the way actions are conditioned on the other agents' compensating transfer offer. In the new equilibrium, each agent $i \in \{A, B\}$ responds to any offer $\tilde{\sigma}_j$ (with $j \neq i$) exactly as he would respond to the offer $\tilde{\sigma}_j + \sigma_i$ in the original equilibrium. ■

Proof of Proposition 7: Recall that both tiers of ex-ante costs are payable simultaneously by the agents and that an agreement (compensating transfer) is feasible only if both agents have paid the first (second) tier of ex-ante costs. Therefore it is obvious that a pair of strategies which prescribe not to pay any ex-ante costs for both agents (and some equilibrium behaviour off-the-equilibrium-path) constitutes an equilibrium. This proves our first claim.

We now move to the construction of a subgame perfect equilibrium of the model with simple compensating transfers in which the parties do negotiate an agreement.

We only deal with the case in which $1 - \lambda < c_B^0$. The case in which $\lambda < c_A^0$ is a simple re-labelling of this one and we omit the details.

Consider the subgame occurring after the transfers (σ_A, σ_B) have been agreed. If $\sigma_B \geq \sigma_A$ the only equilibrium of this subgame is such that both parties do not pay the ex-ante costs (c_A^2, c_B^2) . If instead $\sigma_A \geq \sigma_B$ then by Lemma A.2 we can restrict attention to transfers which satisfy $\sigma_A > 0$ and $\sigma_B = 0$.

If σ_A is such that inequalities (A.1) and

$$1 - \lambda + \sigma_A \geq c_B^0 + c_B^2 \tag{A.3}$$

are satisfied, we assume that the agents play the Pareto-superior of the two equilibria described in Lemma A.1 in which the agreement is successfully negotiated. If instead σ_A is such that inequality (A.3) is *not* satisfied while (A.1) and (A.2) are satisfied we assume that the agents play the Pareto-inferior of the two equilibria described in Lemma A.1 that yields the no-agreement outcome. In case either or both (A.1) and (A.2) are violated then the agents play the unique subgame perfect equilibrium of the subgame.

Proceeding backwards, it is then a best reply for B to accept any offer $\sigma_A > 0$ such that inequality (A.3) is satisfied. Indeed, if B rejects the offer his continuation payoff is zero while by accepting the offer his continuation payoff is non-negative.

It is then optimal for A to make an offer σ_A such that

$$\sigma_A = c_B^2 + c_B^0 - (1 - \lambda) \quad (\text{A.4})$$

This offer is associated with a positive continuation payoff for A . A higher offer is associated with a smaller continuation payoff while a lower offer is associated with a continuation payoff of zero, since the parties expect to play the inefficient equilibrium whenever (A.3) is violated.

Therefore, in equilibrium both parties pay the second tier ex-ante costs (c_A^2, c_B^2) . Paying the cost, B obtains the payoff of zero which coincides with the payoff he gets by not paying. By paying, A gets a strictly positive payoff while he gets a payoff of zero by not paying. This concludes the proof. ■

Proof of Proposition 8: We only deal with the case in which $1 - \lambda < c_B^0$. The case in which $\lambda < c_A^0$ is a simple re-labelling of this one and we omit the details.

Since we are assuming that the parameters of the model yield the no-agreement outcome in the final stage, any renegotiation-proof subgame perfect equilibrium which yields an agreement as an outcome must have both agents paying both tiers of ex-ante costs.

Assume by way of contradiction that such an equilibrium exists and denote by a superscript ‘*’ the equilibrium values of all variables in this equilibrium.

Notice first of all that if $\sigma_B^* \geq \sigma_A^*$ we have an immediate contradiction since in this case $\gamma_B^* > c_B$ and therefore B 's equilibrium payoff must be negative. Since B can guarantee a payoff of zero by not paying any of the ex-ante costs this is a contradiction.

By Lemma A.2, we can then assume without loss of generality that $\sigma_A^* > 0$ and $\sigma_B^* = 0$.

Next, consider the subgame that starts after the transfers (σ_A^*, σ_B^*) have been agreed. We now claim that every renegotiation-proof subgame perfect equilibrium must be such that

$$1 - \lambda + \sigma_A^* - c_B^0 = 0 \quad (\text{A.5})$$

To see this notice that Definition 2 and Lemma A.1 imply that every renegotiation-proof subgame perfect equilibrium must prescribe that in this subgame when

$$\lambda - \sigma_A - c_A^0 > 0 \quad (\text{A.6})$$

$$1 - \lambda + \sigma_A - c_B^0 > 0 \quad (\text{A.7})$$

are satisfied the parties play the Pareto superior equilibrium. This equilibrium involves both agents paying the costs (c_A^0, c_B^0) , negotiating an agreement and obtaining the strictly positive continuation payoffs: $1 - \lambda + \sigma_A - c_B^0$ and $\lambda - \sigma_A - c_A^0$. However, any offer σ_A which satisfies (A.7) cannot be payoff maximizing for A . Therefore the only renegotiation-proof subgame perfect equilibrium offer has to satisfy (A.5).

It follows directly from (A.5) that B 's payoff in this renegotiation-proof subgame perfect equilibrium would be $-c_B^2$. But this is a contradiction since B can guarantee a payoff of zero by not paying any ex-ante costs. This is enough to prove the proposition. ■

Proof of Proposition 9: The proof can be constructed in a way which is completely analogous to the one of Proposition 8 once we observe that when inequalities (A.6) and (A.7) are satisfied the unique subgame perfect equilibrium of the terminal subgame starting at $t = 0$ is for both parties to pay the costs (c_A^0, c_B^0) , negotiated an agreement and obtain payoffs: $1 - \lambda + \sigma_A - c_B^0$ and $\lambda - \sigma_A - c_A^0$.

Indeed, if A has paid his cost c_A^0 it is optimal for B to pay the cost c_B^0 as well given that B obtains a strictly positive continuation payoff by doing so, while he gets a continuation payoff of zero by not paying. On the other hand, if A does not pay his ex-ante cost c_A^0 then it is optimal for B not to pay his cost c_B^0 either. If B does not pay he gets a payoff of zero while if he does pay he gets a negative payoff. Therefore the unique subgame perfect equilibrium of this subgame is for both parties to pay their costs (c_A^0, c_B^0) and negotiate an agreement.

We omit the details of the remaining part of the proof. ■

Lemma A.3. *Let any value of the distribution parameter $\lambda \in [0, 1]$ be given. In equilibrium, in the model with a continuous choice of ex-ante costs, it is not possible for both agents to pay an inefficiently high level of ex-ante transaction costs in the sense that $c_i^* > c_i^E$ for all $i \in \{A, B\}$. This is true irrespective of the sign of the cross partial derivative $x_{A,B}(c_A, c_B)$.*

PROOF: Let $\beta = c_A^* - c_A^E$ and $\delta = c_B^* - c_B^E$. Therefore, as k varies in $[0, 1]$, $(c_A^E + k\beta, c_B^E + k\delta)$ describes the whole segment between (c_A^E, c_B^E) and (c_A^*, c_B^*) . Assume, by way of contradiction, that $\beta > 0$ and $\delta > 0$. The first order conditions (1) and (2) imply that $x_A(c_A^E + \beta, c_B^E + \delta) > x_A(c_A^E, c_B^E)$. Therefore for some k it must be the case that the derivative of $x_A(c_A^E + k\beta, c_B^E + k\delta)$ with respect to k is strictly greater than 0. In other words

$$x_{A,A} \beta + x_{A,B} \delta > 0 \quad (\text{A.8})$$

Since $\beta > 0$, (A.8) implies

$$x_{A,A}\beta^2 + x_{A,B}\delta\beta > 0. \quad (\text{A.9})$$

Proceeding in a completely symmetric way we also obtain that

$$x_{B,B}\delta + x_{B,A}\beta > 0 \quad (\text{A.10})$$

Since $\delta > 0$, (A.10) implies

$$x_{B,B}\delta^2 + x_{B,A}\delta\beta > 0 \quad (\text{A.11})$$

Inequalities (A.9) and (A.11) imply that

$$x_{A,A}\beta^2 + x_{B,B}\delta^2 + 2x_{A,B}\delta\beta > 0 \quad (\text{A.12})$$

which contradicts the fact that, by strict concavity the Jacobian of $x(\cdot, \cdot)$ must be negative definite. ■

Proof of Proposition 10: From Lemma A.3 we know that either $\beta \leq 0$ or $\delta \leq 0$. We only consider the case in which $\beta \leq 0$. The case in which $\delta \leq 0$ can be treated in a symmetric way, and we omit the details.

Assume that the proposition is false. Then $\beta \leq 0$ and $\delta \geq 0$. By concavity $x_{BB} < 0$, and by assumption $x_{AB} > 0$. Therefore

$$x_{BB}\delta + x_{AB}\beta \leq 0 \quad (\text{A.13})$$

Since (A.13) contradicts (A.10) this is enough to prove the claim. ■

References

- ABREU, D., D. PEARCE, AND E. STACCHETTI (1993): “Renegotiation and Symmetry in Repeated Games,” *Journal of Economic Theory*, 60, 217–40.
- AGHION, P., M. DEWATRIPONT, AND P. REY (1994): “Renegotiation Design with Unverifiable Information,” *Econometrica*, 62, 257–82.
- AGHION, P., AND J. TIROLE (1997): “Formal and Real Authority in Organizations,” *Journal of Political Economy*, 105, 1–29.
- ANDERLINI, L., AND L. FELLI (1996): “Costly Contingent Contracts,” Theoretical Economics Discussion Paper TE/96/313, STICERD, London School of Economics.

- (1998): “Costly Coasian Contracts,” Theoretical Economics Discussion Paper TE/98/362, STICERD, London School of Economics.
- (2001a): “Costly Bargaining and Renegotiation,” *Econometrica*, forthcoming.
- (2001b): “Costly Coasian Contracts: General Game Forms,” in preparation.
- BENOÎT, J.-P., AND V. KRISHNA (1993): “Renegotiation in Finitely Repeated Games,” *Econometrica*, 61, 303–323.
- COASE, R. H. (1960): “The Problem of Social Cost,” *The Journal of Law and Economics*, 3, 1–44.
- (1992): “The Institutional Structure of Production,” *American Economic Review*, 82, 713–19.
- DE MEZA, D. (1988): “Coase Theorem,” in *The New Palgrave Dictionary of Economics and the Law*, ed. by P. Newmann, pp. 270–82. Macmillan.
- DIXIT, A., AND M. OLSON (2000): “Does Voluntary Participation Undermine the Coase Theorem?,” *Journal of Public Economics*, 76, 309–335.
- FARRELL, J., AND E. MASKIN (1989): “Renegotiation in Repeated Games,” *Games and Economic Behavior*, 1, 327–360.
- FUDENBERG, D., AND J. TIROLE (1996): *Game Theory*. Cambridge, Massachusetts: M.I.T. Press.
- GROSSMAN, S. J., AND O. D. HART (1986): “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy*, 94, 691–719.
- GROUT, P. (1984): “Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Solution,” *Econometrica*, 52, 449–460.
- HART, O. D., AND J. MOORE (1988): “Incomplete Contracts and Renegotiation,” *Econometrica*, 56, 755–85.
- (1990): “Property Rights and the Nature of the Firm,” *Journal of Political Economy*, 98, 1119–58.

- HOLMSTRÖM, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340.
- MUTHOO, A. (1999): *Bargaining Theory with Applications*. Cambridge: Cambridge University Press.
- NICHOLSON, W. (1989): *Microeconomic Theory: Basic Principles and Extensions*. Chicago: The Dryden Press. Fourth International Edition.
- NÖLDEKE, G., AND K. M. SCHMIDT (1995): "Option Contracts and Renegotiation: A Solution to the Hold-Up Problem," *RAND Journal of Economics*, 26, 163–179.
- RAJAN, R., AND L. ZINGALES (1998): "Power in a Theory of the Firm," *Quarterly Journal of Economics*, 113, 387–432.
- RUBINSTEIN, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97–110.
- RUBINSTEIN, A., AND A. WOLINSKY (1992): "Renegotiation-Proof Implementation and Time Preferences," *American Economic Review*, 82, 600–614.