Contract Design and Non-Cooperative Renegotiation

Robert Evans[1] and Sönje Reiche[2]

This version: November 2014

Abstract

We study a contract design setting in which the contracting parties cannot commit not to renegotiate previous contract agreements. In particular, we characterize the outcome functions that are implementable for an uninformed principal and an informed agent if, having observed the agent's contract choice, the principal can offer a new menu of contracts in its place. An outcome function can be implemented in this setting if and only if it is optimal for the principal for some belief over agent types which is more pessimistic, in the sense of the likelihood ratio order, than the prior. Furthermore, the outcome function cannot be too sensitive to variations in the agent's type. We show that the direct revelation mechanism which implements such a function when renegotiation can be prevented will also implement it in any equilibrium when it cannot, so the standard contract is robust to renegotiation.

Keywords: Renegotiation, Mechanism Design, Contract, Commitment.

JEL Codes: D820, D860.

[1]Corresponding Author. University of Cambridge and St John's College, Cambridge, UK. Address: Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD, UK; robert.evans@econ.cam.ac.uk; +44 (0)1223 338784.

[2]University of Cambridge, UK, and Toulouse School of Economics (CNRS,GREMAQ); Address: Faculty of Economics, University of Cambridge, Sidgwick Avenue, CB3 9DD, UK; skr29@cam.ac.uk.

1 Introduction

One of the main messages of the literature on contract renegotiation is that an inability of contracting parties to commit themselves not to renegotiate hurts those parties because it limits the use of ex ante contracts and prevents desirable outcomes from being implementable.

A leading example is the hold-up problem, in which investment that increases the expected benefit of a relationship is not undertaken because the investing party fears expropriation of the investment benefits by its partner. An ex ante contract specifying the division of ex post surplus between contracting parties can serve to alleviate this problem but its renegotiation will ultimately affect this division and hence damage investment incentives. This problem has been studied in various settings (for example, with symmetric information, asymmetric information, selfish investment, cooperative investment) and has been shown in most cases to limit the scope of contracting and decrease the level of investment; see for instance Segal (1999), Maskin and Moore (1999), Che and Hausch (1999), Reiche (2006) and Goltsman (2011).

Renegotiation can also be harmful in situations in which a trading opportunity is repeated several times and in which parties cannot commit not to renegotiate future trade agreements; see for instance Dewatripont and Maskin(1990), Hart and Tirole (1988) and Laffont and Tirole (1988, 1990). In this context, the ratchet effect implies that parties tend to understate the value of trade in order to avoid more demanding schedules (for example, a higher price) in the future. Spot contracts or long-term contracts which are vulnerable to renegotiation tend to be less efficient in solving the asymmetric information problem between trading partners than long-term contracts which cannot be renegotiated.

We consider a standard contracting problem between an uninformed principal and a privately informed agent, and ask which outcome functions (mappings from the agent's private information into some action and transfer payment) can be implemented when parties cannot commit not to renegotiate their contract. More precisely, we suppose that after the agent has played the initial mechanism, determining a default outcome, the principal can offer a second-stage mechanism which, if the agent accepts it, determines the actual outcome. This second mechanism will of course depend on what the principal has learned from her interaction with the agent in the initial mechanism and we can consequently not assume that the agent's initial message fully reveals his type. That is, the standard revelation principle does not apply.

Most of the literature on contract renegotiation is concerned only with implementing the outcome function which is optimal for the principal at the stage when the mechanism is played (see, for example, Skreta (2006) or Bester and Strausz (2001)). In our setting this outcome function is simple to implement, as is the ex post efficient one. However, in many contexts (such as the hold-up context mentioned above) this is not the best outcome function to implement. Our aim is to characterize all outcome functions that can be implemented subject only to the agent's incentive and participation constraints and the constraint of subsequent renegotiation. What outcome functions are optimal will depend on the particular circumstances; for instance, it will depend on who designs the initial mechanism, whether the principal or an outside agency such as a social planner.

If the designer is the principal she might, as in the hold-up problem, want to propose ex ante a mechanism to improve investment incentives. This will not, in general, be the same as the one which is optimal for her once investment is undertaken and the state of the world is realized. At this point the initial purpose of the mechanism is served and the parties will have an incentive to renegotiate the existing contract. In the next Section we outline our results in the context of a specific example of this kind. Alternatively, the principal might want to propose a mechanism to attract a specific pool of agents, so that the initial mechanism has to satisfy some particular set of participation constraints. If, once an agent is locked in with the principal, those participation constraints have changed, the principal has an incentive to offer a different contract at the second stage. In both cases, the principal's optimal ex ante mechanism taking these considerations into account may not be the same one which she would wish to offer ex post.

If the designer is an outside agency he may be, for example, a regulator, or a higher level of authority in the organization to which the principal belongs, or the designer of a trading platform or a market where sellers (principals) and buyers (agents) who do not know each other are matched. In each of these cases, the designer may have an objective function which differs from those of the players, though the arguments of the function may include the principal's expected payoff and/or the distribution of utilities and decisions across the various types of agent.

The above considerations imply that it is desirable to know which outcome functions can be implemented when renegotiation is taken into account. Our first result says that any r-implementable (i.e., implementable with renegotiation) outcome function must satisfy a simple renegotiation-invariance property, which can be interpreted as a modification of the revelation principle. Namely, if an outcome is r-implementable with some initial mechanism it can also be achieved by giving the parties, at the outset, the same direct revelation mechanism which would implement it in the no-renegotiation case. Moreover, after each announcement by the agent, the principal, in equilibrium, offers the same direct revelation mechanism again. In the no-renegotiation case, the agent would tell the truth in this initial mechanism but, because of the ratchet effect, this is not so in the renegotiation case. Instead, the agent will understate his type, i.e., randomize over announcements of types below his true type. The principal, after a particular announcement $\hat{\theta}$, say, will therefore have post-announcement beliefs that are distributed over types $\hat{\theta}$ and above, and these beliefs are such that the initial direct revelation mechanism now becomes optimal for her. In this second stage mechanism the agent will then tell the truth and obtain the outcome intended for his type. We also show that this equilibrium is unique.

Because the agent always understates his true type in the initial mechanism, the principal's possible beliefs at the renegotiation stage are, in a particular sense, more pessimistic than her prior beliefs. Our second main result says that an outcome function is r-implementable if and only if it is optimal, for the principal, for some distribution which is lower than her prior in the likelihood ratio order.

Since the principal's beliefs are related to the outcome function through the first-order condition of her maximization problem at renegotiation, we can express this result in terms of the slope of the decision function. Since we assume supermodularity of the agent's payoff function, the decision function, as in the no-renegotiation case, cannot be decreasing. We show that, in addition, it cannot be too steep: our third main result is that an outcome function is r-implementable if and only if it does not vary too much with the agent's type, in the sense that the slope of the decision function must be below a certain bound.

In summary, for a large class of decision rules, the standard incentive-compatible mechanism has a strong renegotiation-invariance property - after any message, the principal always wants to offer the initial mechanism again. The designer does not have to be concerned about whether renegotiation might be possible - the same mechanism delivers the desired outcome for every type whether it is possible or not. Our third result can easily be used to verify whether or not, in a particular applied setting, renegotiation poses a problem because the desired decision rule falls outside the above class. Another appealing feature is that an outside designer wishing to implement a particular outcome function does not need to know the principal's prior distribution over the agent's types, only that this distribution is above a certain lower bound in the likelihood ratio order.

The results apply in addition to the case of interim, as opposed to ex post, renegotiation and also to a model in which the renegotiation has finitely many stages. Our analysis does assume that the renegotiation bargaining game is finite. This could be because there is an exogenous deadline such as may arise in many contexts (see, e.g., Fuchs and Skrzypacz (2013)): for example, in a buyer-seller model, the good might be perishable or needed as an input into a production process which cannot be delayed. Alternatively, the agent may face financial constraints which vary over time. Another possibility is that a third party designs the mechanism and, while the principal is able to commit to a mechanism (as is generally assumed in the principal-agent literature) the designer cannot fully commit the principal. We also briefly discuss in subsection 4.5 below possible extensions to a model of infinite-horizon renegotiation with discounting.

Literature

Ex post renegotiation has been studied by Green and Laffont (1987), Forges (1994), and Neeman and Pavlov (2013). In these contributions the concepts employed are variations on the principle that a mechanism is (ex post) renegotiation-proof if, for any outcome x of the mechanism and any alternative outcome y, the players would not vote unanimously for y in preference to x if a neutral third party were to propose it to them. Such definitions of renegotiation-proofness have the merit that, if a given mechanism satisfies it, the mechanism is robust against all possible alternative outcomes. However, it also has the drawback that the implied renegotiation process does not have a non-cooperative character. In contrast, we assume that the renegotiation process is given by an explicit ex post bargaining game, a simple take-it-or-leave it offer by the uninformed party. This one-shot model of renegotiation is close to the one generally used for mechanism design with complete information (Maskin and Moore (1999), Segal and Whinston (2002)), in which, for any inefficient outcome of the mechanism, there is a single renegotiation outcome, which can be predicted by the players.[3]

A recent strand of the literature on the Coase Conjecture (see Strulovici (2014) and Maestri (2013)) is concerned with contract negotiations with limited commitment in which contracts are (re)negotiated using infinite-horizon protocols with frictions. As those frictions vanish the essentially unique equilibrium involves only efficient contracts (see also Beaudry and Poitevin (1995) and Goltsman (2011)). Our focus is different in that we are concerned with what can be achieved when there are non-negligible frictions. Our results concern the case in which there is an exogenous deadline, but, as we discuss in subsection 4.5, we conjecture that a version of our results would apply in an infinite-horizon model with (non-negligible) discounting.

None of the contributions in the literature, to our knowledge, has shown the renegotiation-invariance property of standard incentive-compatible mechanisms, or derived our results about the relation between the prior and the possible post-renegotiation

[3]Rubinstein and Wolinsky (1992) model renegotiation as costly because it involves delay and show that the set of implementable outcomes in a complete information buyer-seller model is larger than those of the standard model of implementation with renegotiation.

beliefs.

Finally, our analysis is indirectly related to the literature on incomplete information bargaining beginning with Fudenberg and Tirole (1983). One interpretation of a mechanism is that it is a device for understanding what can be achieved by non-cooperative bargaining games; see for instance Ausubel and Deneckere (1989a) and (1989b). In contrast to these papers, we consider what a mechanism can achieve when it is played before parties enter into such a bargaining game. In addition, we consider general revelation mechanisms and not simply price offers. It is also related to recent work on organizational theory, stemming from Crawford and Sobel (1982). In Krishna and Morgan (2008) the uninformed decision maker can commit to a contract which pays the informed sender a monetary transfer which depends on the message sent, but cannot commit to the action which she then takes. In our setting the sender is the agent and the decision maker is the principal, who can only partially commit to her action (the renegotiation mechanism). See also Ottaviani (2000) for a model with informed senders, monetary transfers and lack of commitment by the receiver.

Outline

Section 2 contains several examples to motivate and demonstrate our analysis. Section 3 sets out the model formally. Section 4 contains the analysis and results. Subsection 4.1 proves the renegotiation-invariance principle, which is helpful in deriving the necessary and sufficient conditions. Subsection 4.2 derives necessary conditions for implementation. Subsection 4.3 provides sufficient conditions and a discussion of the strong implementation (uniqueness) result. Subsection 4.4 discusses the special case in which utility is linear. Subsection 4.5 contains a discussion of several applications and our main assumptions. Some of the proofs are in the Appendix.

2 Examples

In this section we outline one setting to which our analysis applies and use some simple examples to illustrate our main arguments and results. The setting features ex ante investment and ex post hold-up. The principal is S, the seller of a good, and the agent is B, a potential buyer with ex ante uncertain valuation. B can undertake an unobservable investment that raises his expected valuation but, without an ex ante contract, he will fail to do so because of the hold-up threat.

B's investment costs I. If he does not invest his type is distributed according to F^0 on $\Theta^0 = [\underline{\theta}^0, \overline{\theta}^0]$, and, if he does, it is distributed according to F^1 on $\Theta^1 = [\underline{\theta}^1, \overline{\theta}^1]$, where $\underline{\theta}^0 \leq \underline{\theta}^1 \leq \overline{\theta}^0 \leq \overline{\theta}^1$ and F^1 first-order stochastically dominates F^0. Both distributions satisfy the increasing hazard rate condition. If B is of type θ and buys x units of the good for payment t then his payoff is $\theta u(x) - t$ and S's payoff is $t - cx$, where $u' > 0$, $u'' < 0$, and $\lim_{x \to 0} u'(x) = \infty$. θ, once realized, is B's private information.

Contracting Ex Post

Assume first that there is no ex ante contract. Then S offers an incentive compatible and individually rational mechanism ex post (i.e., after B's type is realized) that maximizes her expected payoff, where her expectation depends on whether she believes B to have invested or not. If S believes that B has invested she will offer a quantity schedule $x^{F^1}(\cdot)$ that pointwise maximizes the virtual surplus

$$\left(\theta - \frac{1 - F^1(\theta)}{f^1(\theta)}\right) u(x(\theta)) - cx(\theta), \tag{1}$$

together with associated utilities $U^1(\theta) = \int_{\underline{\theta}^1}^{\theta} u(x^{F^1}(s))ds$. Of course, B will only invest if his expected utility gain from investment justifies the investment cost.

Contracting Ex Ante

If the investment cost is higher than B's expected gain, given the ex-post contract, and if S wants B to invest, then she will need to design an ex ante contract which takes investment incentives into account. We assume initially that S can commit not to renegotiate. Then she will offer, before the investment stage, an incentive-compatible and interim individually rational contract,[4] $(x^I(\cdot), U^I(\cdot))$, that maximizes S's expected payoff (for belief F^1) subject to the investment constraint

$$\int_{\underline{\theta}^1}^{\overline{\theta}^1} U(\theta)(dF^1 - dF^0) \geq I. \tag{2}$$

One can show that $x^I(\cdot)$ pointwise maximizes "investment adjusted" virtual surplus

$$\left(\theta - \frac{1 - F^1(\theta)}{f^1(\theta)} + \mu \frac{F^0(\theta) - F^1(\theta)}{f^1(\theta)}\right) u(x(\theta)) - cx(\theta), \tag{3}$$

where $\mu \geq 0$ is the Lagrange multiplier associated with (2). Since F^1 first-order stochastically dominates F^0, the term multiplying μ is positive, so $x^I(\theta) \geq x^{F^1}(\theta)$.

Example A Suppose that $u(x) = \sqrt{x}$, $c = \frac{1}{2}$, F^0 is uniform on $\Theta^0 = [1, 3]$, and F^1 is uniform on $\Theta^1 = [2, 3]$. The first-best is $x^*(\theta) = \theta^2$.

Then $x^{F^1}(\theta) = (2\theta - 3)^2$ and $U^1(\theta) = \theta^2 - 3\theta + 2$. Since $\int_{\underline{\theta}^1}^{\overline{\theta}^1} U^1(\theta)(dF^1 - dF^0) = 5/12$, if $I > 5/12$ B will not invest without an ex ante contract. Assume that $I = 0.5$. Then, using (3) and the binding investment constraint (2),

$$x^I(\theta) = \frac{9}{4}(\theta - 1)^2. \tag{4}$$

This is illustrated in Figure 1 below. It can be checked that inducing investment is optimal for S.

The above shows that there exist contexts in which the principal would like to implement an outcome function which is neither efficient nor ex post optimal (in Section 4.5 we discuss other possible contexts for which this is the case). Furthermore she would like to commit herself at an early stage to a contract which might subsequently not be optimal. Our main focus, however, is on the case in which such commitment is not possible.

[4] We maintain the assumption that the ex ante mechanism has to satisfy the participation constraint for each type of B: S cannot insist that B accepts the mechanism before learning his type, for example because she is facing a population of anonymous buyers or because she cannot observe the timing of the realization of B's type.

Renegotiation

Suppose that S has publicly announced, ex ante, the contract associated with (x^I, U^I) for Example A with $I = 0.5$ (a menu of quantity-transfer pairs $\{(x,t)\}$ which we can take to be equivalent to a direct revelation mechanism) and the investment stage has passed. If B now places an order (x,t) this defines the guaranteed reservation outcome. However, in the light of what S has learned about B from his order, S may now prefer a different outcome, which B may also prefer. In that case we would expect some renegotiation to take place. Suppose then that S, after[5] seeing B's choice, can offer a new contract (mechanism) and B can either choose to play the new mechanism or else stick with the reservation outcome (x,t). The question is: what initial contract should S propose if she wants to implement (x^I, U^I), bearing in mind that she should expect the contract to be renegotiated?

We show in the following sections that this three-stage game of incomplete information, beginning just before B sends a message in the ex ante mechanism (x^I, U^I), has a unique equilibrium outcome. Each type of B initially randomizes over all messages up to and including his true type; after any message S offers the initial (ex ante) mechanism again; and B accepts the second mechanism and tells the truth. In other words, B's initial randomization has the effect that S's beliefs always change in such a way that her ex ante contract becomes ex post optimal. Therefore S does not need to be concerned with the fact that renegotiation will take place - she can design the initial contract exactly as if she could commit to it.

The equilibrium works as follows in Example A. Type $\theta \in [\underline{\theta}, \bar{\theta}]$ puts probability $(3-\theta)(3-\hat{\theta})^{-1}$ on messages in the interval $[2, \hat{\theta}]$ (and zero probability on messages $\hat{\theta} > \theta$), for any $\hat{\theta} \leq \theta$. This implies that, given a message $\hat{\theta}$, S knows that B's type is $\hat{\theta}$ or higher and her belief about such types, $G_{\hat{\theta}}(\theta)$, has, by Bayes' Rule, a density of the form $h(\hat{\theta})(3-\theta)$. Therefore these conditional densities (varying the message $\hat{\theta}$)

[5] Our main focus below is on ex post renegotiation. Renegotiation could alternatively take place at the interim stage, i.e. before B plays the pre-announced mechanism but after the type is realized, S could offer a new mechanism to B. As we discuss below (Corollary 1), our results would also apply to this case.

are all scaled versions of each other, the scalar being $h(\hat{\theta})$. Moreover, for any $\hat{\theta} \leq \theta$,

$$\frac{1 - G_{\hat{\theta}}(\theta)}{g_{\hat{\theta}}(\theta)} = \frac{\int_{\theta}^{3} h(\hat{\theta})(3-u)du}{h(\hat{\theta})(3-\theta)} = \frac{(3-\theta)}{2}. \tag{5}$$

This in turn implies, using (1), that if S has belief $G_{\hat{\theta}}(\theta)$ her optimal quantity schedule is $x^I = (9/4)(\theta - 1)^2$. Furthermore, she will want to give utility $U^I(\hat{\theta})$ to type $\hat{\theta}$ since she regards this type as the lowest possible one, and $U^I(\hat{\theta})$ is its reservation utility after announcing $\hat{\theta}$ (telling the truth) in the original mechanism. This shows that, after any message, the optimal ex post contract[6] for S is (x^I, U^I). After sending an initial message $\hat{\theta} < \theta$, type θ will then, in the second mechanism, recontract to $(x^I(\theta), t^I(\theta))$. Finally, B is happy to mix in the way described because he is indifferent between all messages.

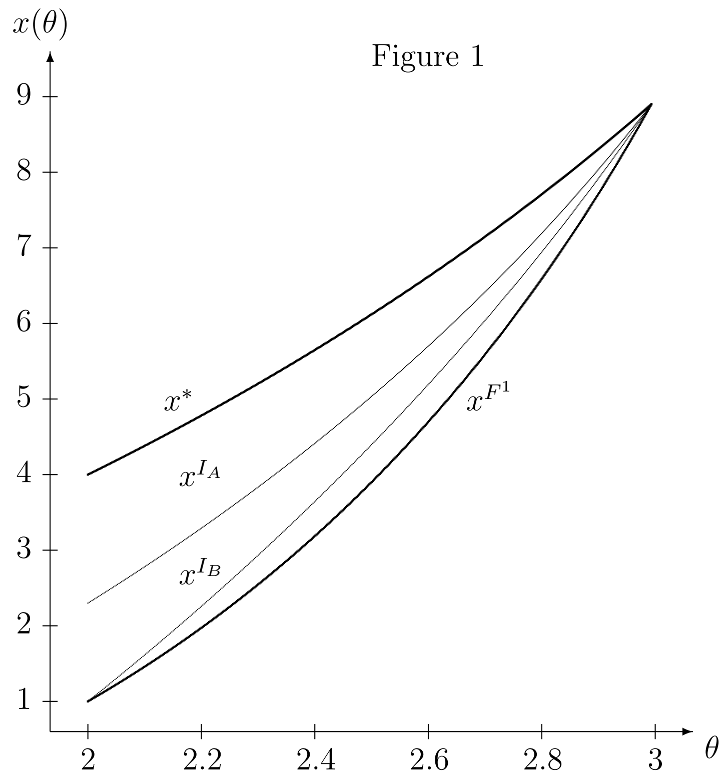
In Example A, therefore, S's optimal ex ante contract can be implemented despite renegotiation. This is not always the case. If a contract can be implemented in our setting then, without loss of generality, B stochastically understates his type and the ex ante mechanism must be optimal for S after any message (Proposition 1). This implies that S's post-initial-message belief must be smaller than her prior F^1 in the likelihood-ratio order (Proposition 2). Consequently, there is an upper bound on the rate at which the quantity schedule x can increase (Proposition 3). The intuition for this is as follows. S's post-message belief must be invariant to the message, up to scaling, so each type of S must randomize in a proportionally similar way (i.e. two types who both send messages $\hat{\theta}$ and $\hat{\theta}'$ must both weight them in the same ratio, $(3 - \hat{\theta}')^2(3 - \hat{\theta})^{-2}$ in our example). This means that, given any message $\hat{\theta}$, and a uniform prior, S's posterior density must be declining; she places more weight on low types than on high ones because high ones randomize over a larger set of messages. In this sense, B is conditionally pessimistic: although a given message implies that B's type is above a certain threshold, above that threshold her belief is shifted downwards compared with the uniform prior. Therefore the slope of S's optimal quantity schedule is reduced: essentially, if low types are relatively more likely, she will want to induce a

[6] There is no loss of generality in assuming that S offers the whole mechanism, including those outcomes intended for types below $\hat{\theta}$. See footnote 11 below.

relatively lower degree of inefficiency for those types (higher x) compared with types just above.

Example B For an example in which S's ex ante optimal contract cannot be implemented with renegotiation, suppose that everything is as in Example A except that F^0 is some cumulative distribution with support $\Theta^0 = [2,3]$, and is first-order stochastically dominated by F^1, which, as before, is uniform on $\Theta^1 = [2,3]$. Since $F^0(2) = F^1(2)$, (3) implies that $x^I(2) = x^{F^1}(2)$. For some higher values of θ the ex ante optimal schedule x^I will be strictly greater than x^{F^1}, the optimum for belief F^1. However, this implies that x^I must somewhere increase faster than x^{F^1} which, as just argued, is incompatible with the form of our equilibrium. Hence x^I cannot be implemented with renegotiation.

In summary, an ex ante contract can be implemented with renegotiation if and only if it is optimal for some possible post-message beliefs of S. However, because B always understates his true value the possible post-message beliefs of S are those which are more pessimistic, that is, lower in the likelihood ratio order, than the prior (see Corollary 2). Fig. 1 shows x^*, x^{F^1} and x^I (labelled x^{I_A} and x^{I_B} respectively for Examples A and B).



3 The Model

A principal (P) and an agent (A) must choose a decision x from the set $[\underline{x}, \bar{x}] \subseteq \Re_+$, and a money transfer t. The agent has a privately known type θ which follows a distribution F, with differentiable density $f > 0$, on the interval $\Theta = [\underline{\theta}, \bar{\theta}]$, where $\underline{\theta} > 0$. In addition, F satisfies the increasing-hazard-rate condition. Both players are expected utility maximizers and have quasi-linear utility for money. If the decision is x and A transfers t to P, then P's payoff is $t - cx$, where $c > 0$, and A's payoff is $u(x, \theta) - t$, where u is a thrice-differentiable function satisfying the conditions $u_x > 0, u_{xx} < 0, u_{x\theta} > 0$, with subscripts denoting derivatives. We make two further assumptions, (a) either $u_{xx\theta} \geq 0$ or $u(x, \theta) = \theta u(x)$, and (b) $\partial(u_{\theta xx}/u_{\theta x})/\partial\theta \geq 0$. Assumption (a) together with the increasing-hazard rate condition for F guarantee that the second-best decision rule x^F is unique and non decreasing.[7] We discuss the use of Assumption (b) at the end of this section. Finally, we make the assumption that $u_x(\underline{x}, \underline{\theta}) > c > u_x(\bar{x}, \bar{\theta})$, which guarantees that, for each type, the ex post efficient decision is interior.

The choice of decision and transfer is governed by a mechanism γ, i.e., a triple (M, x_M, t_M) consisting of a set of messages M, where M is a metric space, and a pair of functions $x_M : M \to [\underline{x}, \bar{x}]$ and $t_M : M \to \Re$. A chooses a message $m \in M$. When message m is sent, $x_M(m)$ is the contracted decision and $t_M(m)$ is the contracted payment to be paid by A to P. We assume throughout that communication is direct (there is no mediator).[8] Denote the set of possible mechanisms by Γ. The mechanism might be chosen either by P as in our hold-up example or by a third party. In Section 4.5 we discuss applications involving third parties to which our analysis can be applied. The reservation utility for each type of A is zero. In the case where there is a third party who designs the initial mechanism we do not model the contracting game and we therefore do not consider P's reservation utility explicitly. Such an analysis would have to include this as an additional constraint.

The parties are not able to commit not to renegotiate the mechanism. We assume that at the renegotiation stage, after the play of the mechanism, all of the bargaining power lies with the principal, the uninformed party.[9] In other words, once the outcome of the initial mechanism, (x, t), is known, the principal chooses a mechanism to offer to the agent. A can either play this new mechanism or obtain the outcome (x, t). Our aim is to characterize the set of outcome functions and corresponding utility schedules which can be implemented by some mechanism taking into account the fact that the mechanism can be renegotiated ex post.

We restrict attention to non-stochastic mechanisms throughout. Assumptions (a) and (b) are sufficient for P's optimal contract offer at renegotiation to be non-stochastic (see Proposition 8 of Jullien (2000)). They also imply that the optimal

[7]See for instance Fudenberg and Tirole (1991), p.263.

[8]For an analysis of contracting with renegotiation and mediated communication, see Goltsman (2011) and Bester and Strausz (2007)

[9]If the agent had the bargaining power results analogous to ours would trivially hold.

ex ante contract in the no-renegotiation case is deterministic in our hold-up example above. There exist mechanism design problems[10] for which the optimal contract is stochastic; however, we do not consider these.

Strategies and Equilibrium

An initial mechanism (M, x_M, t_M) and the post-mechanism stage together define a three-stage game of incomplete information: A sends a message; then P, after observing the message, offers a new mechanism; finally A plays the second mechanism or chooses the default outcome resulting from his message. Call this game $\Phi(M, x_M, t_M)$. We will consider the perfect Bayesian equilibria of this game.

Given an outcome (x, t) of the initial mechanism, and a mechanism $\gamma \in \Gamma$ offered by P, A either chooses the default outcome (x, t) or plays the mechanism γ. In a perfect Bayesian equilibrium A will choose optimally given his type, i.e., will either play the mechanism optimally or, if the default gives a higher payoff, choose the latter. Given her belief, P will, at the preceding stage (i.e., after an initial message), choose a mechanism to offer to A which is optimal for P. We impose a regularity condition on the possible equilibrium post-message beliefs of P: they must lie in $\Delta^1(\Theta)$, the set of distribution functions on Θ which have density functions except possibly at a finite set of jump points (to our knowledge, the literature has not established the nature of the optimal contract for other distributions).

Let $D_{IC}(x, t)$ be the set of incentive-compatible direct revelation mechanisms which dominate the default outcome (x, t) for all types, i.e., mechanisms $(\Theta, x_\Theta, t_\Theta) \in \Gamma$ such that, for all $\theta, \theta' \in \Theta$,

$$u(x_\Theta(\theta), \theta) - t_\Theta(\theta) \geq u(x_\Theta(\theta'), \theta) - t_\Theta(\theta')$$

and

$$u(x_\Theta(\theta), \theta) - t_\Theta(\theta) \geq u(x, \theta) - t.$$

It is straightforward to show, by a revelation principle argument, that we can

[10] See for example Maskin and Riley (1984) and Strausz (2006).

assume without loss of generality that P chooses a mechanism in $D_{IC}(x,t)$ and that, for all $\theta \in \Theta$, type θ of A accepts the mechanism and tells the truth.[11]

Given the above, we can take a pure strategy for P in $\Phi(M, x_M, t_M)$ to be a function $s_P : M \to \Gamma$ such that, for $m \in M$, $s_P(m) \in D_{IC}(x_M(m), t_M(m))$. We only consider equilibria in which P's strategy is pure.[12]

Similarly, we can take a pure strategy for A in $\Phi(M, x_M, t_M)$ to be a function which maps Θ to M. We take a mixed strategy for A to specify a mixed strategy for each type of A, where a mixed strategy[13] for type θ of A is a probability measure $s_A(.|\theta)$ on M.

If P's strategy is s_P and A is type $\theta \in \Theta$ and sends $m \in M$, let the post-renegotiation decision and transfer be denoted by $x(m, s_P, \theta)$ and $t(m, s_P, \theta)$; that is, the mechanism $s_P(m)$ gives this outcome when the agent tells the truth.

Definition 1: A *renegotiation equilibrium* (or *r-equilibrium*) of $\Phi(M, x_M, t_M)$ is a profile of strategies (s_A, s_P), and, for each $m \in M$, a belief $G_m \in \Delta^1(\Theta)$, such that

(i) for each $\theta \in \Theta$, $s_A(\cdot|\theta)$ puts probability 1 on messages m which maximize

$$u(x(m, s_P, \theta), \theta) - t(m, s_P, \theta);$$

(ii) for each $m \in M$, $s_P(m)$ solves

$$max_{(\Theta, x_\Theta, t_\Theta) \in D_{IC}(x(m), t(m))} \int_{\underline{\theta}}^{\bar{\theta}} t_\Theta(\theta) - c x_\Theta(\theta) dG_m(\theta);$$

[11] The standard argument would imply that P can offer a mechanism in $D_{IC}(x,t)$, but restricted to types in the support of her belief. If instead she offers this kind of mechanism for the entire type space, then types in the support of her belief will not want to choose any of the "extra" options, so this makes no difference to her. It is convenient to assume that she offers a mechanism defined on all types in Θ because, as we show below, this implies that in equilibrium she will offer the original mechanism again, rather than offering the original mechanism with gaps in the type space.

[12] Assumptions (a) and (b) imply that P's optimal quantity schedule is unique (see Theorem 4 of Jullien (2000)), and so P never wants to mix over different direct revelation mechanisms at renegotiation.

[13] It is possible to define a continuum of mixed strategies over M via a distributional strategy as in Milgrom and Weber (1985), i.e., a joint distribution on $M \times \Theta$ for which the marginal on Θ corresponds to the prior F. $s_A(.|\theta)$ is then the measure on M conditional on θ. See also Crawford and Sobel (1982).

(iii) for each $m \in supp(s_A) \equiv \bigcup_{\theta\in\Theta} supp(s_A(.|\theta))$, G_m is consistent with Bayes' Rule, given prior belief F and strategy s_A.

Slightly abusing notation, we denote by $(x(s_A, s_P, \theta), t(s_A, s_P, \theta))$ the final outcome, given θ, if the strategy profile is (s_A, s_P). This may be stochastic if s_A is mixed.

Definition 2: (a) A function $(X, T) : \Theta \to [\underline{x}, \bar{x}] \times \Re$ is a *r-implementable outcome function* if (i) $U(\theta) \equiv u(X(\theta), \theta) - T(\theta) \geq 0$ for all $\theta \in \Theta$, and (ii) there exists a mechanism (M, x_M, t_M) such that, for all $\theta \in \Theta$, $x(s_A, s_P, \theta) = X(\theta)$ and $t(s_A, s_P, \theta) = T(\theta)$ with probability 1 for some r-equilibrium $(s_A, s_P, \{G_m\}_{m\in M})$ of $\Phi(M, x_M, t_M)$.

(b) (X, T) is *strongly r-implementable* if, in addition, (X, T) is the outcome of all r-equilibria of $\Phi(M, x_M, t_M)$.

A utility schedule $U : \Theta \to \Re$ is (strongly) r-implementable if $U(\theta) = u(X(\theta), \theta) - T(\theta)$ for some (strongly) r-implementable (X, T). The fact that the utility schedule U must be non-negative reflects the fact that A's outside utility has been normalized to zero and we allow him not to participate in the mechanism.

We refer to an outcome function (and associated utility schedule) as *c-implementable* if it can be implemented in the case in which the players can be committed not to renegotiate the mechanism. By standard results (see Fudenberg and Tirole (1991), Milgrom and Segal (2002)) (X, T) is c-implementable if and only if X is non-decreasing, and for all $\theta \in \Theta$, $U(\theta) - U(\underline{\theta}) = \int_{\underline{\theta}}^{\theta} u_\theta(X(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}$, and $U(\underline{\theta}) \geq 0$.[14]

Remark It is easy to show, using revelation principle arguments, that if (X, T) (resp. U) is r-implementable then (X, T) (resp. U) is c-implementable.

The first-best decision for θ solves the problem $max_{x\in[\underline{x},\bar{x}]} u(x, \theta) - cx$. By our assumptions this has a unique solution which we denote by $x^*(\theta)$. Furthermore, x^* is strictly increasing in θ. We assume that $u(x^*(\theta), \theta) - cx^*(\theta) > 0$ for all θ so that

[14] A c-implementable U is absolutely continuous and a.e. differentiable.

there is strictly positive surplus for each type.

For future reference, we include two standard definitions of orderings of probability distributions (see Shaked and Shanthikumar (1994)).

Definition 3: Given two distribution functions F and G, with density functions f and g respectively, G is smaller than F in the likelihood ratio ordering, denoted $G \preceq_{LR} F$, if, for all $\theta_1 \leq \theta_2$,

$$g(\theta_2)f(\theta_1) \leq g(\theta_1)f(\theta_2).$$

In the case in which f and g are both differentiable and $f, g > 0$, this corresponds to the condition that, for all θ,

$$\frac{g'(\theta)}{g(\theta)} \leq \frac{f'(\theta)}{f(\theta)},$$

i.e., that the proportional rate of growth of g is always less than that of f.

Definition 4: Given two distribution functions F and G, with density functions $f > 0$ and $g > 0$ respectively, G is smaller than F in the hazard rate ordering, denoted $G \preceq_{HR} F$, if

$$\frac{1 - G(\theta)}{g(\theta)} \leq \frac{1 - F(\theta)}{f(\theta)}$$

for all θ.

4 Analysis

4.1 Renegotiation Invariance

It is straightforward to show that the ex post efficient decision schedule x^* is r-implementable. Take an incentive-compatible direct revelation mechanism (x^*, t^*) which would implement it in the no-renegotiation case. There is an equilibrium in which each type tells the truth in this mechanism and, after any message θ, the principal offers (x^*, t^*). This is an optimal offer because A's type is now common knowledge and so the default, $(x^*(\theta), t^*(\theta))$, is known to be efficient. Equally, it is easy to implement P's optimal mechanism given belief F, denoted by (Θ, x^F, t^F), using a null initial mechanism - at the second stage P will choose (Θ, x^F, t^F). The questions we ask are: what other schedules are r-implementable, and how can they be implemented? For examples of situations in which such a schedule would be desirable, see the hold-up example in Section 2, or the discussion in Subsection 4.5. Henceforth (X, T) will refer to an outcome schedule other than the two described above.

Consider P's optimal decision given belief $G \in \Delta^1(\Theta)$ and default outcome (x, t). Denote the minimum and maximum of $supp(G)$ (the support of G) by $\underline{\theta}(G)$ and $\bar{\theta}(G)$ respectively. It is straightforward to show that if an incentive-compatible direct revelation mechanism $(\Theta, x_\Theta, t_\Theta)$ satisfies

$$u(x_\Theta(\underline{\theta}(G)), \underline{\theta}(G)) - t_\Theta(\underline{\theta}(G)) \geq u(x, \underline{\theta}(G)) - t$$

then, for all $\theta > \underline{\theta}(G)$,

$$u(x_\Theta(\theta), \theta) - t_\Theta(\theta) \geq u(x, \theta) - t.$$

It follows that choosing P's optimal $(\Theta, x_\Theta, t_\Theta) \in D_{IC}(x, t)$ is payoff-equivalent to choosing P's optimal incentive-compatible direct revelation mechanism for type space $supp(G)$ subject to the constraint that the payoff of type $\underline{\theta}(G)$ is at least $u(x, \underline{\theta}(G)) - t$. Therefore, by standard results, an optimal mechanism $(\Theta, x_\Theta, t_\Theta)$ satisfies

$$x_\Theta(\bar{\theta}(G)) = x^*(\bar{\theta}(G)),$$

$$x_\Theta(\theta) \leq x^*(\theta) \ \ \forall \theta \in supp(G),$$

and

$$u(x_\Theta(\underline{\theta}(G)), \underline{\theta}(G)) - t_\Theta(\underline{\theta}(G)) = u(x, \underline{\theta}(G)) - t.$$

Furthermore, the downward incentive constraints bind. Therefore, if $\theta \in supp(G)$ and $\theta' \in supp(G)$ for $\theta' > \theta$ but $(\theta, \theta') \subseteq (supp(G))^C$ then $u(x_\Theta(\theta'), \theta') - t_\Theta(\theta') = u(x_\Theta(\theta), \theta') - t_\Theta(\theta)$.

The Lemma below establishes that, in any r-equilibrium of any mechanism, the final (post-renegotiation) decisions satisfy the usual monotonicity property (message by message). This is because the final outcome schedule is incentive-compatible and the utility functions are supermodular. It also establishes that the decisions are less than or equal to the efficient decisions and (in part *(iii)*), using these two properties, that decisions are deterministic - although a given type of A may randomize over messages, each message in the support of his strategy will lead to the same final decision (and transfer). This Lemma, and all subsequent Lemmas and Propositions, are to be understood as referring to almost all θ.

Lemma 1   Suppose that $(s_A, s_P, \{G_m\}_{m\in M})$ *is a* r*-equilibrium of* $\Phi(M, x_M, t_M)$*, where* $(M, x_M, t_M) \in \Gamma$.

(i) For any θ *and* $\theta' > \theta$*, if* $m \in supp(s_A(.|\theta))$ *and* $m' \in supp(s_A(.|\theta'))$ *then* $x(m, s_P, \theta) \leq x(m', s_P, \theta')$*;*

(ii) $x(s_A, s_P, \theta) \leq x^*(\theta)$ *w.pr.1;*

(iii) Suppose m *and* m' *are both in* $supp(s_A(.|\theta))$*. Then* $x(m, s_P, \theta) = x(m', s_P, \theta)$ *and* $t(m, s_P, \theta) = t(m', s_P, \theta)$.

To see why part *(iii)* of the Lemma is true, suppose that $x(m, s_P, \theta) > x(m', s_P, \theta)$. In that case, $x(m', s_P, \theta)$ must be less than the efficient quantity for θ. This implies that some higher types must also send message m', otherwise, after m', θ would be the top type and so would get an efficient quantity. This however, is incompatible with monotonicity: some of these higher types would prefer to send m and obtain $x(m, s_P, \theta)$.

Fix a mechanism (M, x_M, t_M) and a r-equilibrium $(s_A, s_P, \{G_m\}_{m\in M})$ of $\Phi(M, x_M, t_M)$. Lemma 1 implies that for each θ this equilibrium has a deterministic final outcome $(x(s_A, s_P, \theta), t(s_A, s_P, \theta))$. Define an outcome schedule (X, T) by $X(\theta) = x(s_A, s_P, \theta)$ and $T(\theta) = t(s_A, s_P, \theta)$, for $\theta \in \Theta$. This is an incentive-compatible schedule, otherwise some type could profitably deviate by imitating another type over the three-stage game. Furthermore, after any m, the outcome schedule which P proposes in $s_P(m)$ coincides with (X, T) for types in $supp(G_m)$.

The next proposition gives a modified revelation principle. It shows that the same outcome as is achieved in the given equilibrium (namely (X, T)) can also be achieved by giving the parties, at the outset, the direct revelation mechanism (Θ, X, T).

Proposition 1 (Renegotiation Invariance)   Suppose the outcome function (X, T) *is r-implementable. Then* (X, T) *can be implemented by using as the initial mechanism the direct revelation mechanism* (Θ, X, T) *whose outcomes coincide with those specified by the function* (X, T)*. In the equilibrium of the game beginning with this mechanism in place, each type* θ *mixes over a subset of* $[\underline{\theta}, \theta]$*, and, after observing any message in* Θ*,* P *offers the same mechanism,* (Θ, X, T)*, but this time the agent reports truthfully.*

In the equilibrium of Proposition 1, A must randomize over messages in such a way that P's optimal mechanism is always (Θ, X, T), no matter what message A sends. This renegotiation-invariance property is distinct from the renegotiation-proofness principle. In our setting the latter would say that our three-stage game can be regarded as a single mechanism which is not renegotiated. By contrast, renegotiation-invariance means that the outcome of the initial mechanism is in fact renegotiated in equilibrium, but the final outcome is the same as if renegotiation were not possible.

We have assumed that renegotiation takes place ex post. What if P, instead of proposing a new mechanism after the initial one is played, may propose to replace the latter with a new one before it is played but after A has learned his type? Then it follows from Proposition 1 that in equilibrium, given mechanism (Θ, X, T), she will refrain from renegotiation:

Corollary 1   If (X,T) is r-implementable then (X,T) is interim renegotiation-proof.

Proof   In the Appendix.

It remains to discover which outcome functions (X, T) are r-implementable. We examine this question in the following subsections.

4.2 Necessary Conditions for r-implementability

Proposition 1 enables us to establish conditions which r-implementable decision (and hence utility) schedules must satisfy, since the form of the equilibrium described in the Proposition restricts the possible second-stage beliefs.

Together with Lemma 1, Proposition 1 implies that if (X,T) is r-implementable then $X(\bar{\theta}) = x^*(\bar{\theta})$ and $X(\theta) \leq x^*(\theta)$ for all $\theta \in \Theta$. Furthermore, since (X,T) must be incentive-compatible X must be non-decreasing. We maintain henceforth the following assumption about X.

Assumption 1 $X : \Theta \to [\underline{x}, \bar{x}]$ satisfies $X(\theta) < x^*(\theta)$ for all $\theta < \bar{\theta}$ and $X(\bar{\theta}) = x^*(\bar{\theta})$.

The first part of Assumption 1 is made to simplify the exposition and we discuss its relaxation below. In this subsection we also require strict monotonicity:

Assumption 2 $X : \Theta \to [\underline{x}, \bar{x}]$ is strictly increasing.

We relax this requirement in Proposition 5 in the next subsection, on sufficient conditions for r-implementability, to allow non-decreasing functions which have flat sections.

The next Lemma shows that, for (X,T) such that X satisfies these two assumptions, any message $\hat{\theta}$ which is sent in the equilibrium described in Proposition 1 is sent by all types above $\hat{\theta}$.

Lemma 2 *Suppose (X,T) is r-implementable and X satisfies Assumptions 1 and 2. Then (X,T) is r-implemented by an equilibrium $(s_A, s_P, \{G_\theta\}_{\theta\in\Theta})$ of $\Phi(\Theta, X, T)$ in which, for all $\hat{\theta} \in supp(s_A)$, $supp(G_{\hat{\theta}}) = [\hat{\theta}, \bar{\theta}]$.*

Proof In the equilibrium described in Proposition 1, after message $\hat{\theta}$, P will optimally offer a mechanism which gives the efficient outcome for $\bar{\theta}(G_{\hat{\theta}}) = max(supp(G_{\hat{\theta}}))$, by efficiency at the top. If $\bar{\theta}(G_{\hat{\theta}}) < \bar{\theta}$ this implies that she doesn't offer (Θ, X, T). Contradiction. Therefore $\bar{\theta}(G_{\hat{\theta}}) = \bar{\theta}$ for any message $\hat{\theta}$ in the support of s_A.

Assume that the lowest type sending message $\hat{\theta}$ (i.e., $\underline{\theta}(G_{\hat{\theta}})$) is $\theta > \hat{\theta}$. Then, type θ would get zero renegotiation surplus, hence would get payoff $u(X(\hat{\theta}), \theta) - T(\hat{\theta})$. But, since X is strictly increasing and (X, T) is incentive-compatible, type θ could get a higher utility by announcing θ and then declining to renegotiate, a contradiction. Therefore, $\underline{\theta}(G_{\hat{\theta}}) = \hat{\theta}$ for any message $\hat{\theta}$ in the support of s_A.

Suppose that $\theta_1 \in supp(G_{\hat{\theta}})$, $\theta_2 \in supp(G_{\hat{\theta}})$, where $\theta_2 > \theta_1$ but $(\theta_1, \theta_2) \cap supp(G_{\hat{\theta}}) = \emptyset$. Then, since downward incentive constraints bind in $s_P(\hat{\theta})$, type θ_2 is indifferent between $(X(\theta_1), T(\theta_1))$ and $(X(\theta_2), T(\theta_2))$. But this contradicts the fact that (X, T) is incentive-compatible for the type set Θ and X is strictly increasing. Hence, the support of P's posterior belief is an interval. QED

Consider a schedule (X, T) which satisfies the conditions in Lemma 2. (Θ, X, T) r-implements this outcome by means of an equilibrium $(s_A, s_P, \{G_\theta\}_{\theta \in \Theta})$, as in Proposition 1. Since no type puts positive probability on messages above their true type $\underline{\theta}$ is in the support of A's strategy s_A. Let $\underline{G}$ denote P's belief after message $\underline{\theta}$. Lemma 2 implies that $supp(\underline{G}) = \Theta$. Furthermore, (X, T) is optimal for P given belief $\underline{G}$, so (see Myerson (1981), Fudenberg and Tirole (1991)) X must point-wise maximize virtual surplus

$$u(X(\theta), \theta) - \frac{1 - \underline{G}(\theta)}{\underline{g}(\theta)} u_\theta(X(\theta), \theta) - cX(\theta),$$

where $\underline{g}$ is the density of $\underline{G}$. This expression is well-defined because $\underline{G}$ is continuous and has a strictly positive density for all $\theta < \bar{\theta}$. If it had a mass point then the optimal schedule X would be constant to its right;[15] it would also be constant on any interval on which the density were zero. In either case, contrary to our assumption, X could not be strictly increasing. Therefore, for all $\theta \in [\underline{\theta}, \bar{\theta})$,

$$\frac{1 - \underline{G}(\theta)}{\underline{g}(\theta)} = \frac{(u_x(X(\theta), \theta) - c)}{u_{x\theta}(X(\theta), \theta)}. \quad (6)$$

For future reference, note that the RHS of this expression is decreasing in $X(\theta)$ if $X(\theta) < x^*(\theta)$.

[15] See Bergemann and Pesendorfer (2001) for an analysis of mechanism design for the case of type distributions with both densities and mass points (only in the working paper version).

Furthermore, take any other message $\hat{\theta}$ in the support of A's strategy. From Lemma 2, the support of P's belief $G_{\hat{\theta}}$ is $[\hat{\theta}, \bar{\theta}]$. Then it is again optimal for P to offer (Θ, X, T), so for $\theta \in [\hat{\theta}, \bar{\theta}]$,

$$\frac{1 - G_{\hat{\theta}}(\theta)}{g_{\hat{\theta}}(\theta)} = \frac{1 - \underline{G}(\theta)}{\underline{g}(\theta)}. \tag{7}$$

Moreover, $G_{\hat{\theta}}$ must be the same as $\underline{G}$, scaled to the support $[\hat{\theta}, \bar{\theta}]$, i.e.,

$$G_{\hat{\theta}}(\theta) = \frac{\underline{G}(\theta) - \underline{G}(\hat{\theta})}{1 - \underline{G}(\hat{\theta})}. \tag{8}$$

As Lemma 3 below shows, (7) and (8), combined with the fact that each type only sends messages below his true type, imply that the distribution $\underline{G}$ is smaller than the distribution F in the likelihood ratio ordering and, therefore, in the hazard rate ordering.

Lemma 3 *(i)* $\underline{G} \preceq_{LR} F$*; (ii)* $\underline{G} \preceq_{HR} F$*; (iii)* $\underline{g}$ *is continuous.*

The intuition for parts *(i)* and *(ii)* of Lemma 3 is given in Section 2 above. The Lemma is key to deriving necessary and sufficient conditions for r-implementability of an outcome schedule (X, T). We provide two kinds of characterization. One is in terms of the type distribution for which the outcome function would be optimal for the principal. This is given in Propositions 2 and 4 and Corollary 2. The other, given in Propositions 3 and 5 and Corollary 3, is in terms of properties of the decision function X, in particular that the slope of X must not be too high. Recall that x^F is P's optimal decision schedule given belief F. Given $G \in \Delta^1(\Theta)$ and $V \in \Re$, we denote by $\Gamma_P(G, V) \subseteq \Gamma$ the set of incentive-compatible direct revelation mechanisms which are optimal for P if her belief about types is G and A's reservation utility is V.

Proposition 2 *Suppose that* (X, T)*, with associated utility schedule* U*, is* r*-implementable and* X *satisfies Assumptions 1 and 2. Then (i)* $(\Theta, X, T) \in \Gamma_P(G, U(\underline{\theta}))$ *for some* G *such that* $G \preceq_{LR} F$*; (ii)* $X(\theta) \geq x^F(\theta)$ *for all* θ*; and (iii)* X *is continu-*

ous.

Proof *(i)* follows from Lemma 3*(i)* since (X,T) is optimal for type distribution $\underline{G}$.

(ii) follows from Lemma 3*(ii)*, (6), the corresponding equation for F, i.e.,

$$\frac{1-F(\theta)}{f(\theta)} = \frac{u_x(x^F(\theta),\theta)-c}{u_{x\theta}(x^F(\theta),\theta)}$$

and the fact that the RHS of (6) is decreasing in X.

(iii) follows from Lemma 3*(iii)* and (6). QED

In the case in which X is differentiable, we can identify restrictions which r-implementability places on the rate at which X can increase.

Proposition 3 *Suppose that (X,T) is r-implementable and X satisfies Assumptions 1 and 2 and is differentiable. Then*

$$\frac{f'(\theta)}{f(\theta)} + A(X(\theta),\theta) + X'(\theta)B(X(\theta),\theta) \geq 0 \tag{9}$$

for all $\theta \in [\underline{\theta},\bar{\theta})$, where

$$A(x,\theta) = \frac{2u_{x\theta}(x,\theta)}{(u_x(x,\theta)-c)} - \frac{u_{x\theta\theta}(x,\theta)}{u_{x\theta}(x,\theta)}$$

and

$$B(x,\theta) = \frac{u_{xx}(x,\theta)}{(u_x(x,\theta)-c)} - \frac{u_{xx\theta}(x,\theta)}{u_{x\theta}(x,\theta)}.$$

Proof *(i)* By (6), if X is differentiable then $\underline{g}$ is differentiable. By Lemma 3*(i)*,

$$\frac{f'(\theta)}{f(\theta)} - \frac{\underline{g}'(\theta)}{\underline{g}(\theta)} \geq 0$$

for all $\theta \in [\underline{\theta},\bar{\theta})$, f' and g' being understood as the right-hand derivative at the lower bound. Since

$$\frac{\underline{g}'(\theta)}{\underline{g}(\theta)} = -\frac{\underline{g}(\theta)}{1-\underline{G}(\theta)} - \frac{\frac{d}{d\theta}(\frac{1-\underline{G}(\theta)}{\underline{g}(\theta)})}{\frac{1-\underline{G}(\theta)}{\underline{g}(\theta)}} \tag{10}$$

it follows, using (6), that

$$\frac{\underline{g}'(\theta)}{\underline{g}(\theta)} = -A(X(\theta),\theta) - X'(\theta)B(X(\theta),\theta).$$

QED.

The necessary condition (9) places an upper bound on the slope of X, the bound depending locally on the prior and on the level of X. For some priors, this upper bound is negative at certain points; in that case a strictly increasing X cannot be implemented and so X would have to have a flat section there. Consider the case in which $u(x,\theta) = \theta u(x)$. Then the condition becomes

$$X'(\theta) \leq \frac{-u'(X(\theta))(\theta u'(X(\theta)) - c)}{cu''(X(\theta))}[\frac{f'(\theta)}{f(\theta)} + \frac{2u'(X(\theta))}{(\theta u'(X(\theta)) - c)}].$$

Since $X(\theta)$ is strictly below the efficient level, $\theta u'(X(\theta)) - c > 0$. Therefore the right hand side is negative if

$$\frac{f'(\theta)}{f(\theta)} + \frac{2u'(X(\theta))}{(\theta u'(X(\theta)) - c)} < 0,$$

so (9) is harder to satisfy if f is falling fast.

We have assumed (Assumption 1) that X is strictly below the efficient level for all $\theta < \bar{\theta}$. Suppose instead that $(\underline{\theta}, \bar{\theta}]$ is partitioned into open intervals (a_i, b_i) on which $X(\theta) < x^*(\theta)$ and closed intervals $[b_i, a_{i+1}]$ on which $X(\theta) = x^*(\theta)$. In that case, if (X, T) is r-implementable then, in an equilibrium of the type described in Proposition 1, all types in any interval $(b_i, a_{i+1}]$ tell the truth. A type $\theta \in (a_i, b_i]$ randomizes in a way which is similar to the case above but only over messages in $(a_i, \theta]$. A result analogous to Proposition 2 would hold, but separately for each interval (a_i, b_i). We omit the details.

We have also assumed that X is a strictly increasing function. For the case in which X has intervals on which it is constant we do not know if the conditions given in Proposition 2 are necessary. After every message the principal's belief must, as above, be such that X is optimal. However, the optimal solution may involve Myerson ironing, which depends on global properties of the type distribution, and so the local characterization provided by (6) is not available for the flat intervals. Different messages could in principle give rise to the same optimal X despite beliefs which are very different in nature. Nevertheless, such functions can be r-implemented if they satisfy the conditions of Proposition 3, as we show in the next subsection.

4.3 Sufficient Conditions for r-implementability

First we give a sufficient condition which corresponds to Proposition 2, namely that (X, T) would be optimal for P if she had belief G, where G is any distribution smaller than the prior in the likelihood ratio ordering.

Proposition 4 *Suppose that (X, T), with associated utility schedule $U \geq 0$, is such that $(\Theta, X, T) \in \Gamma_P(G, U(\underline{\theta}))$ for some type distribution $G \in \Delta^1(\Theta)$ such that $G \preceq_{LR} F$. Suppose also that X satisfies Assumptions 1 and 2 and is differentiable. Then (X, T) is r-implementable.*

The proof is constructive. In the equilibrium, type θ's message strategy is given by the distribution function

$$s_A([\underline{\theta}, \hat{\theta}]|\theta) = \frac{g(\theta)}{f(\theta)} \frac{f(\hat{\theta})}{g(\hat{\theta})}$$

for $\hat{\theta} \leq \theta$, which ensures that P's posterior belief is proportional to G after every message.

Combining Propositions 2 and 4 gives

Corollary 2 *Given X which is differentiable and satisfies Assumptions 1 and 2, (X, T), with associated utility schedule U, is r-implementable if and only if $U(\underline{\theta}) \geq 0$*

and $(\Theta, X, T) \in \Gamma_P(G, U(\underline{\theta}))$ *for some* G *such that* $G \preceq_{LR} F$.

The next proposition gives sufficient conditions on X which correspond to the necessary conditions in Proposition 3. Again, the proof is constructive. In the constructed equilibrium, the message strategy of type θ is given by the distribution function

$$s_A([\underline{\theta}, \hat{\theta}]|\theta) = \frac{f(\hat{\theta})}{f(\theta)} exp[-\int_{\hat{\theta}}^{\theta} z(v)dv],$$

where $z(v) = A(X(v), v) + X'(v)B(X(v), v)$ for $v \in \Theta$.

Proposition 5 *Suppose an incentive-compatible and individually rational schedule* (X, T) *is such that* X *is differentiable and satisfies Assumption 1 and condition (9). Then* (X, T) *is r-implementable.*

Combining this with Proposition 3,

Corollary 3 *Given* X *which is differentiable and satisfies Assumptions 1 and 2, an incentive-compatible and individually rational schedule* (X, T) *is r-implementable if and only if it satisfies condition (9).*

Proposition 5 establishes that any schedule (X, T) which satisfies the necessary conditions can be implemented by simply giving the parties the incentive-compatible direct revelation mechanism which implements the schedule in the case in which renegotiation is impossible. The next Proposition shows that, in the game defined by this mechanism, the equilibrium described above is essentially unique - in any equilibrium of the game, the outcome is (X, T).

Proposition 6 *Suppose that* (X, T) *is an incentive-compatible and individually rational schedule such that* X *is differentiable and satisfies Assumptions 1 and 2 and condition (9). Then the game* $\Phi(\Theta, X, T)$ *has a unique equilibrium outcome. That is,* (X, T) *is strongly r-implementable.*

A sketch of the proof of Proposition 6 is as follows (the full proof is in the Appendix). Let U be the utility schedule associated with (X, T). Suppose that there is some other equilibrium of $\Phi(\Theta, X, T)$ with equilibrium utility schedule $\tilde{U}$. Call this equilibrium $(\tilde{s}_A, \tilde{s}_P, \tilde{G})$ and let its outcome schedule be $(\tilde{X}, \tilde{T})$. If $\tilde{U} = U$ then it follows by incentive compatibility of $\tilde{U}$ and U that $(\tilde{X}, \tilde{T}) = (X, T)$, so assume that $\tilde{U} \neq U$.

Step 1. $\tilde{U}(\theta) \geq U(\theta)$ for all $\theta \in \Theta$. This is because any type θ has the option to tell the truth in $\Phi(\Theta, X, T)$ and then decline to renegotiate, giving him a payoff of $U(\theta)$.

Suppose, for example, that $\tilde{U}$ and U coincide for types up to θ_1 and for all higher types $\tilde{U}$ is strictly higher than U. This implies that, for $\theta > \theta_1$,

$$\tilde{U}(\theta) - U(\theta) = \int_{\theta_1}^{\theta} u_\theta(\tilde{X}(v), v)dv - \int_{\theta_1}^{\theta} u_\theta(X(v), v)dv > 0. \tag{11}$$

Step 2. No type sends any message in $(\theta_1, \bar{\theta}]$. This follows from the argument in the proof of Lemma 2. The lowest type sending such a message θ would be θ. Having sent this message this type would then get his default payoff $U(\theta)$. But our assumption is that $\tilde{U}(\theta) > U(\theta)$, so it cannot be part of θ's equilibrium strategy to send message θ.

In the equilibrium corresponding to $\tilde{U}$ we can assume without loss of generality that after any message P offers the mechanism $(\Theta, \tilde{X}, \tilde{T})$. This means that it would be optimal for P to offer this mechanism to types above θ_1 if the default outcome were $(X(\theta_1), T(\theta_1))$ and she knew only that the message was in the set $[\underline{\theta}, \theta_1]$. However, P's updated belief, conditional on this set of messages, about types in $(\theta_1, \bar{\theta}]$ is F conditional on $(\theta_1, \bar{\theta}]$. This is because, by Step 2, these types only send messages in $[\underline{\theta}, \theta_1]$. So $\tilde{X}$ must equal x^F (P's ex ante optimal schedule) for types above θ_1. Hence, by Proposition 2*(ii)*, $X(\theta) \geq \tilde{X}(\theta)$ for $\theta \in (\theta_1, \bar{\theta})$. However, by (11), this contradicts our assumption that $\tilde{U} > U$ on the interval $(\theta_1, \bar{\theta}]$.

4.4 The Linear Case

One leading case, treated in an earlier version of this paper, is the bilateral trade model, in which P is the seller of a unit quantity of a divisible good and A is a buyer, type θ of whom has utility θx for quantity x. So $[\underline{x}, \bar{x}] = [0, 1]$ and $u(x, \theta) = \theta x$. Furthermore, assume that $c < \underline{\theta}$, so that the efficient quantity for all types is equal to 1. The seller's optimal schedule x^F is a step function corresponding to a posted price mechanism, where x^F is equal to zero below θ^* and equal to 1 above θ^*, with θ^* maximizing the seller's revenue function $R(\theta) = (\theta - c)(1 - F(\theta))$. Hence, neither x^* nor x^F is strictly increasing.

Our main results above apply also to this case. The necessary condition (9) in Proposition 3 becomes $\theta f'(\theta) + 2f(\theta) \geq 0$. Since this is independent of $X'(\theta)$, any increasing function X such that $X(\theta) = 1$ for all $\theta \geq \theta^*$ can be r-implemented as long as this condition is satisfied by the prior. The condition is equivalent to concavity of P's revenue function $R(\theta)$, which in turn is implied by the increasing hazard rate assumption on F. The density of the mixed strategy defined in the proof of Proposition 5 becomes in this case $(f(\hat{\theta})(\hat{\theta})^2)(f(\theta)\theta^2)^{-1}$ for types θ below a critical value θ', and higher types have the same strategy as type θ', where $\theta' = \{\min \theta | X(\theta) = 1\} \leq \theta^*$. It is straightforward to show that the principal's updated belief $\underline{G}$ is such that[16]

$$\frac{1 - \underline{G}(\theta)}{\underline{g}(\theta)} = \theta - c,$$

and so virtual utility is the same for all types. Therefore P is indifferent between all incentive-compatible mechanisms for which the individual rationality constraint binds at the bottom. Hence it is optimal for her to offer the initial mechanism (Θ, X, T) again. Although, for generic beliefs, only posted price mechanisms are optimal for P, the beliefs which arise endogenously in equilibrium are the non-generic ones which justify the given mechanism.

4.5 Discussion

In Section 2 we presented one setting to which our analysis applies. Here we outline several additional ones and discuss some of our main assumptions. The initial

[16]For $\theta \leq \theta'$: for higher types the game is over, since the initial mechanism has to give quantity 1 to them.

mechanism could be designed by the principal, as in our hold-up example, or by a third party such as a social planner. If the mechanism is chosen by the principal, the question arises why she would not simply have a null contract initially and then implement her optimal schedule x^F. In other words, why should we be interested in r-implementability of any other schedules?

As we have seen in the case of the hold-up example, an ex ante contract serves to give investment incentives that an ex post contract fails to provide. For a different example, consider the case of a firm, the principal, seeking to hire a new employee, the agent. The initial mechanism corresponds to an announced labor contract that is designed to attract applications from potential workers. There are various reasons why a null initial contract might be strictly suboptimal. One possibility is that workers have a fixed cost $a > 0$ of applying to the firm (i.e. of taking part in the mechanism) and the announced contract therefore has to incorporate a type-independent rent. If the initial contract were null, the principal's optimal mechanism after the worker has arrived would leave the lowest types with utility below the reservation level of 0 since a is sunk. In this case the principal's optimal mechanism would be $(\Theta, x^F, t^F - a)$, but she would have to announce it in advance, and be legally obliged to honor it, which introduces the problem that it may be vulnerable to renegotiation after the worker has arrived. A richer set of optimal (for the principal) schedules could arise if workers have type-dependent reservation utilities, given, for example, by employment contracts which they could obtain from another firm. If this outside option is no longer available once the worker has arrived at the principal then, again, the principal would be obliged to announce her mechanism in advance. Her optimal mechanism in this case may be very different from (Θ, x^F, t^F).

Alternatively, the initial mechanism may be chosen by a third party. For example suppose this third party is a regulator, the principal is a firm and the agent a potential buyer. The regulator wishes to maximize the weighted sum of the buyer's expected utility and the seller's expected profits and so chooses a mechanism (Θ, x, t) that maximizes

$$\int_{\underline{\theta}}^{\overline{\theta}} [U(\theta) + \alpha\pi(\theta)]\, dF,$$

for some $\alpha > 1$, where $\pi(\theta)$ is the firm's profit and $U(\theta) = \theta u(x(\theta)) - t(\theta)$ is buyer type θ's utility. Using standard arguments, the optimal schedule x^{R^*} solves

$$\left(\theta + \frac{1-\alpha}{\alpha}\frac{1-F(\theta)}{f(\theta)}\right) u'(x^{R^*}(\theta)) = c,$$

which implies that

$$x^*(\theta) \geq x^{R^*}(\theta) \geq x^F(\theta).$$

As in the previous examples, the optimal initial mechanism is neither the principal's preferred mechanism, nor the efficient one. Nor, in general, is it a convex combination of those two mechanisms.

In another similar example, the planner is the headquarters of the firm. The division (principal) aims to maximize its own profits; the headquarters, however, is interested both in the profit which the division makes from a particular buyer (agent) but also in the profits to be made from this buyer by its other divisions in the future. This profit may depend both on the type of the buyer and, because, say, of learning effects, on the quantity consumed by the buyer, which affects future willingness-to-pay.

The Form of Renegotiation

In the case in which the initial mechanism is chosen by a planner, we do not need to assume that he can oblige the parties to use his mechanism. Instead, both parties have a legal right to take part in it. Would the principal, if she could, offer a different mechanism to be played in stage 1 instead of the planner's mechanism? We have already shown (Corollary 1) that she would not offer one after the agent's type is realized. In the hold-up example of Section 2 the agent does not know his type at the outset, so the question of ex ante renegotiation arises - would the principal want to propose a new mechanism before the type is realized (but after the investment stage)? We rule this out by assumption, on the grounds that it is unlikely in most applications that the principal could observe the timing of this realization, and the agent would have an incentive not to reveal it.

Our formulation assumes that after the initial mechanism is played the subsequent bargaining takes the form of an ultimatum game. Our results extend to a model in which there are finitely many stages at which the principal makes a renegotiation proposal and there is no discounting. In an equilibrium[17] of that game, after the agent's initial message, the principal always offers the initial mechanism again on the equilibrium path and a type-θ agent accepts a new mechanism before the final stage only if it offers a utility strictly higher than $U(\theta)$. At the final stage the agent accepts and reveals his type as in our analysis above.

In these formulations of the bargaining game there is typically some unrealized surplus after the play has ended. For an illustration of the kind of situation for which this is an appropriate model, consider the hold-up example of Section 2. The seller and buyer meet at date t and the buyer places an order, i.e., chooses a contract from the menu which the seller pre-announced before the investment stage, and after date t there is an exogenous finite amount of time by which trade must be completed. This could be because the good is perishable or because at some point it will become obsolete as a result of external competition. Alternatively, it may be that the seller has an exogenous production window available for this buyer after which other more profitable options will appear. In such a situation, when there is incomplete information, it is natural that in the post-mechanism bargaining game some inefficiency remains at the end of the game, i.e., after the deadline.

Could the principal avoid the renegotiation problem by proposing the initial mechanism only at the last minute before the deadline? In our hold-up example, it is easy to see that this is not possible, since the initial mechanism must be announced before the investment stage, otherwise the buyer would have the wrong investment incentives. In the case in which the designer is a third party, it may be that she cannot observe the precise time at which P and A meet, or the date of the subsequent dead-

[17] Other equilibria are outcome-equivalent.

line. This is particularly likely to be the case when the planner is designing a general mechanism for a population of principals and agents. It might be argued that the principal could propose to delay and play the planner's mechanism at the deadline; however, the agent would have no incentive to agree, since renegotiation cannot harm him and in principle could benefit him. A further possibility is that the initial mechanism, though announced some time before the deadline, could itself stipulate, in order to obviate renegotiation, that it must not be played until just before the deadline. This would be difficult to enforce because, as just noted, the exact date of the deadline may not be observable to third parties. Moreover, the mechanism would still be vulnerable to interim renegotiation.

Even if there is no exogenous deadline it may be the case that the principal is able to commit to her second-stage mechanism. This is implicitly assumed in most of the literature on principal-agent theory, in which the ultimatum game is standardly adopted. In that case, if there is a third-party designer, our assumption is that the designer cannot fully commit the principal to her initial mechanism. There are many settings in which a third party finds it harder to commit another person than it is for that person to commit herself.

We assume that the planner cannot prevent renegotiation by, for example, destroying any remaining quantities of the good (in case the principal is a seller and the agent a buyer) or by taxing away the principal's surplus from renegotiation. Physically destroying remaining quantities might be impossible if the planner cannot verify at what point his mechanism has been executed. Similarly, in order to tax the principal's surplus the planner would have to be able to verify if renegotiation has taken place, which might be difficult if parties' renegotiation agreements are silent or can be claimed to form part of an entirely new contractual agreement between the principal and the agent (for a further discussion, see Hart and Moore (1999)).

Finally, we conjecture that our results will extend in some form to models in which the post-mechanism bargaining takes the form of an infinite-horizon bargaining game with discounting. Consider the bilateral trade case with linear utility outlined in subsection 4.4. Suppose that the initial mechanism is a direct revelation mechanism[18] in which the buyer announces his type and the outcome for announcement θ is a contract according to which the buyer receives the good at time $\tau(\theta)$ and pays a price $p(\theta)$. Suppose also that the bargaining is in discrete time and the uninformed party makes all the offers. Our conjecture is that it is possible to r-implement any outcome function which corresponds to an equilibrium σ of the bargaining game for a belief which is lower in the likelihood ratio order than the prior, and that this is achieved, as in our analysis above, by giving the parties the corresponding direct revelation mechanism, and by a mixed strategy of the buyer which, after any message, gives an updated belief which supports, in the subsequent bargaining equilibrium, the outcome of σ. The derivation of the buyer's mixed strategy would have to be substantially more complicated than in our analysis above because the property that beliefs after different messages are scaled versions of each other would no longer hold, otherwise the rate at which the seller would subsequently screen would vary with the buyer's message. We leave the exploration of these generalizations to future work.

Appendix

Proof of Lemma 1 *(i)* Since m is optimal for θ and m' is optimal for θ',

$$u(x(m, s_P, \theta), \theta) - t(m, s_P, \theta)) \geq u(x(m', s_P, \theta'), \theta) - t(m', s_P, \theta')$$

and

$$u(x(m', s_P, \theta'), \theta') - t(m', s_P, \theta') \geq u(x(m, s_P, \theta), \theta') - t(m, s_P, \theta).$$

Therefore, since $u_x > 0$ and $u_{x\theta} > 0$, $x(m', s_P, \theta') \geq x(m, s_P, \theta)$.

(ii) Let $M'(\theta) = \{m \in M | x(m, s_P, \theta) > x^*(\theta)\}$. By standard results, for any m, P's optimal schedule given G_m satisfies $x(m, s_P, \theta) \leq x^*(\theta)$ for all $\theta \in supp(G_m)$.

[18] Cramton (1985) refers to this as a *direct revelation sequential bargaining mechanism.*

Hence, if $m \in M'(\theta)$ then $\theta \notin supp(G_m)$. But

$$pr(\{(\theta, m) \in \Theta \times M | \theta \notin supp\,(G_m)\,\text{and}\,m \in supp\,(s_A(.|\theta))\} = 0,$$

where pr refers to the joint distribution derived from F and s_A. Therefore $pr\{\theta \in \Theta | s_A(M'(\theta)|\theta) > 0\} = 0$.

(iii) Suppose $x(m, s_P, \theta) > x(m', s_P, \theta)$. Then Lemma 1(ii) implies that $x(m', s_P, \theta) < x^*(\theta)$, and so $\theta < \bar{\theta}(G_{m'})$. There are two cases to consider. (a) there exists $\theta_1 = min\{\tilde{\theta} > \theta | \tilde{\theta} \in supp(G_{m'})\}$. (b) there exists a sequence $\{\theta_i\}_{i=1}^{\infty} \subseteq supp(G_{m'})$ and $\{\theta_i\}_{i=1}^{\infty} \downarrow \theta$.

Case (a): downward incentive constraints bind for the mechanism $s_P(m')$ so

$$u(x(m', s_P, \theta_1), \theta_1) - t(m', s_P, \theta_1) = u(x(m', s_P, \theta), \theta_1) - t(m', s_P, \theta). \tag{12}$$

But θ is indifferent between m and m', so

$$u(x(m', s_P, \theta), \theta) - t(m', s_P, \theta) = u(x(m, s_P, \theta), \theta) - t(m, s_P, \theta).$$

Since $\theta_1 > \theta$ and $x(m, s_P, \theta) > x(m', s_P, \theta)$, this last equation, together with supermodularity, implies

$$u(x(m, s_P, \theta), \theta_1) - t(m, s_P, \theta) > u(x(m', s_P, \theta), \theta_1) - t(m', s_P, \theta).$$

So, by (12),

$$u(x(m, s_P, \theta), \theta_1) - t(m, s_P, \theta) > u(x(m', s_P, \theta_1), \theta_1) - t(m', s_P, \theta_1),$$

which contradicts optimality of message m' for θ_1.

Case (b). By Lemma 1(i), $x(m, s_P, \theta) \leq x(m', s_P, \theta_i)$ for all $\theta_i \in \{\theta_i\}_{i=1}^{\infty}$. There is no loss of generality in taking $s_P(m')$ to be right-continuous. This implies $x(m', s_P, \theta) \geq x(m, s_P, \theta)$. Contradiction. QED

Proof of Proposition 1 Let $(s_A, s_P, \{G_m\}_{m \in M})$ be an r-equilibrium of $\Phi(M, x_M, t_M)$ which r-implements the given outcome function (X, T). Take a message m which is in the support of s_A. After this message is sent the default outcome is $(x_M(m), t_M(m))$ and P's belief is G_m. The minimum of the support of G_m is $\underline{\theta}(G_m)$. For brevity we refer to $\underline{\theta}(G_m)$ as $\underline{\theta}_m$.

As argued above, the outcome function which is given by $s_P(m)$ must coincide with (X, T) for types in the support of G_m. Therefore (Θ, X, T) is optimal for P given belief G_m subject to the constraint that type $\underline{\theta}_m$ gets at least $u(x_M(m), \underline{\theta}_m) - t_M(m)$. It follows that

(*) (Θ, X, T) is optimal for P given belief G_m subject to the constraint that type $\underline{\theta}_m$ gets at least $u(X(\underline{\theta}_m), \underline{\theta}_m) - T(\underline{\theta}_m)$.

Now suppose that the initial mechanism is (Θ, X, T). In $\Phi(\Theta, X, T)$, A's strategy is defined by the two-step procedure:

(i) select a message m in M using the strategy s_A;

(ii) given m, announce $\underline{\theta}_m$, the lowest type which sends m according to s_A.

P's strategy is: offer (Θ, X, T) after any message. P's beliefs are derived via Bayes' Rule.

This profile clearly implements the schedule (X, T). To see that it is an equilibrium, note first that A is indifferent between all type announcements since any message leads to the same schedule and all possible defaults belong to this schedule. Therefore A's strategy is optimal. Next, consider P's strategy. Initially, suppose that P can observe the message m chosen by A in stage (i) of his strategy, in addition to his type announcement. Then P's belief is G_m, with lower bound of support $\underline{\theta}_m$. The default outcome is $(X(\underline{\theta}_m), T(\underline{\theta}_m))$. Therefore, by (*), it is optimal for P to offer (Θ, X, T). In fact, P only observes the message θ, not m. However, P knows that m lies in the set $\{m | \underline{\theta}_m = \theta\}$ and, as just shown, (Θ, X, T) is optimal for each such m. Therefore P's strategy is optimal and so the given strategies and beliefs form an equilibrium.

The fact that, for any m chosen at stage (i) of his strategy, A announces the lowest type who would send m implies that he never announces a type above his true type. QED

Proof of Corollary 1 In the interim-renegotiation case, given (X, T) and associated U, P proposes an incentive compatible direct revelation mechanism that satisfies A's participation constraint given by $\hat{U}(\theta) \geq U(\theta)$ for all θ. In the ex-post renegotiation case, in the equilibrium of Proposition 1 P's optimal mechanism is (Θ, X, T) after every message m when the reservation utility is given by $u(X(m), \theta) - T(m)$. Therefore, it is also the optimal mechanism when the reservation utility is $U(\theta) \geq u(X(m), \theta) - T(m)$. Since this is true for every message, the same statement is true before observing the message. Hence (Θ, X, T) solves P's interim problem. QED

Proof of Lemma 3 *(i)* Take $\theta_2 > \theta_1 > \underline{\theta}$. Then the distribution over Θ conditional on messages up to θ_1 is, by Bayes' Rule,

$$G(\theta_2 | [\underline{\theta}, \theta_1]) = \frac{\int_{\underline{\theta}}^{\theta_2} s_A([\underline{\theta}, \theta_1] | \theta) f(\theta) d\theta}{\int_{\underline{\theta}}^{\bar{\theta}} s_A([\underline{\theta}, \theta_1] | \theta) f(\theta) d\theta}.$$

The denominator is strictly positive because types only send messages below them. The integrals are well-defined because of our assumption that mixed strategies are derived from distributional strategies.

Differentiating w.r.t. θ_2,

$$g(\theta_2 | [\underline{\theta}, \theta_1]) = \frac{s_A([\underline{\theta}, \theta_1] | \theta_2) f(\theta_2)}{\int_{\underline{\theta}}^{\bar{\theta}} s_A([\underline{\theta}, \theta_1] | \theta) f(\theta) d\theta},$$

so

$$\frac{g(\theta_2 | [\underline{\theta}, \theta_1])}{g(\theta_1 | [\underline{\theta}, \theta_1])} = \frac{s_A([\underline{\theta}, \theta_1] | \theta_2) f(\theta_2)}{s_A([\underline{\theta}, \theta_1] | \theta_1) f(\theta_1)}. \tag{13}$$

By (8), which applies except on a set of s_A-measure zero,

$$g(\theta_2 | [\underline{\theta}, \theta_1]) = \frac{\int_{\underline{\theta}}^{\theta_1} g_{\hat{\theta}}(\theta_2) ds_A(\hat{\theta})}{s_A([\underline{\theta}, \theta_1])} = \frac{\underline{g}(\theta_2)}{s_A([\underline{\theta}, \theta_1])} \int_{\underline{\theta}}^{\theta_1} (1 - \underline{G}(\hat{\theta}))^{-1} ds_A(\hat{\theta}),$$

where s_A is the measure over messages, i.e. $s_A([\underline{\theta}, \theta_1]) = \int_{\underline{\theta}}^{\bar{\theta}} s_A([\underline{\theta}, \theta_1]|\theta) f(\theta) d\theta$.

Hence

$$\frac{g(\theta_2|[\underline{\theta}, \theta_1])}{g(\theta_1|[\underline{\theta}, \theta_1])} = \frac{\underline{g}(\theta_2)}{\underline{g}(\theta_1)},$$

so, from (13),

$$\frac{\underline{g}(\theta_2)}{\underline{g}(\theta_1)} = \frac{s_A([\underline{\theta}, \theta_1]|\theta_2) f(\theta_2)}{s_A([\underline{\theta}, \theta_1]|\theta_1) f(\theta_1)}.$$

This proves *(i)* since $s_A([\underline{\theta}, \theta_1]|\theta_2) \leq s_A([\underline{\theta}, \theta_1]|\theta_1) = 1$.

(ii) follows because if $\underline{G}$ is smaller than F in the likelihood ratio ordering then it is also smaller in the hazard rate ordering (see Shaked and Shanthikumar (1994)).

(iii) Since $\underline{G} \preceq_{LR} F$ and f is continuous, $\underline{g}$ cannot jump up. If it jumps down then, by (6), X must jump down because the RHS of (6) is decreasing in X. This contradicts the fact that X is non-decreasing. QED

Proof of Proposition 4 $(\Theta, X, T) \in \Gamma_P(G, U(\underline{\theta}))$ implies that G has a strictly positive density g. We construct an equilibrium of the type described in Proposition 1 which implements (X, T). The initial mechanism is (Θ, X, T). (i) P's strategy is to offer (Θ, X, T) again after any message. (ii) Each type θ has a mixed strategy with support $[\underline{\theta}, \theta]$ and a mass point on $\underline{\theta}$, given by the distribution function

$$s_A([\underline{\theta}, \hat{\theta}]|\theta) = \frac{g(\theta)}{f(\theta)} \frac{f(\hat{\theta})}{g(\hat{\theta})}$$

for $\hat{\theta} \leq \theta$ and $s_A([\underline{\theta}, \hat{\theta}]|\theta) = 1$ for $\hat{\theta} > \theta$. (iii) After message $\hat{\theta}$, P's belief about types is given by G conditional on $\theta \geq \hat{\theta}$.

The distribution in (ii) is well-defined because $G \preceq_{LR} F$ implies that s_A is non-decreasing and $s_A([\underline{\theta}, \theta]|\theta) = 1$. Furthermore, by the argument in the proof of Lemma 3*(iii)*, s_A is continuous.

Since X is differentiable, (6) applied to G implies that g is differentiable. The density of the mixed strategy of type θ is

$$\sigma_A(\hat{\theta}|\theta) = \frac{g(\theta)}{f(\theta)} \frac{f(\hat{\theta})}{g(\hat{\theta})} \left[\frac{f'(\hat{\theta})}{f(\hat{\theta})} - \frac{g'(\hat{\theta})}{g(\hat{\theta})} \right].$$

The fact that F has no mass points and s_A is continuous implies that $G_{\hat{\theta}}$ is continuous for every message $\hat{\theta}$. Bayes' Rule implies that P's belief over types in $[\hat{\theta}, \bar{\theta}]$ after message $\hat{\theta}$ is given by

$$G_{\hat{\theta}}(\theta) = \frac{\int_{\hat{\theta}}^{\theta} \sigma_A(\hat{\theta}|v) f(v) dv}{\int_{\hat{\theta}}^{\bar{\theta}} \sigma_A(\hat{\theta}|v) f(v) dv} = \frac{\int_{\hat{\theta}}^{\theta} g(v) dv}{\int_{\hat{\theta}}^{\bar{\theta}} g(v) dv} = \frac{G(\theta) - G(\hat{\theta})}{1 - G(\hat{\theta})}.$$

This shows that the principal's beliefs are correct after every message. Given those beliefs, applying (6), it is optimal for P to offer (Θ, X, T) again after any message. A's strategy is optimal because every message leads to the same offered schedule (X, T), so he is indifferent between all messages. This shows that the strategies form an equilibrium. QED.

Proof of Proposition 5 As in the proof of Proposition 4, we construct an equilibrium for initial mechanism (Θ, X, T) which r-implements (X, T).

Let $z(\theta) = A(X(\theta), \theta) + X'(\theta) B(X(\theta), \theta)$. (i) P's strategy is to offer (Θ, X, T) again after any message. (ii) The mixed strategy of type θ of A, $s_A(.|\theta)$, is given by the distribution function

$$s_A([\underline{\theta}, \hat{\theta}]|\theta) = \frac{f(\hat{\theta})}{f(\theta)} exp[-\int_{\hat{\theta}}^{\theta} z(v) dv]$$

for $\hat{\theta} \leq \theta$ and $s_A([\underline{\theta}, \hat{\theta}]|\theta) = 1$ for $\hat{\theta} > \theta$. By (9) $-z(\theta)$ is bounded, so the integral is well-defined. The density is then

$$\sigma_A(\hat{\theta}|\theta) = \frac{1}{f(\theta)} [exp(-\int_{\hat{\theta}}^{\theta} z(v) dv)][f'(\hat{\theta}) + f(\hat{\theta}) z(\hat{\theta})].$$

This distribution is well-defined because $f'(\hat{\theta}) + f(\hat{\theta}) z(\hat{\theta}) \geq 0$ by (9).

(iii) Given message $\hat{\theta} \in \Theta$, P's belief (c.d.f) is

$$G_{\hat{\theta}}(\theta) = \frac{\int_{\hat{\theta}}^{\theta} exp[-\int_{\hat{\theta}}^{v} z(w) dw] dv}{\int_{\hat{\theta}}^{\bar{\theta}} exp[-\int_{\hat{\theta}}^{v} z(w) dw] dv}$$

for $\hat{\theta} \leq \theta$ and $G_{\hat{\theta}}(\theta) = 0$ for $\hat{\theta} > \theta$.

Note that if $\theta_1 < \theta_2 < \theta$

$$\frac{s_A([\underline{\theta}, \theta_1]|\theta)}{s_A([\underline{\theta}, \theta_2]|\theta)} = \frac{f(\theta_1)}{f(\theta_2)} exp[- \int_{\theta_1}^{\theta_2} z(v)dv],$$

which is independent of θ, so that any two types θ and θ' randomize in the same way, proportionally, over the set of messages below $min[\theta, \theta']$. This is the property which ensures that the principal's posterior distribution is invariant, apart from scaling, to the message.

To see that this is an equilibrium, note first that, by Bayes' rule, the conditional density after message $\hat{\theta}$ of type $\theta \geq \hat{\theta}$ is

$$g_{\hat{\theta}}(\theta) = \frac{\sigma_A(\hat{\theta}|\theta) f(\theta)}{\int_{\hat{\theta}}^{\bar{\theta}} \sigma_A(\hat{\theta}|v) f(v) dv} = \frac{exp[- \int_{\hat{\theta}}^{\theta} z(w)dw]}{\int_{\hat{\theta}}^{\bar{\theta}} exp[- \int_{\hat{\theta}}^{v} z(w)dw]dv},$$

so P's beliefs are correct given A's strategy. A's strategy is optimal because every message leads to the same offered schedule (X, T), so he is indifferent between all messages. It remains to show that P's optimal mechanism is (Θ, X, T) after every message, i.e., that

$$\frac{1 - G_{\hat{\theta}}(\theta)}{g_{\hat{\theta}}(\theta)} = \frac{(u_x(X(\theta), \theta) - c)}{u_{x\theta}(X(\theta), \theta)}$$

for every message $\hat{\theta} \in \Theta$ and every $\theta \geq \hat{\theta}$.

Let for all $v \geq \hat{\theta}$, $k_{\hat{\theta}}(v) = \int_{\hat{\theta}}^{v} z(w)dw$. Then

$$\frac{1 - G_{\hat{\theta}}(\theta)}{g_{\hat{\theta}}(\theta)} = \frac{\int_{\theta}^{\bar{\theta}} exp[-k_{\hat{\theta}}(v)]dv}{exp[-k_{\hat{\theta}}(\theta)]},$$

so we need to show that

$$\int_{\theta}^{\bar{\theta}} exp[-k_{\hat{\theta}}(v)]dv = exp[-k_{\hat{\theta}}(\theta)] \frac{(u_x(X(\theta), \theta) - c)}{u_{x\theta}(X(\theta), \theta)}. \tag{14}$$

For $\theta = \bar{\theta}$, the LHS of (14) is zero, and the RHS is also zero since $u_x(X(\bar{\theta}), \bar{\theta}) - c = 0$ by efficiency at the top. The derivative of the LHS with respect to θ is $-exp[-k_{\hat{\theta}}(\theta)]$.

The derivative of the RHS is

$$e^{-k_{\hat{\theta}}(\theta)}(-k'_{\hat{\theta}}(\theta))\frac{(u_x - c)}{u_{x\theta}} + e^{-k_{\hat{\theta}}(\theta)}\frac{u_{x\theta}[u_{xx}X'(\theta) + u_{x\theta}] - (u_x - c)[u_{x\theta\theta} + u_{xx\theta}X'(\theta)]}{(u_{x\theta})^2},$$

where arguments $(X(\theta), \theta)$ have been omitted for brevity. Since $k'_{\hat{\theta}}(\theta) = z(\theta)$, this is equal to $-exp[-k_{\hat{\theta}}(\theta)]$ and so (14) is true for all θ. This shows that P's strategy is optimal. QED.

Proof of Proposition 6 Let U be the utility schedule associated with (X, T).

By standard results,

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u_\theta(X(v), v)dv. \tag{15}$$

Therefore, if every equilibrium of $\Phi(\Theta, X, T)$ has the same utility schedule then every equilibrium gives the same outcome, namely $(X(\theta), T(\theta))$, to each type θ, since $u_{x\theta} > 0$. Suppose then that there is an equilibrium with utility schedule $\tilde{U} \neq U$. Call this equilibrium $(\tilde{s}_A, \tilde{s}_P, \tilde{G})$. Since any type θ is able to tell the truth in $\Phi(\Theta, X, T)$ and decline to renegotiate, giving $U(\theta)$, it must be that $\tilde{U}(\theta) \geq U(\theta)$ for all $\theta \in \Theta$.

By Proposition 1, we can assume without loss of generality that in the strategy profile $(\tilde{s}_A, \tilde{s}_P)$ P offers $(\Theta, \tilde{X}, \tilde{T})$ after any message, where $(\tilde{X}, \tilde{T})$ is the outcome implemented by $(\tilde{s}_A, \tilde{s}_P, \tilde{G})$.

Let $\theta_1 = inf(\theta | \tilde{U}(\theta) > U(\theta))$ and let $\theta_2 = inf(\theta | \theta > \theta_1, \tilde{U}(\theta) = U(\theta))$ unless $\tilde{U}(\theta) > U(\theta)$ for all $\theta > \theta_1$, in which case let $\theta_2 = \bar{\theta}$.

(a) Assume that $\theta_2 < \bar{\theta}$.

Then $\tilde{U}(\theta) > U(\theta)$ for all $\theta \in (\theta_1, \theta_2)$, $\tilde{U}(\theta_1) = U(\theta_1)$ and $\tilde{U}(\theta_2) = U(\theta_2)$, by continuity of $\tilde{U}$ and U. Since, for $\theta \in supp(\tilde{s}_A)$, $min[supp(\tilde{G}_\theta)] = \theta$ it follows that, for $\theta \in (\theta_1, \theta_2)$, $\theta \notin supp(\tilde{s}_A)$, otherwise θ would be the lowest type to send message θ, hence $\tilde{U}(\theta) = U(\theta)$. So no type in (θ_1, θ_2) sends any message in (θ_1, θ_2).

Since P offers $(\Theta, \tilde{X}, \tilde{T})$ after any message, $(\tilde{X}, \tilde{T})$ is optimal for P conditional on the set of messages $[\underline{\theta}, \theta_1]$. Let P's probability distribution conditional on this set be denoted by $\tilde{G}_1$. Then, for $\theta \in (\theta_1, \theta_2)$, $\tilde{G}_1$ must have a density $\tilde{g}_1$ and $\tilde{g}_1(\theta)$ is proportional to $f(\theta)$ since types in (θ_1, θ_2) only send messages in $[\underline{\theta}, \theta_1]$. Hence, by (6) and the argument in the proof of Proposition 3, $\tilde{X}$ is differentiable on (θ_1, θ_2) and

$$\frac{f'(\theta)}{f(\theta)} = -A(\tilde{X}(\theta), \theta) - \tilde{X}'(\theta) B(\tilde{X}(\theta), \theta)$$

for $\theta \in (\theta_1, \theta_2)$.

By Lemma 3,

$$\frac{\underline{g}'(\theta)}{\underline{g}(\theta)} \leq \frac{f'(\theta)}{f(\theta)}.$$

So

$$-A(\tilde{X}(\theta), \theta) - \tilde{X}'(\theta) B(\tilde{X}(\theta), \theta) \geq -A(X(\theta), \theta) - X'(\theta) B(X(\theta), \theta)$$

for $\theta \in (\theta_1, \theta_2)$. Hence, if $\tilde{X}(\theta) = X(\theta), \tilde{X}'(\theta) \geq X'(\theta)$. For small enough $\varepsilon > 0$, $\tilde{U}(\theta) > U(\theta)$ for $\theta \in (\theta_1, \theta_1 + \varepsilon)$. Therefore $\tilde{X}(\theta) > X(\theta)$ for $\theta \in (\theta_1, \theta_1 + \varepsilon)$ by (15). Therefore, since $\tilde{X}' \geq X'$ whenever $\tilde{X} = X$,

$$\int_{\theta_1}^{\theta_2} u_\theta(\tilde{X}(\theta), \theta) d\theta > \int_{\theta_1}^{\theta_2} u_\theta(X(\theta), \theta) d\theta,$$

which contradicts $\tilde{U}(\theta_2) = U(\theta_2)$.

(b) Now assume that $\theta_2 = \bar{\theta}$, so that $\tilde{U}(\theta) > U(\theta)$ for all $\theta \in (\theta_1, \bar{\theta}]$.

According to the equilibrium strategy $\tilde{s}_A$, types in $(\theta_1, \bar{\theta}]$ only send messages in $[\underline{\theta}, \theta_1]$, so, conditional on this set of messages, P's belief $\tilde{G}_1$ satisfies

$$\frac{1 - \tilde{G}_1(\theta)}{\tilde{g}_1(\theta)} = \frac{1 - F(\theta)}{f(\theta)}$$

for $\theta > \theta_1$. Also $(\tilde{X}, \tilde{T})$ is optimal for P given this belief so

$$\frac{1 - F(\theta)}{f(\theta)} = \frac{u_x(\tilde{X}(\theta), \theta) - c}{u_{x\theta}(\tilde{X}(\theta), \theta)}.$$

From Lemma 3

$$\frac{1 - F(\theta)}{f(\theta)} \geq \frac{1 - \underline{G}(\theta)}{\underline{g}(\theta)} = \frac{u_x(X(\theta), \theta) - c}{u_{x\theta}(X(\theta), \theta)},$$

so $X(\theta) \geq \tilde{X}(\theta)$ for $\theta \in (\theta_1, \bar{\theta})$ since $u_{\theta x} > 0$. By (15) this contradicts the fact that $\tilde{U}(\theta) > U(\theta)$ on this interval. QED

Acknowledgements

This is a revised version of a paper previously circulated under the title "Bilateral Trading and Renegotiation". We are grateful to the editor and three referees whose comments greatly improved the paper. We also benefited from valuable conversations with Andrea Attar, Bruno Biais, Johannes Hörner, Bruno Jullien and Thomas Mariotti and from the comments of various seminar audiences. Part of this research was conducted while Evans was a Fernand Braudel Fellow at the European University Institute and he is grateful for their hospitality.

References

Ausubel, L. and R. Deneckere (1989a), "Reputation in Bargaining and Durable Goods Monopoly", *Econometrica* 57, 511-531.

Ausubel, L. and R. Deneckere (1989b), "A Direct Mechanism Characterization of Sequential Bargaining with One-Sided Incomplete Information", *Journal of Economic Theory* 48, 18-46.

Beaudry, P. and M. Poitevin (1995), "Contract Renegotiation: A Simple Framework and Implications for Organization Theory", *Canadian Journal of Economics* 28, 302-35.

Bergemann, D. and M. Pesendorfer (2001) "Information Structures in Optimal Auctions" *Cowles Foundation Discussion Paper* 1323.

Bester , H. and R. Strausz (2001) "Contracting with Imperfect Commitment and the Revelation Principle: The Single-Agent Case", *Econometrica* 69, 1077-98.

Bester , H. and R. Strausz (2007) "Contracting with Imperfect Commitment and Noisy Communication", *Journal of Economic Theory* 136, 236-259.

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

Cramton, P. (1985) "Sequential Bargaining Mechanisms", in *Game Theoretic Models of Bargaining*, Ed. A. Roth, pp. 149-180, CUP, NY.

Crawford, V. and J. Sobel (1982) "Strategic Information Transmission", *Econometrica* 50, 1431-1451.

Dewatripont, M. and E. Maskin (1990) "Contract Renegotiation in Models of Asymmetric Information", *European Economic Review* 34, 311-321.

Forges, F. (1994) "Posterior Efficiency", *Games and Economic Behavior* 6, 238-261.

Fuchs, W. and A. Skrzypacz (2013) "Bargaining with Deadlines and Private Information", *American Economic Journal: Microeconomics* 5, 219-43.

Fudenberg, D. and J. Tirole (1983) "Sequential Bargaining with Incomplete Information", *Review of Economic Studies* 50, 221-47.

Fudenberg, D. and J. Tirole (1991) *Game Theory*, Cambridge, MA: MIT Press.

Goltsman, M. (2011) "Optimal Information Transmission in a Hold-Up Problem", *RAND Journal of Economics* 42(3), 495-526.

Green, J. R. and J.-J. Laffont (1987) "Posterior Implementability in a Two-Person Decision Problem", *Econometrica* 55, 69-94.

Hart, O. and J. Moore (1999) "Foundations of Incomplete Contracts", *Review of Economic Studies* 66, 115-138.

Hart. O and J. Tirole (1988) "Contract Renegotiation and Coasian Dynamics", *Review of Economic Studies* 55, 509-40.

Jullien, B. (2000) "Participation Constraints in Adverse Selection Models", *Journal of Economic Theory* 93, 1-47.

Krishna, V. and J. Morgan (2008) "Contracting for Information under Imperfect

Commitment", *RAND Journal of Economics* 39(4), 905-925.

Laffont, J.J. and J. Tirole (1988) "The Dynamics of Incentive Contracts", *Econometrica* 56, 1153-75.

Laffont, J.J. and J. Tirole (1990) "Adverse Selection and Renegotiation in Procurement", *Review of Economic Studies* 57, 597-625.

Maestri, L. (2013) "Dynamic Contracting under Adverse Selection and Renegotiation", working paper.

Maskin, E. and J. Moore (1999) "Implementation and Renegotiation", *Review of Economic Studies* 66, 39-56.

Maskin, E. and J. Riley (1984) "Optimal Auctions with Risk-Averse Buyers", *Econometrica* 52, 1473-1518.

Milgrom, P. and I. Segal (2002) "Envelope Theorems for Arbitrary Choice Sets", *Econometrica* 70, 583-601.

Milgrom, P. and R. Weber (1985) "Distributional Strategies for Games with Incomplete Information", *Mathematics of Operations Research* 10, 619-632.

Myerson, R. (1981) "Optimal Auction Design", *Mathematics of Operations Research* 6, 58-73.

Neeman, Z. and G. Pavlov (2013) "Ex Post Renegotiation-Proof Mechanism Design", *Journal of Economic Theory* 148, 473-501.

Ottaviani, M. (2000) "The Economics of Advice", working paper .

Rubinstein, A. and A. Wolinsky (1992) "Renegotiation-Proof Implementation and Time Preferences", *American Economic Review* 82, 600-614.

Reiche, S. (2006) "Ambivalent Investment and the Hold-Up Problem", *Journal of the European Economic Association* 4, 1148-1164.

Segal, I. (1999) "Complexity and Renegotiation: A Foundation for Incomplete Contracts", *Review of Economic Studies* 66, 57-82.

Shaked, M. and J. G. Shanthikumar (1994) *Stochastic Orders and their Applications*, Academic Press.

Skreta, V. (2006) "Sequentially Optimal Mechanisms", *Review of Economic Studies* 73, 1085-1111.

Strausz, R. (2006) "Deterministic versus Stochastic Mechanisms in Principal-Agent Models", *Journal of Economic Theory* 128, 306-314.

Strulovici, B. (2014) "Contract Negotiation and the Coase Conjecture", working paper.