Mechanism Design and Non-Cooperative Renegotiation

Robert Evans[1] and Sönje Reiche[2]

This version: December 2013

Abstract

We characterize decision rules which are implementable in mechanism design settings when, after the play of a mechanism, the uninformed party can propose a new mechanism to the informed party. The necessary and sufficient conditions are, essentially, that the rule would be implementable if parties could commit not to renegotiate the mechanism, that for each type the decision is at least as high as if there were no mechanism, and that the slope of the decision function is not too high. The direct mechanism which implements such a rule when renegotiation can be prevented will also implement it in any equilibrium when it cannot, so the standard mechanism is robust to renegotiation.

This is a revised version of a paper previously circulated under the title "Bilateral Trading and Renegotiation". We are grateful to various seminar audiences for helpful comments. 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.

Keywords: Renegotiation, Mechanism Design

---

[1]St. John's College and University of Cambridge, UK. robert.evans@econ.cam.ac.uk
[2]Toulouse School of Economics (CNRS, GREMAQ)

1 Introduction

Suppose that the interaction between a number of asymmetrically informed parties is governed by a mechanism. However, the parties cannot fully commit to the outcome of this mechanism - once the outcome is known, it may be renegotiated by the parties. What is the set of allocations that can be achieved in this environment and how can they be achieved?

We address these questions in the context of a model with two players and one-sided asymmetric information: one player's (the principal's) payoff function is common knowledge, but the other's (the agent's) is private information.

In the mechanism the agent sends a message to the principal, determining some contracted decision and money payment. However, the two players cannot be obliged to stick to this decision. We assume that, at this point, the principal is able to design a second-stage mechanism to determine the actual decision and transfer. Her optimal mechanism will depend on what she has learned from her interaction with the agent in the initial mechanism. Consequently, we cannot assume that the agent's message in the initial mechanism reveals his type because the principal, knowing the truth, would subsequently extract all the remaining surplus. This in turn would give the agent an incentive to understate his type.

To determine what can be achieved in this setting we characterize the implementable decision and utility schedules: that is, functions mapping the agent's type to, respectively, decision and expected utility, taking renegotiation into account. As in the case in which the mechanism cannot be renegotiated, once the implementable decision schedules have been determined, the implementable expected utility schedules can be derived by integration.

The initial mechanism may be designed either by the principal or by an outside agency, such as a planner. If the designer is the principal she might want to propose a mechanism to attract a particular pool of agents or to induce relationship specific investment. Her 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. An alternative application is the design of a trading platform or a market where sellers and buyers 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.

If a decision schedule (mapping types of agent to decisions) is renegotiation-implementable (i.e., implementable taking into account renegotiation as described above) then it is easy to see that it must, as in the no-rengotiation case (i.e. when parties can commit not to renegotiate), be an increasing function. It must also give the efficient decision to the top type and a weakly lower-than-efficient decision to all types. We derive (in Proposition 2) two further conditions which a strictly increasing, differentiable decision schedule must satisfy if it is renegotiation-implementable. One puts an upper bound on the slope of the function, which depends on the prior distribution over types. The second condition is that, for every type, the decision must be at least as high as it would be if there were no initial mechanism and the principal simply offered her ex-post optimal mechanism.

Moreover, one mechanism which implements a particular implementable schedule is simply the same truth-telling direct revelation mechanism which would implement it in the no-renegotiation case, although the equilibrium is very different. In equilibrium, rather than tell the truth with probability 1, the agent uses a mixed strategy - a type θ of the agent randomizes over messages below θ, so that the principal, given announcement θ', has a posterior belief distributed over types θ' and above. The principal's equilibrium strategy is to offer the initial mechanism again after any message. The agent then selects the decision and transfer which he would have chosen had the two players been committed not to renegotiate this mechanism in the first place.

In Proposition 3 we show, by construction, that any decision schedule which satisfies the necessary conditions can be renegotiation-implemented in this way. In Proposition 4, we show that the equilibrium is unique. In other words, we have the striking result that, 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. A further appealing feature is that an outside designer wishing to implement a particular outcome function does not need to know the prior distribution over the agent's types (the principal's prior belief).

These results can be regarded as contributing to the bargaining literature as well as to the mechanism design literature. Given a fixed bargaining game of incomplete information, one can ask: in what ways is it possible to alter the outcome of the game by obliging the parties to sign a contract beforehand? Our framework can also be interpreted from this point of view.

Related Literature

Various notions of renegotiation-proofness for mechanisms have been proposed. In the incomplete information case, much of the literature concerns interim renegotiation, i.e., the parties have an opportunity to renegotiate before they play the mechanism. For example, Holmström and Myerson (1983) define a decision rule (or mechanism) M as *durable* if, given any type profile, and any alternative mechanism $\tilde{M}$, the players would not vote unanimously to replace M by $\tilde{M}$ if a neutral third party were to propose it to them (see also Crawford (1985), Palfrey and Srivastava (1991) and Lagunoff (1995)). 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. Under an alternative modeling of this process, a renegotiation proposal would be made by one of the parties to the mechanism.

In this paper we use the latter notion of renegotiation. This is closer to the one generally used for the complete information case (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] It also corresponds to the approach used in the literature on contract renegotiation (e.g. Dewatripont and Maskin (1990), Hart and Tirole (1988), Laffont and Tirole (1988,1990)) in which a trading opportunity is repeated a number of times and the focus is on comparing the outcomes of long-term contracts, sequences of short-term contracts, and long-term contracts which can be renegotiated (i.e., in the two-period case, the parties are committed for one period, but in the second period there is an opportunity to change the contract). The contract renegotiation literature is mostly concerned with analyzing the mechanism that maximizes the payoff of one of the contracting parties, for example, the principal. The same applies to Skreta (2006), who considers a buyer-seller model with T periods and discounting, and shows that it is optimal for the principal to offer a price in each period. Our paper is different in that we are concerned with characterizing the set of all outcome functions which could in principle be implemented by the parties. This is important either if the contracting parties have objectives that differ from surplus or profit maximization, for example because one of them can make a relationship specific investment, or, alternatively, if a third party, such as a social planner, wants to implement an outcome in accordance with some broader objectives.

Our analysis is also related to the literature on incomplete information bargaining beginning with Fudenberg and Tirole (1983). Firstly, 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

[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.

contrast, we consider what a mechanism can achieve when it is played before parties enter into such a bargaining game. Also, in our paper bargaining is finite, though we have some discussion of the infinite horizon case in Subsection 3.6. Finite-time bargaining is considered in Fuchs and Skrzypacz (2012): the bargainers face an exogenous deadline at which they receive a fixed share of total surplus. The authors study how the period length of bargaining rounds affects patterns of trade. In contrast we keep the timing fixed and consider how an initial mechanism can affect what is implemented at the deadline. We also consider a more general class of utility functions and therefore need to study a larger set of possible mechanisms, whereas in Ausubel and Deneckere (1989a) and (1989b) and in Fuchs and Skrzypacz (2012) parties need only make price offers.

A recent strand of the literature on the Coase Conjecture is concerned with contract negotiations with limited commitment. In Strulovici (2013) contracting parties negotiate over the contract they wish to adopt over a given status-quo, using an explicit infinite time horizon bargaining protocol with a fixed break-down probability. He shows that, when the probability of break-down approaches zero, all equilibria converge to an equilibrium with efficient contracting. Maestri (2013) studies an infinite time horizon non-durable goods monopoly problem, in which parties can write long-term contracts that are renegotiated before each new trading round. He shows that as parties become infinitely patient the essentially unique equilibrium converges to one in which contracts are efficient subject to incentive compatibility.

The above papers on bargaining and the Coase Conjecture characterize contracting outcomes when negotiation frictions disappear in situations with an infinite time horizon. We, on the other hand, provide a characterization of the outcomes that arise if a mechanism constitutes the status quo of an exogenously given bargaining game. None of the above papers derives our renegotiation-invariance results.

Finally, the paper is related to recent work in 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 sets out the model. Section 3 contains the analysis and results. Subsection 3.1 proves the renegotiation-invariance principle, which is helpful in deriving the necessary conditions. It shows that without loss of generality we can consider equilibria of the form which we construct later. Subsection 3.2 derives necessary conditions for implementation. In Subsection 3.3 we construct an equilibrium for an arbitrary decision schedule which satisfies the necessary conditions and proves the strong implementation (uniqueness) result. Subsection 3.4 discusses the special case in which utility is linear. Subsection 3.5 contains a dicussion of several applications and our main assumptions. Subsection 3.6 has a discussion of infinite horizon bargaining. Section 4 concludes. Some of the proofs are in the Appendix.

2 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 $\underline{\theta} > 0$. Both players are expected utility maximizers and have quasi-linear utility for money. If the decision is $x \in [\underline{x}, \bar{x}]$ 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 $u_x > 0, u_{xx} < 0, u_{x\theta} > 0$ and $u_{xx\theta} > 0$, with subscripts denoting derivatives. 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.

We denote by $\Delta(\Theta)$ the set of distribution functions on Θ. The reservation utility for P and for each type of A is zero.

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) and that mechanisms are non-stochastic. Denote the set of possible mechanisms by Γ.

The parties, however, are not able to commit not to renegotiate the mechanism. After the play of the mechanism they have the option of choosing a further mechanism to play in order to arrive at an outcome which they both prefer. We assume that at this renegotiation stage all of the bargaining power lies with the principal, the uninformed party.[4] 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). Clearly P's optimal mechanism at the renegotiation stage will depend on her updated belief about A which the play of the initial mechanism has generated.

Our aim is to characterize the set of utility schedules which can be implemented by some mechanism taking into account the fact that the mechanism can be renegotiated ex post. The main complication, of course, arises from the fact that the agent, anticipating the renegotiation, will alter his behavior when he plays the initial mechanism.

We include a discussion of our assumptions about renegotiation and the timing of it in Subsection 3.5. We also present there several examples to which our analysis can be applied.

Strategies and Equilibrium

An initial mechanism (M, x_M, t_M) and the post-mechanism stage together define a two-stage game of incomplete information. Call this game $\Phi(M, x_M, t_M)$. We will consider the perfect Bayesian equilibria of this game.

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

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 $G \in \Delta(\Theta)$ over A's types, P will, at the preceding stage, choose a mechanism to offer to A which is optimal for P.

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 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.

Given the above, we can take a pure strategy for P in $\Phi(M, x_M, t_M) = \phi$ 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. Denote by S_P^ϕ the set of P's pure strategies in the game ϕ.

Similarly, we can take a pure strategy for A in $\Phi(M, x_M, t_M) = \phi$ 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[5] for type θ of A is a probability measure $s_A(.|\theta)$ on M. Let the set of these strategies be denoted by S_A^ϕ.

If P's strategy is $s_P \in S_P^\phi$ and A is type $\theta \in \Theta$ and sends $m \in M$, let the post-

[5]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).

renegotiation decision and transfer be denoted by $(x^{\phi}(m, s_P, \theta), t^{\phi}(m, s_P, \theta))$; that is, the mechanism $s_P(m)$ gives this outcome. Then the expected payoff of type θ if he sends message m is $U^{\phi}(m, s_P, \theta) = u(x^{\phi}(m, s_P, \theta), \theta) - t^{\phi}(m, s_P, \theta)$.

For $(x, t) \in [\underline{x}, \bar{x}] \times \Re$ and $G \in \Delta(\Theta)$, let $P((x, t), G) \subseteq D_{IC}(x, t)$ be the set of solutions to the problem

$$max_{(\Theta, x_\Theta, t_\Theta) \in D_{IC}(x,t)} \int_{\underline{\theta}}^{\bar{\theta}} t_\Theta(\theta) - cx_\Theta(\theta) dG(\theta),$$

in which $x_\Theta(.)$ is a right-continuous function.[6] That is, these are the optimal direct revelation mechanisms for P when she has belief G and the default outcome is (x, t).

Definition 1: A *renegotiation equilibrium* (or *r-equilibrium*) of $\Phi(M, x_M, t_M) = \phi$ is a profile of strategies $(s_P, s_A) \in S_P^{\phi} \times S_A^{\phi}$, and, for each $m \in M$, a belief $G(.|m) \in \Delta(\Theta)$ such that

(i) for each $\theta \in \Theta$ $s_A(.|\theta)$ puts probability 1 on messages which maximize $U^{\phi}(m, s_P, \theta)$;

(ii) for each $m \in M$, $s_P(m) \in P((x_M(m), t_M(m)), G(.|m))$;

and

(iii) for each $m \in M$, $G(.|m)$ is consistent with Bayes' Rule, where appropriate, given prior belief F and strategy s_A.

If the strategy profile is (s_A, s_P) then the expected payoff of type θ of A is $U^{\phi}(s_A, s_P, \theta) = \int_m U^{\phi}(m, s_P, \theta) ds_A(m|\theta)$. Let $x^{\phi}(s_A, s_P, \theta)$ be the final decision if the strategy profile is (s_A, s_P). This will be stochastic if s_A is mixed.

Definition 2: (i) A function $U : \Theta \to \Re_+$ is a *r-implementable utility schedule* if there exists a mechanism $(M, x_M, t_M) \in \Gamma$ such that $\Phi(M, x_M, t_M) = \phi$ has a renegotiation equilibrium $(s_A, s_P, \{G(.|m)\}_{m \in M})$ for which, for all $\theta \in \Theta$, $U(\theta) = U^{\phi}(s_A, s_P, \theta)$.

(ii) A function $U : \Theta \to \Re_+$ is *strongly r-implementable* if there exists a mechanism

[6] For any solution in which $x(.)$ is not right-continuous, there is a payoff-equivalent one in which it is.

(M, x_M, t_M) such that, for all $\theta \in \Theta$, $U(\theta) = U^{\phi}(s_A, s_P, \theta)$ for every renegotiation equilibrium $(s_A, s_P, \{G(.|m)\}_{m \in M})$ of $\Phi(M, x_M, t_M) = \phi$.

Definition 3: A function $X : \Theta \to [\underline{x}, \bar{x}]$ is a *r-implementable decision schedule* if there exists a mechanism (M, x_M, t_M) and a renegotiation equilibrium $(s_A, s_P, \{G(.|m)\}_{m \in M})$ of $\Phi(M, x_M, t_M) = \phi$ such that, for all $\theta \in \Theta$, $x^{\phi}(s_A, s_P, \theta) = X(\theta)$ with probability 1.

Associated with a r-implementable decision schedule X is the corresponding transfer schedule $T : \Theta \to \Re$. We refer to the pair (X, T) as a r-implementable outcome function.

The fact that 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 a utility schedule or decision 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 (1993), Milgrom and Segal (2002)) X is c-implementable if and only if $X(\cdot)$ is non-decreasing, and $U \geq 0$ is c-implementable if and only if, for all $\theta \in \Theta$, $U(\theta) - U(\underline{\theta}) = \int_{\underline{\theta}}^{\theta} u_{\theta}(X(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta}$ for some non-decreasing function $X : \Theta \to [\underline{x}, \bar{x}]$. A c-implementable U is absolutely continuous and a.e. differentiable.

Remark It is easy to show, using revelation principle arguments, that if U (resp. X) is r-implementable then U (resp. X) 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 there is strictly positive surplus for each type.

3 Analysis

3.1 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 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, leading to default $(x^*(\theta), t^*(\theta))$, the principal offers the default again, as a fixed outcome. This is an optimal offer because A's type is common knowledge and so the default is known to be efficient. Equally, it is easy to implement P's optimal mechanism given belief F, denoted by $(\Theta, x^F(.), t^F(.))$, that guarantees each type of A at least his reservation utility of 0, using a null initial mechanism - at the second stage the principal will choose $(\Theta, x^F(.), t^F(.))$. The questions we ask are: what other schedules are r-implementable, and how can they be implemented?

Consider P's optimal decision given belief $G \in \Delta(\Theta)$ 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

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

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

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 $\Phi(M, x_M, t_M) = \phi$, where $(M, x_M, t_M) \in \Gamma$.

(i) Take any θ and $\theta' > \theta$. If $m \in supp(s_A(.|\theta))$ and $m' \in supp(s_A(.|\theta'))$ then $x^\phi(m, s_P, \theta) \leq x^\phi(m', s_P, \theta')$;

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

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

Fix a mechanism (M, x_M, t_M) and a r-equilibrium $(s_A, s_P, \{G(.|m)\}_{m \in M})$ of $\Phi(M, x_M, t_M) = \phi$. Lemma 1 implies that for each θ this equilibrium has a deterministic final outcome $(x^{\phi}(s_A, s_P, \theta), t^{\phi}(s_A, s_P, \theta))$. Define an outcome schedule (X, T) by $X(\theta) = x^{\phi}(s_A, s_P, \theta)$ and $T(\theta) = t^{\phi}(s_A, s_P, \theta)$, for $\theta \in \Theta$. This is an incentive-compatible schedule, otherwise some type could profitably deviate by imitating another type over the two-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). In the equilibrium of $\Phi(\Theta, X, T)$ which achieves this the agent will not tell the truth, as he would in the no-renegotiation case, unless X is efficient; this is because, after A has spoken truthfully and revealed himself to be type θ, P would offer an efficient outcome $x^*(\theta)$. Instead, as we show, A randomizes over messages below his true type and, whatever message he sends, the principal will always offer the initial mechanism (Θ, X, T) again.

Proposition 1 (Renegotiation Invariance) For any r-implementable outcome function (X, T) it is possible to implement it by means of the direct revelation mechanism (Θ, X, T) and an equilibrium in which, for each type θ of A, the support of the mixed strategy is a subset of $[\underline{\theta}, \theta]$, and, after any message, P offers the same mechanism, (Θ, X, T).

Proof 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 $G(.|m)$ as G_m and 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 announcement 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 announcement 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

In the equilibrium of Proposition 1, A must randomize over announcements in such a way that P's optimal mechanism is always (Θ, X, T), no matter what announcement A makes. This renegotiation-invariance property is distinct from the renegotiation-proofness principle. In our setting the latter would say that our two-stage game ϕ 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. It remains to discover which mechanisms (Θ, X, T) have the property that this is possible. We examine this question in the following subsections.

3.2 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 restrict attention to decision schedules $X(.)$ which are strictly increasing, differentiable and satisfy $X(\theta) < x^*(\theta)$ for all $\theta < \bar{\theta}$. The next Lemma shows that, for such schedules, any message θ which is sent in the equilibrium described in Proposition 1 is sent by all types above the lowest type which sends θ - after any message, the support of P's belief is of the form $[\theta', \bar{\theta}]$.

Lemma 2 *Suppose (X, T) is r-implementable and X is strictly increasing and satisfies $X(\theta) < x^*(\theta)$ for all $\theta < \bar{\theta}$. 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 $\theta \in supp(s_A)$, $supp(G(.|\theta)) = [\underline{\theta}(G(.|\theta), \bar{\theta}]$ and, if $\theta' \in supp(s_A(\theta_1))$ then $\theta' \in supp(s_A(\theta_2))$ for all $\theta_2 > \theta_1$.*

Consider a schedule (X, T) which satisfies the assumptions of Lemma 2, and such that X is differentiable. (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, $\underline{\theta}$ must put probability 1 on $\underline{\theta}$, i.e., tell the truth, so $\underline{\theta}$ is in the support of A's strategy s_A. Denote $G(.|\underline{\theta})$ by G^X. Then Lemma 2 implies that $supp(G^X) = \Theta$. Furthermore, (X, T) is optimal for P given belief G^X, so (see Myerson (1981), Fudenberg and Tirole (1991)) X must point-wise maximize virtual surplus

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

where g^X is the density of G^X (it can be shown, using a sequence of approximating models with finitely many types, that G^X has a density, and so that, if G^X has an atom, it must be at $\underline{\theta}$). Therefore, for all $\theta > \underline{\theta}$,

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

Since $X(.)$ is differentiable, this implies that g^X is differentiable.

Furthermore, take any other message θ_1 in the support of A's strategy. Let the support of P's belief $G(.|\theta_1)$ be $[\underline{\theta}_1, \bar{\theta}]$. Then it is again optimal for P to offer (Θ, X, T), so $G(.|\theta_1)$ must be the same as G^X, scaled to the support $[\underline{\theta}_1, \bar{\theta}]$, i.e., for $\theta' \in [\underline{\theta}_1, \bar{\theta}]$,

$$G(\theta'|\theta_1) = \frac{G^X(\theta') - G^X(\underline{\theta}_1)}{1 - G^X(\underline{\theta}_1)}$$

and

$$\frac{1 - G(\theta'|\theta_1)}{g(\theta'|\theta_1)} = \frac{1 - G^X(\theta')}{g^X(\theta')}.$$

As Lemma 3 below shows, this, combined with the fact that each type only sends messages below his true type, implies that the hazard rate of G^X is everywhere greater than that of the prior F and that the proportional growth rate of g^X is everywhere less than that of f. Essentially, all types must randomize in a proportionally similar way, in order for P to want to offer the same mechanism no matter what message she receives. However, lower types randomize over a smaller set of messages, so any message θ' is more likely to have been sent by lower types in $[\theta', \bar{\theta}]$ than by higher ones.

Lemma 3 *Let G^X and g^X be defined as above. For all $\theta \in \Theta$, (i)*

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

and (ii)

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

The next proposition gives necessary conditions for $X(\theta)$ to be r-implementable. Recall that x^F is P's optimal decision schedule given belief F.

Proposition 2 *Suppose that $(X(.), T(.))$ is r-implementable and X is strictly increasing and differentiable and satisfies $X(\theta) < x^*(\theta)$ for all $\theta < \bar{\theta}$. Then (i)*

$$\frac{f'(\theta)}{f(\theta)} + A(X(\theta), \theta) + X'(\theta) B(X(\theta), \theta) \geq 0 \tag{2}$$

for all $\theta \in \Theta$, 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

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

(ii) $X(\theta) \geq x^F(\theta)$ for all θ; and (iii) (X, T) gives non negative utilities to the principal and the lowest agent type.

Proof (i) By Lemma 3(ii),

$$\frac{f'(\theta)}{f(\theta)} - \frac{(g^X)'(\theta)}{g^X(\theta)} \geq 0.$$

Since

$$\frac{(g^X)'(\theta)}{g^X(\theta)} = -\frac{g^X(\theta)}{1 - G^X(\theta)} - \frac{\frac{d}{d\theta}\left(\frac{1-G^X(\theta)}{g^X(\theta)}\right)}{\frac{1-G^X(\theta)}{g^X(\theta)}} \tag{3}$$

it follows, using (1), that

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

(ii) follows from Lemma 3(i), (1), the corresponding equation for F and the fact that

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

is decreasing in x if $x < x^*(\theta)$. (iii) must be true to satisfy the participation constraints. QED

The necessary condition (2) 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

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

$u' > 0, u'' < 0$ and, 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 (2) is harder to satisfy if f is falling fast.

In the linear case,[7] in which $u(x,\theta)=\theta x$ and the set of decisions $[\underline{x},\bar{x}]=[0,1]$, $B(x,\theta)=0$ and $A(x,\theta)=2(\theta-c)^{-1}$. Therefore the necessary condition (2) becomes $\theta f'(\theta)+2f(\theta)\geq 0$. Since this is independent of $X'(\theta)$, any increasing function which is above x^F can be implemented as long as the condition is satisfied. The condition is equivalent to concavity of P's revenue function $R(\theta)=\theta(1-F(\theta))$, which in turn is implied by the increasing hazard rate assumption on F.

3.3 Sufficient Conditions for r-implementability

Suppose that an incentive-compatible schedule (X,T) satisfies the conditions of

[7] We discuss the linear case in subsection 3.4 below.

Proposition 2. Is it possible to r-implement it? In this subsection we show that it is. We construct an equilibrium of the type described in Proposition 1. The initial mechanism is (Θ, X, T). Each type θ has a mixed strategy with support $[\underline{\theta}, \theta]$ and a mass point on $\underline{\theta}$. After any announcement, P offers (Θ, X, T) again.

Let $z(\theta) = A(X(\theta), \theta) + X'(\theta)B(X(\theta), \theta)$. The mixed strategy of type θ of A, $s_A(.|\theta)$, is given by the distribution function

$$s_A([\underline{\theta}, \theta']|\theta) = \frac{f(\theta')}{f(\theta)} exp[-\int_{\theta'}^{\theta} z(u)du]$$

for $\theta' \leq \theta$ and $s_A([\underline{\theta}, \theta']|\theta) = 1$ for $\theta' > \theta$. By (2) $-z(\theta)$ is bounded, so the integral is well-defined. The density is then

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

This distribution is well-defined because $f'(\theta') + f(\theta')z(\theta') \geq 0$ by (2).

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

$$G(\theta'|\theta) = \frac{\int_{\theta}^{\theta'} exp[-\int_{\theta}^{u} z(w)dw]du}{\int_{\theta}^{\bar{\theta}} exp[-\int_{\theta}^{u} z(w)dw]du}$$

for $\theta' \geq \theta$ and $G(\theta'|\theta) = 0$ for $\theta' < \theta$.

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

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

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

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

$$\frac{f(\theta')\sigma_A(\theta|\theta')}{\int_\theta^{\bar{\theta}} f(u)\sigma_A(\theta|u)du} = \frac{exp[-\int_\theta^{\theta'} z(w)dw]}{\int_\theta^{\bar{\theta}} exp[-\int_\theta^u z(w)dw]du}$$

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

$$\frac{1 - G(\theta'|\theta)}{g(\theta'|\theta)} = \frac{(u_x(X(\theta'), \theta') - c)}{u_{x\theta}(X(\theta'), \theta')}$$

for every message $\theta \in \Theta$ and every $\theta' \geq \theta$.

Let $k(v) = \int_\theta^v z(w)dw$ for $v \geq \theta$. Then

$$\frac{1 - G(\theta'|\theta)}{g(\theta'|\theta)} = \frac{\int_{\theta'}^{\bar{\theta}} exp[-k(v)]dv}{exp[-k(\theta')]}$$

so we need to show that

$$\int_{\theta'}^{\bar{\theta}} exp[-k(v)]dv = exp[-k(\theta')]\frac{(u_x(X(\theta'), \theta') - c)}{u_{x\theta}(X(\theta'), \theta')}. \quad (4)$$

For $\theta' = \bar{\theta}$, the LHS of (4) 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 $-exp[-k(\theta')]$. The derivative of the RHS is

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

where arguments $(X(\theta'), \theta')$ have been omitted for brevity. Since $k'(\theta') = z(\theta')$, this is equal to $-exp[-k(\theta')]$ and so (4) is true for all θ'. This shows that P's strategy is optimal. Therefore we have:

Proposition 3 *Any incentive-compatible schedule (X, T) such that X is strictly increasing and differentiable, and satisfies $x^F(\theta) \leq X(\theta) < x^*(\theta)$ for $\theta < \bar{\theta}$, $X(\bar{\theta}) =$*

$x^*(\theta)$ and condition (2), is r-implementable.

Proposition 3 establishes that any schedule (X, T) which satisfies the necessary conditions can be implemented by simply giving the parties the incentive-compatible DRM 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 4 *Suppose that (X, T) is an incentive-compatible schedule such that X is strictly increasing and differentiable and satisfies $x^F(\theta) \leq X(\theta) < x^*(\theta)$ for $\theta < \bar{\theta}$, $X(\bar{\theta}) = x^*(\bar{\theta})$ and condition (2). Then the game $\Phi(\Theta, X, T)$ has a unique equilibrium outcome.*

Proof Let $U(\theta)$ be the payoff schedule of the equilibrium of Proposition 3.

By standard results,

$$U(\theta) = U(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u_\theta(X(\tilde{\theta}), \tilde{\theta}) d\tilde{\theta} \tag{5}$$

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 $u(X(\theta), \theta) - T(\theta) = U(\theta)$, it must be that $\tilde{U}(\theta) \geq U(\theta)$ for all $\theta \in \Theta$.

Given $\theta' \in supp(\tilde{s}_A)$, let $\theta'' = min[supp(\tilde{G}(.|\theta')]$. Suppose that $\theta'' \neq \theta'$. Since the lowest type in the support gets zero renegotiation surplus, the equilibrium payoff of type θ'' is the default payoff $u(X(\theta'), \theta'') - T(\theta') < u(X(\theta''), \theta'') - T(\theta'') = U(\theta'')$. Contradiction. Therefore the lowest type which sends message θ' is θ', and this type's equilibrium payoff $\tilde{U}(\theta') = U(\theta')$. This implies that no type sends messages above their true type.

By Lemma 1(iii), 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 $min(supp(\tilde{G}(\cdot|\theta))) = \theta$ if $\theta \in supp(\tilde{s}_A)$ it follows that $\theta \notin supp(\tilde{s}_A)$ if $\theta \in (\theta_1, \theta_2)$, 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 $\tilde{g}_1(\theta) = f(\theta)$ since types in (θ_1, θ_2) only send messages in $[\underline{\theta}, \theta_1]$. Hence, by (1) and the argument in the proof of Proposition 2, $\tilde{X}$ 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,

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

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 (5). 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

$$\frac{1-\tilde{G}_1(\theta)}{g_1(\theta)} = \frac{1-F(\theta)}{f(\theta)}$$

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

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

so $X(\theta) \geq \tilde{X}(\theta)$ for $\theta \in (\theta_1, \bar{\theta})$ since $u_{\theta x} > 0$. By (5) this contradicts the fact that $\tilde{U}(\theta) > U(\theta)$ on this interval. QED

3.4 The Linear Case

One leading case, treated in an earlier version of this paper, is the bilateral trade model, in which the principal is a seller of a unit quantity of a divisible good and the agent is a buyer, type θ of whom has utility θx for quantity x. So $[\underline{x}, \bar{x}] = [0, 1]$ and $u(x, \theta) = \theta x$. x^F is a step function corresponding to a posted price mechanism, equal to zero below some $\hat{\theta}$ and equal to 1 above $\hat{\theta}$. The efficient quantity is 1 (assuming $c < 1$), hence not strictly increasing as in our model above.

Our results above apply also to this case. The density of the mixed strategy defined in the argument leading to Proposition 3 becomes in this case $(f(\theta')(\theta')^2)(f(\theta)\theta^2)^{-1}$ for types θ below a critical value θ^*, and higher types have the same strategy as type θ^*. It is straightforward to show that the principal's updated belief G^X is such that[8]

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

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

and so virtual utility is zero for all types. Therefore P is indifferent between all mechanisms and it is optimal for her to offer the initial mechanism 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.

3.5 Discussion

Here we outline several settings in which our analysis applies and discuss some of its main assumptions. The initial mechanism could be either designed by the principal or by a third party such as a social planner. The choice of the initial mechanism may be modelled by a variety of extensive forms and the appropriate participation constraints will vary accordingly.

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?

To answer this question, consider an example in which the principal is a firm seeking to hire new employees. 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).

Another class of cases arises when the principal would like to induce ex ante investment by the agent in the relationship, leading to a potential hold-up problem. Consider the case in which the agent's utility function is $u(\theta, x) = \theta u(x)$, where $u(\cdot)$ is a strictly increasing and strictly concave function with $u'(0) = \infty$.

Assume that before the realization of the agent's type he can make an unobservable investment costing I that raises his type in the sense of first-order stochastic dominance. That is, his type distribution will be F^0 with continuous density f^0 if he does not make the investment, and it will be F^1 with corresponding density f^1 if he makes the investment, where F^1 first-order stochastically dominates F^0. The support is $[\underline{\theta}, \overline{\theta}]$ in both cases.

If there is no ex-ante mechanism the principal will ex-post offer $(\Theta, x^{F^i}, t^{F^i})$, $(i = 0, 1)$ depending on whether she believes investment has taken place or not. Let $U^i(\theta) = \theta u(x^{F^i}(\theta)) - t^{F^i}(\theta)$.

Assume that the principal's profit is higher with investment than without but that the agent will not invest given mechanism $(\Theta, x^{F^1}, t^{F^1})$, that is,

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

Consequently, the principal would like to design a mechanism ex-ante that in addition to incentive compatibility and individual rationality also takes the agent's investment incentives into account.[9]

It can be shown that the optimal schedule $x^{1^*}(\theta)$ solves

$$\left(\theta - \frac{1 - F^1(\theta)}{f^1(\theta)} + \mu \frac{F^0(\theta) - F^1(\theta)}{f^1(\theta)}\right) u'(x^{1^*}(\theta)) = c,$$

[9]We maintain the assumption that the principal's mechanism has to satisfy interim individual rationality: the principal cannot insist that the agent accepts the mechanism before learning his type, for example because she is facing a population of anonymous agents or because she cannot observe the timing of the realization of agents' types.

where $\mu > 0$ is the Lagrange multiplier associated with the constraint that the agent must want to invest. Since F^1 first-order stochastically dominates F^0, the last term in brackets is positive, which implies that

$$x^*(\theta) \geq x^{1^*}(\theta) \geq x^{F^1}(\theta).$$

In summary, the principal's optimal initial contract is neither the null contract, nor her ex post optimal one.

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 $(\Theta, x(\cdot), t(\cdot))$ 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 $x^{R^*}(\cdot)$ 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.

Our formulation assumes that the principal is able to commit to her second-stage mechanism. One possible reason for this is that from the point at which P and A meet there is, for exogenous reasons such as perishability, a finite time available in which to complete the transaction. A third party, on the other hand, is not able to exploit this deadline because he cannot observe the precise times at which principals and agents meet, or their horizons. More generally, there are many settings in which it is harder for a third party to commit other agents than it is for those agents to commit themselves. Moreover, we conjecture that our results would generalize to other extensive forms, such as bargaining games in which both players discount the future and the principal makes offers at each discrete period over an infinite horizon. We include some remarks on this in the following subsection.

We do not need to assume that the planner can oblige the parties to use his mechanism. Rather 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? Since, in our equilibrium, after every message of the planner's mechanism the principal offers this mechanism again, she would also choose to offer it before the agent plays it, i.e. the planner's mechanism is interim renegotiation-proof. It might be argued that the principal could propose, in order to circumvent renegotiation, to delay and play the planner's mechanism at the deadline, which would solve the renegotiation problem. However, the agent would have no incentive to agree, since renegotiation cannot harm him and in principle could benefit him.

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)).

3.6 Infinite-Horizon Bargaining

In the model above we have assumed that the principal, at the second stage, is able to make a take-it-or-leave-it offer of a mechanism to the agent. We conjecture that our results will generalize in some form to other extensive forms, including those in which the principal has much less commitment power.

For example, consider the following extensive form for the bilateral trade case. As in subsection 3.4, the utility of type θ of the buyer is θx but now the seller has a single indivisible good for sale, with c normalized to zero. Time is discrete and extends over an infinite horizon ($\tau = 0, 1, 2, ...$). The buyer plays the initial mechanism at time 0. If the outcome is that the good is unsold then, beginning at time 0, seller and buyer play a standard infinite-horizon bargaining game in which the seller makes all the price offers, as in Fudenberg, Levine and Tirole (1985). That is, at any time $\tau \geq 0$, if the buyer has not yet accepted any offer, the seller makes a price offer p_τ which the buyer either accepts or rejects. Both players have discount factor $\delta < 1$. Extending Definition 2, we can regard a utility schedule U as r-implementable if there exists a mechanism and a Perfect Bayesian Equilibrium of the induced game (consisting of the mechanism plus subsequent bargaining) in which each type θ of buyer has utility $U(\theta)$.

Take a r-implementable schedule U. The analog of Lemma 1 continues to apply for this model, so the outcome for any type θ is deterministic. It takes the form $(\tau(\theta), t(\theta))$, meaning that trade takes place at time $\tau(\theta)$ and the discounted value of the transfers from buyer to seller is $t(\theta)$, where $U(\theta) = \theta\delta^{\tau(\theta)} - t(\theta)$. In the natural direct revelation mechanism[10] associated with $U(.)$, which we denote by $\gamma(U)$, 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 $p(\theta)$, where $t(\theta) =$

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

$\delta^{\tau(\theta)}p(\theta)$. The types are partitioned into intervals $[\theta_n, \theta_{n-1}), [\theta_{n-1}, \theta_{n-2}), ..., [\theta_0, \theta_{-1}]$, where $\theta_n = \underline{\theta}$ and $\theta_{-1} = \bar{\theta}$, with associated prices $p_n, p_{n-1}, ..., p_1, p_0$, in such a way that for all $\theta \in [\theta_i, \theta_{i-1})$, $\tau(\theta) = i$ and $p(\theta) = p_i$.

Our conjecture is that a version of the Renegotiation Invariance Principle holds in this model. That is, U can be implemented by $\gamma(U)$; the buyer randomizes over all announcements below his true type in such a way that, for any announcement, the seller's beliefs are such that the sequence of prices $p_0, p_1, p_2, ...p_n$ and the sequence of acceptance sets $[\theta_0, \theta_{-1}), [\theta_1, \theta_0), ..., [\theta_n, \theta_{n-1})$ are the equilibrium path of a Perfect Bayesian Equilibrium continuation. If so, a planner who would like to implement a particular utility schedule U which is different from the one resulting from the bargaining game alone can do so by giving the parties the mechanism $\gamma(U)$ and thus induce the required beliefs for the seller. Developing this analysis, and establishing which outcomes can be r-implemented, are left for future work.

4 Conclusion

In this paper we have analyzed the impact of non-cooperative ex-post renegotiation on the set of implementable outcomes in a general mechanism design problem. When parties can commit not to renegotiate a mechanism, any increasing decision rule can be implemented by using a direct revelation mechanism that is designed to elicit the truth from privately informed parties. When this commitment is not possible, the set of implementable rules is restricted because a direct revelation mechanism cannot fully extract all information from the parties. Nevertheless, we have shown that the restriction takes a very simple form - essentially, no type's decision can be reduced by the mechanism, and the slope of the decision function cannot be too high. Furthermore, the direct revelation mechanism which is appropriate for the no-renegotiation case implements the desired outcome in the renegotiation case too.

Appendix

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

$$u(x^{\phi}(m, s_P, \theta), \theta) - t^{\phi}(m, s_P, \theta)) \geq u(x^{\phi}(m', s_P, \theta'), \theta) - t^{\phi}(m', s_P, \theta')$$

and

$$u(x^{\phi}(m', s_P, \theta'), \theta') - t^{\phi}(m', s_P, \theta') \geq u(x^{\phi}(m, s_P, \theta), \theta') - t^{\phi}(m, s_P, \theta).$$

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

(ii) Let $M'(\theta) = \{m \in M | x^{\phi}(m, s_P, \theta) > x^*(\theta)\}$. 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 $x^{\phi}(m, s_P, \theta) > x^{\phi}(m', s_P, \theta)$. Then Lemma 1(ii) implies that $x^{\phi}(m', s_P, \theta) < x^*(\theta)$, 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

$$u(x^{\phi}(m', s_P, \theta_1), \theta_1) - t^{\phi}(m', s_P, \theta_1) = u(x^{\phi}(m', s_P, \theta), \theta_1) - t^{\phi}(m', s_P, \theta) \quad (6)$$

But θ is indifferent between m and m', so

$$u(x^{\phi}(m', s_P, \theta), \theta) - t^{\phi}(m', s_P, \theta) = u(x^{\phi}(m, s_P, \theta), \theta) - t^{\phi}(m, s_P, \theta).$$

Therefore, since $\theta_1 > \theta$ and $x^\phi(m, s_P, \theta) > x^\phi(m', s_P, \theta)$,

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

So, by (6),

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

which contradicts optimality of message m' for θ_1.

Case (b). By Lemma 1(i), $x^\phi(m, s_P, \theta) \leq x^\phi(m', s_P, \theta_i)$ for all $\theta_i \in \{\theta_i\}_{i=1}^\infty$. Right-continuity of $s_P(m')$ implies $x^\phi(m', s_P, \theta) \geq x^\phi(m, s_P, \theta)$. Contradiction. QED

Proof of Lemma 2 In the equilibrium described in Proposition 1, after message θ, P will optimally offer a mechanism which gives the efficient outcome for $\bar{\theta}(G) = max(supp(G(.|\theta)))$, by efficiency at the top. If $\bar{\theta}(G(.|\theta))) < \bar{\theta}$ this implies that she doesn't offer (Θ, X, T). Contradiction. Therefore $\bar{\theta}(G(.|\theta))) = \bar{\theta}$ for any message θ in the support of A's strategy.

Suppose that $\theta_1 \in supp(G(.|\theta))$, $\theta_2 \in supp(G(.|\theta))$, where $\theta_2 > \theta_1$ but $(\theta_1, \theta_2) \cap supp(G(.|\theta)) = \emptyset$. Then, since downward incentive constraints bind in $s_P(\theta)$, 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 IC for the type set Θ and X is strictly increasing. Hence, the support of P's posterior belief is an interval. QED

Proof of Lemma 3 We can take $s_A(.|\theta)$ to have a density on $(\underline{\theta}, \bar{\theta}]$. Denote this density by $\sigma_A(.|\theta)$. Take any θ_1 in the support of s_A and any $\theta_2 > \theta_1$. By Bayes' Rule,

$$[\frac{1 - G(\theta_2|\theta_1)}{g(\theta_2|\theta_1)}] = [\frac{1 - F(\theta_2)}{f(\theta_2)}]\frac{\int_{\theta_2}^{\bar{\theta}} \sigma_A(\theta_1|\theta)h(\theta)d\theta}{\sigma_A(\theta_1|\theta_2)},$$

where

$$h(\theta) = \frac{f(\theta)}{1 - F(\theta_2)}.$$

Hence

$$\sigma_A(\theta_1|\theta_2)(\frac{1-G^X(\theta_2)}{g^X(\theta_2)}) = (\frac{1-F(\theta_2)}{f(\theta_2)})\int_{\theta_2}^{\bar{\theta}} \sigma_A(\theta_1|\theta)h(\theta)d\theta.$$

If $\theta_1 = \underline{\theta}$, the same applies, with s_A replacing σ_A, i.e. probability mass rather than density. Integrating over $\theta_1 \in [\underline{\theta}, \theta_2]$,

$$(\frac{1-G^X(\theta_2)}{g^X(\theta_2)})[s_A(\underline{\theta}|\theta_2) + \int_{\underline{\theta}}^{\theta_2} \sigma_A(\theta|\theta_2)d\theta]$$

$$= (\frac{1-F(\theta_2)}{f(\theta_2)})[\int_{\theta_2}^{\bar{\theta}} s_A(\underline{\theta}|\theta)h(\theta)d\theta + \int_{\underline{\theta}}^{\theta_2}\int_{\theta_2}^{\bar{\theta}} \sigma_A(\theta_1|\theta)h(\theta)d\theta d\theta_1].$$

But

$$s_A(\underline{\theta}|\theta_2) + \int_{\underline{\theta}}^{\theta_2} \sigma_A(\theta|\theta_2)d\theta = 1$$

and

$$s_A(\underline{\theta}|\theta) + \int_{\underline{\theta}}^{\theta_2} \sigma_A(\theta_1|\theta)d\theta_1 \leq 1$$

for $\theta \in (\theta_2, \bar{\theta}]$. Hence $\frac{1-G^X(\theta)}{g^X(\theta}) \leq \frac{1-F(\theta)}{f(\theta)}$. This proves (i).

(ii) Take $\theta' \geq \underline{\theta}$ in the support of s_A, $\theta > \theta'$ and $\delta > 0$. Then

$$\frac{g(\theta+\delta|\theta')}{g(\theta|\theta')} = \frac{f(\theta+\delta)}{f(\theta)}\frac{\sigma_A(\theta'|\theta+\delta)}{\sigma_A(\theta'|\theta)},$$

so

$$\frac{g^X(\theta+\delta)}{g^X(\theta)} = \frac{f(\theta+\delta)}{f(\theta)}\frac{\sigma_A(\theta'|\theta+\delta)}{\sigma_A(\theta'|\theta)},$$

Therefore

$$\frac{\sigma_A(\theta'|\theta+\delta)}{\sigma_A(\theta'|\theta)}$$

is independent of θ' and equal to, say, $\nu(\theta, \delta)$. Similarly,

$$\frac{s_A(\underline{\theta}|\theta+\delta)}{s_A(\underline{\theta}|\theta)} = \nu(\theta, \delta).$$

However,

$$s_A(\underline{\theta}|\theta) + \int_{\underline{\theta}}^{\theta} \sigma_A(\theta'|\theta)d\theta' = 1$$

and

$$s_A(\underline{\theta}|\theta+\delta) + \int_{\underline{\theta}}^{\theta} \sigma_A(\theta'|\theta+\delta)d\theta' \leq 1.$$

Hence

$$\frac{g^X(\theta+\delta)}{g^X(\theta)} \leq \frac{f(\theta+\delta)}{f(\theta)}.$$

Letting $\delta \to 0$, this implies

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

QED

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.

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

Crawford, V. (1985) "Efficient and Durable Decision Rules: A Reformulation", *Econometrica* 53, 817-835.

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

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

Forges, F. (1993) "Some Thoughts on Efficiency and Information", in *Frontiers of Game Theory*, Ed. K. Binmore, A. Kirman, and P. Tani, MIT Press.

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

261.

Fuchs, W. and A. Skrzypacz (2012), "Bargaining with Deadlines and Private Information", forthcoming in *American Economic Journal: Microeconomics*.

Fudenberg, D., D. Levine, and J. Tirole (1985), "Infinite Horizon Models of Bargaining with One-Sided Incomplete Information", in *Game Theoretic Models of Bargaining*, Ed. A. Roth, pp. 73-98, CUP, NY.

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.

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.

Holmström, B. and R. Myerson (1983) "Efficient and Durable Decision Rules with Incomplete Information", *Econometrica* 51, 1799-1819.

Kamenica, E. and M. Gentzkow (2011) "Bayesian Persuasion", *American Economic Review* 101, 2590-2615.

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.

Lagunoff, R. D. (1995) "Resilient Allocation Rules for Bilateral Trade", *Journal* *of Economic Theory* 66, 463-487.

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.

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.

Mussa, M. and S. Rosen (1978) “Monopoly and Product Quality”, *Journal of Economic Theory* 18, 301-17.

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”, mimeo.

Palfrey, T. R. and S. Srivastava (1991) “Efficient Trading Mechanisms with Pre-Play Communication”, *Journal of Economic Theory* 55, 17-40.

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

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

Strulovici, B. (2013) “Contract Negotiation and The Coase Conjecture”, working paper.