UNIVERSITY OF
CAMBRIDGE

Cambridge Working Papers in Economics

Mechanism Design and Non-Cooperative Renegotiation

Robert Evans and Sonje Reiche

CWPE 1331

Mechanism Design and Non-Cooperative Renegotiation

Robert Evans[1] and Sönje Reiche[2]

This version: September 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 be implementable with commitment, 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 with commitment will also implement it in any equilibrium without commitment, 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 (GREMAQ) and University of Cambridge, UK. skr29@cam.ac.uk

1 Introduction

Suppose that the interaction between a number of asymmetrically informed parties is governed by a mechanism which is designed by an outside agency, or planner, in accordance with his objectives. However, the planner is not able to commit the parties fully to the outcome of his mechanism - once the outcome is known, it may be renegotiated by the parties. What is the set of allocations which the planner is able to achieve 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. A third party designs a mechanism to govern their relationship. We have in mind situations in which this designer (the planner) is 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 planner 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 planner puts in place a mechanism in which 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 planner's mechanism. Consequently, we cannot assume that the agent's message in the planner's 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 the planner can achieve 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 commitment case, once the implementable decision schedules have been determined, the implementable expected utility schedules can be derived by integration.

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 commitment case, 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 3) 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 planner's mechanism and the principal simply offered her prior 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 commitment 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 planner's mechanism again after any message. The agent then selects the decision and transfer which he would have chosen had the two players been committed to the mechanism in the first place.

In Proposition 4 we show, by construction, that any decision schedule which satisfies the necessary conditions can be renegotiation-implemented in this way. In Proposition 5, 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-proofness property - after any message, the principal never wants to offer a new mechanism. The planner 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 fea-

ture is that the planner 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 for an uninformed outsider 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. 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 concerned with analyzing the optimal mechanism from the point of view of the principal (one of the two parties to the contract). The same applies to Skreta (2006), who considers a buyer-seller model similar in some ways to ours, but 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 outcome functions which could in principle be implemented by a third party, the planner, whose objectives differ from those of the insiders.

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 and secondly, as noted above, our analysis can be understood as a characterization of what can be achieved by imposing a contract on two bargainers before they begin an exogenous non-cooperative bargaining game.

Another strand of literature to which the paper is related is 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 buyer and the decision maker is the seller, who can only partially commit to her action (the renegotiation price offer). 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 a modified revelation 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. 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 $X = [\underline{x}, \bar{x}] \subseteq \Re_+$, 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 $x \in X$ and A transfers t to P, then P's payoff is $V(x,t) = t - cx$, where $c > 0$, and A's payoff is $U(x,t,\theta) = u(x,\theta) - t$, 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 following assumption about u.

Assumption 1 $u_x(\underline{x}, \bar{\theta}) < c < u_x(\bar{x}, \underline{\theta})$.

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.

A third party (the planner) chooses a mechanism to govern the choice of decision and transfer. A mechanism γ is a triple (M, x, t) consisting of a set of messages M, where M is a metric space, and (abusing notation slightly) a pair of functions $x : M \to X$ and $t : M \to \Re$. A chooses a message $m \in M$. When message m is sent,

$x(m)$ is the contracted decision and $t(m)$ 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 planner, however, is not fully able to commit the parties to the mechanism. Although A and P are obliged to play the planner's mechanism if they want to interact, 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[3]. In other words, once the outcome of the planner's mechanism, $(\tilde{x}, \tilde{t})$, is known, the principal chooses a mechanism to offer to the agent. A can either play this new mechanism or obtain the outcome $(\tilde{x}, \tilde{t})$. Clearly P's optimal mechanism at the renegotiation stage will depend on her updated belief about A which the play of the planner's mechanism has generated.

Discussion

If the planner cared only about maximizing the expected payoff of the principal then he could simply choose a null mechanism (no decision and no payment). At the second stage P would then select a mechanism which is optimal for herself, hence for the planner. More generally, however, the planner, in designing the mechanism, cares about the distribution of utilities and/or decisions across the different types of agent, rather than solely the principal's expected payoff.

For example, consider a case in which A is a buyer, P is a division of a firm, the planner is the headquarters of the firm and the decision x is the quantity of production of a good to be sold to A. The division aims to maximize its own profits; the headquarters, however, is interested both in the profit which the division makes from a particular buyer 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

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

affects future willingness-to-pay. The planner, in order to further its wider objectives, chooses a mechanism (price schedule) for all its retailers to employ. When a customer arrives at a shop and places an order which is not optimal from the salesperson's perspective, i.e. not profit-maximizing given her updated beliefs, the salesperson may have an incentive to offer to sell more at a discounted price.

For another example, the planner may be a government which is regulating the firm (P). The government's objective function is increasing in the firm's profit but it is also interested in the distribution of consumers' utilities, perhaps because willingness-to-pay for the good in question is related to income.

An alternative formulation is that the mechanism designer is the principal, who is interested only in her own expected profit, but expected profit is a function of the distribution of utilities across agent types. For example, as in the hold-up literature, there may be a prior investment stage. Suppose, for example, that the agent first chooses a level of costly unverifiable investment and that higher investment will lead, on average, to a higher marginal utility of x for the agent. P chooses a mechanism which determines a utility schedule (mapping agent type to utility). This determines the agent's prior investment which in turn determines the distribution of types and hence P's expected profit.

In all of the above cases, the first step is to characterize the set of utility schedules which can be implemented by some mechanism. This implementation problem, for the case in which the planner cannot prevent the parties from renegotiating the mechanism ex post, is the subject of this paper. The main complication, of course, arises from the fact that the agent, anticipating the renegotiation, will alter his behavior when he plays the planner's mechanism.

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. The planner, 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. 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 subsection 3.5 below.

We assume above that the principal and agent have to choose between taking their zero-utility outside option and, at least initially, using the planner's mechanism. This is appropriate to the examples described above and also to many others: for example, the case in which the planner is a market designer who constructs an environment in which buyers and sellers meet and interact. Our assumption that the principal, at the second stage, can choose any mechanism that the planner could have chosen seems strong, but, as will become clear, our results will apply as long as the set of mechanisms from which the principal can choose includes the mechanism which the planner has given them - i.e., it is always an option for the principal to offer the same mechanism again.

Strategies and Equilibrium

A planner's mechanism (M, x, t) and the post-mechanism stage together define a game of incomplete information. Call this game $\Phi(M, x, t)$.

Given an outcome $(\tilde{x}, \tilde{t})$ of the planner's mechanism, and a mechanism $\gamma \in \Gamma$ offered by P, A chooses either the default outcome $(\tilde{x}, \tilde{t})$ 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 $D_{IC}(\tilde{x}, \tilde{t})$ be the set of incentive-compatible direct revelation mechanisms which dominate the default outcome $(\tilde{x}, \tilde{t})$ for all types, i.e., mechanisms (Θ, x, t)

such that, for all $\theta, \theta' \in \Theta$,

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

and

$$u(x(\theta), \theta) - t(\theta) \geq u(\tilde{x}, \theta) - \tilde{t}.$$

By the revelation principle, we can assume without loss of generality that P chooses a mechanism in $D_{IC}(\tilde{x}, \tilde{t})$ 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 $\Phi(M, x, t)$ to be a function $s_P : M \to \Gamma$ such that, for $m \in M$, $s_P(m) \in D_{IC}(x(m), t(m))$. We only consider equilibria in which P's strategy is pure. Denote by S_P the set of P's pure strategies.

Similarly, we can take a pure strategy for A in $\Phi(M, x, t)$ 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[4] for type θ of A is a distribution function $s_A(.|\theta)$ on M. Let the set of these strategies be denoted by S_A.

If P's strategy is $s_P \in S_P$ and A is type $\theta \in \Theta$ and sends $m \in M$, let the post-renegotiation decision and transfer be denoted by $(x(m, s_P, \theta), t(m, s_P, \theta))$; that is, the mechanism $s_P(m)$ gives this outcome. Then the expected payoff of type θ if he sends message m is $U(m, s_P, \theta) = u(x(m, s_P, \theta), \theta) - t(m, s_P, \theta)$.

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

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

in which $x(.)$ is a right-continuous function[5].

[4]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 distribution on M conditional on θ. See also Crawford and Sobel (1982).

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

Definition 1: A *renegotiation equilibrium* (or *r-equilibrium*) of $\Phi(M, x, t)$ is a profile of strategies $(s_P, s_A) \in S_P \times S_A$, 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 $U(m, s_P, \theta)$;

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

and

(iii) for each $m \in M$, $G(.|m)$ is the conditional distribution over Θ given F and s_A.

If the strategy profile is (s_A, s_P) then the expected payoff of type θ of A is $U(s_A, s_P, \theta) = \int_m U(m, s_P, \theta) ds_A(m|\theta)$. Let the random variable $x(s_A, s_P, \theta)$ be the final decision if the strategy profile is (s_A, s_P).

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

(ii) A function $U : \Theta \to \Re_+$ is *strongly r-implementable* if there exists a mechanism (M, x, t) such that, for all $\theta \in \Theta$, $U(\theta) = U(s_A, s_P, \theta)$ for every renegotiation equilibrium $(s_A, s_B, \{G(.|m)\}_{m \in M})$ of $\Phi(M, x, t)$.

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

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 to 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 func-

tion $x : \Theta \to X$. A c-implementable U is absolutely continuous and a.e. differentiable. It is easy to show, using revelation principle arguments, that r-implementability and c-implementability are related as follows.

Proposition 1 *If U (resp. x) is r-implementable then U (resp. x) is c-implementable.*

The first-best decision for θ solves the problem

$$max_{x \in X} u(x, \theta) - cx.$$

By our assumptions this has a unique solution which we denote by $x^*(\theta)$. Furthermore, $x^*(.)$ is strictly increasing in θ.

3 Analysis

3.1 A Revelation Principle

It is easy to show that the efficient decision schedule $x^*(.)$ is r-implementable. Take an incentive-compatible direct revelation mechanism which implements it in the commitment 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, as we noted in the Discussion above, it is easy to implement P's ex ante optimal mechanism (i.e., given belief F), which we denote by $(x_F(.), t_F(.))$, using a null mechanism. 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 $(\tilde{x}, \tilde{t})$. Denote the minimum and maximum of $supp(G)$ by $\underline{\theta}(G)$ and $\bar{\theta}(G)$ respectively. If an

incentive-compatible direct revelation mechanism $\{x(\theta), t(\theta)\}_{\theta\in\Theta}$ satisfies

$$u(x(\underline{\theta}(G)), \underline{\theta}(G)) - t(\underline{\theta}(G)) \geq u(\tilde{x}, \underline{\theta}(G)) - \tilde{t}$$

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

$$u(x(\theta), \theta) - t(\theta) \geq u(\tilde{x}, \theta) - \tilde{t}.$$

It follows that choosing P's optimal $(\Theta, x, t) \in D_{IC}(\tilde{x}, \tilde{t})$ 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 $u(\tilde{x}, \underline{\theta}(G)) - \tilde{t}$. Therefore, by standard results, an optimal mechanism $\{x(\theta), t(\theta)\}_{\theta\in\Theta}$ satisfies

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

$$x(\theta) < x^*(\theta) \quad \forall \theta \in supp(G)/\theta^*,$$

and

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

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

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) and are less than or equal to the efficient decisions. It also establishes, 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* $\Phi(M, x, t)$*, where* $(M, x, t) \in \Gamma$*.*

(i) Take any θ *and* $\theta' > \theta$*. If* $m \in supp(s_A(.|\theta))$ *and* $m' \in supp(s_A(.|\theta'))$ *then*

$x(m, s_P, \theta) \leq x(m', s_P, \theta')$;

(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* $t(m, s_P, \theta) = t(m', s_P, \theta)$.

Fix a mechanism $(M, \tilde{x}, \tilde{t})$ and a r-equilibrium $(\tilde{s}_A, \tilde{s}_P, \{\tilde{G}(.|m)\}_{m\in M})$ of $\Phi(M, \tilde{x}, \tilde{t})$. Lemma 1 implies that for each θ this equilibrium has a deterministic final outcome $(x(\tilde{s}_A, \tilde{s}_P, \theta), t(\tilde{s}_A, \tilde{s}_P, \theta))$. Abusing notation slightly, denote this outcome function by $\{(x(\theta), t(\theta))\}_{\theta\in\Theta}$. This is an incentive-compatible schedule, otherwise some type could profitably deviate by imitating another type over the two-stage game. So, for any m, P's proposed mechanism coincides with $\{(x(\theta), t(\theta))\}_{\theta\in supp(\tilde{G}(.|m))}$ for types in $supp(\tilde{G}(.|m))$.

The next proposition gives a modified revelation principle. It shows that the same outcome as is achieved in the given equilibrium (namely $(\{x(\theta), t(\theta)\}_{\theta\in\Theta})$) can also be achieved by giving the parties the direct revelation mechanism (Θ, x, t). It is clear that in the equilibrium of $\Phi(\Theta, x, t)$ which achieves this the agent will not tell the truth, as he would in the commitment case. Instead, he randomizes over messages below his true type and, whatever message he sends, the principal will always offer the planner's mechanism again.

Proposition 2 *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 As above, let $(\tilde{s}_A, \tilde{s}_P, \{\tilde{G}(.|m)\}_{m\in M})$ be an r-equilibrium of $\Phi(M, \tilde{x}, \tilde{t})$ which r-implements the given outcome function (x, t). It is convenient to construct the argument in two steps: first define a mechanism $(\tilde{M}, \hat{x}, \hat{t})$ which r-implements (x, t) and then construct the required direct revelation mechanism.

Let $\tilde{M} = \bigcup_\Theta supp(\tilde{s}_A(.|\theta))$. Define mechanism $(\tilde{M}, \hat{x}, \hat{t})$ by $\hat{x}(m) = x(\underline{\theta}(\tilde{G}(.|m)))$ and $\hat{t}(m) = t(\underline{\theta}(\tilde{G}(.|m)))$. That is, the new mechanism, for message m, gives as

default the final outcome in the original mechanism and equilibrium for the lowest type which sends that message. Then $(\tilde{s}_A, s_P, \tilde{G}(.|m)_{m\in\tilde{M}})$ is an r-equilibrium of $\Phi(\tilde{M}, \hat{x}, \hat{t})$, where, for all $m \in \tilde{M}$, $s_P(m)$ is the direct revelation mechanism (Θ, x, t). To see this, note first that, since A's strategy is the same as in the first equilibrium, $(\tilde{s}_A, \tilde{s}_P)$ and so are P's beliefs, the beliefs satisfy Bayesian updating. For any message, the offered menu is the same, and all possible defaults belong to this menu, so any type of A is indifferent between all messages. Therefore A's strategy is optimal. To see that P's strategy is optimal, consider a message $m \in \tilde{M}$. In the first equilibrium, P chooses an optimal incentive-compatible direct revelation mechanism (IC-DRM) subject to the constraint that the lowest type, $\underline{\theta}(G(.|m))$, gets at least

$$u(\tilde{x}(m), \underline{\theta}(\tilde{G}(.|m))) - \tilde{t}(m).$$

In the new game, P chooses an optimal IC-DRM such that the lowest type gets at least

$$u(x(\underline{\theta}(\tilde{G}(.|m))), \underline{\theta}(\tilde{G}(.|m))) - t(\underline{\theta}(\tilde{G}(.|m))).$$

Since, in the first problem, the lowest type gets zero rent, these two expressions are equal. Also, the beliefs are the same, so the two problems have the same set of solutions. For types outside $supp(\tilde{G}(.|m)$ the offered schedule is arbitrary, as long as overall incentive-compatibility is satisfied. Therefore P's strategy is optimal.

Now suppose, for the second step, that the first-stage mechanism is (Θ, x, t). There is an equilibrium of $\Phi(\Theta, x, t)$ in which, for any message $\theta' \in \Theta$, $s_P(\theta') = (\Theta, x, t)$. That is, after any message, P offers the same mechanism again, and so the decision schedule is x, as in the given equilibrium of $\Phi(M, \tilde{x}, \tilde{t})$. A's strategy s_A in this equilibrium is given by

$$\mu_A(B|\theta) = \tilde{\mu}_A(\{m \in \tilde{M} | \underline{\theta}(\tilde{G}(.|m)) \in B\}|\theta),$$

for all $\theta \in \Theta$ and $B \subseteq \Theta$, where $\mu_A(.|\theta)$ is the measure over Θ corresponding to s_A and $\tilde{\mu}(.|\theta)$ corresponds similarly to $\tilde{s}_A$. In effect, type θ of A randomizes over $\tilde{M}$

according to $\tilde{s}_A(.|\theta)$ and then, given m, reports $\underline{\theta}(\tilde{G}(.|m))$, the lowest type who would send m, according to $\tilde{s}_A$. To see that this is an equilibrium, observe first that A's strategy is optimal because he is indifferent between all messages. To see that P's strategy is optimal, take $\theta \in supp(s_A)$. Let $m(\theta) = \{m \in \tilde{M} | \underline{\theta}(\tilde{G}(.|m)) = \theta\}$. For any $m \in m(\theta)$, we know that P finds it optimal to offer (Θ, x, t) given default $(x(\theta), t(\theta))$ when A's strategy is $\tilde{s}_A$ and A has sent message m. Hence, given default $(x(\theta), t(\theta))$ and announcement θ (which is equivalent to the knowledge that $m = m(\theta)$), P still finds (Θ, x, t) optimal.

Note that for any $m \in supp(\tilde{s}_A(.|\theta))$, $\underline{\theta}(\tilde{G}(.|m)) \leq \theta$, so in this equilibrium type θ only randomizes over types weakly below his true type. QED.

3.2 Necessary Conditions for r-implementability

Proposition 2 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 2 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 2 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 $\theta_2 > \theta_1$.*

Proof In the equilibrium described in Proposition 2, 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

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 2. 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

$$u(x(\theta), \theta) - \frac{1 - G_x(\theta)}{g_x(\theta)} u_\theta(x(\theta), \theta) - cx,$$

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

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

This, together 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 3 *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

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

and (ii) $x(\theta) \geq x_F(\theta)$ *for all* θ.

Proof (i) By Lemma 3(ii),

$$\frac{f'(\theta)}{f(\theta)} - \frac{g'_x(\theta)}{g_x(\theta)} \geq 0.$$

Since

$$\frac{g'_x(\theta)}{g_x(\theta)} = -\frac{g_x(\theta)}{1-G_x(\theta)} - \frac{\frac{d}{d\theta}(\frac{1-G_x(\theta)}{g_x(\theta)})}{\frac{1-G_x(\theta)}{g_x(\theta)}} \tag{3}$$

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 $x < x^*(\theta)$. QED

The necessary condition in Proposition 3(i) 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 an 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 must be negative if

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

so the necessary condition is harder to satisfy if f is falling fast.

In the linear case[6], in which $u(x, \theta) = \theta x$ and the set of decisions $X = [0, 1]$, then $B(x(\theta), \theta) = 0$ and $A(x(\theta), \theta) = 2(\theta - c)^{-1}$. Therefore the necessary condition 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 the 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 Proposition 3. 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 2. The planner's 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

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

for $\theta' \leq \theta$ and $s_A(\theta'|\theta) = 1$ for $\theta' > \theta$. By (2) $-z(\theta)$ is bounded, so this 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 $f'(\theta') + f(\theta')z(\theta') > 0$ by (2).

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

Given message $\theta \in \Theta$, P's belief 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 $\theta' < \theta$.

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 $\theta' \geq \theta$.

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

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

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 4 *Any incentive-compatible schedule (x,t) such that x is strictly increasing and differentiable, 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 4 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 when they can be committed to the mechanism. 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 5 *Suppose given an incentive-compatible schedule (x,t) such that x is strictly increasing and differentiable and satisfies $x_F(\theta) \leq x(\theta) < x^*(\theta)$ for $\theta < \bar{\theta}$, $x(\bar{\theta}) = x^*(\bar{\theta})$ 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 described above.

By standard results,

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

Therefore, if every equilibrium of $\Phi(\Theta, x, t)$ has the same utility schedule then it 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 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 the 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.

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 $min(supp(\tilde{G}(\cdot|\theta))) = \theta$ 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 $\tilde{G}_1$ with density $\tilde{g}_1$. Then, for $\theta \in (\theta_1, \theta_2)$, $\tilde{g}_1(\theta) = f(\theta)$ since types in (θ_1, θ_2) only send messages in $[\underline{\theta}, \theta_1]$. Hence, by the argument in the proof of Proposition 3,

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

for $\theta \in (\theta_1, \theta_2)$, which also implies that $\tilde{x}$ is differentiable on (θ_1, θ_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 the previous 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 $X = [0, 1]$ 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 4 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[7]

$$\frac{1 - G_x(\theta)}{g_x(\theta)} = \theta - c,$$

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 planner's 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 No Commitment by the Principal

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, or no, commitment power.

One such extensive form, in the linear case of subsection 3.4, would be one in which there is an infinite horizon, discrete time, and at each period, if trade has not concluded, P offers a posted price for the whole amount of good available. Suppose that in period 1 A plays the planner's mechanism and, if some of the good remains unsold, P then offers a price, which A either accepts or rejects. In subsequent periods 2,3,... P similarly makes an offer, and A responds. The parties both discount the future. To each possible belief which P might have after the play of the planner's mechanism, we can associate an equilibrium of the bargaining game, and hence a schedule of discounted expected utilities for the various types of buyer. Suppose that the planner wants to implement one of these schedules. Our conjecture is that, as in the model above, it is possible to do so in many cases by giving the parties the direct revelation mechanism which corresponds to that schedule. This is left for future work.

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

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 full commitment is possible, 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 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 commitment case implements the desired outcome in the non-commitment case too.

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, by supermodularity, $x(m', s_P, \theta') \geq x(m, s_P, \theta)$.

(ii) Let $M'(\theta) = \{m \in M | x(m, s_P, \theta) > x^*(\theta)\}$. If $m \in M'(\theta)$ then $\theta \notin supp(G(.|m))$. But

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

Therefore $pr\{\theta \in \Theta | s_A(M'(\theta)) > 0\} = 0$.

(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

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

But θ is indifferent between m and m', so

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

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

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

So, by (6),

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

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 $\theta_i \in \{\theta_i\}_{i=1}^{\infty}$. Right-continuity of $s_P(m')$ implies $x(m', s_P, \theta) \geq x(m, s_P, \theta)$. Contradiction. QED

Proof of Lemma 3 Since G_x has an atom only at $\underline{\theta}$, the same must be true of $s_A(.|\theta)$, which we can take 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

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.

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. 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.

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*

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(2), 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.

Segal, I., and M. Whinston (2002) "The Mirrlees Approach to Mechanism Design with Renegotiation (with Applications to Hold-Up and Risk Sharing)", *Econometrica* 70, 1-45.

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