Mechanism design without commitment* †

Hannu Vartiainen‡
HECER and University of Helsinki

Mar 5, 2013

Abstract

This paper identifies mechanisms that are implementable even when the planner cannot commit to the rules of the mechanism. The standard approach is to require mechanism to be robust against redesign. This often leads nonexistence of acceptable mechanisms. The novelty of this paper to require robustness against redesigns that are themselves robust against redesigns that are themselves robust against... . That is, we allow the planner to costlessly redesign the mechanism any number of times, and identify redesign strategies that are both optimal and dynamically consistent. A mechanism design strategy that credibly implements a direct mechanism after all histories is shown to exist. The framework is applied to bilateral bargaining situations. We demonstrate that a welfare maximizing second best mechanism can be implemented even without commitment.

Keywords: mechanisms, commitment, consistency, optimality, bilateral bargaining.
JEL: C72, D44, D78.

1 Introduction

Mechanism design theory provides powerful tools for the planner to implement desired outcomes in collective choice situations with incomplete information. However, the theory relies on an assumption that is both limiting and, at times, unreasonable: that the planner herself can commit to the mechanism. This assumption

*Preliminary and incomplete.

†I thank Klaus Kultti, Juuso Välimäki, Pauli Murto, and Hannu Salonen for useful discussions and good comments.

‡E-mail: hannu.vartiainen@helsinki.fi. Address: HECER, Arkadiankatu 7, 00014 University of Helsinki, Finland.

is crucial since the incentive compatibility of the mechanism requires that it is played as planned. What exacerbates the problem is that the optimality properties of the mechanism may change when information is being revealed during the play. Hence, given ex post information, another continuation mechanism may begin to dominate the original mechanism and the planner is tempted to change the rules of the game.

In the literature on commitment in mechanism design, the usual approach is to appeal to the *incrutability principle* (Myerson 1991, 1979) by assuming that parties anticipate how the mechanism will be redesigned. As in Neeman and Pavlov (2012), the foreseen renegotiation can then be incorporated into the original mechanism, and the attention can be limited to mechanisms that are robust against renegotiation ex post, given the posterior information.[1] The problem is that not all set-ups admit such a renegotiation-proof mechanism, as the next example demonstrates.

Consider the case of bilateral bargaining. There is a single indivisible good, a buyer, and a seller. Agents' privately known valuations are independently drawn from an interval. By the remarkable result of Myerson and Satterthwaite (1983), there is no incentive compatible, individually rational, and budget balanced mechanism that allocates the good to the agent with the highest valuation. Thus any feasible mechanism occasionally implements the inefficient no-trade outcome. But then the agents are tempted to renegotiate the mechanism rather than follow its instructions whenever no-trade outcome should materialize.

Renegotiation-proofness may thus be thought as a sufficient but not necessary condition for mechanisms that are implementable without commitment. To avoid the existence problems, the criterion of admissible mechanisms must be relaxed. The natural way to do that is to restrict redesigns to new mechanisms that are themselves robust to renegotiation, when exposed to the same criterion as the original mechanism (see also the discussion in Neeman and Pavlov, 2012). This approach - the theme of this paper - provides a way to close the gap between the necessary and sufficient conditions for mechanisms without commitment.

This paper develops a framework to identify implementable mechanisms when the planner cannot commit to the mechanism. Instead, she is permitted to redesign the mechanism *any* number of times without a cost. The key idea is to require robustness against redesigns that are themselves robust against redesigns that are themselves robust against... . The framework is portable to any mechanism design scenario. The structural assumptions that guarantee the existence of the solution are that the agents' type sets are finite and that their preferences exhibit value

[1] See also Forges (1995) and Dewatripont (1989). Other contributions on mechanism design without commitment include Segal and Whinston (2002),Freixas et al. (1985), McAfee and Vincent (1997), Baliga and Sjöström (1997), Bester and Strautz (2001), Skreta (2006, 2011), and Vartiainen (2012).

distinction (no pure belief types). No restrictions (except continuity) are put on the preferences of the planner.

As the starting point we take the observation that potential redesigns take place in sequential order and, hence, can be thought as a strategy. We identify redesign strategies that are *dynamically consistent*. By appealing to the inscrutability principle, our research strategy is to reduce, after each history, the continuation equilibria to a single direct incentive compatible mechanism.[2] In order to do this, we separate the two tasks of a mechanism: information processing and implementation. An information processing device generates a public signal on the basis of the agents' reports, and simulates the information flow in the continuation game.[3] An implementation device then reflects what outcomes are implemented on the basis of revealed information. That is, after communication has been taken place via an information processing device, the planner reconsiders whether to implement the outcome suggested by the implementation device, or to design a new mechanism given the posterior information. Hence she cannot commit to the implementation device. However, no restrictions are put on how she coordinates communication between the parties through the information processing device.

The central question is what conditions should we put on the sequences of direct mechanism that reflect dynamically consistent redesign strategy. In the bilateral bargaining example above, the conditions should embody the intuition that a feasible mechanism is not renegotiated ex post, after the outcome has been revealed, to a new mechanism that is *itself* not subject to renegotiation, and so forth. More generally, after each history, the designer must be able to commit to the mechanism that the strategy assigns to her, given the counterfactual of not doing so.

The planner's mechanism selection strategy must be specified for all histories, compactly summarized by sequences of beliefs. Our solution concept guarantees that, after each history, the chosen mechanism gives the agents the incentives to play truthfully the information processing device and planner the incentives to obediently follow the implementation device. The two conditions that are necessary and sufficient for the mechanism design strategy to meet these desiderata are *optimality* and *consistency*. The former implies that, after all histories, the prescribed mechanism must maximize the planner's preferences among all the mechanisms that are feasible. This condition is dubbed as *Bellman optimality*. The latter condition requires that the mechanism prescribed by the strategy today must not be in conflict with the mechanism prescribed to her in the future.

[2] Incrutability principle: any equilibrium of the mechanism selection can be represented as a direct single stage mechanism that is truthfully played and obediently implemented.

[3] Assuming public signals restrics away private communication. This is a simplification. See Skreta (2006, 2010) for analyses of mechanism design without commitment but with private communication.

Our main result is that a Bellman optimal and consistent mechanism design strategy always exists. The proof, which relies on a fixed point argument, uses history dependent mechanism design strategies. Indeed, there may be no history *in*dependent design strategy that meets the two desiderata.

Our approach highlights the central aspect of the mechanism design problem when the mechanism can be redesigned or renegotiated: it is not only the a priori incentives to reveal information that matter for the design but also how information flows within the mechanism are managed. Information that is revealed along the play may adversely affect the incentives at later stages (in Freixas et al. 1985, this property is called the "ratchet effect") which, given farsighted agents, affects the incentives already at the information revelation stage. How the information processing device should optimally be designed is the central - but difficult - question. The information processing device must be informative enough to allow implementing the desired outcome. But this still leaves much freedom for the designer, and effective solutions often exists. We demonstrate the power of designing information processing devices in the bilateral bargaining context.

Our second result studies mechanisms that can be implemented wihtout commitment in the canonical bargaining set up of Myerson and Satterthwaite (1983). The central question we ask whether the commitment inability rule out the possibility to implement the second best mechanism (Pareto-optimal in the class of incentive compatible, individually rational, and budget balanced mechanisms)? Our answer to this question is the affirmitive: there is a Bellman optimal and consistent mechanism design strategy that implements the *incentive efficient* bargaining mechanism even if the agents do not have any external ways to commit to the inefficient no-trade outcomes. The driving force behind this result is that, by managing what information is being revealed during the bargaining process, the planner can induce a situation ex post where the buyer and the seller can commit not to continue bargaining any further even if they know that mutually beneficial transactions would still be possible. Interestingly, this calls for an information structure that is not as coarse as possible nor as fine as possible, but rather something in the middle. Specifically, the information structure that permits this the one in which the agents conceive it possible that the agents' valuations are equally high, the agents valuations are equally low, or the buyer has the high valuation and the seller the low valuation. We demonstrate that, under such occurrences, the bargainers cannot reliably execute trade as it would require no trade in both the cases where the valuations are equal which cannot be committed to.

The novelty of our approach is that renegotiated mechanism is subjected to the same criticism than the original mechanism but otherwise possible mechanism/communication structures are not restricted. The key difference to Neeman and Pavlov (2012) and Forges (2995) is that they only focus on one-step counterfactuals whereas we account for the infinite hierarchy of counterfactuals. As a consequence, their solutions have more cutting power but suffer from existence problems.

Bester and Strausz (2001) study the one-agent scenario where the principal cannot commit to a certain action after the agent has communicated his type. Their main achievement is in showing that implementable outcomes can still be characterized via a version of the revelation principle. This result, however, heavily relies on the restricted form of the commitment problem. The principal can commit not to employ another mechanism once the agent has communicated his information. In particular, she can commit not to add another layer of mechanism on top of the old one. In contrast, we allow the planner to change the mechanism without restrictions.

Commitment is critical question in the context of bargaining. The famous Coase Theorem asserts that, in the absence of commitment, the uniformed seller cannot commit to selling the good above her own reservation valuation. A mechanism design version of this theorem is provided by Ausubel and Deneckere (1989). McAfee and Vincent (1997) focus on a related question of designing an auction in a multi-agent environment when the seller cannot commit to the reserve price. They obtain a version of the Coase Theorem: when the opportunity cost of waiting vanishes, the seller is forced to sell without a reserve price. Skreta (2006, 2011) studies more auction design when the seller has more flexibility in changing the rules of the game. Allowing remarkably rich strategy set for the seller, she is able to characterize the equilibrium mechanism. Her analysis relies on the assumption that redesigning the game is costly for the seller. Vartiainen (2011) approaches auction design without commitment from another angle. No waiting or other redesign costs are assumed. Applying the same solution as this paper, the key assumption in Vartiainen (2011) is that the information processing device prevents private communication between the seller and any individual agents. It is shown that the unique mechanism that is implementable by using a stationary mechanism design strategy implements the English auction in all cases.

General analyses of mechanism design without commitment include Holmström and Myerson (1983), Green and Laffont (1985), Baliga et al. (1997), and Lagunoff (1992). None of these does, however, address the main question of this paper: how to design mechanism when the planner can change the rules of the game as many times she wishes. The focus of Holmstöm and Myerson (1983) is in the question of ex ante committing to a particular rule. Their criterion "durability" excludes mechanisms that are not robust against a subset of types revealing that they belong to this set by designing a new mechanism for the types in this set. The posterior implementability concept of Green and Laffont (1985) demands that the incentives of the agents must not be sensitive to them understanding which outcome becomes implemented. As in this paper, Baliga et al. (1997) study mechanism design when planner is also a player. However, their focus is in Nash implementation which renders the informational processing property of the mechanism quite different. Lagunoff (1992) studies repeated redesign of complete information mechanism. He aim is to show that, under rather mild conditions, the any outcome that can potentially become implemented is Pareto optimal.

This paper is organized as follows. Section 2 specifies the set up and introduces the solution concept. Section 3 proves the existence of the solution. Section 4 applies the solution to the bilateral bargaining set up, and Section 5 provides concluding discussion.

2 Set up

Preferences There is a set $\{1, ..., n\}$ of agents, a planner, and a *finite* set of physical outcomes X. Agent i's privately known type θ_i is drawn from a *finite* set Θ_i. Write $\Theta = \times_{i \in N} \Theta_i$ with a typical element $\theta = (\theta_i)_{i \in N}$, and $\Theta_{-i} = \times_{j \neq i} \Theta_i$ with a typical element $\theta_{-i} = (\theta_j)_{j \neq i}$.[4] Denote the set of probability distributions on a (countable) set A by ΔA. Denote a typical element of $\Delta\Theta$ by p and by $p_i(\cdot : \theta_i)$ is the conditional distribution over Θ_{-i} given p and the agent i's type θ_i. The support of the probability distribution p is denoted by $\text{supp}(p)$.[5]

Agent i's vNM payoff function is of form $u_i : X \times \Theta_i \to \mathbb{R}$. We assume that the agents' preferences exhibit *value distinction*: For any $\theta_i, \theta_i' \in \Theta_i$ and for any $\theta_{-i} \in \Theta_{-i}$, there are $x, y \in X$ such that $u_i(x, \theta) > u_i(y, \theta)$ and $u_i(x, \theta_i', \theta_{-i}) \leq u_i(y, \theta_i', \theta_{-i})$. This assumption precludes pure belief types (see e.g. Bergemann and Morris, 2005).

The agents and the planner want to maximize their expected payoff. Both expectations are defined with respect to the *outcome function* $f : \Theta \to \Delta X$ that specifies the lottery of outcomes for each type profile. Denote by

$$\Theta^{\Delta X} = \{f : \Theta \to \Delta X\}$$

the set of all outcome functions. Endowed with the uniform metric, $\Theta^{\Delta X}$ is a compact metric space.

Given a common prior over the agents' types $p \in \Delta\Theta$ and an outcome function $f : \Theta \to \Delta X$, agent i's expected payoff is

$$\sum_{\theta_{-i}} \sum_{x} p(\theta_{-i} : \theta_i) \, u_i(x, \theta_i) f(x : \theta) \, .$$

[4] That is, $p_i(\theta_i) = \sum_{\theta_{-i}} p(\theta_i, \theta_{-i})$.

[5] $\text{supp}(p) = \{\theta : p(\theta) > 0\}$.

Payoffs span a complete, transitive, and continuous preference ordering over the elements of $\Theta^{\Delta X} \times \Delta$.

The planners's preferences are captured by a transitive and continuous relation $\succ$ on $\Theta^{\Delta X} \times \Delta$, i.e. the set

$$\{(f', p') \in \Theta^{\Delta X} \times \Delta\Theta : (f', p') \succ (f, p)\}$$

is open for all (f, p) in the topology of the uniform metric.[6] Possible planner's preferences include, but are not restricted to, the following examples.

Example 1 *Pareto preferences:* $(f, p) \succ (f', p')$ *if*

$$\sum_{\theta_{-i}} \sum_{x} p(\theta)\, u_i(x, \theta) f(x : \theta) \geq \sum_{\theta_{-i}} \sum_{x} p'(\theta)\, u_i(x, \theta) f'(x : \theta), \quad \text{for all } i \in N, \text{ for all } \theta_i \in \Theta_i,$$

with at least one strict inequality.

Example 2 *Private preferences: there is a continuous Bernoulli utility function* $v : X \times \Theta \to \mathbb{R}$ *such that* $(f, p) \succ (f', p')$ *if*

$$\sum_{\theta} \sum_{x} p(\theta)\, v(x, \theta) f(x : \theta) > \sum_{\theta} \sum_{x} p'(\theta)\, v(x, \theta) f'(x : \theta).$$

Mechanism Since the information of the agent is private, the planner invokes a *mechanism* to implement an outcome function f. A mechanism does two things: processes information and implements an outcome. In order to study commitment problems of the planner, we separate these tasks. A mechanism, denoted by $\phi = (g, r)$, is then a composite function

$$g \circ r : \Theta \to \Delta X,$$

consisting of an information processing device r and an implementation device g such that

$$r : \Theta \to \Delta S \quad \text{and} \quad g : S \to X,$$

where ΔS is the set of probability distributions over a set S which we assume to be *countably infinite*.

A composite mechanism works as follows. After receiving the agents' messages $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_n)$, the information processing device r generates a (possibly random)

[6] That is, $(f, p) \succ (f', p')$ implies that there is an open neighborhoods B and B' of (f, p) and (f', p'), respectively, such that any element in B is preferred to any element in B'.

public signal $s \in S$ such that $r(s : \hat{\theta}) > 0$. This signal is used by the outcome function g to implement the outcome $g(s) \in X$. The signal s is the only information anyone - including the planner - obtains from r.[7]

Letting $\Theta^{\Delta S} = \{r : \Theta \to \Delta S\}$ and $S^X = \{g : S \to X\}$ denote the sets of information processing devices and implementation devices, respectively, the set of all composite mechanisms is

$$\Phi = \Theta^{\Delta S} \times S^X.$$

A composite mechanism (g, r) *induces* an outcome function f if $f = g \circ r : \Theta \to \Delta X$. That is, for any $x \in X$,

$$f(x : \theta) = \sum_{s \in g^{-1}(x)} r(s : \theta), \qquad \text{for all } \theta \in \Theta.$$

There are many composite mechanism that induce the same outcome function. A particular example is the *direct* mechanism where g is a one-to-one function. This mechanism reveals the least amount of information necessary to implement the outcome specified by the outcome function f. In the other extreme there is the *fully revealing* mechanism that has the property that r is one-to-one. Under such mechanism, the agents fully reveal their types to the designer who then takes an action. It is clear that a fully revealing mechanism is likely to suffer from the planner's commitment problems.. Once the planner becomes informed of all the relevant information, she often is no longer interested in implementing the planned outcome. However, as we demonstrate in Section 5, committing to a mechanism may mean that some information should be induced – the direct mechanism is, in general, not the right mechanism either.

The mechanism (g, r) is *constant under* p if $r(\Theta)$ is singleton. A constant mechanism does not affect the beliefs and implements the same outcome with probability one, i.e. $g(r(\Theta)) = \{x\}$ for some x. Denote the constant mechanism that implements x simply by

$$1_x \in \Phi.$$

Also write $r(\theta) = \{s : r(s : \theta) > 0\}$ and $r(\text{supp}(p)) = \{s : s \in r(\theta)$ and $\theta \in \text{supp}(p)\}$. Given p, a signal $s \in r(\Theta)$ of the information processing device r induces a posterior distribution $p(\cdot : r, s)$ such that

$$p(\theta : r, s) = \frac{p(\theta) r\,(s : \theta)}{\sum_{\theta' \in \Theta} p(\theta') r\,(s : \theta')}, \qquad \text{whenever } s \in \text{supp}(p).$$

To economize on notation, write $p(\cdot : r, s) = p(r, s) \in \Delta\Theta$.

[7] That the implementation device g is deterministic reflects the idea that the designer cannot make partial commitment, e.g. in the probabilistic sense, concerning implementation before the outcome is actually implemented. However, allowing random implementation device would not affect our results.

3 Solution

The planner's problem is that she cannot commit to the implementation device g once the signal s has been produced by the information processing device r. Rather, she may be tempted to design a new mechanism under her post-signal belief. In this section, we identify conditions that the mechanism needs to satisfy for the seller to credibly commit to it.

We solve the mechanisms that can be committed to in two nested parts. First we specify conditions under which, by the revelation principle, the agents could commit to the mechanism. Then we identify conditions under which the planner can commit to the mechanism given that the agents can. This requires defining which mechanism the planner would implement under different ex post beliefs, if they were to materialize.

Agents' incentives In order to study mechanisms that are consistent with the agents' incentives, let us assume that the planner can commit . Then what matters is the outcome function associated to the mechanisms. Given p, player i's payoff from a mechanism $\phi = (g, r)$ depends on the outcome function $g \circ r$ the mechanism induces. The *interim* payoff of i with type θ_i when he reports type θ_i' is then

$$\sum_{\theta_{-i}} \sum_{x} p\left(\theta_{-i} : \theta_i\right) u_i(g(s), \theta_i) r\left(s : \theta_{-i}, \theta_i'\right).$$

By the *revelation principle* (Myerson, 1979), an implementable mechanism (g, r) must be *incentive compatible* (IC):

$$\sum_{\theta_{-i}} \sum_{s} p\left(\theta\right) u_i(g(s), \theta_i) \left[r\left(s : \theta\right) - r\left(s : \theta_{-i}, \theta_i'\right)\right] \geq 0, \quad \text{for all } \theta_i, \theta_i' \in \Theta_i, \text{ for all } i \in N.$$

Denote the set of incentive compatible mechanisms $\phi = (g, r)$ under p by

$$IC(p) = \{\phi : \phi \text{ is incentive compatible under } p\}$$

Truthful announcements form a Bayes-Nash equilibrium in an incentive compatible mechanism $\phi = (g, r)$ *if* the designer can *commit* to follow g after r has performed the information processing task, i.e., the produced signal s. Thus a mechanism maximizing the designer's payoff in $IC(p)$ can be interpreted as the designer's full commitment benchmark. Since incentive compatibility concerns only the payoffs, any signal structure - even the one that fully reveals the buyers' types - is consistent with incentive compatibility. However, such mechanism may not be consistent with the designer's ex post incentives.

Designer's incentives Any implementable mechanism $(g, r) \in \Phi$ must be robust against the designer's temptation to redesign it after observing the signal s from the information processing device r. That is, of replacing the outcome $g(s)$ with *another* mechanism in Φ that is preferred to the outcome $g(s)$ under the posterior belief generated by the signal s. Our task is to identify the conditions under which she will not do that.

We say that a mechanism $(g, r) \in \Phi$ is *ex post dominated* under p by a mechanism $\phi \in \Phi$ if there is a signal $s \in r(\text{supp}(p))$ such that, given the posterior belief $p(s, r)$,

$$(\phi, p(s, r)) \succ (1_{g(s)}, p(s, r)).$$

That is, the designer prefers ϕ over the recommended outcome $g(s)$, given the ex post beliefs due to signal s. In such a case, the original mechanism (g, r) is tempted to redesign the mechanism. The question is whether she can indeed redesign the mechanism by useing another mechanism ex post depends on whether she can commit to the new mechanism. In order to understand this, we need to specify what the designer will do after all possible sequences of redesigns.

The designer can condition her design strategy on the past design history. It is convenient to summarize the public history by a sequence of beliefs

$$(p^0, p^1, ..., p^t),$$

where stage $k+1$ belief p^{k+1} is obtained from stage k belief p^k by updating after observing the signal s^k of the stage k mechanism (g^k, r^k) such that $s^k \in r^k(\text{supp}(p^k))$, i.e.

$$p^{k+1} = p^k(r^k, s^k).$$

Denote by $H = \cup_{t=0}^{\infty}(\Delta\Theta)^t$ the set of all finite public histories.

Let the designer's mechanism design strategy be captured by a *choice rule* σ that specifies her mechanism choice for each history $h \in H$. Since it is without loss of generality to focus on mechanisms that the designer can commit to, we may let the choice rule satisfy

$$\sigma : H \to \Phi \quad \text{such that} \quad \sigma(p^0, ..., p^t) \in IC(p^t), \quad \text{for all } h \in H. \tag{1}$$

Then $\sigma(h)$ represents the mechanism that the designer implements after history h. The *function* $\sigma(\cdot)$ represents the dynamic mechanism selection strategy of the seller.

We now identify properties that the choice rule σ should satisfy. We argue that the sequential rationality of the designer, and the players' knowledge of this, requires that σ reflect internal consistency and optimization. Now we define the set of mechanisms that the designer can commit to today given that σ is followed in the future. Under history h, denote by $C^\sigma(h)$ the designer's *choice set* at history h, given σ. That is, the set of incentive compatible mechanisms that are *not* subject to redesign under the hypothesis that σ is followed ex post:

$$C^{\sigma}(p^0, ..., p^t) = \left\{ (g, r) \in IC(p^t) : \begin{array}{l} \text{not ex post dominated by } \sigma(p^0, ..., p^t, p^t(r, s)), \\ \text{for any } s \in r(\text{supp}(p^t)) \end{array} \right\}.$$

Hence, by the revelation principle, and under the hypothesis that the designer can commit to the choice rule σ :

- A mechanism ϕ is truthfully playable if $\phi \in C^{\sigma}(h)$, since then it will *not* be redesigned ex post.

- A mechanism ϕ is *not* truthfully playable if $\phi \notin C^{\sigma}(h)$, since then it will be redesigned ex post.

Choice set $C^{\sigma}(h)$ is defined with respect to the assumed rule σ. We now formally specify conditions that sequential rationality imposes on the choice rule σ itself. The first condition requires consistency in the sense that employing σ *ex ante* should not contradict σ being employed *ex post*.

Definition 1 (Consistency) *Choice rule σ is* consistent *if $\sigma(p^0, ..., p^t) \in C^{\sigma}(p^0, ..., p^t)$, for all $(p^0, ..., p^t) \in H$.*

The second condition reflects optimality. Given σ and p, the designer should choose a mechanism that maximizes her payoff in the set $C^{\sigma}(h)$.

Definition 2 (Bellman optimality) *Choice rule σ is* Bellman optimal *if, for any $(p^0, ..., p^t) \in H$, there is no $\phi \in C^{\sigma}(p^0, ..., p^t)$ such that $(\phi, p^t) \succ (\sigma(p^0, ..., p^t), p^t)$.*

Under the hypothesis that σ can be committed to in the future, the planner does not want to make a one-shot deviation to σ after any history. Conversely, without Bellman optimality, σ could not be convincingly committed to since the planner is able to make a reliable deviation to it after some history.

Consistency, Bellman-optimality, and ex post dominance play different roles in the solution. Bellman-optimality together with consistency reflect optimization: $\sigma[p]$ maximizes the designer's objective function in $C^{\sigma}[p]$. Ex post dominance in turn guarantees that this act of optimization is consistent with farsightedness. That is, since σ is obeyed in the future, not being ex post dominated with respect to $\sigma[\cdot]$ guarantees that a mechanism can be committed to. The role of ex post dominance is to test whether the designer can commit to a particular mechanism under the hypothesis that σ is followed in the future. In particular, the one-deviation property is *not* implied by ex post dominance.

This solution concept imposes, implictly, restricts beliefs by assuming that the off-equilibrium beliefs are not dependent on the history but only on the current prior and the employed mechanism. Removing this restriction, which could be done in the obvious way, would potentially permit new choice rules to be consistent and Bellman optimal. The reason for the omission is simplicity.

4 Existence

We now state the main result of the paper.

Theorem 1 *A consistent and Bellman optimal mechanism selection strategy* σ *exists.*

The rest of the section is devoted to proving the theorem. To construct a mechanism design strategy σ that meets the two desiderata, we need middle results that say something about the structure of mechanisms that can, even in principle, committed to.

Denote the *graph* of the IC correspondence by $A \subseteq \Phi \times \Delta\Theta$, i.e.

$$A = \{(\phi, p) : \phi \in IC(p) \text{ and } \quad p \in \Delta\Theta\}.$$

We say that $((g, r), p)$ is *ex post dominated in* $D \subseteq A$ if there is $s \in r(\text{supp}(p))$ and $\phi \in IC(p(s, r))$ such that $(\phi, p(s, r)) \in D$, and such that $(\phi, p(s, r)) \succ (1_{g(s)}, p(s, r))$.

We now divide the mechanism/prior pairs to subsets in which the the mechanism could, in principle, be committed to under the respective prior and those that the cannot be committed to, even in principle, under the respective prior. Partition A into sets via the following transfinite recursion:

$$\begin{aligned}
G^0 &= \{(\phi, p) : \text{not ex post dominated in } A\}, \\
B^1 &= \left\{(\phi, p) : \text{ex post dominated in } G^0\right\}, \\
G^1 &= \left\{(\phi, p) : \text{not ex post dominated in } A \backslash B^1\right\}, \\
&\vdots \\
B^\alpha &= \left\{(\phi, p) : \text{ex post dominated in } \cup_{\beta<\alpha} G^\beta\right\}, \\
G^\alpha &= \left\{(\phi, p) : \text{not ex post dominated in } A \backslash \cup_{\beta\leq\alpha} B^\beta\right\}, \\
&\vdots
\end{aligned}$$

Note that $G^\beta \subseteq G^\alpha$ and $B^\beta \subseteq B^\alpha$ for all ordinals $\beta \leq \alpha$. By transfinite recursion, the two sets

$$\begin{aligned}
B &= \cup_\alpha B^\alpha, \\
G &= \cup_\alpha G^\alpha.
\end{aligned}$$

are well defined and disjoint subsets of A. These sets have the fixed point property that

$$\begin{aligned} B &= \{(\phi, p) : \text{ex post dominated in } G\}, \\ G &= \{(\phi, p) : \text{not ex post dominated in } A \backslash B\}. \end{aligned}$$

Not much can be said about the structure of G and B. In particular, while $G \cup B \subseteq A$, nothing guarantees that the inclusion holds as equality.

4.1 An illustrative subcase

Intuitively, mechanism/prior pairs in B can never be committed to and in G can always be committed to, since anything that ex post dominates an element in G^* cannot be committed to. there are two problems in constructing a desired mechanism design strategy. In this subsection, we *assume*, for the reasons of illustration, that G^* *is a closed set.* This simpifies considerably the construction of a mechanism design strategy that meets consistency and Bellman optimality. The assumption is relaxed in the next subsection, where the general existence is proven.

Construct a function $x : \Delta\Theta \to \Delta X$ such that $1_{x(p)}$ maximizes $\succ$ in the class of constant mechanisms under p. That is,

$$(1_x, p) \not\succ (1_{x(p)}, p), \text{ for all } x \in \Delta X. \tag{2}$$

Since $\succ$ is a continuous relation, and X finite, $x(p)$ exists for all p.

Denote by P^G and P^B the sets of probability distributions under which $1_{x(p)}$ *is not* ex post dominated by any element in $A \backslash B$, and *is* ex post dominated by some element in G^*, respectively. Formally,

$$\begin{aligned} P^G &= \{p : (1_{x(p)}, p) \in G\}, & (3) \\ P^B &= \{p : (1_{x(p)}, p) \in B\}. & (4) \end{aligned}$$

By the construction of P^G and P^B, if $p \notin P^G \cup P^B$, then $x(p)$ is ex post dominated under p by some ϕ such that $(\phi, p) \in A \backslash (G \cup B)$.

Now we construct a strategy $\sigma : \cup_{k=1}^{\infty} \Delta^k \to \Phi$ that meets our desiderata. While it is necessary to condition σ on the history, not all the information hidden in the history is needed - it will be sufficient to summarize the history beyond the last two stages in a four element state space $Q = \{0, 1\}$. That is, we describe the history $(p^0, ..., p^k)$ by an array (q, p', p), where the state $q \in Q$ summarizes the information contained by $(p^0, ..., p^{k-2})$, p' is the belief that preceded the current belief, and $p = p^k$ is the current belief.

Before defining the mechanism design strategy on $Q\times\Delta\Theta\times\Delta\Theta$, let us construct two functions. Given sets G and B, let function $\delta:\Delta\Theta\backslash(P^G\cup P^B)\to\Phi$ satisfy

$$(\delta(p),p)\quad\text{ex post dominates}\quad(1_{x(p)},p),\qquad\text{for all } p\in\Delta\Theta\backslash(P^G\cup P^B).\tag{5}$$

By the construction of P^G and P^B, such a function δ does exist. For any $D\subseteq A$, denote by $\mu(p,D)$ a mechanism that maximizes the designer's payoff when she is capable to choose any mechanism in D, given p:

$$\mu(p,D)\in\{\phi:(\phi',p)\not\succ(\phi,p),\ \text{for all } (\phi',p)\in D\}.\tag{6}$$

Continuity of $\succ$ implies that $\mu(p,D)$ is well defined whenever D is compact and there is some ϕ such that $(\phi,p)\in D$.

For any $(g,r)\in\Phi$ and p, denote by $\mathcal{P}((g,r),p)$ the collection of possible posterior beliefs

$$\mathcal{P}((g,r),p):=\{p(s,r)\}_{s\in r(\mathrm{supp}(p))}.$$

Describe the *transition function* between states $\tau:Q\times\Delta\Theta\times\Delta\Theta\to Q$ such that

$$\tau(q,p',p)=\begin{cases}1, & \text{if } p\in P^G,\\ 1, & \text{if } p\in P^B,\\ 0, & \text{if } p'\in\Delta\Theta\backslash(P^G\cup P^B),\ p\in\mathcal{P}(\delta(p'),p'),\ \text{and } q=1,\\ 1, & \text{otherwise.}\end{cases}\tag{7}$$

Construct now a *mechanism design strategy* $\sigma:\cup_{k=1}^{\infty}\Delta\Theta^k\to\Phi$ that is measurable with respect to $Q\times\Delta\Theta\times\Delta\Theta$. Hence we may write $\sigma:Q\times\Delta\Theta\times\Delta\Theta\to\Phi$. Let

$$\sigma(q,p',p)=\begin{cases}\mu(p,G), & \text{if } p\in P^B,\\ 1_{x(p)}, & \text{if } p\in P^G,\\ 1_{x(p)}, & \text{if } p\in\Delta\Theta\backslash(P^G\cup P^B)\text{ and } q=0,\\ \delta(p), & \text{if } p\in\Delta\Theta\backslash(P^G\cup P^B)\text{ and } q=1.\end{cases}\tag{8}$$

Note that this construction is well defined since P^G and P^B do not have common elements.

Lemma 1 *The constructed mechanism design strategy σ is consistent and Bellman optimal.*

Proof. Given p, any $((g,r),p)\in G$ is not ex post dominated by any element in G. Thus, by the transitivity of $\succ$, $(1_{x(p(r,s))},p(r,s))$ is not ex post dominated by any element in G, for any $s\in r(\mathrm{supp}(p))$, and hence $(1_{x(p(r,s))},p(r,s))\in G$. Equivalently, $p(r,s)\in P^G$. By construction, $(1_{x(p(r,s))},p(r,s))=\sigma(q,p,p(r,s))$. Since $((g,r),p)\in G$ is not ex post dominated by $(\sigma(q,p,p(r,s),p(r,s))$, for any

$s \in r(\text{supp}(p))$, it follows that $\{\phi : (\phi, p) \in G\} \subseteq C^\sigma(q, p', p)$, for any (q, p', p). Thus the choice sets are

$$C^\sigma(q,p',p) = \begin{cases} \{\phi : (\phi,p) \in G\}, & \text{if } p \in P^B, \\ \{\phi : (\phi,p) \in G\}, & \text{if } p \in P^G, \\ \{1_{x(p)}\} \cup \{\phi : (\phi,p) \in G\}, & \text{if } \begin{cases} p \in \Delta\Theta \backslash (P^G \cup P^B), \text{ and} \\ q = 0, \end{cases} \\ \{\delta(p)\} \cup \{\phi : (\phi,p) \in G\}, & \text{if } \begin{cases} p \in \Delta\Theta \backslash (P^G \cup P^B), \text{ and} \\ q = 1. \end{cases} \end{cases}$$

It now follows directly from the construction of σ that

$$\sigma(q,p',p) \in C^\sigma(q,p',p), \quad \text{for all } (q,p',p) \in Q \times \Delta\Theta \times \Delta\Theta.$$

We check Bellman optimality of σ in four distinct cases:

- By the definition of $\mu(p,G)$, no element in $\{\phi : (\phi,p) \in G\}$ payoff dominates $(\mu(p,G),p)$ when $p \in P^B$.

- By the definition of P^G, no element in $\{\phi : (\phi,p) \in G\}$ payoff dominates $1_{x(p)}$ when $p \in P^G$.

- By the definition of P^G and P^B, no element in $\{\phi : (\phi,p) \in G\}$ payoff dominates $(1_{x(p)},p)$ when $p \in \Delta\Theta \backslash (P^G \cup P^B)$ and $q = 0$.

- By the construction of δ, $\delta(p)$ payoff dominates $1_{x(p)}$ while, by the definition of P^G and P^B, no element in $\{\phi : (\phi,p) \in G\}$ payoff dominates $(1_{x(p)},p)$ when $p \in \Delta\Theta \backslash (P^G \cup P^B)$. Thus, by the transitivity of $\succ$, no element in $\{\phi : (\phi,p) \in G\}$ payoff dominates $\delta(p)$ when $p \in \Delta\Theta \backslash (P^G \cup P^B)$ and $q = 1$. ■

The first is that we cannot guarantee that the designer can optimize in G as nothing in the construction implies that G is a closed set. The key part of our proof is to show that also B and the *closure* of G are disjoint.

4.2 General case

In this subsection, we prove the existence of the colution in the general case. This amounts to extending the construction (7) - (8) to the case where G is no longer compact and, hence, $\mu(p,G)$ as defined in (6) may not be well defined. Our strategy is to use the closure of G instead. This requires, on the one hand, that the mechanism design strategy is conditioned on a more detailed information. On the other hand, it requires showing that the closure of G and B can be separated. Towards this end, we state two intermediary results.

After fixing the implementation device, the set of information processing devices that, together with this implementation device, constitute an incentive compatible mechanism is simply a subset of a finitely dimensional Euclidean space.

Since this subset is defined by finitely many independent linear inequalities, it is easy to see that the set of incentive compatible information processing devices is both upper and lower hemicontinuous as a correspondence of p at points where p has a full support. However, when p does not have a full support, a more delicate argument is needed, as the incentive constraints on θ_i are suddency removed when $p(\{\theta_i\} \times \Theta_{-i})$ approaches 0.

We say that a sequence of information processing devices $\{r^t\}$ converges to the information processing device r under p if

$$\max_{\theta \in \mathrm{supp}(p)} \left| r^t(s:\theta) - r(s:\theta) \right| \to 0, \quad \text{for all } s \in S.$$

That is, the uniform distance between r^t and r when restricted to the support of $p^{'}$ vanishes as t becomes large.

Lemma 2 *If a sequence $\{p^t\}$ converges to $p \in \Delta\Theta$, and $(g, r) \in IC(p)$, then there is a sequence $\{r^t\}$ of information processing devices such that $(g, r^t) \in IC(p^t)$ for all t that converges to r under p.*

Proof. Take $p \in \Delta\Theta$ and $(g, r) \in IC(p)$. Define a correspondence $R_p(\cdot)$ such that

$$R_p(p') = \left\{ r' \in \Theta^{\Delta S} : \begin{array}{c} \sum_{\theta_{-i}} \sum_s p'(\theta)\, u_i(g(s), \theta_i) \left[r'(s:\theta) - r'(s:\theta_{-i}, \theta_i') \right] \geq 0, \\ \text{for all } \theta_i \in \mathrm{supp}(p_i), \text{ for all } \theta_i' \in \Theta_i, \text{ for all } i \in N \end{array} \right\}.$$

Thus $R_p(p')$ restricts information processing devices only in $\times_i \mathrm{supp}(p_i)$. Since, by the value distinction assumption, the definining constraints are linearly independent, $R_p(\cdot)$ is lower hemicontinuous at any p' such that $\mathrm{supp}(p) \subseteq \mathrm{supp}(p')$.

By construction, $r \in R_p(p)$. Let $\{p^t\}$ converge to p. Since $\mathrm{supp}(p) \subseteq \mathrm{supp}(p^t)$ for all high enough t, the lower hemicontinuity of $R_p(\cdot)$ implies that there is a sequence $\{r^t\}$ that converges to r under p and has the property that $r^t \in R_p(p^t)$ for all t. Since no restrictions are put on the incentives of the types outside $\times_i \mathrm{supp}(p_i)$, we can, by the revelation principle, assume that r^t induces truthtelling also there. Hence, $(g, r^t) \in IC(p^t)$, for all t. ■

We now show that, for all practical purposes, it is without loss of generality to focus on information processing devices r that have a *bounded* support.

Lemma 3 *For any $((g, r), p)$ there is $((g, \bar{r}), p)$ such that (i) $\mathcal{P}((g, \bar{r}), p) \subseteq \mathcal{P}((g, r), p)$, (ii) $\sum_{s \in g^{-1}(x)} r(s:\theta) = \sum_{s \in g^{-1}(x)} \bar{r}(s:\theta)$ for all x, and for all θ, and (iii) $\bar{r}(\Theta)$ contains at most $|X| (|\Theta| + 1) + 1$ elements.*

Proof. Take $((g, r), p)$. Since X is a finite set, if $\{s : r(s : \Theta) > 0,\ g(s) = x\}$ contains at most $|\Theta| + 1$ elements for all $x \in X$, then $r(\Theta)$ itself contains at most $|X|\,(|\Theta|+1)+1$ elements and we are done. Thus suppose that there is x such that $\{s : r(s : \Theta) > 0,\ g(s) = x\}$ contains more than $|\Theta| + 1$ elements.

We construct $\bar{r}$ that meets the desiderata (i) - (iii) with respect to a specific x, i.e. (i) $\{p(\bar{r}, s)\}_{s\in g^{-1}(x)} \subseteq \{p(r, s)\}_{s\in g^{-1}(x)}$, (ii) $\sum_{s\in g^{-1}(x)} r(s : \theta) = \sum_{s\in g^{-1}(x)} \bar{r}(s : \theta)$ for all θ, and (iii) $\{s : \bar{r}(s : \Theta) > 0,\ g(s) = x\}$ contains at most $|\Theta|+1$ elements. Applying the argument to all elements of X proves the lemma.

Given a prior $p \in \Delta\Theta$ denote by $p_x \in \Delta\Theta$ the conditional probability distribution in the event $g^{-1}(x) = \{s : x = g(s)\}$, i.e.

$$p_x(\theta) = \frac{p(\theta) r\left(g^{-1}(x)\} : \theta\right)}{\sum_{\theta'} p(\theta') r\left(g^{-1}(x) : \theta'\right)}, \quad \text{for all } \theta \in \Theta. \tag{9}$$

The posterior probability after observing a signal $s \in g^{-1}(x)$ is

$$p\left(\theta : r, s\right) = \frac{p(\theta) r\left(s : \theta\right)}{\sum_{\theta'} p(\theta') r\left(s : \theta'\right)}, \quad \text{for all } \theta \in \Theta.$$

Then we may write (9) in the form

$$p_x = \sum_{s\in g^{-1}(x)} p\left(r, s\right) \sum_{\theta\in\Theta} p(\theta) r\left(s : \theta\right) \in \Delta\Theta.$$

Since p_x lies in a convex hull of $\{p(r, s)\}_{s\in g^{-1}(x)} \subset \Delta\Theta$, it follows, by Carathéodory's Theorem, that there is $\bar{S} \subseteq g^{-1}(x)$, consisting of $|\Theta| + 1$ or fewer elements, such that p_x lies in the convex hull of $\{p(r, s)\}_{s\in\bar{S}}$. That is, there is a vector $(\lambda_s)_{s\in\bar{S}}$ of nonnegative weights summing to one such that

$$p_x(\theta) = \sum_{s\in\bar{S}} p\left(\theta : r, s\right) \lambda_s, \text{ for all } \theta \in \Theta. \tag{10}$$

Construct $\bar{r}$ such that, for all $\theta \in \Theta$,

$$\bar{r}\left(s : \theta\right) = \begin{cases} r\left(s : \theta\right), & \text{for all } s \in S\backslash g^{-1}(x), \\ \dfrac{r\left(g^{-1}(x) : \theta\right) p\left(\theta : r, s\right) \lambda_s}{p_x(\theta)}, & \text{for all } s \in \bar{S}, \\ 0, & \text{for all } s \in g^{-1}(x)\backslash\bar{S}. \end{cases}$$

To verify that $\bar{r}$ meets (i), it suffices that

$$p(\bar{r}, s) = p(r, s), \quad \text{for all } s \in \bar{S}.$$

To see this, note that, for all $s \in \bar{S}$,

$$\begin{aligned} p\left(\theta : \bar{r}, s\right) &= \frac{p(\theta)\bar{r}(s:\theta)}{\sum_{\theta'} p(\theta')\bar{r}(s:\theta')} \\ &= \frac{p_x(\theta)\bar{r}(s:\theta)/r\left(g^{-1}(x):\theta\right)}{\sum_{\theta'} p_x(\theta')\bar{r}(s:\theta')/r\left(g^{-1}(x):\theta'\right)} \\ &= \frac{p\left(\theta : s, r\right)\lambda_s}{\sum_{\theta'} p\left(\theta' : s, r\right)\lambda_s} \\ &= p\left(\theta : r, s\right), \quad \text{for all } \theta \in \Theta, \end{aligned}$$

where the first equality is by definition of $p\left(\theta : \bar{r}, s\right)$, the second equality by (9), the third equality by the definition of $\bar{r}\left(s : \theta\right)$ on $\bar{S}$, and the final equality from $\sum_{\theta'} p\left(\theta' : s, r\right) = 1$.

To see that $\bar{r}$ meets (ii), note that

$$\begin{aligned} \bar{r}\left(g^{-1}(x):\theta\right) &= \sum_{s\in g^{-1}(x)} \bar{r}\left(s:\theta\right) \\ &= r\left(g^{-1}(x):\theta\right)\frac{\sum_{s\in\bar{S}} p\left(\theta : r, s\right)\lambda_s}{p_x(\theta)} \\ &= r\left(g^{-1}(x):\theta\right), \end{aligned}$$

where the second equality follows from $\bar{r}\left(s:\theta\right) = r\left(s:\theta\right)$ for all $s \notin g^{-1}(x)$ and the third by (10). Moreover, $\{s : \bar{r}(s:\Theta) > 0,\ g(s) = x\} = \bar{S}$, which consists of $|\Theta| + 1$ elements, and hence $\bar{r}$ meets the desideratum (iii). ■

Lemma 4 $\mathrm{cl}G \cap B = \emptyset$.

Proof. Before we prove the claim, note that if a sequence of implementation devices $\{g^t\}$ converges to g, then, since the range of g is finite, $g^t = g$, for high enough t. Moreover, if a sequence of information processing devices $\{r^t\}$ converges to r under p, then there is $s \in$supp(p) such that $s \in r^t(\mathrm{supp}(p))$, for high enough t.

To prove a contradiction, suppose that α is the least ordinal such that there is a sequence $\{((g, r^t), p^t)\} \subseteq G^\alpha$ that converges to $((g, r), p) \in B^\alpha$. We prove a contradiction via two subclaims.

Claim 1: Let sequence $\{((g, r^t), p^t)\} \subseteq G^\alpha$ converge to $((g, r), p) \in B^\alpha$. Then there is an ordinal $\beta \le \alpha$, a signal $s \in r(\mathrm{supp}(p))$, and a sequence $\{(\phi^t, p^t(s, r^t))\} \subseteq B^\beta$ such that $\{(\phi^t, p^t(s, r^t))\}$ converges to $(\phi, p(s, r)) \in G^\beta$.

Proof: By the construction of B^α, there is $\beta \leq \alpha$ and $(\phi, p(s,r)) \in G^\beta$ such that $(\phi, p(s,r)) \succ (1_{g(s)}, p(s,r))$ for some $s \in r(\text{supp}(p))$. Since $\{((g, r^t), p^t)\}$ converges to $((g,r),p)$, it follows that $s \in r^t(\text{supp}(p^t))$ for any high enough t. Then $\{p^t(s, r^t)\}$ converges to $p(s,r)$. By the continuity of $\succ$, it follows that $(\phi, p^t(s, r^t)) \succ (1_{g(s)}, p^t(s, r^t))$, for high enough t. By Lemma 2, there is $\{\phi^t\}$ such that $\phi^t \in IC(p^t(s^t, r^t))$ and such that the information processing device of ϕ^t converges to the information processing device of ϕ under $p(s,r)$. By the continuity of $\succ$, it follows that $(\phi^t, p^t(s, r^t)) \succ (1_{g(s)}, p^t(s, r^t))$, for high enough t. By the construction of G^β, $\{(\phi^t, p^t(s, r^t))\} \subseteq B^\beta$. This proves the claim.

Claim 2: Let sequence $\{((g, r^t), p^t)\} \subseteq B^\beta$ converge to $((g,r),p) \in G^\beta$. Then there is an ordinal $\gamma < \beta$ and a sequence $\{(\phi^t, p^t(s, r^t))\} \subseteq G^\gamma$ such that $s \in r^t(\text{supp}(p))$ for all t, that converges to $(\phi, p(s,r)) \in B^\gamma$ such that $s \in r(\text{supp}(p))$.

Proof: Since $\{((g, r^t), p^t)\}$ converges to $((g,r),p)$, $\{p^t(s, r^t)\}$ converges to $p(s,r)$, for all $s \in S$. By the construction of B^β, there is $\{(\phi^t, p^t(s, r^t))\} \subseteq G^\gamma$ such that $s \in r^t(\text{supp}(p))$ for high enough t, and such that $(\phi^t, p(s, r^t)) \succ (1_{g(s)}, p(s, r^t))$ for all t. By Lemma 3, it is without loss of generality to assume that ϕ^t is drawn from a set $\Theta^{\Delta \bar{S}} \times \bar{S}^X$, where $\bar{S}$ contains finitely many elements. Hence, $\Theta^{\Delta \bar{S}} \times \bar{S}^X$ is a compact set. This implies that the sequence $\{\phi^t\}$ has a subsequence, also denoted by $\{\phi^t\}$, which converges to some $\phi \in \Phi$. Since $\phi^t \in IC(p^t(s, r^t))$ for all t, it follows by the closed graph theorem that ϕ is incentive compatible under $p(s,r)$. By the continuity of $\succ$, $(\phi, p(s,r)) \succ (1_{g(s)}, p(s,r))$. By the construction of G^β, $(\phi, p(s,r)) \in B^\gamma$ for some γ such that $\gamma < \beta$. This proves the claim.

Combining Claims 1 and 2, α cannot be the least ordinal for which it holds that $\{((g, r^t), p^t)\} \subseteq G^\alpha$ converges to $((g,r),p) \in B^\alpha$. The contradiction proves the lemma. ■

Construct sets P^G and P^G as in (3) and (4), respectively, and define functions δ and μ as in (5) and (6), respectively. Describe the *transition function* between states $\tau : Q \times \Delta\Theta \times \Delta\Theta \to Q$ such that :

$$\tau(q, p', p) = \begin{cases} 1, & \text{if } p \in P^G, \\ 1, & \text{if } p \in P^B, \\ 1, & \text{if } p' \in P^B,\ p \in \mathcal{P}(\mu(p', \text{cl}G), p'), \\ 0, & \text{if } \begin{cases} p' \in \Delta\Theta \backslash (P^G \cup P^B), \\ p \in \mathcal{P}(\delta(p'), p'), \\ q = 1, \end{cases} \\ 1, & \text{otherwise.} \end{cases}$$

Note that, starting from some initital probability distrbution p^0, the transition rule τ partitions the set of histories $\cup_{k=1}^{\infty} \Delta\Theta^k$ into two sets which we may associate to the elements of Q. Construct now a *mechanism design strategy* σ that is measurable with respect to this parition such that we may write $\sigma : Q \times \Delta\Theta \times \Delta\Theta \to \Phi$.

Let

$$\sigma(q,p',p) = \begin{cases} \mu(p,\text{cl}G), & \text{if } p \in P^B, \\ 1_{x(p)}, & \text{if } p \in P^G, \\ 1_{x(p)}, & \text{if } p' \in P^B,\ p \in \mathcal{P}(\mu(p',\text{cl}G),p'), \\ 1_{x(p)}, & \text{if } \begin{cases} \text{not } [p' \in P^B \text{ and } p \in \mathcal{P}(\mu(p',\text{cl}G),p')], \\ p \in \Delta\Theta\backslash(P^G \cup P^B), \text{ and} \\ q = 0 \end{cases} \\ \delta(p), & \text{if } \begin{cases} \text{not } [p' \in P^B \text{ and } p \in \mathcal{P}(\mu(p',\text{cl}G),p')], \\ p \in \Delta\Theta\backslash(P^G \cup P^B), \text{ and} \\ q = 1 \end{cases} \end{cases}.$$

Note that this construction is well defined since cl$G$ and B do not have common elements.

Lemma 5 *The constructed mechanism design strategy σ is consistent and Bellman optimal.*

Proof. The corresponding the choice sets are

$$C^\sigma(q,p',p) = \begin{cases} \{\phi : (\phi,p) \in \text{cl}G\}, & \text{if } p \in P^B, \\ \{\phi : (\phi,p) \in \text{cl}G\}, & \text{if } p \in P^G, \\ \{\phi : (\phi,p) \in \text{cl}G\}, & \text{if } p' \in P^B,\ p \in \mathcal{P}(\mu(p',\text{cl}G),p'), \\ \{1_{x(p)}\} \cup \{\phi : (\phi,p) \in G\}, & \text{if } \begin{cases} \text{not } [p' \in P^B \text{ and } p \in \mathcal{P}(\mu(p',\text{cl}G),p')], \\ p \in \Delta\Theta\backslash(P^G \cup P^B), \text{ and} \\ q = 0, \end{cases} \\ \{\delta(p)\} \cup \{\phi : (\phi,p) \in G\}, & \text{if } \begin{cases} \text{not } [p' \in P^B \text{ and } p \in \mathcal{P}(\mu(p',\text{cl}G),p')], \\ p \in \Delta\Theta\backslash(P^G \cup P^B), \text{ and} \\ q = 1. \end{cases} \end{cases}.$$

By the construction of G and B,

$$\sigma(q,p',p) \in C^\sigma(q,p',p), \quad \text{for all } (q,p',p) \in Q \times \Delta \times \Delta.$$

To prove Bellman optimality, let us focus the case $p' \in P^B$, $p \in \mathcal{P}(\mu(p',\text{cl}G),p')$. Since $(\mu(p',\text{cl}G),p')$ is a limit point of G, there is a sequence $\{(\phi^t,p^t)\}$ in G that converges to $(\mu(p',\text{cl}G),p')$. Then $p'' \in \mathcal{P}(\phi^t,p^t)$ implies $(1_{x(p'')},p'') \in G$, for all t. By the continuity of the conditional probability, also $(1_{x(p)},p)$ is a limit point of G. Hence $(1_{x(p)},p) \in$cl$G$. By Lemma 4, $(1_{x(p)},p) \notin B$. Thus $p \notin P^B$ and, by the construction of P^B, no element in $\{\phi : (\phi,p) \in G\}$ ex post dominates $1_{x(p)}$. The remaining cases follow as in Lemma 1. ■

4.3 A note on participation constraints

Players may be permitted to exit the game. However, modeling choices need to be made as regards to when this will be possible. There are two primary possibilities: (i) Once the agents enter a mechanism, they commit to it until the designer changes its rules. At this point, they choose again whether or not enter the new mechanism. (ii) Agents can exit the mechanism at any time.

As is clear from the analysis above, what is sufficient for the existence is the continuity of the relevant set of mechanisms. Previously, this was guaranteed by the value distinction assumption. With participation constraints, this condition to be strenghtened.

Interim individual rationality The first alternative leads to the standard interim participation constraint. Normalizing the value of the outside option of a player to zero, a mechanism $\phi = (g, r)$ is *interim individually rational* if

$$\sum_{\theta_{-i}} \sum_{x} p\,(\theta)\, u_i(g(s), \theta) r\,(s : \theta) \geq 0, \quad \text{for all } \theta_i \in \Theta_i, \text{ for all } i \in N,$$

and it is *ex post individually rational* if[8]

$$u_i(g(s), \theta) \geq 0, \quad \text{for all } s \in r(\theta), \quad \text{for all } \theta \in \Theta, \text{ for all } i \in N.$$

Here, the appropriate strenghtening of the value distinction assumption is *individually rational value distinction*: For any $\theta_i, \theta_i' \in \Theta_i$ and for any $\theta_{-i} \in \Theta_{-i}$, there are $x, y \in X$ such that $u_i(x, \theta) > u_i(y, \theta) \geq 0$ and $u_i(y, \theta_i', \theta_{-i}) \geq u_i(x, \theta_i', \theta_{-i}) \geq 0$. Again, the incentive constraints and the participations constraints are linearly independent for all p, and hence guarantee the continuity of the incentive compatible outcome functions.

Veto-incentive compatibility The problem is that incentive compatibility and ex post individual rationality are not independent: an agent might exercise the veto right after off-equilibrium histories. The following simple extension to incentive compatibility resolves the problem by allowing i to veto the outcome even after his untruthful announcements.[9] Denote by

$$\tilde{u}_i(x, \theta) := \max\{u_i(x, \theta), 0\}$$

[8] *Interim* individual rationality requires that participation be weakly profitable before the output has been realized. Ex post constraint has been analysed e.g. by Forges (1993, 1998) and Gresik (1991, 1996).

[9] Veto-incentive compatibility is due to Forges (1998), and is closely related to IC* of Matthews and Postlewaite (1989).

Given p, a mechanism $\phi = (g, r)$ is *veto-incentive compatible* if

$$\sum_{\theta_{-i} \in \Theta_{-i}} p(\theta) \left[\sum_{s \in S} \tilde{u}_i(g(s), \theta_i) r(s : \theta) - \sum_{s \in S} \tilde{u}_i(g(s), \theta_i) r(s : \theta_{-i}, \theta_i') \right] \geq 0, \quad (11)$$

$$\text{for all } \theta_i, \theta_i' \in \Theta_i, \text{ for all } i \in N.$$

Veto-incentive compatibility requires that truthful reporting forms a Bayes-Nash equilibrium even if vetoing is possible after an untruthful announcement. Any implementable mechanism must thus be veto-incentive compatible. For any p, denote the set of veto-incentive compatible mechanisms by $VIC(p)$. It is easy to see that any veto-incentive compatible mechanism is incentive compatible and ex post individually rational (but not vice versa).[10]

Now individually rational value distinction may no longer suffice

It is clear that, under the assumption that Θ is a finite set, adding participation constraints either in the form of interim individual rationality, or in the form of veto-incentive compatibility would not affect Lemma 2:

5 Application: Bilateral Bargaining

Since Myerson and Satterthewaite (1983), it has been well known that committing to bilateral bargaining mechanisms is difficult. Consider a situation where two agents, a buyer (agent 1) and a seller (agent 2), are about to trade a good. Agents' valuations θ_1 and θ_2 are drawn from the finite set $T = \left\{0, \frac{1}{K}, ..., \frac{K-1}{K}, 1\right\}$, for some $K \in \mathbb{N}$.

Our focus is on mechanisms that satisfy budget balance. The set of possible outcomes is $X = \{0,1\}^2 \times \mathbb{R}$ with a typical element (a, m) where $a = 1$ if trade takes place and $a = 0$ if not, and m is a transfer from the buyer to the seller. Given valuations θ_1, θ_2, the payoffs of the agents from the outcome (a, m) are

$$\begin{aligned} u_1(a, m, \theta_1) &= a\theta_1 - m, \\ u_2(a, m, \theta_2) &= m - a\theta_2. \end{aligned}$$

Let the agents' types be independently distributed according to $p_1 \in \Delta T$ and $p_2 \in \Delta T$. Assume that p_1 and p_2 have a full support.

Assume that the planner maximizes the joint surplus of the agents, as specified in Example 1. Then she has an incentive to continue negotiation whenever there is room for mutually beneficial trading.

An direct mechanism or an outcome function is a mapping $(a, m) : T \times T \to \{0,1\} \times \mathbb{R}$. A mechanism is *ex post efficient* if $a(\theta_1, \theta_2) = 1$ whenever $\theta_1 \geq \theta_2$ and $a = 0$ otherwise. A mechanism is inefficient if it is not ex post efficient.

[10]Choose $\theta_i = \theta_i'$ in (11). We only need EXP-IR and IC in the remainder of the paper.

By construction

$$\begin{aligned} U_p\,(a,m),\theta_1:\theta_1') &= \sum_{\theta_2} p_2(\theta_2)[a(\theta_1',\theta_2)\theta_1 - m(\theta_1',\theta_2)], \\ U_p\,((a,m),\theta_2:\theta_2') &= \sum_{\theta_2} p_1(\theta_1)[m(\theta_1,\theta_2') - a(r(\theta_1,\theta_2')\theta_2]. \end{aligned}$$

A direct mechanism is incentive compatible if

$$U_p\,((a,m),\theta_i:\theta_i) \geq U_p\,((a,m),\theta_1:\theta_1')\,,\ \text{for all}\ \theta_i,\theta_i' \in T,\ \text{for all}\ i=1,2,$$

and it is (interim) individually rational if

$$U_p\,((a,m),\theta_i:\theta_i) \geq 0,\ \text{for all}\ \theta_i \in T,\ \text{for all}\ i=1,2$$

A direct mechanism (a,m) is *incentive efficient* if there is no other incentive compatible, individually rational, and budget balanced mechanism that generates higher expected payoffs to both the agents.

The classic result due to Myerson and Satterthwaite (1983) is that incentive constraints prevent full efficiency: any incentive compatible, individually rational, and budget balanced mechanism implements an inefficient outcome with strictly positive probability. In particular, any incentive efficient mechanism is inefficient. The inefficiency of any incentive feasible mechanism raises the question of renegotiation. Would the parties stop bargaining once they know that all mutually beneficial transactions are not exhausted?

The aim of this section is to show that the agents' inability to commit to the mechanism does not prevent them implementing an incentive efficient contract, i.e. there is a consistent and Bellman optimal mechanism design rule that allows committing even to the inefficient outcome. This requires a carefully designed information processing device.

One problem with our endeavour is that the classic characterization results of Myerson and Satterthwaite (1983) are derived in the context where the set of valuations is a continuum (an interval). In our set up, however, the set of valuations T is discrete. This is not a completely innocent modification of the model since the original Myerson-Satterthwaite (1993) result famously relies on the envelope argument, and hence requires the set of types to be connected. Thus we have to develop an analogous argument which is applicable to the discrete case, and which permits the characterization. This is what we do next.

Given p_i, denote the cumulative distribution by

$$P_i(\theta_i) = \sum_{t\leq\theta_i} p_i(t), \qquad \text{for}\ i=1,2,$$

and, for any $\gamma \in [0,1]$,

$$\begin{aligned} c_1(\theta_1,\gamma) &= \theta_1 - \gamma\frac{1-P_1(\theta_1)}{p_1(\theta_1)}, \quad \text{for all } \theta_1 \in T, \\ c_2(\theta_2,\gamma) &= \theta_2 + \gamma\frac{P_2(\theta_2)}{p_2(\theta_2)}, \quad \text{for all } \theta_2 \in T. \end{aligned}$$

We say that the two distributions p_1 and p_2 are *regular* if $c_1(\cdot,1)$ and $c_2(\cdot,1)$ are increasing.

We are now ready to establish the finite versions of the classic result of Myerson and Satterthwaite (1983). The proof of the proposition is relegated to the appendix.

Proposition 1 *If p_1 and p_2 are regular, then there is an* incentive efficient *direct mechanism (a^γ, m^γ) such that, for some $\gamma \in (0,1]$,*

$$\begin{aligned} &\text{if} \quad c_1(\theta_1,\gamma) - c_2(\theta_2,\gamma) \geq 0, \quad \text{then} \quad a^\gamma(\theta_1,\theta_2) = 1 &(12) \\ &\text{if} \quad c_1(\theta_1,\gamma) - c_2(\theta_2,\gamma) < 0, \quad \text{then} \quad a^\gamma(\theta_1,\theta_2) = 0. &(13) \end{aligned}$$

From this result it is clear that, with sufficiently fine grid in T, the incentive efficient direct mechanism (a^γ, m^γ) will be inefficient: an inefficient no-trade outcome will materialize whenever

$$\gamma\frac{1-P_1(\theta_1)}{p_1(\theta_1)} + \gamma\frac{P_2(\theta_2)}{p_2(\theta_2)} > \theta_1 - \theta_2 > 0.$$

We make two observations on the incentive efficient mechanism. These properties will be used to construct a mechanism on which the planner can commit to.

Remark 1 *Let p_1 and p_2 be regular distributions. Let (a^γ, m^γ) be an incentive efficient direct mechanism as defined in (12)-(13). Then, for any $(\theta_1,\theta_2) \in T \times T$,*

$$a^\gamma(\theta_1,\theta_2) = 0 \quad \text{implies} \quad \begin{cases} a^\gamma(\theta_1',\theta_2) = 0, & \text{for all } \theta_1' \leq \theta_1, \\ a^\gamma(\theta_1,\theta_2') = 0, & \text{for all } \theta_2' \geq \theta_2. \end{cases}$$

In particular, $\theta_1 > \theta_2$ and $a(\theta_1,\theta_2) = 0$ imply $a(\theta_1,\theta_1) = a(\theta_2,\theta_2) = 0$.

Our aim is to construct a mechanism that allows the parties to commit not to continue negotiation even when trade does not take place. To this end, the information processing device of the mechanism must be designed such a way that the prescribed outcome can be committed to under the posterior information.

Since the information structure with respect to the outcome function (a, m) is measurable is at most as coarse than that of r, we need to verify that that the outcome of the optimal mechanism does itself reveal unintended information. For our purposes, it suffices that there is an efficient mechanism that prescribes zero monetary transfer when trade does not take place. The no-trade outcome then only reveals that the types of the agents (θ_1, θ_2) satisfy (13).

This guarantees that, when trade does not take place, only this information is revealed. Gresik (1991) establishes the existence of such transfers in the continuous type sets case. For completeness, we construct such schemes in the current case when the types sets are finite. The proof of the following lemma appears in the appendix.

Lemma 6 *Let p_1 and p_2 be regular distributions. Then there is an incentive efficient direct mechanism (a^γ, m^γ) as defined in (12)-(13) such that the transfer rule m^γ prescribes zero monetary transfer when trade does not take place, i.e.*

$$a^\gamma(\theta_1, \theta_2) = 0 \quad \textit{implies} \quad m^\gamma(\theta_1, \theta_2) = 0.$$

Our question is whether there is a Bellman optimal and consistent mechanism choice rule that permits implementation of a compound mechanism that is outcome equivalent with the incentive efficient mechanism (a^γ, m^γ). We shall show that this is the case.

We are now ready to state the desired result: the agents can commit to implementing the Myerson-Satterthwaite incentive efficient mechanism in the bilateral bargaining context even in the absence of external commitment devices. This entails that the agents design an information processing device through which their communication takes place in a way that they cannot commit not to continue bargaining after it becomes clear that the inefficient no-trade outcome will become implemented.

Theorem 2 *Let p_1 and p_2 be regular distributions. Then there is a Bellman optimal and consistent mechanism design strategy σ that implements the incentive efficient mechanism under (p_1, p_2).*

The remainder of this section proves the result. Our key task is to construct an information processing device which provides just the right amount of information for the agents to commit to the inefficient no-trade outcome.

There are many ways to for the information precessing device r to provide enough information for the mechanism to work properly. Our central task is to design r in such a way that it blocks further negotiation but still permits implementation the outcomes prescribed by the incentive efficient mechanism (a^γ, m^γ).

Let the signal space be defined by ordered pairs

$$S^* = \{\langle \theta_1, \theta_2 \rangle : \theta_1 \geq \theta_2\} \cup \{0\}.$$

Consider the following information processing device r^*. For any t, let $\kappa(t) = \#\{t' : t \geq t' \text{ and } c_1(t,\gamma) \leq c_2(t',\gamma) \text{ or } t \geq t' \text{ and } c_1(t',\gamma) \leq c_2(t,\gamma)\}$. Then

$$r^*(\cdot : \theta_1, \theta_2) = \begin{cases} 1_{\langle \theta_1, \theta_2 \rangle}, & \text{if } \theta_1 > \theta_2, \\ \frac{1}{\kappa(t)} \left(\sum_{t' : t' \leq t \text{ and } c_1(t,\gamma) \leq c_2(t',\gamma)} 1_{\langle t, t' \rangle} + \sum_{t' : t' \geq t \text{ and } c_1(t',\gamma) \leq c_2(t,\gamma)} 1_{\langle t', t \rangle}) \right), & \text{if } \theta_1 = \theta_2 = t. \\ 1_0 & \text{if } \theta_1 < \theta_2. \end{cases}$$

That is, a signal $\langle \theta_1, \theta_2 \rangle$ such that $c_1(\theta_1, \gamma) \geq c_2(\theta_2, \gamma)$ may be sent only by the type pair (θ_1, θ_2), and a signal $\langle \theta_1, \theta_2 \rangle$ such that $c_1(\theta_1, \gamma) < c_2(\theta_2, \gamma)$ and $\theta_1 \geq \theta_2$ may be sent by type pairs $(\theta_1, \theta_2), (\theta_1, \theta_1)$,or (θ_2, θ_2). A signal "0" may only be send by a type pair (θ_1, θ_2) such that $\theta_1 < \theta_2$.

Further, define an implementation device (a^*, m^*) such that, for any $s \in S^*$,

$$(a^*, m^*)(s) = \begin{cases} (1, m^\gamma(\theta_1, \theta_2)), & \text{if } s = \langle \theta_1, \theta_2 \rangle \text{ and } c_1(\theta_1, \gamma) - c_2(\theta_2, \gamma) \geq 0, \\ (0, 0), & \text{if } s = \langle \theta_1, \theta_2 \rangle \text{ and } c_1(\theta_1, \gamma) - c_2(\theta_2, \gamma) < 0, \\ (0, 0), & \text{if } s = 0. \end{cases}$$

By construction, the compound mechanism $((a^*, m^*), r^*)$ satisfies

$$\begin{array}{llll} \text{if} & c_1(\theta_1, \gamma) - c_2(\theta_2, \gamma) \geq 0, & \text{then} & (a^*, m^*)(r^*(\theta_1, \theta_2)) = (1, m^\gamma(\theta_1, \theta_2)), \\ \text{if} & c_1(\theta_1, \gamma) - c_2(\theta_2, \gamma) < 0, & \text{then} & (a^*, m^*)(r^*(\theta_1, \theta_2)) = (0, 0). \end{array}$$

By Lemma 6,

$$(a^*, m^*)(r^*(\cdot)) = (a^\gamma, m^\gamma)(\cdot).$$

By Proposition 1, (a^γ, m^γ) is an incentive efficient mechanism when p_1 and p_2 are regular, and hence we conclude:

Lemma 7 *Let p_1 and p_2 be regular distributions. Then the compound mechanism $((a^*, m^*), r^*)$ is incentive efficient.*

Our aim is to show that the mechanism $((a^*, m^*), r^*)$ can be committed under regular distributions (p_1, p_2). To show this, we have to construct a mechanism design strategy σ that is consistent and Bellman optimal, and implements $((a^*, m^*), r^*)$ under (p_1, p_2).

Note first that when $s = \langle \theta_1, \theta_2 \rangle$ such that $c_1(\theta_1, \gamma) \geq c_2(\theta_2, \gamma)$ or $s = 0$, the implemented outcome $a^*(s)$ is ex post efficient. Since there is no mechanism that ex post dominates such an outcome, the only issue is whether the agents can commit to the inefficient no-trade outcome, i.e. when $s = \langle\theta_1, \theta_2\rangle$ such that $\theta_1 > \theta_2$ and $c_1(\theta_1, \gamma) < c_2(\theta_2, \gamma)$. We need to consider the posterior belief that is induced by such a signal.

Note that an information processing device r^* may send a signal $s = \langle\bar{t}, \underline{t}\rangle$ such that $\bar{t} > \underline{t}$ and $c_1(\bar{t}, \gamma) < c_2(\underline{t}, \gamma)$ under the following ordered pairs of types (θ_1, θ_2) : $(\bar{t}, \bar{t}), (\bar{t}, \underline{t}), (\underline{t}, \underline{t})$. This implies that the signal $\langle\bar{t}, \underline{t}\rangle$ induces a posterior belief $p(r^*, \langle\bar{t}, \underline{t}\rangle)$ such that

$$\text{supp}(p(r^*, \langle\bar{t}, \underline{t}\rangle)) = \{(\bar{t}, \bar{t}), (\bar{t}, \underline{t}), (\underline{t}, \underline{t})\}.$$

Our task is to construct a mechanism selection rule σ^* such that there is no credible way to continue bargaining under the belief $p(r^*, \langle\bar{t}, \underline{t}\rangle)$ even though a mutually profitable trading opportunity exists with strictly positive probability.

Lemma 8 *Let supp$(p) = \{(\bar{t}, \bar{t}), (\bar{t}, \underline{t}), (\underline{t}, \underline{t})\}$ with $\bar{t} > \underline{t}$. Then there is a Bellman optimal and consistent choice rule σ^* such that $\sigma^*(p)$ implements $(0, 0)$ with probability one.*

Proof. We construct a σ^* that satisfies Bellman optimality and consistency on $\text{supp}(p) = \{(\bar{t}, \bar{t}), (\bar{t}, \underline{t}), (\underline{t}, \underline{t})\}$. Construct a strategy $\sigma^*(\cdot)(q, p')$ that depends on the distribution $p' \in \Delta\{(\bar{t}, \bar{t}), (\bar{t}, \underline{t}), (\underline{t}, \underline{t})\}$ and a state variable $q \in \{0, 1\}$. If σ^* is independent on q under p', i.e. $\sigma^*(\cdot)(0, p') = \sigma^*(\cdot)(1, p')$, then we write simply $\sigma^*(\cdot)(p')$.

(i) Consider first the cases where p' is degenerate. Let σ^* be defined as follows:

$$\text{if} \quad \text{supp}(p') = \begin{cases} \{(\bar{t}, \bar{t})\}, & \text{then } \sigma^*(\bar{t}, \bar{t})(p') = 1_{(1,\bar{t})}, \\ \{(\bar{t}, \underline{t})\}, & \text{then } \sigma^*(\bar{t}, \underline{t})(p') = 1_{(1,(\bar{t}+\underline{t})/2)}, \\ \{(\underline{t}, \underline{t})\}, & \text{then } \sigma^*(\underline{t}, \underline{t})(p') = 1_{(1,\underline{t})}. \end{cases} \tag{14}$$

Then σ^* satisfies consistency and Bellman optimality at any degenerate p'.

(ii) Consider then the case such that $\text{supp}(p') \subsetneq \{(\bar{t}, \bar{t}), (\bar{t}, \underline{t}), (\underline{t}, \underline{t})\}$. By (14), there is a unique incentive compatible and ex post individually rational mechanism $\sigma^*(\cdot)(p')$ that is not ex post dominated by σ^* at cases where the posterior is degenerate. More specifically,

$$\text{if } \text{supp}(p') = \begin{cases} \{(\bar{t}, \underline{t}), (\underline{t}, \underline{t})\}, & \text{then } \sigma^*(\theta)(p) = 1_{(1,\underline{t})}, \text{ for all } \theta \in \{\underline{t}\} \times \{\underline{t}, \bar{t}\}, \\ \{(\bar{t}, \bar{t}), (\bar{t}, \underline{t})\}, & \text{then } \sigma^*(\theta)(p) = 1_{(1,\bar{t})}, \text{ for all } \theta \in \{\underline{t}, \bar{t}\} \times \{\bar{t}\}, \\ \{(\bar{t}, \bar{t}), (\underline{t}, \underline{t})\}, & \text{then } \sigma^*(\theta)(p') = \begin{cases} 1_{(1,\underline{t})}, & \text{for } \theta = (\underline{t}, \underline{t}), \\ 1_{(1,\bar{t})}, & \text{for } \theta = (\bar{t}, \bar{t}), \\ 1_{(0,0)}, & \text{for all } \theta \in \{(\underline{t}, \bar{t}), (\bar{t}, \underline{t})\}. \end{cases} \end{cases}$$

(iii) Consider finally the case such that $\text{supp}(p') = \{(\bar{t}, \bar{t}), (\bar{t}, \underline{t}), (\underline{t}, \underline{t})\}$. In this case, the mechanism is dependent on the state $q \in \{0, 1\}$. Let first a transition function be defined by $\tau : \{0,1\} \times \Delta\{(\bar{t},\bar{t}),(\bar{t},\underline{t}),(\underline{t},\underline{t})\} \to \{0,1\}$ such that $\tau(q,p') \neq q$ whenever $\text{supp}(p') = \{(\bar{t},\bar{t}),(\bar{t},\underline{t}),(\underline{t},\underline{t})\}$, and $\tau(q,p') = q$ otherwise. To define the mechanism design rule, let $\lambda \in (0,1)$ and choose $\{\sigma^*(\cdot)(q,p')\}_{q\in\{0,1\}}$ such that

$$\begin{aligned}
\sigma^*(\theta)(0,p') &= 1_{(0,0)}, \quad \text{for all } \theta \in \{\underline{t},\bar{t}\} \times \{\underline{t},\bar{t}\}, \\
\sigma^*(\theta)(1,p') &= \begin{cases} \lambda \cdot 1_{(1,(\bar{t}+\underline{t})/2)} + (1-\lambda)\cdot 1_{(0,0)}, & \text{for } \theta = (\bar{t},\underline{t}) \\ 1_{(0,0)}, & \text{for all } \theta \in \{(\bar{t},\bar{t}),(\underline{t},\bar{t}),(\underline{t},\underline{t})\}. \end{cases}
\end{aligned}$$

We claim that the constructed choice rule (τ,σ^*) is consistent and Bellman optimal. By (i) and (ii), we only need to check the case $\text{supp}(p') = \{(\bar{t},\bar{t}),(\bar{t},\underline{t}),(\underline{t},\underline{t})\}$. We first claim that there is no ex post efficient choice function f that is also incentive compatible and ex post individually rational.

Suppose the above claim is false. Note that ex post efficiency and ex post individual rationality imply that $f(\bar{t},\bar{t}) = 1_{(1,\bar{t})}$, $f(\underline{t},\underline{t}) = 1_{(1,\underline{t})}$, and $f(\bar{t},\underline{t}) = 1_{(1,m)}$, for some $\underline{t} \leq m \leq \bar{t}$. The buyer's incentive compatibility implies that it cannot be the case that $m < \bar{t}$ and by the seller's incentive compatibility that it cannot be the case that $\underline{t} < m$. But this contradicts $\underline{t} < \bar{t}$.

Thus any σ^* that is defined by (i) and (ii), consistency and Bellman optimality now imply that $\sigma^*(p')$ must have at least one posterior belief p'' such that $\text{supp}(p'') = \{(\bar{t},\bar{t}),(\bar{t},\underline{t}),(\underline{t},\underline{t})\}$. ■

6 Concluding remarks

A Appendix

Let T be as in Section 5. We denote by $t' \in T$ the valuation immediately preceding $t \in T$, i.e. $t' = \max\{s \in T : s < t\}$.

Proof of Proposition 1. Denote

$$\begin{aligned}
a_1(\theta_1) &= \sum_{\theta_2} p_2(\theta_2) a(\theta_1,\theta_2), \\
a_2(\theta_2) &= \sum_{\theta_1} p_1(\theta_1) a(\theta_1,\theta_2),
\end{aligned}$$

and use the shorthand

$$\begin{aligned}
U_p\left(a,m,\theta_1 : \theta_1\right) &= V_1(\theta_1), \\
U_p\left(a,m,\theta_2 : \theta_2\right) &= V_2(\theta_2).
\end{aligned}$$

Denoting t' the immediate predecessor of t, incentive compatibility of a mechanism implies

$$\begin{aligned} a_1(\theta_1')(\theta_1-\theta_1') &\leq V_1(\theta_1)-V_1(\theta_1') \leq a_1(\theta_1)(\theta_1-\theta_1'). \\ a_2(\theta_2')(\theta_2-\theta_2') &\geq V_2(\theta_2')-V_2(\theta_2) \geq a_2(\theta_2)(\theta_2-\theta_2'). \end{aligned}$$

Thus a_1 is increasing, a_2 is decreasing, and

$$\begin{aligned} V_1(\theta_1) &\geq \sum_{t\leq\theta_1'} a_1(t)+V_1(0), \\ V_2(\theta_2) &\geq \sum_{t\geq\theta_2'} a_2(t)+V_2(1). \end{aligned}$$

Let

$$\begin{aligned} P_i(\theta_i) &= \sum_{t\leq\theta_i} p_i(t), \\ A_i(\theta_i) &= \sum_{t\leq\theta_i} a_i(t). \end{aligned}$$

Then

$$\begin{aligned} P_1(\theta_1)A_1(\theta_1) &= \sum_{t\leq\theta_1} P_1(t)[A_1(t)-A_1(t')]+\sum_{t\leq\theta_1}[P_1(t)-P_1(t')]A_1(t') \\ &= \sum_{t\leq\theta_1} P_1(t)a_1(t)+\sum_{t\leq\theta_1} p_1(t)A_1(t'). \end{aligned}$$

Thus

$$\begin{aligned} \sum_t p_1(t)A_1(t') &= P_1(1)A_1(1)-\sum_t P_1(t)a_1(t) \\ &= \sum_t a_1(t)(1-P_1(t)). \end{aligned} \tag{15}$$

And similarly for the agent 2 :

$$\begin{aligned} \sum_t p_2(t)[A_2(1)-A_2(t')] &= A_2(1)-\sum_t p_2(t)A_2(t') \\ &= \sum_t a_2(t)P_2(t). \end{aligned} \tag{16}$$

The planner's problem can be written

$$\max_{a(r(\cdot))} \sum_{\theta_1}\sum_{\theta_2} p_1(\theta_1)p_2(\theta_2)(\theta_1-\theta_2)a(r(\theta_1,\theta_2))$$

s.t.

$$\sum_{\theta_1}\sum_{\theta_2} p_1(\theta_1)p_2(\theta_2)(\theta_1-\theta_2)a(r(\theta_1,\theta_2)) = \sum_{\theta_1} p_1(\theta_1)V_1(\theta_1) + \sum_{\theta_2} p_2(\theta_2)V_2(\theta_2) \tag{17}$$

$$A_1(\theta_1) \geq V_1(\theta_1) - V_1(0) \geq A_1(\theta_1'), \text{ for all } \theta_1 \tag{18}$$

$$A_2(1) - A_2(\theta_2) \geq V_2(\theta_2) - V_2(1) \geq A_2(1) - A_2(\theta_2'), \text{ for all } \theta_2 \tag{19}$$

$$V_1(\theta_1) \geq 0 \text{ for all } \theta_1,\ V_2(\theta_2) \geq 0 \text{ for all } \theta_2 \tag{20}$$

where (17) is the ex ante budget balance condition, (18) and (19) are the incentive compatibility constraints, and (20) is the participation constraint.

Since the right hand side inequalities of (18) and (19) imply

$$\begin{aligned}
\sum_{\theta_1} p_1(\theta_1)V_1(\theta_1) &\geq \sum_{\theta_1} p_1(\theta_1)A_1(\theta_1') + V_1(0), \\
\sum_{\theta_2} p_2(\theta_2)V_2(\theta_2) &\geq \sum_{\theta_2} p_2(\theta_2)[A_2(1) - A_2(\theta_2')] + V_2(1),
\end{aligned}$$

(15), (16), and (17) result in

$$\begin{aligned}
& V_1(0) + V_2(1) + \sum_{\theta_1} a_1(\theta_1)(1 - P_1(\theta_1)) + \sum_{\theta_2} a_2(\theta_2)P_2(\theta_2) \\
\leq & \sum_{\theta_1}\sum_{\theta_2} p_1(\theta_1)p_2(\theta_2)(\theta_1-\theta_2)a(r(\theta_1,\theta_2)),
\end{aligned}$$

or, more compactly,

$$\sum_{\theta_1}\sum_{\theta_2} p_1(\theta_1)p_2(\theta_2)\left[\left(\theta_1 - \frac{1-P_1(\theta_1)}{p_1(\theta_1)}\right) - \left(\theta_2 + \frac{P_2(\theta_2)}{p_2(\theta_2)}\right)\right] a(\theta_1,\theta_2) \geq 0. \tag{21}$$

Maximizing the objective function with respect to (21), and interpreting $\gamma/(1-\gamma)$ as the Lagrange multiplier, gives the desired programme. Since, at the optimum, (21) holds as equality, the solution to the programme also meets the left hand side inequalities of (18) and (19). Since this implies that a_1 is increasing and a_2 is decreasing, it also follows that the participation constraint (20) is met whenever $V_1(0) \geq 0$ and $V_2(1) \geq 0$ which hold as equality at the optimum. Finally, the optimality of a^γ under regular p_1 and p_2 follows by maximizing the objective function pointwisely. ■

Proof of Lemma 6. Our task is to construct an $m(\cdot)$ that prescribes zero monetary transfer when trade does not take place. That is

$$m(r(\theta_1,\theta_2)) = 0 \quad \text{whenever} \quad c_1(\theta_1,\gamma) - c_2(\theta_2,\gamma) < 0.$$

Denote by a^γ the incentive efficient allocation rule under Lagrange multiplier γ. Denote by m_1^γ and m_2^γ the implied expected transfers from 1 and to 2 :

$$\begin{aligned} m_1^\gamma(\theta_1) &= a_1^\gamma(\theta_1)\theta_1 - \sum_{t<\theta_1'} a_1^\gamma(t), \quad \text{for all } \theta_1 \in T \\ m_2^\gamma(\theta_2) &= a_2^\gamma(\theta_2)\theta_2 + \sum_{t>\theta_2'} a_2^\gamma(t), \quad \text{for all } \theta_2 \in T. \end{aligned}$$

The ex ante budget balance of the incentive efficient mechanism implies

$$\sum_{\theta_1} p_1(\theta_1) m_1^\gamma(\theta_1) = \sum_{\theta_2} p_2(\theta_2) m_2^\gamma(\theta_2). \tag{22}$$

Construct $m(\cdot)$ such that

$$\begin{aligned} m(\theta_1, 0) &= m_1^\gamma(\theta_1), \quad \text{for all } \theta_1 < 1, \\ m(1, \theta_2) &= m_2^\gamma(\theta_2), \quad \text{for all } \theta_2 > 0, \\ m(\theta_1, \theta_2) &= 0, \quad \text{for all } (\theta_1, \theta_2) \text{ such that } \theta_1 < 1 \text{ and } \theta_2 > 0. \end{aligned}$$

To complete the description of m, let $m(1,0)$ satisfy

$$p_1(1)m(1,0) + \sum_{t<1} p_1(t) m_1^\gamma(t) = m_2^\gamma(0), \tag{23}$$

$$p_2(0)m(1,0) + \sum_{t>0} p_2(t) m_2^\gamma(t) = m_1^\gamma(1). \tag{24}$$

Then $m(\cdot)$ prescribes zero transfer under no-trade and

$$\begin{aligned} m_1(\theta_1) &= m_1^\gamma(\theta_1), \quad \text{for all } \theta_1 \in T, \\ m_2(\theta_2) &= m_2^\gamma(\theta_2), \quad \text{for all } \theta_2 \in T. \end{aligned}$$

Thus m is consistent with the incentive efficient allocation a^γ rule. However, since a single variable $\bar{m}$ is determined by two equations (23) and (24), we need to verify that a desired $m(1,0)$ does exist. The remainder of the proof establishes this.

First, fix any $m(1,0)$ that completes the description of m. Since the order of summation does not matter,

$$\sum_{\theta_1} p_1(\theta_1) \sum_{\theta_2} p_2(\theta_2) m(\theta_1, \theta_2) = \sum_{\theta_2} p_2(\theta_2) \sum_{\theta_1} p_1(\theta_1) m(\theta_1, \theta_2). \tag{25}$$

By construction,

$$\begin{aligned} \sum_{\theta_1} p_1(\theta_1) \sum_{\theta_2} p_2(\theta_2) m(\theta_1, \theta_2) &= \sum_{t<1} p_1(t) m_1^\gamma(t) + p_1(1) \left(p_2(0) m(1,0) + \sum_{t>0} p_2(t) m_2^\gamma(t) \right) \\ \sum_{\theta_2} p_2(\theta_2) \sum_{\theta_1} p_1(\theta_1) m(\theta_1, \theta_2) &= p_2(0) \left(p_1(1) m(1,0) + \sum_{t<1} p_1(t) m_1^\gamma(t) \right) + \sum_{t>0} p_2(t) m_2^\gamma(t) \end{aligned}$$

Now letting $m(1,0)$ be defined by (23), it follows that

$$\sum_{\theta_2} p_2(\theta_2) \sum_{\theta_1} p_1(\theta_1) m(\theta_1, \theta_2) = \sum_{\theta_2} p_2(\theta_2) m_2^{\gamma}(\theta_2).$$

By (25) and (22),

$$\sum_{\theta_1} p_1(\theta_1) \sum_{\theta_2} p_2(\theta_2) m(\theta_1, \theta_2) = \sum_{\theta_1} p_1(\theta_1) m_1^{\gamma}(\theta_1).$$

Thus $m(1,0)$ also satisfies (24). ■

References

[1] Aghion, P., Dewatripont, M., and P. Rey (1994), Renegotiation design with unverifiable information, *Econometrica* **62**, 257-282.

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

[3] Ausubel, L. and Deneckere, R. (1993),, Efficient Sequential Bargaining, *Review of Economic Studies* **60**, 435-461.

[4] Baliga, S., Corchón, L. and Sjöström, T. (1997). The theory of implementation when the planner is a player, *Journal of Economic Theory* **77**, 15-33.

[5] Bergemann, D and Morris, S. (2005), Robust Mechanism Design, *Econometrica* **73**, 1771-1813

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

[7] Bester, H. and Strausz, R. (2000), Imperfect Commitment and the Revelation Principle: the Multi-Agent Case, *Economics Letters* **69**, 165–171

[8] Coase. R. (1972), Durability and Monopoly, *Journal of Law and Economics* **15**, 143-9.

[9] Evans, R. (2012), Mechanism design with renegotiation and costly messages, *Econometrica*.**80**, 2089–2104

[10] Forges, F. (1994), Posterior Efficiency, *Games and Economic Behavior* **6**, 238-61.

[11] Forges, F. (1998), Ex Post Individually Rational Trading Mechanisms, manuscript, THEMA, Universite of Cergy-Pontoise.

[12] Freixas, X., Guesnerie, R., and Tirole, J. (1985), Planning under Incomplete Information ans the Ratchet Effect, *Rewiev of Economic Studies* **52,** 173-92

[13] Gerardi, D., Hörner, J., and L. Maestri (2011), The Role of Commitment in Bilateral Trade, manuscript, Collegio Carlo Alberto

[14] Green, J. and Laffont, J.-J.. (1987), Posterior Implementability in a Two-Player Decision Problem, *Econometrica* **55,** 69-94

[15] Gresik, T. (1991), Ex Ante Efficient, Ex Post Individually Rational Trade, *Journal of Economic Theory* **53,** 131-45.

[16] Gresik, T. (1996), Incentive Efficient Equilibria of Two-Party Sealed-Bid Bargaining Games, *Journal of Economic Theory* **68,** 26-48

[17] Holmstrom, B. and Myerson, R. (1983), Efficient and Durable Decision Rules with Incomplete Information, *Econometrica* **51**, 1799-819,

[18] Kiyotaki, N. (2011), A Mechanism Design Approach to Financial Frictions, manuscript, Princeton University

[19] Krishna, V. and Perry, M. (1998), Efficient Mechanism Design, working paper, Penn State University.

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

[21] Lagunoff, R. (1995), Resilient Allocation Rules for Bilateral Trade, *Journal of Economic Theory* **66**, 463-87

[22] Lagunoff, R. (1992), Fully Endogenous Mechanism Selection on Finite Outcome Sets, *Economic Theory* **2**, 465-80

[23] McAdams, D. and Schwarz, M. (2006), Credible Sales Mechanisms and Intermediaries, *American Economic Review* forthcoming

[24] McAfee, P. and Vincent, D. (1997), Sequentially Optimal Auctions, *Games and Economic Behavior* **18,** 246-76

[25] MYERSON, R. (1991), *Game Theory, Analysis of Conflict*, Cambridge MA, Harvard University Press

[26] MYERSON, R., (1979), Incentive compatibility and the bargaining problem, *Econometrica* **47**, 61–73.

[27] MYERSON, R. (1982), Optimal Coordination Mechanisms in Generalized Principle-Agent Problems, *Journal of Mathematical Economics* **28,** 67–81.

[28] NEEMAN Z. AND G. PAVLOV (2012), Renegotiation-Proof Mechanism Design, forthcoming in *Journal of Economic Theory.*

[29] SEGAL, I., AND M. WHINSTON (2002), The Mirrlees approach to mechanism design with renegotiation *Econometrica* **70**, 1-45.

[30] SKRETA V. (2011), Optimal Auction Design under Non-Commitment, manuscript, NYU Stern School of Business

[31] SKRETA V., (2006), Sequentially Optimal Mechanisms, *Review of Economic Studies* **73**

[32] VARTIAINEN, H. (2010), Auction Design without Commitment, forthcoming in *Journal of the European Economic Association*