Renegotiation-proof Mechanism Design

by

Zvika Neeman and Gregory Pavlov

Research Report # 2010-1 **March 2010**

Department of Economics
Research Report Series

Department of Economics
Social Science Centre
The University of Western Ontario
London, Ontario, N6A 5C2
Canada

This research report is available as a downloadable pdf file on our website
http://economics.uwo.ca/econref/WorkingPapers/departmentresearchreports.html.

Renegotiation-proof mechanism design*

Zvika Neeman† Gregory Pavlov‡

February 3, 2009

Abstract

We study a mechanism design problem under the assumption that renegotiation cannot be prevented. We investigate what kind of equilibria of which mechanisms are renegotiation-proof under a variety of renegotiation procedures, and which social choice functions can be implemented in a way that is renegotiation-proof. In complete information environments, we show that the set of ex post renegotiation-proof implementable social choice functions contains all ex post efficient allocations when there at least three agents, but only budget balanced Groves allocations when there are two agents. In incomplete information environments with correlated beliefs and at least three agents, every ex post efficient social choice function can be implemented in the presence of ex post renegotiation, but with independent private values only social choice functions that are given by budget balanced "Groves in expectations" mechanisms are implementable in such a way. We further show that the requirement of interim renegotiation-proofness does not impose additional restrictions on implementable social choice functions under complete information, but is likely to impose additional restrictions under incomplete information.

JEL classification: D02; D70; D82; D86

Keywords: Mechanism design; Implementation; Ex post renegotiation; Interim renegotiation

1 Introduction

One of the practical concerns of mechanism design theory is that players might often have incentives to change the rules of the game they are playing. Although in some cases the mechanism designer can prevent such changes, in many situations it is impossible or nearly impossible to do so, especially when a change in the rules of the game, contract, or mechanism is mutually beneficial for the players. Such mutually consensual changes, which are known as *renegotiation*, can occur at different

*We would like to thank Eddie Dekel, Françoise Forges, Maria Goltsman, Johannes Hörner, Bart Lipman, Dilip Mookherjee, Michel Poitevin, Pasquale Schiraldi and the seminar participants at Arizona State University, Boston University, Hebrew University, Université de Montréal, Tel Aviv University, Toulouse, Vienna, University of Western Ontario, CETC (Montreal, 2007), Decentralization conference (Ann Arbor, 2007), World Congress of the Game Theory Society (Evanston, 2008), and mechanism design conference (Bonn, 2009) for helpful comments and conversations. All remaining errors are ours.

†The Eitan Berglas School of Economics, Tel-Aviv University, Tel-Aviv, Israel 69978; Email zvika@post.tau.ac.il, http://www.tau.ac.il/~zvika/.

‡Department of Economics, University of Western Ontario, Social Science Centre, London, Ontario N6A 5C2, Canada; Email gpavlov@uwo.ca, http://economics.uwo.ca/faculty/pavlov/.

stages of the contractual process. *Interim renegotiation* takes place before the mechanism is played and involves a change of the mechanism and the equilibrium the players intend to play. *Ex post renegotiation* takes place after the mechanism is played and involves a change of the outcome or recommendation proposed by the mechanism. The consequences of both interim and ex post renegotiation crucially depend on the details of the renegotiation process: what alternative outcomes or mechanisms are considered? How do the players communicate with each other, and how do they select among the alternative proposals? How is the surplus that is generated by renegotiation shared among the players? Etc.

We study the problem of a mechanism designer who is unable to prevent renegotiation and is ignorant of the exact way in which renegotiation will proceed. More specifically, we are interested in the question of what kind of equilibria of which mechanisms are *renegotiation-proof*, i.e., can be expected to be "stable" in the sense of surviving in their original form under a variety of different renegotiation procedures. We also want to know which social choice functions can be implemented in a way that is renegotiation-proof.

Our notion of renegotiation-proofness differs from the one commonly used in mechanism design and contract theory literature. The standard approach to renegotiation that is used in the literature assumes that the mechanism designer can anticipate exactly how any mechanism will be renegotiated. The anticipated renegotiation can then be incorporated into the original contract, and (in many cases) without loss of generality attention can then be limited to renegotiation-proof mechanisms.[1] Hence, under the standard approach the renegotiation-proofness of a mechanism is defined with respect to a given renegotiation procedure, while our notion of renegotiation-proofness requires that an equilibrium of a mechanism survives under all plausible renegotiation procedures.[2,3]

For the most part of the paper we focus on ex post renegotiation.[4] There is a small literature that studies ex post renegotiation under multiple possible renegotiation procedures.[5] We adopt a more comprehensive approach to this problem and consider general quasi-linear environments with any number of agents under both complete and incomplete information. There are also differences between our notion of renegotiation-proofness and the notions of renegotiation-proofness that are employed by the other papers in the literature. But we postpone further discussion of the relationship to the literature to the end of the paper. See Section 7 for additional details.

[1] This result is known as the "renegotiation-proofness principle." It was first introduced by Dewatripont (1989).

[2] For additional discussion of renegotiation-proofness under the standard approach see Bolton (1990) and Beaudry and Poitevin (1995).

[3] Our approach is also different from the recent approach to renegotiation that emphasizes the "technological details" of renegotiation (Lyon and Rasmusen (2004), Watson (2007)). This literature maintains the assumption that the renegotiation procedure is known by the mechanism designer, and examines the robustness of the results obtained by the papers that employ the standard approach with respect to changes in the assumptions about the details of trade and renegotiation technology.

[4] In most of the literature that employs the standard approach to renegotiation ex post renegotiation imposes more constraints on the attainable outcomes than interim renegotiation. Segal and Whinston (2002) and Watson (2007) observe that this is the case in complete information environments, and Beaudry and Poitevin (1995) make a similar observation in the case of incomplete information environments.

[5] In environments with complete information Sjöström (1999) and Amoros (2004) study the case of three or more agents and Rubinstein and Wolinsky (1992) study a specific buyer and seller problem. Forges (1993, 1994) studies ex post renegotiation under incomplete information.

In Section 3 we study ex post renegotiation in complete information environments. After commonly observing the state of the world, the agents play the game that is induced by a given mechanism. Then, they have a chance to renegotiate the outcome prescribed by the mechanism to some alternative outcome. The prospect of ex post renegotiation can undermine an equilibrium of the mechanism in two ways. First, the equilibrium outcome that is reached by the mechanism can be renegotiated. And second, the incentive compatibility of the original equilibrium may be destroyed because the agents may find it profitable to deviate from the original equilibrium when they play the mechanism in anticipation of subsequent renegotiation. We show that ex post renegotiation-proof equilibria always result in ex post efficient outcomes (Lemma 1), and that for the case of three or more agents every ex post efficient social choice function can be implemented in a way that is ex post renegotiation-proof by a simple mechanism that requires two agents to report the state of the world and requires them to pay high penalties to other agents if they disagree (Proposition 1). When there are only two agents it is impossible to penalize both agents simultaneously because such penalties would be undone by renegotiation. Consequently, we show that the set of implementable social choice functions is strictly smaller than in the case of three or more agents: all ex post renegotiation-proof implementable social choice functions are outcome equivalent to budget balanced Groves mechanisms (Proposition 2). Many mechanism design environments do not admit the existence of budget balanced Groves mechanisms. When this is the case, it is impossible to implement any social choice function in a way that is ex post renegotiation-proof.

Modeling ex post renegotiation in incomplete information environments is a more challenging task than in the case of complete information. In Section 4 we develop one possible approach under which the alternative outcomes are exogenously proposed to the agents who then vote whether to switch to the proposed alternative. If agents' values are interdependent then there may exist ex post inefficient renegotiation-proof equilibria (Example 3), but in the case of private values ex post renegotiation-proof equilibria are always ex post efficient (Lemma 2). In the context of an example of bilateral trade with independent private values we show that ex post renegotiation may also impose restrictions on the relative monetary transfers across type profiles, but these restrictions are weaker than in the case of complete information with two players (Example 2' and Lemma 3). Nevertheless, we show that in this example there are no social choice functions that are both implementable in a way that is ex post renegotiation-proof and satisfy interim individual rationality constraints. When there are three or more agents and the agents' private information is correlated, we show that every ex post efficient social choice function can be implemented in a way that is ex post renegotiation-proof (Proposition 3).

In Section 5 we present a second approach to ex post renegotiation under incomplete information that is meant to capture the idea that in the course of renegotiation the agents may choose to reveal some private information in addition to what was already revealed by the outcome of the mechanism. We model this possibility by introducing an external "oracle" device that reveals the state of the world to the agents at the ex post renegotiation stage. For the case of independent private values we show that every ex post oracle renegotiation-proof implementable social choice function is payoff

equivalent to some budget balanced "Groves in expectations" mechanism (Proposition 4).[6]

In Section 6 we discuss the additional restrictions on implementable social choice functions that are imposed by interim renegotiation-proofness.[7] We find that under complete information interim renegotiation-proofness imposes no additional restrictions on implementability beyond what is already imposed by ex post renegotiation-proofness. Every social choice function that can be implemented by some mechanism in a way that is ex post renegotiation-proof would also survive interim renegotiation (Claims 1 and 2). Under incomplete information we argue that a similar result is not likely to hold, and provide simple sufficient conditions on social choice functions for interim and ex post renegotiation-proof implementability for environments with independent private values and environments with correlated beliefs (Claims 3 and 4).

The related literature is discussed in Section 7, and in Section 8 we offer a few concluding remarks. All proofs are relegated to the Appendix.

2 Setup

A group of n agents must reach an agreement that involves the choice of a *social alternative* a from a set A and *monetary transfers* to the agents, $t = (t_1, ..., t_n)$. An *outcome* (a, t) of the process of negotiation among the agents is said to be *feasible* if $a \in A$ and $t \in \mathbb{T} = \{t \in \mathbb{R}^n : \sum_{i=1}^n t_i \leq 0\}$.[8]

The agents' preferences over outcomes depend on the state of the world θ that is chosen from a finite set Θ.[9] Each agent i is an expected utility maximizer with a quasi-linear payoff function $v_i(a, \theta) + t_i$, where $v_i : A \times \Theta \to \mathbb{R}$ describes his preferences over social alternatives for different states of the world, and t_i denotes a monetary transfer given to him. We assume that for every state θ there is a single alternative $a^*(\theta)$ that maximizes the total surplus $\sum_{i=1}^n v_i(a, \theta)$; this alternative is called *ex post efficient in state* θ.[10] A vector of transfers t that satisfies $\sum_{i=1}^n t_i = 0$ is called *budget balanced*. An outcome (a, t) is called *ex post efficient in state* θ if $a = a^*(\theta)$ and t is budget balanced.

A *social choice function* is a mapping $f : \Theta \to A \times \mathbb{T}$ from the set of states into feasible outcomes. A social choice function f is said to be *ex post efficient* if for every $\theta \in \Theta$ the outcome $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient in state θ.

We distinguish between complete and incomplete information environments as follows:

Complete information. Complete information environments are analyzed in Sections 3 and 6. The state of the world θ is commonly known among the agents. For some results we also assume

[6] "Groves in expectations" mechanisms are equivalent to Groves mechanisms from the agents' interim perspective. See, for example, Williams (1999).

[7] See Holmström and Myerson (1983) for a classic approach to interim renegotiation-proofness under incomplete information.

[8] In some environments the right-hand-side of the feasibility constraint on the sum of monetary transfers may depend on the chosen social alternative. Using a standard argument it is possible to redefine the agents' payoffs in a way that is consistent with our model. See, for example, Chapter 7 in Fudenberg and Tirole (1991).

[9] We assume that Θ is finite to simplify the exposition, but our results continue to hold if Θ is infinite.

[10] This assumption is generically satisfied if Θ is finite.

that the set of states has a structure of a Cartesian product, $\Theta = \prod_{i=1}^{n} \Theta_i$, where Θ_i describes agent i's set of payoff relevant types.[11]

Condition 1 ***(Product domain)*** *$\Theta = \prod_{i=1}^{n} \Theta_i$, and the payoff of each agent depends only on his type, i.e., with a slight abuse of notation:*

$$v_i(a, \theta_i, \theta_{-i}) = v_i(a, \theta_i) \qquad \forall i \in \{1, ..., n\},\ a \in A,\ \theta_i \in \Theta_i,\ \theta_{-i} \in \Theta_{-i} = \prod_{j \neq i} \Theta_j$$

A *mechanism* $\langle S, m \rangle$ specifies a set of messages S_i for each agent i and an outcome rule $m : S \to \Delta(A \times \mathbb{T})$ from the set of message profiles $S = \prod_{i=1}^{n} S_i$ into the set of lotteries over feasible outcomes. A mechanism $\langle S, m \rangle$ together with a state of the world θ defines a complete information game, where a strategy of agent i is $\sigma_i(\theta) \in \Delta(S_i)$. We say that $\sigma = (\sigma_1, ..., \sigma_n)$ is an *equilibrium* if for every $\theta \in \Theta$ the strategy profile $(\sigma_1(\theta), ..., \sigma_n(\theta))$ is a Nash equilibrium of the complete information game in state θ. Equilibrium σ is said to be *ex post efficient* if the equilibrium outcome is ex post efficient in every state $\theta \in \Theta$.

A social choice function f is said to be *implementable* if there exists a mechanism $\langle S, m \rangle$ that has an equilibrium σ such that the equilibrium outcome in every state θ coincides with $f(\theta)$.[12]

Incomplete information. Incomplete information environments are analyzed in Sections 4, 5 and 6. The set of states of the world is $\Theta = \prod_{i=1}^{n} \Theta_i$, there is a common prior distribution P over Θ, and each agent i privately observes his type $\theta_i \in \Theta_i$. We consider both private and interdependent values environments; by private values we mean the following.[13]

Condition 2 ***(Private values)*** *The payoff of each agent depends only on his type, i.e., with a slight abuse of notation:*

$$v_i(a, \theta_i, \theta_{-i}) = v_i(a, \theta_i) \qquad \forall i \in \{1, ..., n\},\ a \in A,\ \theta_i \in \Theta_i,\ \theta_{-i} \in \Theta_{-i}$$

Like in the case of complete information, a *mechanism* $\langle S, m \rangle$ specifies a set of messages S_i for each agent i and an outcome rule $m : S \to \Delta(A \times \mathbb{T})$ from the set of message profiles $S = \prod_{i=1}^{n} S_i$ into the set of lotteries over feasible outcomes. A mechanism $\langle S, m \rangle$ together with a common prior P defines a Bayesian game, where a strategy of agent i is $\sigma_i : \Theta_i \to \Delta(S_i)$. We say that $\sigma = (\sigma_1, ..., \sigma_n)$ is an *equilibrium* if such strategy profile is a Bayesian Nash equilibrium of the Bayesian game. Equilibrium σ is said to be *ex post efficient* if the equilibrium outcome is ex post efficient in every state $\theta \in \Theta$.

A social choice function f is said to be *implementable* if there exists a mechanism $\langle S, m \rangle$ that has an equilibrium σ such that the equilibrium outcome is payoff equivalent to f from the agents'

[11] This is a standard domain restriction in implementation theory (sometimes called "independent domain"). See, for example, Moore (1992).

[12] Thus we employ a weak notion of implementation.

[13] Condition 2 is formally identical to Condition 1. We give them two different names so as to be consistent with the literature.

interim perspective. That is, the equilibrium expected payoff of agent i of type $\theta_i \in \Theta_i$ is equal to

$$E_{\theta_{-i}|\theta_i}\left[v_i\left(a\left(\theta\right),\theta\right)+t_i\left(\theta\right)\right]=\sum_{\theta_{-i}\in\Theta_{-i}}P\left(\theta_{-i}\mid\theta_i\right)\left(v_i\left(a\left(\theta\right),\theta\right)+t_i\left(\theta\right)\right)$$

where $\left(a\left(\theta\right),t\left(\theta\right)\right)=f\left(\theta\right)$ for every $\theta\in\Theta$.

Although we do not discuss the subject of individual rationality in this paper (except for Sections 4.2 and 5.2), it is straightforward to incorporate into the analysis an appropriate notion of individual rationality as an additional set of constraints on the social choice functions.

3 Ex post renegotiation-proofness under complete information

3.1 Preliminaries

In a complete information environment a state of the world $\theta\in\Theta$ is realized and becomes commonly known among the agents, but is not known to the mechanism designer. The agents then play the game induced by a given mechanism $\langle S,m\rangle$. After the outcome of playing the mechanism is determined, the agents have an opportunity to change it to some other outcome according to some renegotiation procedure.[14] Our goal is to determine which equilibria of which mechanisms are robust against such renegotiation.

Rather than modeling various renegotiation procedures as games, we take the following shortcut. Suppose an outcome (a,t) was reached by play of the mechanism in state θ. We say that the outcome (a,t) *can be renegotiated* if there exists another feasible outcome (a',t') that Pareto dominates (a,t) in state θ. That is, in this state of the world all the agents weakly prefer, and at least one agent strictly prefers, the alternative outcome (a',t') to the original outcome (a,t) that was prescribed by the mechanism.[15,16]

Consider an equilibrium of some mechanism. If we allow the possibility of ex post renegotiation then an equilibrium outcome may be undermined in two ways. First, the equilibrium outcome reached by the mechanism may be renegotiated. Second, the original equilibrium strategy profile may cease to be an equilibrium once ex post renegotiation is allowed, since deviating when playing the mechanism may bring about an outcome which may itself be beneficially renegotiated. These two possibilities lead to the following definition.

[14] Alternatively one could allow renegotiation to take place before the lotteries from $\Delta\left(A\times\mathbb{T}\right)$ prescribed by the outcome rule m are carried through. Since the agents have quasi-linear payoffs such an approach would not affect the results, but it would make the arguments more cumbersome.

[15] A standard approach to modeling renegotiation, like in Maskin and Moore (1999), assumes that ex post renegotiation proceeds according to some given process (or a game). This process induces a mapping from outcomes and states into (possibly different) outcomes, and this mapping is commonly known by the agents and the mechanism designer. In contrast, we allow more flexibility in the ways renegotiation may proceed, and assume that the mechanism designer is ignorant of the exact way in which renegotiation will proceed. See Section 7.1 for discussion of the relation of our approach to the existing literature.

[16] One may want also to require that the renegotiation proposal (a',t') be "credible", so that (a',t') cannot in turn be renegotiated to some other outcome (a'',t''). In complete information environments the "credibility" of proposed outcomes can be ensured by restricting them to be ex post efficient for a given state. Such a change would not affect our results.

Definition 1 *An equilibrium σ of a mechanism $\langle S, m\rangle$ is* ex post renegotiation-proof (EPRP) *if both of the following conditions hold:*

(i) An outcome that is obtained under the equilibrium play of the mechanism cannot be renegotiated.

(ii) No agent can improve upon his equilibrium payoff in any state by a unilateral deviation from σ followed by renegotiation of the resulting outcomes.

Part (i) of Definition 1 immediately implies the following result that is given without proof.

Lemma 1 *In a complete information environment every EPRP equilibrium is ex post efficient.*

The next example illustrates part (ii) of Definition 1. It demonstrates that an equilibrium may fail to be EPRP in spite of being ex post efficient and in dominant strategies.

Example 1 *A buyer and a seller can trade a single good. The buyer values the good at V that can be either 0 or 2, the seller values the good at 1. The realization of V and the seller's valuation are commonly known between the agents. Consider a mechanism where the buyer is asked to report his value: after a report "$V = 2$" the good is transferred from the seller to the buyer at a price p_2, and after a report "$V = 0$" there is no trade and the buyer pays p_0 to the seller. It is easy to see that the buyer has a dominant strategy to report his true valuation if $p_2 - p_0 \in (0, 2)$, and the resulting outcome is ex post efficient. However, as we show below, this equilibrium is not EPRP unless $p_2 - p_0 = 1$.*

Suppose $p_2 - p_0 \in (1, 2)$. If the buyer with $V = 2$ reports "$V = 0$" then the payoffs of the buyer and the seller (without renegotiation) would be $-p_0$ and p_0, respectively. This outcome is Pareto dominated by a decision to trade at a new price $\widehat{p}$ that satisfies $\widehat{p} - p_0 \in (1, 2)$. Hence, for any such $\widehat{p} < p_2$, the buyer would prefer to misreport and then renegotiate the outcome to trade at the price $\widehat{p}$ rather than report his true valuation. Thus, the original equilibrium is not EPRP.[17]

The EPRP requirement in this example not only implies ex post efficiency, but also restricts the relative transfers across different states of the world. Of course, this example deals with just one particular simple mechanism. In order to study the range of implementable outcomes, we introduce a notion of EPRP implementable social choice functions.

Definition 2 *A social choice function is* ex post renegotiation-proof (EPRP) implementable *if there exists a mechanism $\langle S, m\rangle$ that has an EPRP equilibrium σ such that the equilibrium outcome coincides with $f(\theta)$ in every state θ.*

[17] The argument for the case $p_2 - p_0 \in (0, 1)$ is similar. The buyer with $V = 0$ will find it profitable to report "$V = 2$" and then renegotiate to "no trade" as long as long a new payment $\widehat{p}$ is smaller than p_0.

3.2 Results

As often happens in implementation theory, the case of three or more agents is very different from the case of two agents.

Proposition 1 *Consider a complete information environment with $n \geq 3$ agents. A social choice function f is EPRP implementable if and only if $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient for every θ.*

The "only if" part of the result follows from Lemma 1. The idea for the "if" part is standard.[18] Agents 1 and 2 are asked to report the state of the world. If their reports coincide then the outcome prescribed by f for the reported state is implemented. If their reports differ then an arbitrary alternative a_0 is implemented, and both agents 1 and 2 pay large penalties to agent 3. There exists an equilibrium where agents 1 and 2 report the state truthfully if the penalties are large enough. This equilibrium is EPRP: (i) the outcome is ex post efficient if the agents tell the truth; (ii) the penalties can be chosen to be large enough so that any strategy of sending a wrong report followed by renegotiation would be unprofitable.

When there are only two agents, however, it is impossible to penalize both of them simultaneously when their reports conflict. This is because there is no third agent to whom transfers can be made and so any penalties must involve ex post inefficient outcomes, which can be undone by renegotiation. Thus, it is harder to provide incentives to tell the truth, and the class of EPRP implementable social choice functions is smaller than in Proposition 1. Our result for the case of two agents is for environments that satisfy the *product domain* restriction (Condition 1). The role of this restriction is discussed at the end of this section.

Proposition 2 *Consider a complete information environment with product domain and two agents. A social choice function f is EPRP implementable if and only if $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient for every θ and transfers t satisfy*

$$t_i(\theta_i, \theta_j) = v_j(a^*(\theta_i, \theta_j), \theta_j) + H_i(\theta_j) \qquad \forall i \in \{1,2\},\ \theta_i \in \Theta_i,\ \theta_j \in \Theta_j, \tag{1}$$

for some function $H_i : \Theta_j \to \mathbb{R}$.

For the "if" part of the result we consider the following mechanism. Each agent i reports his own payoff relevant type $\theta_i \in \Theta_i$; an ex post efficient outcome rule $(a^*, t) : \Theta_1 \times \Theta_2 \to A \times \mathbb{T}$ that satisfies (1) is implemented given the reports. We show that this mechanism has a truthful equilibrium that is EPRP. In the mechanism design literature such mechanisms (without budget balance requirement) are called *Groves mechanisms*. Such mechanisms implement ex post efficient alternatives in dominant strategies by making each agent's transfer (as a function of his reported payoff relevant type) equal to the externality he imposes on other agents. Our result shows that

[18] Maskin and Moore (1999) used the same idea to characterize implementable social choice functions under their notion of renegotiation.

Groves mechanisms are also EPRP: no agent can gain by submitting a wrong report and then renegotiating the mechanism's outcome, because he already internalizes the marginal effect that his report has on the total surplus.

The "only if" part shows that all EPRP implementable social choice functions are outcome equivalent to truthful equilibria of budget balanced Groves mechanisms. Note that in Groves mechanisms each agent is asked to report only his own payoff relevant type, even though in a complete information environment he also knows the other agent's type as well. Why doesn't there exist an EPRP mechanism that requires agents to report their entire private information, which consists of the whole state of the world $\theta = (\theta_1, \theta_2) \in \Theta$, and then implements transfers that do not satisfy (1)? A short answer is that a mechanism could indeed require each agent to report the type of the other agent, but the renegotiation-proofness constraints would not allow the mechanism's outcome rule to depend on this additional information.

Example 1 (continued) *By Proposition 2 every EPRP implementable social choice function is such that trade takes place if and only if* $V = 2$*, the buyer's transfers satisfy* $t_b(2) - t_b(0) = -1$ (that is, the buyer pays 1 more when he obtains the object)*, and the seller's transfers are given by* $t_s(V) = -t_b(V)$ *for every* V*, which ensures budget balance. Note that this is exactly the allocation that is induced by the mechanism that is described in the first part of the example above for the case where* $p_2 - p_0 = 1$*. In particular note that: (i) allowing the seller to report the state of the world does not expand the set of EPRP implementable social choice functions; (ii) the mechanisms described above are the budget balanced Groves mechanisms for this environment.*

It is easy to construct budget balanced Groves mechanisms if the preferences of one of the agents are independent of the state of the world. The argument is the same as in Example 1, so the result is stated without proof.

Corollary 1 *Consider a complete information environment with two agents such that for (at least) one of the agents* $v_i(a, \theta) = v_i(a)$ *for every* $a \in A$*,* $\theta \in \Theta$*. Then EPRP implementable social choice functions exist.*

However, the next example shows that existence is not guaranteed when the preferences of both agents depend on the state of the world.

Example 2 *A buyer and a seller can trade a single good. The buyer values the good at* V*, which can be either* 0 *or* 2*, and the seller values the good at* C*, which can be either* 1 *or* 3*. The realization of* (V, C) *is commonly known between the agents. By Proposition 2 every EPRP implementable social choice function is such that trade takes place if and only if the state is* $(2, 1)$*, the buyer's transfers satisfy* $t_b(2,1) - t_b(0,1) = -1$ *and* $t_b(0,3) - t_b(2,3) = 0$*, and the seller's transfers satisfy* $t_s(2,1) - t_s(2,3) = 2$ *and* $t_s(0,3) - t_s(0,1) = 0$*. In addition, budget balance requires that* $t_b(V, C) + t_s(V, C) = 0$ *for every* (V, C)*. This implies that*

$$0 = (t_b(2,1) + t_s(2,1)) + (t_b(0,3) + t_s(0,3)) - (t_b(0,1) + t_s(0,1)) - (t_b(2,3) + t_s(2,3)) = 1.$$

Hence, there are no EPRP implementable social choice functions in this environment.

To get a better sense of the scope of this nonexistence suppose that each agent's set of types is a connected open subset of an Euclidean space and that payoffs are continuously differentiable in types. Then any mechanism that implements ex post efficient alternatives in dominant strategies is outcome equivalent to a truthful equilibrium of a Groves mechanism.[19] If we require in addition budget balance, then Groves mechanisms often fail to exist.[20] Whenever this is the case, Proposition 2 implies that there are no EPRP implementable social choice functions either.

Finally, the product domain condition is necessary for the characterization result in Proposition 2. In environments where it is not satisfied, the requirements of ex post renegotiation-proofness are weaker, and so, as shown below, there may exist EPRP mechanisms that are not budget balanced Groves mechanisms.

Example 2 (continued) *Consider again the environment in Example 2, but suppose there are just two possible states: in state L the buyer's value is* 0 *and the seller's value is* 1*, and in state H the buyer's value is* 2 *and the seller's value is* 3*. The previous analysis implies that there are no budget balanced Groves mechanisms in this environment. However, a social choice function that prescribes no trade together with a constant payment from one agent to another is EPRP implementable by a mechanism that simply ignores the agents' messages.*

4 Ex post renegotiation-proofness under incomplete information

4.1 Preliminaries

In an incomplete information environment the state of the world $\theta = (\theta_1, ..., \theta_n)$ is drawn according to a prior probability distribution P, each agent i learns his private information θ_i and employs Bayes' rule to update his beliefs about the other agents' types. Then, the agents play the Bayesian game that is induced by a given mechanism $\langle S, m \rangle$. The agents then have an opportunity to change the outcome that is determined by the mechanism to some other outcome according to some renegotiation procedure. Our goal is to determine which equilibria of which mechanisms are robust against various renegotiation scenarios of this form. In this section we develop one possible approach to ex post renegotiation under incomplete information, and in Section 5 we present an alternative approach.[21]

Suppose that the mechanism produced the outcome (a, t), and that an alternative outcome (a', t') is proposed to the agents. Under complete information each agent knows which of the two

[19] See, for example, Holmström (1979).

[20] See, for example, Green and Laffont (1979) and Walker (1980).

[21] As in the case of complete information, the standard approach to modeling renegotiation assumes that ex post renegotiation proceeds according to some given process (or a game). This process induces a mapping from outcomes, states, and the agents' beliefs into outcomes and the agents' beliefs, and this mapping is commonly known among the agents and the mechanism designer. In contrast, we permit renegotiation to proceed in many different ways, and assume that the mechanism designer is ignorant of the exact way in which it will proceed. See Section 7.2 for discussion of the relation of our approach to the existing literature.

outcomes he prefers, so in Section 3.1 we wrote that "(a,t) can be renegotiated to (a',t') if the latter Pareto dominates the former given the state of the world." Under incomplete information, however, there are two difficulties: (i) the agents' preferences over (a,t) vs. (a',t') may depend on other agents' private information; and (ii) each agent may have several different types, and it may not be clear with respect to which of these types Pareto dominance should be defined.[22] To address these issues we pay special attention to the agents' beliefs conditional on the observables and their private information, and we build into the definition of renegotiation a simple voting procedure that the agents rely upon in order to decide whether to switch to the alternative outcome.

Fix an equilibrium strategy profile σ of a given mechanism $\langle S, m\rangle$. For every outcome (a,t) that may be potentially reached we can derive the agents' beliefs conditional on this outcome being realized. First suppose agent i of type θ_i has played his equilibrium strategy σ_i. If outcome (a,t) is on the equilibrium path, then his beliefs are derived by Bayes' rule: he takes into account the reached outcome (a,t), the equilibrium strategy profile σ, and his type θ_i. If outcome (a,t) was not supposed to happen under σ, then he may hold arbitrary beliefs.[23] Next suppose agent i of type θ_i has played σ_i' instead of his equilibrium strategy σ_i. For every outcome (a,t) the beliefs of this agent are derived by Bayes' rule: he takes into account the reached outcome (a,t), the strategy profile (σ_i', σ_{-i}), and his type θ_i.

Consider the following voting procedure. Agents vote simultaneously on whether to switch from the original outcome (a,t) to the new outcome (a',t'); if all agents vote in favor of switching then (a',t') is implemented, otherwise (a,t) is implemented. Note that the agents face an additional inference problem when they decide how to vote. If agent i believes that only a subset $\widehat{\Theta}_{-i}$ of the types of his opponents votes for (a',t'), then he needs to compare (a,t) vs. (a',t') conditional on $\widehat{\Theta}_{-i}$.

Summarizing, here is our definition of renegotiation under incomplete information.[24]

Definition 3 *Fix an equilibrium σ of a mechanism $\langle S, m\rangle$.*

(i) Suppose that an outcome (a,t) can be reached with positive probability under σ. We say that (a,t) can be renegotiated with positive probability on the equilibrium path *if there exists another feasible outcome (a',t') and a set of types $\widehat{\Theta} = \prod_{i=1}^{n} \widehat{\Theta}_i$ that has a positive probability (given (a,t) and σ) such that all types in $\widehat{\Theta}$ (and only types in $\widehat{\Theta}$) weakly prefer, and at least one type of one agent strictly prefers, (a',t') to (a,t), where the beliefs of each agent i are derived by Bayes' rule (conditional on (a,t), σ, θ_i and $\widehat{\Theta}_{-i}$).*

(ii) Suppose that an outcome (a,t) can be reached with positive probability following agent j's

[22] Holmström and Myerson (1983) discuss various notions of efficiency under incomplete information.

[23] Different types may hold different beliefs if the outcome (a,t) is inconsistent with the equilibrium σ.

[24] One may want to require the renegotiation proposals to be "credible", so that the resulting post-renegotiation state contingent allocation could not be renegotiated again in some states of the world. We have chosen not to restrict attention to "credible" renegotiation proposals, because we think that allowing a more permissive notion of renegotiation goes well with our goal of determining which equilibria of which mechanisms are robust against various renegotiation scenarios. For example, in some cases the agents might have time for just one round of renegotiation, and then some "non-credible" renegotiation agreements may take place.

unilateral deviation σ'_j *from* σ. *We say that* (a,t) can be renegotiated with positive probability off the equilibrium path *if there exists another feasible outcome* (a',t') *and a set of types* $\widehat{\Theta} = \prod_{i=1}^{n} \widehat{\Theta}_i$ *that has a positive probability (given* (a,t) *and* $\left(\sigma'_j, \sigma_{-j}\right)$*) such that all types in* $\widehat{\Theta}$ *(and only types in* $\widehat{\Theta}$*) weakly prefer, and at least one type of one agent strictly prefers,* (a',t') *to* (a,t)*, where the beliefs of agent* j *are derived by Bayes' rule (conditional on* (a,t)*,* $\left(\sigma'_j, \sigma_{-j}\right)$*,* θ_j *and* $\widehat{\Theta}_{-j}$*), and the beliefs of each agent* $i \neq j$ *are derived by Bayes' rule (conditional on* (a,t)*,* σ*,* θ_i *and* $\widehat{\Theta}_{-i}$*) whenever possible and are arbitrary otherwise.*

As in the case of complete information, there are two ways in which ex post renegotiation may undermine the equilibrium of a given mechanism. First, the equilibrium outcome that is reached by the mechanism may be renegotiated. Second, the original equilibrium strategy profile may cease to be an equilibrium once ex post renegotiation is allowed, since a deviation in the mechanism would bring about an outcome which may be potentially beneficially renegotiated. These two possibilities lead to the following incomplete information analog of Definition 1 (Section 3.1).

Definition 4 *An equilibrium* σ *of a mechanism* $\langle S, m \rangle$ *is* ex post renegotiation-proof (EPRP) *if both of the following conditions hold:*

(i) The outcomes obtained under the equilibrium play of the mechanism cannot be renegotiated with positive probability on the equilibrium path.

(ii) No agent can improve upon his interim equilibrium payoff by a unilateral deviation from σ *followed by renegotiation of the resulting outcomes with positive probability off the equilibrium path.*

Here is an incomplete information analog of Definition 2 (Section 3.1).

Definition 5 *A social choice function is* ex post renegotiation-proof (EPRP) implementable *if there exists a mechanism* $\langle S, m \rangle$ *that has an EPRP equilibrium* σ *such that the equilibrium outcome is payoff equivalent to* f *from the agents' interim perspective.*

4.2 Private values

In this section we study environments with private values. Note that each agent's preferences over any pair of outcomes are independent of his beliefs over the opponents' types. Thus, in the process of renegotiation each agent's decision whether to vote for the alternative outcome or for the original outcome is independent of his beliefs and depends only on his type.

Definition 3 can then be simplified as follows. An outcome (a,t) *can be renegotiated with positive probability* (on or off the equilibrium path) if there exists another feasible outcome (a',t') and a profile of types $\theta = (\theta_1, ..., \theta_n)$ that has a positive probability such that (a',t') Pareto dominates (a,t) in state θ. In other words, in state θ all agents prefer the new outcome (a',t') to the original outcome (a,t), and thus all of them would vote to switch to (a',t'). This observation leads to the next result.

Lemma 2 *In an incomplete information environment with private values every EPRP equilibrium is ex post efficient.*

As in the case of complete information, the requirement of ex post renegotiation-proofness may impose further restrictions on the relative transfers across different states of the world. This is illustrated by an incomplete information version of Example 1 (Section 3.1).

Example 1' *A buyer and a seller can trade a single good. The buyer values the good at V which can be either 0 or 2 with equal probability, and the seller values the good at 1. Unlike in Example 1 the buyer privately observes the realization of V. Consider a mechanism where the buyer is asked to report his value: after a report "$V = 2$" the good is transferred from the seller to the buyer at a price p_2, after a report "$V = 0$" there is no trade and the buyer pays p_0 to the seller. The buyer has a dominant strategy to report his true valuation if $p_2 - p_0 \in (0, 2)$, and the resulting outcome is ex post efficient. The same argument used in Example 1 implies that this equilibrium is not EPRP unless $p_2 - p_0 = 1$.*

Example 2 (Section 3.2) demonstrated that some complete information environments admitted no EPRP implementable social choice functions. The next example revisits the same setting under incomplete information.

Example 2' *A buyer and a seller can trade a single good. The buyer values the good at V which can be either 0 or 2 with equal probability, and the seller values the good at C which can be either 1 or 3 with equal probability independently of the realization of V. Unlike in Example 2 each agent privately observes the realization of his value. Below we characterize EPRP implementable social choice functions in this environment.*

Lemma 3 *In the environment of Example 2' a social choice function is EPRP implementable if and only if it prescribes trade in state $(V, C) = (2, 1)$ and no trade in the other states, the buyer's transfers satisfy*

$$t_b(2,1) - t_b(0,3) = -\frac{3}{2} \text{ and } t_b(2,3) - t_b(0,1) = \frac{1}{2} \tag{2}$$

and the seller's transfers are given by

$$t_s(V,C) = -t_b(V,C) \text{ for every } (V,C). \tag{3}$$

The reason for discrepancy between Examples 2 and 2' is as follows. Consider an agent who contemplates misreporting (followed by renegotiation) in a direct mechanism. Unlike in the complete information case, he has to make a report without observing the types of the other agents. Hence, under incomplete information there are fewer renegotiation-proofness constraints to be satisfied.[25]

It should be noted, however, that in Example 2' no EPRP implementable social choice function satisfies interim individual rationality.

[25] See Section 5.2 for futher discussion.

Example 2' (continued) *The interim expected payoff of a seller of type* 3 *is* $\frac{1}{2}t_s(0,3)+\frac{1}{2}t_s(2,3)$*, the interim expected payoff of a buyer of type* 0 *is* $\frac{1}{2}t_b(0,1)+\frac{1}{2}t_b(0,3)$*. Summing up and using the expressions for the transfers in Lemma 3, it follows that*

$$-\frac{1}{2}t_b(2,3)+\frac{1}{2}t_b(0,1)=-\frac{1}{4}.$$

Hence, the payoffs of a seller of type 3 *and of a buyer of type* 0 *cannot both be nonnegative, which implies that there are no EPRP implementable social choice functions that are interim individually rational.*

4.3 Correlated and interdependent values

In this section we study the case of interdependent valuations and possibly correlated beliefs. Unlike in the case of environments with private values, each agent's preferences over a given pair of outcomes may depend on his beliefs over the other agents' types.

In environments with interdependent valuations ex post renegotiation-proofness no longer implies ex post efficiency. Suppose every agent prefers the outcome (a',t') to the outcome (a,t) when the state of the world is $\theta=(\theta_1,...,\theta_n)$. Still, agent i of type θ_i may be reluctant to switch to (a',t') if there exists another state $\left(\theta_i,\widetilde{\theta}_{-i}\right)$ in which he prefers (a,t) to (a',t'). The next example illustrates this possibility.

Example 3 *A buyer and a seller can trade a single good. There are two equally likely states: in state* L *the buyer's value is* 0 *and the seller's value is* 1*, and in state* H *the buyer's value is* 2 *and the seller's value is* $\frac{3}{2}$*. The seller privately observes the realization of the state, while the buyer is uninformed. Consider a mechanism that prescribes no trade and no payment (independently of the agents' messages). The trivial equilibrium of this mechanism is EPRP despite being ex post inefficient in state* H*.*

Suppose the agents are offered to renegotiate this outcome to a trade at a price $\widehat{p}$*. Note that* $\widehat{p}$ *must be at least* $\frac{3}{2}$ *to give incentives for the seller to trade in state* H*. But then the seller would like to trade in state* L *as well. So, no trade will take place since the buyer is not willing to pay more than his expected value that is equal to* $\frac{1}{2}\cdot 0+\frac{1}{2}\cdot 2=1$*.*

In mechanism design theory the case of incomplete information with correlated agents' beliefs is very similar to the complete information case. In the latter case the mechanism designer can ask the agents to report the state of the world and verify their reports against each other. In the former case the mechanism designer can ask the agents to report their types and "stochastically compare" them against each other. For our next result we introduce the following condition from Kosenok and Severinov (2008).

Condition 3 *The prior probability distribution* P *over* Θ *satisfies:*

(i) For every agent i and every type θ_i':

$$P\left(\cdot \mid \theta_i'\right) \neq \sum_{\theta_i \in \Theta_i} c_{\theta_i,\theta_i'} P\left(\cdot \mid \theta_i\right) \qquad \forall c_{\theta_i,\theta_i'} \in \mathbb{R}_+,\ \theta_i, \theta_i' \in \Theta_i$$

(ii) For every probability distribution $P' \neq P$ there exists agent i of type θ_i' (with a positive P'-probability) such that:

$$P'\left(\cdot \mid \theta_i'\right) \neq \sum_{\theta_i \in \Theta_i} c_{\theta_i,\theta_i'} P\left(\cdot \mid \theta_i\right) \qquad \forall c_{\theta_i,\theta_i'} \in \mathbb{R}_+,\ \theta_i, \theta_i' \in \Theta_i$$

Part (i) of Condition 3 requires the conditional beliefs of each type of each agent to be linearly independent (in a restricted sense). It allows to design personalized transfers for each agent (as a function of reports) that ensure truthtelling.[26] If we want the system of personalized transfers for the agents to satisfy budget balance, then this may distort the agents' incentives to tell the truth. Part (ii) of Condition 3 takes care of this issue.[27]

Given this condition we obtain the following incomplete information analog of Proposition 1.

Proposition 3 *Consider an incomplete information environment with $n \geq 3$ agents with a prior distribution P that satisfies Condition 3. A social choice function f is EPRP implementable if $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient for every θ.*

Implementation is obtained by the following mechanism. Each agent is asked to report his type, and the outcome prescribed by the social choice function for the reported types is implemented. In addition the agents exchange auxiliary transfers that ensure that each agent tells the truth, and these auxiliary transfers are redistributed between agents in a way that is budget balanced. This mechanism has a truthful equilibrium that induces an equilibrium outcome that is payoff equivalent to f from the agents' interim perspective. This equilibrium is EPRP: (i) the outcome is ex post efficient if the agents tell the truth; and (ii) the auxiliary transfers can be designed in a way that ensures that any strategy of sending the wrong report followed by renegotiation is unprofitable.

5 Ex post oracle renegotiation-proofness under incomplete information

5.1 Preliminaries

In this section we develop an alternative approach to ex post renegotiation under incomplete information. Though we believe that the notion of renegotiation described in Definition 3 (Section

[26] See, for example, Crémer and McLean (1988).

[27] Kosenok and Severinov (2008) show that Condition 3 is generically satisfied if $n > 3$; if $n = 3$ then it is generically satisfied if at least one agent has more than two types (Theorem 2, p.135).

4.1) is a natural first take on the problem, it has several shortcomings. First, it is natural to assume that alternative outcomes are proposed by the agents, rather than exogenously offered to the agents. Alternative proposals that are suggested by the agents may reveal some of their private information in addition to what has already been revealed by the mechanism's outcome. Second, an alternative outcome may be a result of a conversation among the agents in the course of which some additional private information is revealed. The notion of ex post renegotiation presented in the previous section implicitly rules out such possibilities for additional information revelation. So it may happen that an equilibrium of some mechanism is EPRP according to Definition 4 (Section 4.1), but not robust to ex post renegotiation that is preceded by communication among the agents.

Suppose that the mechanism produced the outcome (a,t). Then, an external device, which we call an *oracle*, suggests to the agents an alternative outcome (a',t') *and truthfully reveals the true state of the world* to the agents.[28] This implies that the agents' beliefs conditional on the outcome produced by the play of the mechanism and their private information no longer play any role, and the notion of an outcome being renegotiated is simplified. Thanks to the oracle's revelation of the state each agent knows which of the two outcomes he prefers. So (a,t) can be renegotiated to (a',t') if the latter Pareto dominates the former given the state of the world, as in the case of complete information (Section 3.1).

The oracle device is a shortcut that we take in order to model the possibility that the alternative proposals that arise in the process of renegotiation may depend on the agents' private information beyond what is revealed by the mechanism's outcome. The assumption that the oracle perfectly reveals the state of the world to the agents is strong. Clearly, communication among the agents would not necessarily result in a full revelation of information; rather, the agents would reveal their private information in a way that is incentive compatible.[29] Nonetheless, we believe that such a strong notion of renegotiation accords well with our goal of determining which equilibria of which mechanisms are robust against various renegotiation scenarios. Indeed, we conjecture that (at least for the case of private values) if an equilibrium of a mechanism is oracle ex post renegotiation-proof, then it is also robust against renegotiation in *any* model with an explicit renegotiation/communication protocol. Furthermore, below we show that in some environments the requirements of ex post renegotiation-proofness and ex-post oracle renegotiation-proofness coincide.[30]

Definition 6 *Fix an equilibrium σ of a mechanism $\langle S,m\rangle$.*

(i) Suppose that an outcome (a,t) can be reached with positive probability under σ. We say that (a,t) can be oracle renegotiated with positive probability on the equilibrium path *if there exists another feasible outcome (a',t') and a state $\widehat{\theta}=\left(\widehat{\theta}_1,...,\widehat{\theta}_n\right)$ that has a positive probability (given (a,t) and σ) such that all agents in state $\widehat{\theta}$ weakly prefer, and at least one agent strictly prefers, (a',t') to (a,t).*

[28] Our notion of an oracle differs from the one used in computer science, where an oracle is an abstract machine that can solve certain decision problems in a single operation (http://en.wikipedia.org/wiki/Oracle_Turing_machine).

[29] In some settings the state of the world may become commonly known at the ex post stage. Then ex post oracle renegotiation-proofness is unambiguously required to ensure renegotiation-proofness of an equilibrium of a mechanism.

[30] This is the case for mechanisms in Examples 1' and 2' that are revisited in Section 5.2 below.

(ii) Suppose that an outcome (a,t) can be reached with a positive probability following a unilateral deviation of agent j, denoted σ'_j, from σ. We say that (a,t) can be oracle renegotiated with positive probability off the equilibrium path *if there exists another feasible outcome (a',t') and a state $\widehat{\theta} = \left(\widehat{\theta}_1, ..., \widehat{\theta}_n\right)$ that has a positive probability (given (a,t) and $\left(\sigma'_j, \sigma_{-j}\right)$) such that all agents in state $\widehat{\theta}$ weakly prefer, and at least one agent strictly prefers, (a',t') to (a,t).*

It is straightforward to adapt Definitions 4 and 5 (Section 4.1) to get a notion of *ex post oracle renegotiation-proof (EPRP*) equilibrium* of a given mechanism and a notion of *ex post oracle renegotiation-proof (EPRP*) implementable social choice function.* One immediate implication of ex post oracle renegotiation-proofness is that the outcome must be ex post efficient. This result is stated without proof.

Lemma 4 *In an incomplete information environment every EPRP* equilibrium is ex post efficient.*

In environments with private values ex post renegotiation-proofness implies ex post efficiency even without the oracle device (Lemma 2 in Section 4.2). However, this is not the case for environments with interdependent values as demonstrated in Example 3 (Section 4.3) where the equilibrium is EPRP but not EPRP*.

5.2 Independent private values

In this section we return to the case of private values and assume in addition that the agents' types are distributed independently. We present below a characterization of the class of social choice functions that are EPRP* implementable.

Proposition 4 *Consider an incomplete information environment with independent private values. A social choice function f is EPRP* implementable if and only if $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient and transfers t satisfy*

$$E_{\theta_{-i}}\left[t_i\left(\theta_i, \theta_{-i}\right)\right] = E_{\theta_{-i}}\left[\sum_{j \neq i} v_j\left(a^*\left(\theta_i, \theta_{-i}\right), \theta_j\right)\right] + H_i \qquad \forall i \in N,\ \theta_i \in \Theta_i, \tag{4}$$

for some $H_i \in \mathbb{R}$.

For the "if" part of the proposition we consider the following direct revelation mechanism. Each agent i reports his type $\theta_i \in \Theta_i$; an ex post efficient outcome rule $(a^*, t) : \Theta \to A \times \mathbb{T}$ that satisfies (4) is implemented given the reports. We show that this mechanism has a truthful equilibrium that is EPRP*. This mechanism can be called a *Groves in expectation mechanism*, because the interim expected equilibrium transfers of each agent are identical to those of some Groves mechanism.[31]

The "only if" part shows that all social choice functions that are EPRP* implementable are interim payoff equivalent to truthful equilibria of budget balanced Groves in expectation mechanisms.

[31] The class of "Groves in expectation" mechanisms includes Groves mechanisms, AGV mechanisms (Arrow (1979), d'Aspremont and Gerard-Varet (1979)), as well as other mechanisms.

While this result looks similar to Proposition 2 under complete information (Section 3.2), there are two important differences. First, this result is true for any number of agents, compared to only two agents in the case of Proposition 2, and second, this result is in terms of Groves in expectation mechanisms, while Proposition 2 is in terms of Groves mechanisms.

Groves in expectations mechanisms are known to exist.[32] Hence, by Proposition 4 we get the following result.

Corollary 2 *Consider an incomplete information environment with independent private values. Then EPRP* implementable social choice functions exist.*

It is possible to verify that the EPRP implementable social choice functions in Examples 1' and 2' (Section 4.2) are also EPRP* implementable. Even without the revelation of the state by the oracle, it may be possible for an agent who misreports his type in the mechanism to capture through renegotiation the full marginal total surplus associated with his report. Hence, EPRP implementable social choice functions must assign transfers that make each agent internalize the expected externality on the other agents associated with his report, i.e., the EPRP social choice functions must be interim payoff equivalent to Groves in expectations allocations.[33]

In Section 4.2 we noted that no EPRP implementable social choice functions in Example 2' satisfy interim individual rationality. To get a better sense of the scope of this nonexistence in other settings suppose that each agent's set of types is a connected open subset of an Euclidean space and that payoffs are continuously differentiable in types. Then any mechanism that is both ex post efficient and interim incentive compatible is payoff equivalent to a Groves in expectations mechanism.[34] It is well known that in several economically important environments with incomplete information there are no individually rational ex post efficient mechanisms.[35] Whenever this is the case, Proposition 4 implies that there are no interim individually rational EPRP* mechanisms either.

5.3 Correlated and interdependent values

Lemma 4 (Section 5.1) implies that every EPRP* equilibrium must be ex post efficient. Thus the inefficient EPRP equilibrium of the mechanism in Example 3 (Section 4.3) is not EPRP*. As a matter of fact, the environment described in Example 3 admits no EPRP* implementable social choice functions.

Example 3 (continued) *Ex post efficiency requires trade if and only if the state is H. But this is not incentive compatible for the seller: he is more eager to trade in state L, when his value is*

[32] This follows from existence of AGV mechanisms mentioned in footnote 31.

[33] This is not always the case. In an earlier version of the paper we have an example where an ex post efficient social choice function is EPRP implementable but not EPRP* implementable.

[34] See, for example, Williams (1999).

[35] For example, bilateral trade (Myerson and Satterthwaite (1983)), public good provision (Mailath and Postlewaite (1990)), litigation and settlement (Spier (1994), Klement and Neeman (2005)).

1, *than in state* H*, when his value is* $\frac{3}{2}$*. Hence there are no EPRP* implementable social choice functions in this environment.*

EPRP* implementable social choice functions exist if there are at least three agents with correlated types. It is straightforward to verify that the mechanism described in the proof of Proposition 3 (Section 4.3) is EPRP*.[36]

6 Interim renegotiation-proofness

6.1 Complete information

In this section we discuss the notion of interim renegotiation-proofness under complete information. We keep the discussion somewhat informal since the observations we make are rather straightforward.

Interim renegotiation may take place at the interim stage, that is after the state of the world θ is realized and becomes known among the agents, but before the agents play the mechanism. Interim renegotiation proceeds according to some communication procedure among the agents, through which they may choose some alternative mechanism $\langle S', m' \rangle$ to replace the original mechanism $\langle S, m \rangle$ by unanimous consent. We are interested in the question of what it takes for a given equilibrium σ of a mechanism $\langle S, m \rangle$ to be immune against renegotiation before it is played, as well as what kind of social choice functions f can be implemented in a way that is interim renegotiation-proof.

If a given equilibrium (or a social choice function) is not ex post efficient in some state of the world, then there is obviously a scope for interim renegotiation. Since interim renegotiation takes place after the state of the world is commonly observed, every agent would prefer an alternative mechanism that implements (independently of the agents' messages) an ex post efficient alternative and makes all the agents strictly better off. On the other hand, if a given equilibrium (or a social choice function) is ex post efficient, then there can be no successful renegotiation at the interim stage, because any alternative allocation would make some agents worse off and so will be blocked. Thus using Proposition 1 (Section 3.2) we have the following result.

Claim 1 *Consider a complete information environment with* $n \geq 3$ *agents. A social choice function* f *is interim and ex post renegotiation-proof implementable if and only if* $f(\theta) = (a(\theta), t(\theta))$ *is ex post efficient for every* θ*.*

When there are two agents then all EPRP implementable social choice functions are outcome equivalent to truthful equilibria of budget balanced Groves mechanisms (Proposition 2 in Section 3.2), which implies the following result.

Claim 2 *Consider a complete information environment with product domain and two agents. A social choice function* f *is interim and ex post renegotiation-proof implementable if and only if*

[36] Moreover, Lemma 4 implies that the "only if" statement of Proposition 3 can be strengthened to "if and only if."

$f(\theta) = (a(\theta), t(\theta))$ is ex post efficient for every θ and transfers t satisfy the formula for Groves transfers given in (1).

6.2 Incomplete information

In this section we discuss the notion of interim renegotiation-proofness under incomplete information. As in the case of complete information (Section 6.1) we keep the discussion informal.

Interim renegotiation may take place at the interim stage, that is, after the state of the world $\theta = (\theta_1, ..., \theta_n)$ has been realized and each agent i has learnt his private information θ_i, but before the agents play the mechanism. Interim renegotiation proceeds according to some communication procedure among the agents, through which they may choose some alternative mechanism $\langle S', m' \rangle$ to replace the original mechanism $\langle S, m \rangle$ by unanimous consent. We are interested in the question of what it takes for a given equilibrium σ of mechanism $\langle S, m \rangle$ to be immune against renegotiation before it is played, as well as what kind of social choice functions f can be implemented in a way that is interim renegotiation-proof.

It is natural to suppose that (similarly to the case of complete information) a given equilibrium in some mechanism is interim renegotiation-proof if and only if the equilibrium allocation is Pareto efficient at the interim stage given the agents' incentive constraints, or is interim (incentive) efficient in Holmström and Myerson's (1983) terminology. While interim efficiency is clearly a necessary condition for interim renegotiation-proofness, since otherwise every type of every agent could be made better off by some alternative mechanism that implements an interim efficient allocation, the converse is not true. Holmström and Myerson (1983) show that the agents may choose to reveal some private information at the interim stage and renegotiate away from an interim efficient equilibrium. For the case of quasi-linear private value environments, Palfrey and Srivastava (1993) show that every interim efficient social choice function can be uniquely implemented in a way that is robust to pre-play communication and interim renegotiation.[37] However, their construction relies on inefficient punishments off the equilibrium path and so can be undermined by the possibility of ex post renegotiation.

So far we have not been able to fully characterize the set of social choice functions that can be implemented in a way that is both interim and ex post renegotiation-proof. We are able, however, to provide simple sufficient conditions for such implementation for environments with independent private values and for environments with correlated beliefs.

Claim 3 *Consider an incomplete information environment with independent private values. A social choice function f is interim and ex post (both with and without the oracle) renegotiation-proof implementable if $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient for every θ and transfers t are Groves*

[37] Maskin and Tirole (1992) prove a similar result for interdependent values environments when only one agent has private information.

transfers, or such that

$$t_i(\theta_i, \theta_{-i}) = \sum_{j \neq i} v_j(a^*(\theta_i, \theta_{-i}), \theta_j) + H_i \qquad \forall i \in N,\ \theta_i \in \Theta_i, \tag{5}$$

for some $H_i \in \mathbb{R}$.

Note that the requirement on the allowed transfers here is stronger than that for ex post renegotiation-proofness: this result is for Groves mechanisms while Proposition 4 (Section 5.2) is in terms of Groves in expectations mechanisms. Each agent has a weakly dominant strategy to report the truth, which together with ex post efficiency of the equilibrium outcome ensures that the agents would not want to change their behavior or want to switch to another mechanism at the interim stage.[38]

In order to get a result that is analogous to Proposition 3 we introduce the following condition from Crémer and McLean (1988).

Condition 4 *Suppose that*

(i) The prior probability distribution P *over* Θ *satisfies for every agent* $i \neq 1$ *and every type* θ_i'*:*

$$P(\cdot \mid \theta_i') \neq \sum_{\theta_i \in \Theta_i} c_{\theta_i, \theta_i'} P(\cdot \mid \theta_i) \qquad \forall c_{\theta_i, \theta_i'} \in \mathbb{R},\ \theta_i, \theta_i' \in \Theta_i$$

(ii) Agent 1 *has no private information:* $v_1(a, \theta) = v_1(a)$ *for every* $a \in A$, $\theta \in \Theta$.

Note that part (i) of this condition is stronger than part (i) of Condition 3 (Section 4.3) in that it requires the agents' conditional beliefs to be fully linearly independent. Part (ii) of this condition implies that agent 1 can be utilized as a "budget breaker."

Claim 4 *Consider an incomplete information environment with* $n \geq 3$ *agents such that Condition 4 is satisfied. A social choice function* f *is interim and ex post (both with and without the oracle) renegotiation-proof implementable if* $f(\theta) = (a(\theta), t(\theta))$ *is ex post efficient for every* θ.

7 Related literature

7.1 Complete information

Most of the papers on renegotiation under complete information are from the hold-up and incomplete contracts literature, and the majority of these papers deal with ex post rather than with interim renegotiation. This literature can be divided into three parts: the papers following a standard approach to renegotiation, the papers considering multiple renegotiation scenarios, and the "technological detail" literature.

[38] In Section 3.2 we noted that the existence of budget balanced Groves mechanisms cannot, in general, be guaranteed, but such mechanisms exist if (at least) one of the agents has no private information such as in the case of an auction with an uninformed seller.

Standard approach. The standard approach to modeling renegotiation assumes that ex post renegotiation proceeds according to some given process (or a game). This process is commonly known among the agents, and is also known to the mechanism designer who may take it into account when designing the mechanism. It is often convenient to treat this process as exogenous, that is to assume, like Maskin and Moore (1999), that ex post renegotiation is described by some renegotiation function h that maps outcomes and states into (possibly other) outcomes.[39]

As an illustration of this approach consider the model studied by Edlin and Reichelstein (1996) and Che and Hausch (1999).[40] A buyer and a seller make relation-specific investments that determine their payoffs from subsequent trade. An investment of an agent is called *selfish* if it affects only the agent's own payoff, and is called *cooperative* if it affects the other agent's payoff as well. Following the investment decisions the state of the world is realized and commonly observed, and then the agents play the game that is induced by a given mechanism. The outcomes determined by the mechanism are renegotiated as follows: (i) the post-renegotiation outcome is ex post efficient; (ii) the renegotiation surplus, i.e., the difference between the efficient total surplus and the total surplus from the outcome reached by the mechanism, is commonly known to be shared between the buyer and seller in proportions $1-\alpha$ and α (where $\alpha \in [0,1]$). These assumptions uniquely determine the renegotiation function h mentioned in the previous paragraph. Our approach to ex post renegotiation in such an environment is roughly equivalent to assuming that the sharing rule α is not known to the mechanism designer, and our notion of ex post renegotiation-proofness of an equilibrium of a mechanism is roughly equivalent to a requirement that the equilibrium is not renegotiated for any $\alpha \in [0,1]$.[41]

Edlin and Reichelstein (1996) show that if the agents' investments are selfish, then there exist mechanisms that result in first best trade and investments regardless of the value of α. Che and Hausch (1999) show that if the agents' investments are cooperative, then the existence of such first best mechanisms is not guaranteed, and the form of the second best mechanisms depends crucially on α. To see how these results are consistent with our model recall that all EPRP implementable social choice functions are outcome equivalent to truthful equilibria of budget balanced Groves mechanisms (Proposition 2 in Section 3.2). It is known that Groves mechanisms provide the first best incentives for ex ante investment in the case of selfish investments, but in the case of cooperative investments may lead to either over- or underinvestment.[42] Hence, whenever budget balanced Groves mechanisms exist, our results are consistent with the results of Edlin and Reichelstein (1996) and Che and Hausch (1999): if investments are selfish, then there exist mechanisms that

[39] Some papers, like Chung (1991) and Aghion et al. (1994) assume that the mechanism designer has the power to affect the way the renegotiation will proceed.

[40] Segal and Whinston (2002) discuss other papers that use a similar approach.

[41] In addition: (i) we do not assume that ex post renegotiation is always ex post efficient; and (ii) we allow the renegotiation surplus to be shared differently in different circumstances (i.e., α may depend on the state of the world, the reached outcome in the mechanism, etc.). We also require no renegotiation on the equilibrium path, while these papers allow for renegotiation on the equilibrium path.

[42] See, for example, Rogerson (1992). Recently, Bergemann and Valimaki (2002) showed that Groves mechanisms provide first best incentives for ex ante information acquisition whenever each agent can only improve the precision of his own signal about his payoff. Otherwise Groves mechanisms may lead to either over- or underinvestment.

result in the first best trade and investments and are robust against various renegotiation scenarios; if investments are cooperative, then the existence of such mechanisms is not guaranteed.[43]

Multiple renegotiation scenarios. The approach of Maskin and Moore (1999) that is based on renegotiation function h can be used to study situations when there are many possible ex post renegotiation scenarios. Suppose the agents commonly learn about the exact way ex post renegotiation will proceed after the state of the world is realized. Then this information can be included in the description of the state of the world, e.g. states θ and θ' may describe the same preferences of the agents, but differ in how renegotiation is expected to proceed.[44]

The mechanism designer may then try to extract information about anticipated renegotiation from the agents through the course of play of the mechanism, and to condition the outcomes of the mechanism on this information. Such mechanisms would be quite complex because they must provide the agents with the incentives to reveal all their private information – their payoff relevant private information and their beliefs about the expected renegotiation scenario.[45] Here, we study a simpler class of mechanisms that do not attempt to extract any information about expected renegotiation from the agents, but nonetheless implement the desired social choice function. Studying such mechanisms is a natural first step towards understanding of how to do mechanism design in the presence of multiple renegotiation scenarios.[46]

Sjöström (1999) obtains close analogs of our results for $n \geq 3$ agents (Proposition 1 in Section 3.2 and Claim 1 in Section 6.1). He studies an exchange economy with agents that may be risk-neutral or risk-averse, and assumes that the renegotiation function h is only known to belong to a given class of admissible renegotiation functions. The agents are not asked to report h. Theorem 1 in Sjöström (1999) shows that a social choice function is uniquely implementable in undominated Nash equilibrium if and only if it is Pareto efficient.

Rubinstein and Wolinsky (1992) develop an alternative approach to mechanism design when there are multiple renegotiation scenarios. Suppose there is a buyer and a seller who can trade a single good, and that trade is efficient in every state of the world. In Section III Rubinstein and Wolinsky (1992) study social choice functions that can be uniquely implemented in a subgame-perfect equilibrium that is renegotiation-proof in a sense that it is robust against possible rene-

[43] In environments with selfish investments where budget balanced Groves mechanisms do not exist, Edlin and Reichelstein (1996) are still able to achieve the first best trade and investments. We believe that the differences between our approaches mentioned in footnote 41 are responsible for these discrepances in our results.

[44] See Section 2 in Maskin and Moore (1999) for details.

[45] We are aware of only two papers that allow such complex mechanisms. Green and Laffont (1994) consider a quasi-linear model with two players and costly renegotiation. They are interested in constructing a renegotiation-proof mechanism for implementing ex post efficient alternatives (regardless of the transfers), while our goal is to characterize all renegotiation-proof social choice functions. Amoros (2004) studies unique implementation in general environments where the renegotiation function h is only known to belong to a given class of admissible renegotiation functions. His characterization results require the mechanism designer to be able to destroy the agents' endowments, which would lead to renegotiation in our model.

[46] Conceptually, the simpler mechanisms we study relate to the complex mechanisms outlined above in a way that is similar to the way that dominant strategy (and ex post) mechanisms relate to Bayesian mechanisms. Dominant strategy (and ex post) mechanisms are "belief-free" and rely only on the agents' payoff relevant information, while Bayesian mechanisms may in addition rely on the agents' higher-order beliefs (Bergemann and Morris (2005), Chung and Ely (2007)). The simple mechanisms we study are "belief-free" with respect to anticipated renegotiation scenarios, while complex mechanisms may rely on the agents' beliefs about future renegotiation.

gotiation *both on and off the equilibrium path* in every state of the world. Hence, their notions of implementation and renegotiation-proofness are stronger than ours.[47] Rubinstein and Wolinsky (1992) show that only social choice functions that prescribe trade at a constant price for every state of the world are implementable in a way that is renegotiation-proof. It is easy to verify that in such an environment these are the only EPRP implementable social choice functions as well.[48]

Technological detail. The approach of Maskin and Moore (1999) that is based on a renegotiation function h implicitly assumes that there is a certain moment in time at which the outcome that is reached by the mechanism is enforced by the mechanism designer, and that no further changes are possible.[49] There is a recent literature that reexamines sensitivity of the results of the models with ex post renegotiation to changes in this implicit assumption, as well as to introducing frictions into the renegotiation process.

For example, Lyon and Rasmusen (2004) demonstrate that the outcomes of renegotiation games are very sensitive to specific assumptions about the extensive form.[50] Watson (2007) shows that the effect of ex post renegotiation may be significantly reduced when the actions that consummate trade are modelled as inalienable actions of some agents and trade opportunities are nondurable.[51] In related work, Evans (2009) shows that ex post renegotiation is no stronger than interim renegotiation if sending a message in a mechanism entails an individual cost for the agent who sends it.[52]

In general, Watson (2007) and other papers in this literature argue in favor of paying close attention to the details of trade and renegotiation technology, because seemingly innocuous details can make a big difference. Since the goal of our paper is to study which equilibria of which mechanisms are robust against various renegotiation scenarios, it makes sense for us to allow for the most permissive notion of renegotiation, that is, to work with a standard setting like in Maskin and Moore (1999).[53]

[47] We rely on non-unique implementation in Nash equilibrium; our EPRP equilibria are required to be robust against renegotiation on the equilibrium path, but renegotiation off the equilibrium path is permitted as long as it does not tempt any agent to deviate from equilibrium play (Definition 1 in Section 3.1).

[48] By Proposition 2 (Section 3.2) all EPRP implementable social choice functions are outcome equivalent to truthful equilibria of budget balanced Groves mechanisms, which in such an environment prescribe trade at a constant price for every state of the world.

[49] See Section IX in Moore (1992) for details.

[50] In particular, Lyon and Rasmusen (2004) argue in favor of one particular way of modeling option contracts (like in Noldeke and Schmidt (1998)), so that the option contracts can be used to overcome the hold up problem and be immune to the criticisms put foward in Edlin and Hermalin (2000). Wickelgren (2007) and Watson and Wignall (2007) reexamine this argument in the context of durable trading opportunities.

[51] Watson and Wignall (2007) show that the assumption that the trade opportunities are nondurable is not essential for the result.

[52] The construction in Evans (2009) is similar to the one in Section IV in Rubinstein and Wolinsky (1992). They introduce an explicit time dimension, and show that a possibility of costly delays significantly weakens the effects of ex post renegotiation.

[53] In the terminology of the "technological detail" literature, we model the trade outcomes as "public actions", and we assume that renegotiation is costless and there is sufficient time for it (after the formal provisions of the contract are carried out and before the trade actually occurs).

7.2 Incomplete information

7.2.1 Ex post renegotiation

Like in the case of complete information, the related literature on ex post renegotiation under incomplete information can also be divided into three: the literature that follows a standard approach to renegotiation, the literature that considers multiple renegotiation scenarios, and the "technological detail" literature.

Standard approach. As in the case of complete information, the standard approach to modeling renegotiation under incomplete information assumes that ex post renegotiation proceeds according to some given process (or game). This process is commonly known among the agents, and is also known to the mechanism designer who can take it into account when designing the mechanism. Ex post renegotiation can thus be described by a renegotiation function h that maps outcomes, states and the agents' beliefs into outcomes and the agents' beliefs.

For example, Beaudry and Poitevin (1995) study a model with two agents, one informed and one uninformed. After the agents have played the game induced by a given mechanism, they have a chance to renegotiate the prescribed outcomes as follows. One of the agents proposes a new mechanism; if the other agent agrees then the agents play the new mechanism, which determines the final outcome; if the other agent disagrees then the outcome prescribed by the original mechanism is implemented. The proposer of the new mechanism takes into account the prescribed outcome, his beliefs about the state of the world, beliefs about the opponent's beliefs, etc. The equilibrium of the consequent game (consisting of the ratification procedure and the new mechanism) determines new outcomes and revised beliefs for every state of the world. This procedure uniquely determines the renegotiation function h mentioned in the previous paragraph.

Beaudry and Poitevin (1995) show that depending on who makes the renegotiation proposal, ex post renegotiation may have different consequences both in terms of payoffs and the amount of information revealed.[54] For example, in the case of private values ex post renegotiation results in ex post efficient outcomes when the uninformed agent is the proposer, but it may result in ex post inefficient outcomes when the proposal is made by the informed agent.

Multiple renegotiation scenarios. Holmström and Myerson (1983) were the first to raise the issue of the robustness of the mechanism to ex post renegotiation when there are many possible renegotiation scenarios.[55] Forges (1993, 1994) introduced several alternative notions of *posterior efficiency* that addressed this issue.[56] An equilibrium in a given mechanism is posterior efficient if the outcomes obtained by the equilibrium play cannot be renegotiated with positive probability on the equilibrium path, where the approach to modeling renegotiation of an outcome is very similar

[54] This seems to be a general feature of such models. For example, Fudenberg and Tirole (1990) and Ma (1994) study a moral hazard model with renegotiation and obtain qualitatively different results depending on who initiates renegotiation.

[55] See "Concluding comments" in Holmström and Myerson (1983).

[56] Green and Laffont (1987) introduced and studied a related but distinct notion of *posterior implementability*. It addresses a concern that the agents may want to change their individual actions in the original mechanism once they observe the actions of the other agents, but before the mechanism's outcome is realized. In contrast we are concerned with a possibility that the agents may jointly want to change the mechanism's outcome after it is realized.

to ours. Hence the definition of posterior efficient equilibrium requires an analog of part (i) of our definition of EPRP equilibrium (Definition 4 in Section 4.1). However, the concept of posterior efficiency assumes that ex post renegotiation is not anticipated by the agents when they play in the mechanism and thus cannot undermine the incentive compatibility of the original equilibrium. Thus the definition of posterior efficient equilibrium does not have an analog to part (ii) of our definition of EPRP equilibrium. Forges (1993, 1994) showed that in case of private values every posterior equilibrium is ex post efficient (like our Lemma 2 in Section 4.2), but that this is not true in the case of interdependent valuations (like our Example 3 in Section 4.3).[57]

Forges (1993, 1994) provided examples where a posterior efficient equilibrium is undermined if ex post renegotiation is anticipated by the agents.[58] To address this issue Forges (1993) introduced and briefly discussed a notion of renegotiation-proof equilibrium of a given mechanism which is very similar to our notion of EPRP equilibrium in that it contains analogs of part (i) and (ii) of Definition 4, but did not provide a characterization like we do here.[59]

Krasa (1999) introduced and studied another related notion of ex post renegotiation-proofness called *unimprovability* in the setting of an exchange economy.[60] Unimprovability is a property of a state contingent allocation rather than a property of an equilibrium of some mechanism, and the process by which the agents arrive at a given allocation is not modeled. The allocation is called unimprovable if the agents do not want to change it by re-trading goods or revealing further information. In this respect the definition of unimprovable allocation is similar to the requirement on EPRP equilibrium outcomes given in part (i) of our Definition 4. Unlike in our modeling approach however, the notion of unimprovability requires the alternative allocation proposals to be "credible", in a sense that they could not in turn be improved upon by some other alternative allocation.[61]

Technological detail. As in the case of complete information, this literature reexamines the sensitivity of the results of the models with ex post renegotiation using a standard approach to changes in various assumptions. For example, Evans (2009) shows that the set of ex post renegotiation-proof implementable social choice functions becomes quite large if sending a message in a mechanism entails an individual cost for the agent who sends the message.

7.2.2 Interim renegotiation

Holmström and Myerson (1983) were the first to address the issue of interim renegotiation-proofness by introducing the concept of durability. Consequent literature introduced related concepts: reformulated durability in Crawford (1985), ratifiability in Cramton and Palfrey (1995), resilience in Lagunoff (1995), renegotiation-proofness in the informed principal model in Maskin and Tirole (1992), and communication-proofness and interim renegotiation-proofness in Palfrey and Srivastava

[57] This is discussed on p.139 in Forges (1993) and on p.249 in Forges (1994).

[58] See Example 3.3 in Forges (1993) and the last example in Forges (1994).

[59] This definition and its discussion are on p.143-146 in Forges (1993).

[60] The model in Krasa (1999) is closely related to models in the literature on cooperative solution concepts with incomplete information. See Krasa (1999) and Forges, Minelli and Vohra (2002) for discussion of this literature. See also Sen (2008) for an alternative approach to studying renegotiation-proof allocations.

[61] See footnote 24 for the rationale for our modeling choice.

(1993).

These different concepts were either only shown to exist in a strict subset of environments, or, where shown to always exist as in the case of Holmström and Myerson's durability, were arguably too weak. Durability for example is robust against renegotiation that selects one nontrivial equilibrium in the renegotiation game but not robust against renegotiation that selects the best equilibrium in the renegotiation game.[62]

8 Conclusion

We have analyzed a mechanism design problem under the assumption that there are multiple renegotiation scenarios, and have investigated the properties of renegotiation-proof equilibria of mechanisms and the set of renegotiation-proof implementable social choice functions. We have shown that ex post renegotiation-proof equilibria are often required to be ex post efficient, and that there are no additional restrictions on the implementable social choice functions when there are at least three agents and their reports can be verified against each other. When there are two agents under complete information, or when the agents' private information cannot be cross-checked against the opponents' reports, there are additional restrictions on the set of implementable transfers: each agent has to internalize the effect on total surplus that is associated with his report. We found that the requirements of interim renegotiation-proofness do not impose additional restrictions on the implementable social choice functions in the case of complete information, and are likely to impose additional restrictions in the case of incomplete information.

The notion of renegotiation-proofness we have chosen to work with is strong. We believe that for many specific applications the scope of renegotiation can be significantly reduced through careful contract design that pays special attention to "technological details". We hope that our theory of renegotiation-proof mechanisms would be viewed as theory of "stable" institutions. If a mechanism retains its original form under a wide variety of circumstances, then such a mechanism must be renegotiation-proof in our strong sense. But if such a renegotiation-proof mechanism does not exist in a given environment, then one should not expect to observe an institution for governing transactions whose rules remain stable and not subject to perpetual renegotiation.

9 Appendix

Proof of Proposition 1. <**If**> Suppose $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient in every $\theta \in \Theta$, i.e. $a(\theta) = a^*(\theta)$ and $t(\theta)$ is budget balanced. Consider a mechanism where agents 1 and 2 report the state of the world $\widehat{\theta}^i \in \Theta$, and the outcome as a function of reports is determined as follows:

$$\left(\alpha\left(\widehat{\theta}^1, \widehat{\theta}^2\right), \tau_1\left(\widehat{\theta}^1, \widehat{\theta}^2\right), ..., \tau_n\left(\widehat{\theta}^1, \widehat{\theta}^2\right)\right) = \begin{cases} (a^*(\theta), t_1(\theta), ..., t_n(\theta)) & \text{if } \widehat{\theta}^1 = \widehat{\theta}^2 = \theta \in \Theta \\ (a_0, -L, -L, 2L, 0, ..., 0) & \text{if } \widehat{\theta}^1 \neq \widehat{\theta}^2 \end{cases}$$

[62] See "Concluding comments" in Holmström and Myerson (1983).

where $a_0 \in A$ is some fixed alternative, and a penalty payment L is chosen to be sufficiently large.[63]

There exists a truthful equilibrium since the penalty is chosen to be large enough. To see that this equilibrium is EPRP first note that there will be no renegotiation on the equilibrium path since the outcome $(a^*(\theta), t(\theta))$ is ex post efficient for every θ. If either of the agents misreports, then both pay penalty L to agent 3. The penalty is chosen to be large enough to make such deviations unprofitable, even if a deviator consequently renegotiates the outcome and captures the implied increase in the total surplus.

<**Only If**> Follows from Lemma 1. ■

Proof of Proposition 2. <**If**> Suppose $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient in every $\theta \in \Theta$, i.e. $a(\theta) = a^*(\theta)$ and $t(\theta)$ is budget balanced, and transfers t satisfy (1). Consider a mechanism where each agent i reports his own payoff relevant type $\theta_i \in \Theta_i$, and the outcome rule as a function of reports is given by f. In this mechanism truthtelling is a dominant strategy, since for each agent i of type θ_i

$$\begin{aligned} v_i(a^*(\theta_i,\theta_j),\theta_i) + t_i(\theta_i,\theta_j) &= v_i(a^*(\theta_i,\theta_j),\theta_i) + v_j(a^*(\theta_i,\theta_j),\theta_j) + H_i(\theta_j) \\ &\geq v_i\left(a^*\left(\theta_i',\theta_j\right),\theta_i\right) + v_j\left(a^*\left(\theta_i',\theta_j\right),\theta_j\right) + H_i(\theta_j) \\ &= v_i\left(a^*\left(\theta_i',\theta_j\right),\theta_i\right) + t_i\left(\theta_i',\theta_j\right) \qquad \forall \theta_i' \neq \theta_i \end{aligned}$$

where the equalities use (1), and the inequality uses the fact that $a^*(\theta_i,\theta_j)$ is ex post efficient in state (θ_i,θ_j). Hence, the truthful equilibrium of this mechanism implements f.

Suppose the above equilibrium is not EPRP. First, note that there will be no renegotiation on the equilibrium path since the outcome $(a^*(\theta), t(\theta))$ is ex post efficient for every θ. Suppose agent 1 in state $\theta = (\theta_1,\theta_2)$ can improve upon his truthtelling payoff by sending a wrong report $\theta_1' \neq \theta_1$ and consequent renegotiation of the outcome from $\left(a^*\left(\theta_1',\theta_2\right), t\left(\theta_1',\theta_2\right)\right)$ to another outcome $\left(\widehat{a},\widehat{t}\right)$. This strategy is beneficial for agent 1 if

$$v_1(\widehat{a},\theta_1) + \widehat{t}_1 > v_1(a^*(\theta_1,\theta_2),\theta_1) + t_1(\theta_1,\theta_2) \tag{6}$$

agent 2 agrees to such renegotiation if

$$v_2(\widehat{a},\theta_2) + \widehat{t}_2 \geq v_2\left(a^*\left(\theta_1',\theta_2\right),\theta_2\right) + t_2\left(\theta_1',\theta_2\right) \tag{7}$$

Sum up (6) and (7):

$$\begin{aligned} v_1(\widehat{a},\theta_1) + v_2(\widehat{a},\theta_2) + \widehat{t}_1 + \widehat{t}_2 &> v_1(a^*(\theta_1,\theta_2),\theta_1) + v_2\left(a^*\left(\theta_1',\theta_2\right),\theta_2\right) + t_1(\theta_1,\theta_2) + t_2\left(\theta_1',\theta_2\right) \\ &= v_1(a^*(\theta_1,\theta_2),\theta_1) + v_2\left(a^*\left(\theta_1',\theta_2\right),\theta_2\right) + t_1(\theta_1,\theta_2) - t_1\left(\theta_1',\theta_2\right) \\ &= v_1(a^*(\theta_1,\theta_2),\theta_1) + v_2(a^*(\theta_1,\theta_2),\theta_2) \end{aligned}$$

[63] We can take

$$L > 2n\left(\max_{i,a,\theta}|v_i(a,\theta)| + \max_{i,\theta}|t_i(\theta)|\right).$$

where the first equality uses budget balance ($t_2\left(\theta_1', \theta_2\right) = -t_1\left(\theta_1', \theta_2\right)$), and the second equality uses (1). But $a^*(\theta_1, \theta_2)$ is ex post efficient in state (θ_1, θ_2) and $\widehat{t}_1 + \widehat{t}_2 \leq 0$:

$$v_1\left(a^*\left(\theta_1, \theta_2\right), \theta_1\right) + v_2\left(a^*\left(\theta_1, \theta_2\right), \theta_2\right) \geq v_1\left(\widehat{a}, \theta_1\right) + v_2\left(\widehat{a}, \theta_2\right) + \widehat{t}_1 + \widehat{t}_2$$

which gives a contradiction. Hence, the truthful equilibrium of this mechanism is EPRP.

<**Only If**> Suppose a mechanism $\langle S, m\rangle$ has an EPRP equilibrium σ that implements the social choice function f. For every state $\theta = (\theta_1, \theta_2) \in \Theta$ define a maximized total surplus: $S(\theta_1, \theta_2) = \sum_{i=1}^{2} v_i\left(a^*\left(\theta_1, \theta_2\right), \theta_i\right)$.

First, by Lemma 1 note that $f(\theta) = (a(\theta), t(\theta))$ must be ex post efficient: $a(\theta) = a^*(\theta)$ and $t_1(\theta) + t_2(\theta) = 0$ for every θ. Next, suppose agent 1 in state θ uses strategy $\sigma_1\left(\theta'\right)$ where $\theta' = \left(\theta_1', \theta_2'\right) \in \Theta$. Since σ is EPRP, agent 1 cannot improve upon his equilibrium payoff even if the resulting outcomes are renegotiated in the most profitable way. That is, every outcome $\left(\widetilde{a}, \widetilde{t}\right)$ is renegotiated to an ex post efficient outcome $\left(a^*(\theta), \widehat{t}\right)$ such that agent 2 receives a payoff that leaves him indifferent between the two outcomes:

$$v_2\left(a^*(\theta), \theta_2\right) + \widehat{t}_2 = v_2\left(\widetilde{a}, \theta_2\right) + \widetilde{t}_2$$

The payoff of agent 1 in this case is

$$v_1\left(a^*(\theta), \theta_1\right) - \widehat{t}_2 = v_1\left(a^*(\theta), \theta_1\right) + v_2\left(a^*(\theta), \theta_2\right) - \left(v_2\left(\widetilde{a}, \theta_2\right) + \widetilde{t}_2\right)$$

The renegotiation-proofness constraint to prevent such deviations by agent 1 is

$$v_1\left(a^*(\theta), \theta_1\right) + t_1(\theta) \geq v_1\left(a^*(\theta), \theta_1\right) + v_2\left(a^*(\theta), \theta_2\right) - E_{m\circ(\sigma_1(\theta'), \sigma_2(\theta))}\left[v_2\left(\widetilde{a}, \theta_2\right) + \widetilde{t}_2\right] \tag{8}$$

where the expectation for the last term is taken with respect to a probability measure over outcomes $\left(\widetilde{a}, \widetilde{t}\right)$ induced by the outcome rule m and the strategy profile $\left(\sigma_1\left(\theta'\right), \sigma_2(\theta)\right)$. Rewrite (8):

$$t_1(\theta) + E_{m\circ(\sigma_1(\theta'), \sigma_2(\theta))}\left[\widetilde{t}_2\right] \geq v_2\left(a^*(\theta), \theta_2\right) - E_{m\circ(\sigma_1(\theta'), \sigma_2(\theta))}\left[v_2\left(\widetilde{a}, \theta_2\right)\right] \qquad \forall \theta' \neq \theta \tag{9}$$

Repeat the same argument for agent 2:

$$t_2(\theta) + E_{m\circ(\sigma_1(\theta), \sigma_2(\theta'))}\left[\widetilde{t}_1\right] \geq v_1\left(a^*(\theta), \theta_1\right) - E_{m\circ(\sigma_1(\theta), \sigma_2(\theta'))}\left[v_1\left(\widetilde{a}, \theta_1\right)\right] \qquad \forall \theta' \neq \theta \tag{10}$$

Sum up (9) and (10), and use $t_1(\theta) + t_2(\theta) = 0$:

$$\begin{aligned} & E_{m\circ(\sigma_1(\theta), \sigma_2(\theta'))}\left[\widetilde{t}_1\right] + E_{m\circ(\sigma_1(\theta'), \sigma_2(\theta))}\left[\widetilde{t}_2\right] \\ \geq\ & S(\theta_1, \theta_2) - E_{m\circ(\sigma_1(\theta), \sigma_2(\theta'))}\left[v_1\left(\widetilde{a}, \theta_1\right)\right] - E_{m\circ(\sigma_1(\theta'), \sigma_2(\theta))}\left[v_2\left(\widetilde{a}, \theta_2\right)\right] \end{aligned} \tag{11}$$

Switch the roles of θ and θ' in (11):

$$\begin{aligned} & E_{m\circ(\sigma_1(\theta'),\sigma_2(\theta))}\left[\widetilde{t}_1\right] + E_{m\circ(\sigma_1(\theta),\sigma_2(\theta'))}\left[\widetilde{t}_2\right] \\ \geq\ & S\left(\theta_1',\theta_2'\right) - E_{m\circ(\sigma_1(\theta'),\sigma_2(\theta))}\left[v_1\left(\widetilde{a},\theta_1'\right)\right] - E_{m\circ(\sigma_1(\theta),\sigma_2(\theta'))}\left[v_2\left(\widetilde{a},\theta_2'\right)\right] \end{aligned} \tag{12}$$

Sum up (11) and (12), and use the fact that transfers $\widetilde{t}$ are feasible:

$$\begin{aligned} & E_{m\circ(\sigma_1(\theta),\sigma_2(\theta'))}\left[v_1\left(\widetilde{a},\theta_1\right) + v_2\left(\widetilde{a},\theta_2'\right)\right] + E_{m\circ(\sigma_1(\theta'),\sigma_2(\theta))}\left[v_1\left(\widetilde{a},\theta_1'\right) + v_2\left(\widetilde{a},\theta_2\right)\right] \\ \geq\ & S\left(\theta_1,\theta_2\right) + S\left(\theta_1',\theta_2'\right) \end{aligned} \tag{13}$$

Thus a necessary condition for ex post renegotiation-proofness is

$$S\left(\theta_1,\theta_2'\right) + S\left(\theta_1',\theta_2\right) \geq S\left(\theta_1,\theta_2\right) + S\left(\theta_1',\theta_2'\right). \tag{14}$$

A reverse inequality in (14) is obtained if we repeat the same argument for states $\left(\theta_1,\theta_2'\right)$ and $\left(\theta_1',\theta_2\right)$ instead of $\left(\theta_1,\theta_2\right)$ and $\left(\theta_1',\theta_2'\right)$. Hence, a necessary condition for ex post renegotiation-proofness is

$$S\left(\theta_1,\theta_2'\right) + S\left(\theta_1',\theta_2\right) = S\left(\theta_1',\theta_2'\right) + S\left(\theta_1,\theta_2\right) \tag{15}$$

and thus all inequalities used to arrive at (15) must hold as equalities. In particular, the left side of (13) must be equal to the left side of (14), which implies that $m\circ\left(\sigma_1\left(\theta\right),\sigma_2\left(\theta'\right)\right)$ is equal to $a^*\left(\theta_1,\theta_2'\right)$ with certainty, and $m\circ\left(\sigma_1\left(\theta'\right),\sigma_2\left(\theta\right)\right)$ is equal to $a^*\left(\theta_1',\theta_2\right)$ with certainty. Thus (9) becomes

$$t_1\left(\theta\right) + E_{m\circ(\sigma_1(\theta'),\sigma_2(\theta))}\left[\widetilde{t}_2\right] = v_2\left(a^*\left(\theta\right),\theta_2\right) - v_2\left(a^*\left(\theta_1',\theta_2\right),\theta_2\right) \qquad \forall\theta'\neq\theta \tag{16}$$

and if we switch the roles of θ and θ' in (10):

$$t_2\left(\theta'\right) + E_{m\circ(\sigma_1(\theta'),\sigma_2(\theta))}\left[\widetilde{t}_1\right] = v_1\left(a^*\left(\theta'\right),\theta_1'\right) - v_1\left(a^*\left(\theta_1',\theta_2\right),\theta_1'\right) \qquad \forall\theta\neq\theta' \tag{17}$$

Take $\theta=\left(\theta_1,\theta_2\right)$ and $\theta'=\left(\theta_1',\theta_2\right)$, and add up (16) and (17):

$$t_1\left(\theta_1,\theta_2\right) + t_2\left(\theta_1',\theta_2\right) + E_{m\circ(\sigma_1(\theta'),\sigma_2(\theta))}\left[\widetilde{t}_1+\widetilde{t}_2\right] = v_2\left(a^*\left(\theta_1,\theta_2\right),\theta_2\right) - v_2\left(a^*\left(\theta_1',\theta_2\right),\theta_2\right) \tag{18}$$

Another implication of the argument above is that all inequalities used to arrive at (15) must hold as equalities is $\widetilde{t}_1+\widetilde{t}_2=0$. Rewrite (18) using budget balance ($t_2\left(\theta_1',\theta_2\right)=-t_1\left(\theta_1',\theta_2\right)$):

$$t_1\left(\theta_1,\theta_2\right) - t_1\left(\theta_1',\theta_2\right) = v_2\left(a^*\left(\theta_1,\theta_2\right),\theta_2\right) - v_2\left(a^*\left(\theta_1',\theta_2\right),\theta_2\right)$$

Hence, the transfer rule satisfies (1). The argument for agent 2 is identical. ■

Proof of Lemma 2. Let σ be an EPRP equilibrium of a mechanism $\langle S,m\rangle$. Suppose σ is not ex post efficient in some state $\theta=(\theta_1,...,\theta_n)$. That is, there exists some outcome (a,t) reached with

positive probability by σ when the profile of types is θ, such that (a,t) is not ex post efficient in state θ. Suppose the agents are offered an alternative $a^*(\theta)$, which maximizes the total surplus in state θ, together with transfers $\widehat{t}$ defined as

$$\widehat{t_i} = v_i(a,\theta_i) - v_i(a^*(\theta),\theta_i) + t_i + \frac{1}{n}\left(\sum_{j=1}^{n} v_j(a^*(\theta),\theta_j) - \sum_{j=1}^{n} v_j(a,\theta_j)\right)$$

Note that $\widehat{t}$ is feasible since t is feasible: $\sum_{i=1}^{n}\widehat{t_i} = \sum_{i=1}^{n} t_i \leq 0$. Then each agent i strictly prefers the alternative outcome $\left(a^*(\theta),\widehat{t}\right)$ in state θ:

$$v_i(a^*(\theta),\theta_i) + \widehat{t_i} = v_i(a,\theta_i) + t_i + \frac{1}{n}\left(\sum_{j=1}^{n} v_j(a^*(\theta),\theta_j) - \sum_{j=1}^{n} v_j(a,\theta_j)\right) > v_i(a,\theta_i) + t_i$$

which gives a contradiction. Hence σ is ex post efficient. ■

Proof of Lemma 3. <**If**> Suppose f is as described in the statement of the lemma. Consider a mechanism where each agent reports his own payoff relevant type, and the outcome rule as a function of reports is given by f.

First, we verify that there is a truthtelling equilibrium of this mechanism. Denote $\tau := -\frac{1}{2}t_b(0,3) - \frac{1}{2}t_b(2,3)$.

If the seller reports type 1 then there is trade with probability $\frac{1}{2}$ and his expected transfer is $-\frac{1}{2}t_b(0,1) - \frac{1}{2}t_b(2,1) = 1+\tau$; if the seller reports type 3 then there is no trade and his expected transfer is τ. The payoff of the seller of type 1 from telling the truth is $\frac{1}{2}+\tau$, while his payoff from misreporting is τ; the payoff of the seller of type 3 from telling the truth is τ, while his payoff from misreporting is $-\frac{1}{2}+\tau$. If the buyer reports type 0 then there is no trade and his expected transfer is $\frac{1}{2}t_b(0,1) + \frac{1}{2}t_b(0,3) = -\frac{1}{4}-\tau$; if the buyer reports type 2 then there is trade with probability $\frac{1}{2}$ and his expected transfer is $\frac{1}{2}t_b(2,1) + \frac{1}{2}t_b(2,3) = -\frac{3}{4}-\tau$. The payoff of the buyer of type 0 from telling the truth is $-\frac{1}{4}-\tau$, while his payoff from misreporting is $-\frac{3}{4}-\tau$; the payoff of the buyer of type 2 from telling the truth is $\frac{1}{4}-\tau$, while his payoff from misreporting is $-\frac{1}{4}-\tau$. Thus truthful reporting is incentive compatible when there is no renegotiation.

Next, we verify that this equilibrium is EPRP. Note that there can be no renegotiation on the equilibrium path: the mechanism's outcomes are ex post efficient in every state, and thus no other outcome(s) can Pareto dominate the prescribed outcome(s).

Suppose the buyer of type 0 can improve upon his truthtelling payoff by reporting type 2 and consequent renegotiation of the outcome. When the seller's type is 1, the mechanism prescribes trade and a transfer $t_b(2,1) = -\frac{3}{2} + t_b(0,3)$. Such outcome can be renegotiated because it is Pareto dominated in this state by no trade together with a payment $\widehat{p}$ from the buyer to the seller if $\widehat{p} \in [-2-t_b(2,1), -1-t_b(2,1)] = \left[-\frac{1}{2}-t_b(0,3), \frac{1}{2}-t_b(0,3)\right]$. When the seller's type is 3, the mechanism prescribes no trade and a transfer $t_b(2,3)$. Such outcome cannot be renegotiated because it is ex post efficient in this state. Thus, the expected payoff of the buyer of type 0 from such

deviation is (at most) $\frac{1}{2}\left(-\frac{1}{2}+t_b(0,3)\right)+\frac{1}{2}\left(t_b(2,3)\right)=-\frac{1}{4}-\tau$, which coincides with his equilibrium payoff. Hence, such deviation is unprofitable for the buyer of type 0.

Suppose the buyer of type 2 can improve upon his truthtelling payoff by reporting type 0 and consequent renegotiation of the outcome. Following such report of the buyer the mechanism prescribes no trade, and a transfer is $t_b(0,1)=-\frac{1}{2}+t_b(2,3)$ if the seller's type is 1, and $t_b(0,3)$ if the seller's type is 3. In state $(V,C)=(2,1)$ the prescribed outcome can be renegotiated because it is Pareto dominated by trade at a price $\widehat{p}$ if $\widehat{p}\in[1-t_b(0,1),2-t_b(0,1)]=\left[\frac{3}{2}-t_b(2,3),\frac{5}{2}-t_b(2,3)\right]$. In state $(V,C)=(2,3)$ the prescribed outcome cannot be renegotiated because it is ex post efficient. Thus, the expected payoff of the buyer of type 2 from such deviation is (at most) $\frac{1}{2}\left(2-\frac{3}{2}+t_b(2,3)\right)+\frac{1}{2}\left(t_b(0,3)\right)=\frac{1}{4}-\tau$, which coincides with his equilibrium payoff. Hence, such deviation is unprofitable for the buyer of type 2.

Similar analysis reveals that neither type of the seller can improve upon his equilibrium payoff by misreporting and consequent renegotiation. Thus the truthtelling equilibrium is EPRP.

<**Only If**> Suppose a mechanism $\langle S,m\rangle$ has an EPRP equilibrium σ that implements the social choice function f.

First, by Lemma 2 note that f must be ex post efficient. Thus it must prescribe trade in state $(V,C)=(2,1)$ and no trade in the other states. Moreover, all assigned transfers on the equilibrium path must satisfy budget balance, which implies $t_b(V,C)+t_s(V,C)=0$ for every (V,C).

Now consider the buyer of type 0. His interim equilibrium payoff is $\frac{1}{2}t_b(0,1)+\frac{1}{2}t_b(0,3)$. Suppose the buyer of type 0 instead uses the equilibrium strategy of type 2 followed by renegotiation. When the seller's type is 1, the mechanism prescribes trade. For any assigned transfers $\left(\widetilde{t}_b,\widetilde{t}_s\right)$, such outcome can be renegotiated because it is Pareto dominated in this state by no trade. In particular, it can be renegotiated to no trade together with a payment $\widehat{p}=-1+\widetilde{t}_s$ from the buyer to the seller, which leaves the seller exactly indifferent. When the seller's type is 3, the mechanism prescribes no trade. For any assigned $\left(\widetilde{t}_b,\widetilde{t}_s\right)$, such outcome cannot be renegotiated because it is ex post efficient in this state. The renegotiation-proofness constraint to prevent such deviation by the buyer of type 0 is

$$\frac{1}{2}t_b(0,1)+\frac{1}{2}t_b(0,3)\geq\frac{1}{2}E_{m\circ(\sigma_b(2),\sigma_s(1))}\left[1-\widetilde{t}_s\right]+\frac{1}{2}E_{m\circ(\sigma_b(2),\sigma_s(3))}\left[\widetilde{t}_b\right] \tag{19}$$

where the expectation for the terms on the right hand side is taken with respect to a probability measure over outcomes transfers induced by the outcome rule m and equilibrium σ. Using budget balance ($\widetilde{t}_b=-\widetilde{t}_s$) and the expression for the interim equilibrium transfer of the buyer of type 2 we can rewrite (19) as follows

$$\frac{1}{2}t_b(0,1)+\frac{1}{2}t_b(0,3)\geq\frac{1}{2}+\frac{1}{2}t_b(2,1)+\frac{1}{2}t_b(2,3) \tag{20}$$

Next consider the buyer of type 2. His interim equilibrium payoff is $\frac{1}{2}\left(2+t_b(2,1)\right)+\frac{1}{2}t_b(2,3)$. Suppose the buyer of type 2 instead uses the equilibrium strategy of type 0 followed by renegotiation. When the seller's type is 1, the mechanism prescribes no trade. For any assigned transfers $\left(\widetilde{t}_b,\widetilde{t}_s\right)$, such outcome can be renegotiated because it is Pareto dominated in this state by trade. In particular,

it can be renegotiated to trade together with a payment $\widehat{p} = 1 + \widetilde{t}_s$ from the buyer to the seller, which leaves the seller exactly indifferent. When the seller's type is 3, the mechanism prescribes no trade. For any assigned $\left(\widetilde{t}_b, \widetilde{t}_s\right)$, such outcome cannot be renegotiated because it is ex post efficient in this state. The renegotiation-proofness constraint to prevent such deviation by the buyer of type 2 is

$$\begin{aligned} \frac{1}{2}\left(2 + t_b\left(2,1\right)\right) + \frac{1}{2}t_b\left(2,3\right) &\geq \frac{1}{2}\left(2 + E_{m\circ(\sigma_b(0),\sigma_s(1))}\left[-1 - \widetilde{t}_s\right]\right) + \frac{1}{2}E_{m\circ(\sigma_b(0),\sigma_s(3))}\left[\widetilde{t}_b\right] \quad (21) \\ &= \frac{1}{2} + \frac{1}{2}t_b\left(0,1\right) + \frac{1}{2}t_b\left(0,3\right) \end{aligned}$$

where the equality uses budget balance ($\widetilde{t}_b = -\widetilde{t}_s$) and the expression for the interim equilibrium transfer of the buyer of type 0. Combining (20) and (21) it follows that the buyer's interim expected transfers satisfy

$$\left(\frac{1}{2}t_b\left(2,1\right) + \frac{1}{2}t_b\left(2,3\right)\right) - \left(\frac{1}{2}t_b\left(0,1\right) + \frac{1}{2}t_b\left(0,3\right)\right) = -\frac{1}{2} \quad (22)$$

Using a similar argument for the seller we find that his interim expected transfers must satisfy

$$\left(\frac{1}{2}t_s\left(0,1\right) + \frac{1}{2}t_s\left(2,1\right)\right) - \left(\frac{1}{2}t_s\left(0,3\right) + \frac{1}{2}t_s\left(2,3\right)\right) = 1 \quad (23)$$

Subtract (23) from (22) using budget balance to get $t_b\left(2,1\right) - t_b\left(0,3\right) = -\frac{3}{2}$. Add (22) and (23) using budget balance to get $t_b\left(2,3\right) - t_b\left(0,1\right) = \frac{1}{2}$. Hence the transfers satisfy (2) and (3). ■

Proof of Proposition 3. Suppose $f\left(\theta\right) = \left(a\left(\theta\right), t\left(\theta\right)\right)$ is ex post efficient in every $\theta \in \Theta$, i.e. $a\left(\theta\right) = a^*\left(\theta\right)$ and $t\left(\theta\right)$ is budget balanced. Consider a mechanism where each agent i reports his type $\theta_i \in \Theta_i$, and the outcome rule as a function of reports $\theta = \left(\theta_1, ..., \theta_n\right)$ specifies $\alpha\left(\theta\right) = a^*\left(\theta\right)$ and budget balanced transfers $\tau\left(\theta\right)$ such that

$$E_{\theta_{-i}|\theta_i}\left[\tau_i\left(\theta_i, \theta_{-i}\right)\right] = E_{\theta_{-i}|\theta_i}\left[t_i\left(\theta_i, \theta_{-i}\right)\right] \qquad \forall i,\, \theta_i \quad (24)$$

and

$$E_{\theta_{-i}|\theta_i}\left[\tau_i\left(\theta_i', \theta_{-i}\right)\right] < -L \qquad \forall i,\, \theta_i' \neq \theta_i \quad (25)$$

where $L \in \mathbb{R}$ is chosen to be sufficiently large.[64] The existence of such transfer rule τ follows from the results in Kosenok and Severinov (2008).[65]

There exists a truthful equilibrium since the expected penalty L is chosen to be large enough. To see that this equilibrium is EPRP first note that there will be no renegotiation on the equilibrium path since the outcome $\left(\alpha\left(\theta\right), \tau\left(\theta\right)\right)$ is ex post efficient for every θ. If agent i misreports, then his expected payment is more than L, which is chosen to be large enough to make such a deviation

[64] We can take the same L as in the proof of Proposition 1.

[65] The result follows from their Corollary 1 (p.134) to Theorem 1 (p.132-133) if we note that: (i) the uniqueness of the ex post efficient social alternative for every state implies that the agents' equilibrium interim transfers (given by (24)) are uniquely determined; (ii) the agents' interim transfers from misreporting (given by the left side of (25)) can be made arbitrarily large (p.149).

unprofitable, even if the deviator consequently renegotiates the outcome and captures the implied increase in the total surplus. ■

Proof of Proposition 4. <**If**> Suppose $f(\theta) = (a(\theta), t(\theta))$ is ex post efficient in every $\theta \in \Theta$, i.e. $a(\theta) = a^*(\theta)$ and $t(\theta)$ is budget balanced, and transfers t satisfy (4). Consider a direct revelation mechanism where each agent i reports his type $\theta_i \in \Theta_i$, and the outcome rule as a function of reports is given by f. There is a truthtelling equilibrium of this mechanism, since for each agent i of type θ_i

$$\begin{aligned} E_{\theta_{-i}}\left[v_i\left(a^*(\theta_i,\theta_{-i}),\theta_i\right)+t_i(\theta_i,\theta_{-i})\right] &= E_{\theta_{-i}}\left[v_i\left(a^*(\theta_i,\theta_{-i}),\theta_i\right)+\sum_{j\neq i}v_j\left(a^*(\theta_i,\theta_{-i}),\theta_j\right)\right]+H_i \\ &\geq E_{\theta_{-i}}\left[v_i\left(a^*\left(\theta_i',\theta_{-i}\right),\theta_i\right)+\sum_{j\neq i}v_j\left(a^*\left(\theta_i',\theta_{-i}\right),\theta_j\right)\right]+H_i \\ &= E_{\theta_{-i}}\left[v_i\left(a^*\left(\theta_i',\theta_{-i}\right),\theta_i\right)+t_i\left(\theta_i',\theta_{-i}\right)\right] \qquad \forall\theta_i'\neq\theta_i \end{aligned}$$

where the equalities use (4), and the inequality uses the fact that $a^*(\theta_i,\theta_{-i})$ is ex post efficient in state (θ_i,θ_{-i}). Hence, the truthful equilibrium of this mechanism implements f.

Suppose the above equilibrium is not EPRP*. First, note that there will be no renegotiation on the equilibrium path since the outcome $(a^*(\theta), t(\theta))$ is ex post efficient for every θ. Suppose agent 1 of type θ_1 can improve upon his interim payoff from truthtelling by sending a wrong report $\theta_1' \neq \theta_1$ and consequent renegotiation of the outcomes from $\left(a^*\left(\theta_1',\theta_{-1}\right), t\left(\theta_1',\theta_{-1}\right)\right)$ to $\left(\widehat{a}(\theta_{-1}), \widehat{t}(\theta_{-1})\right)$ for every $\theta_{-1} \in \Theta_{-1}$, where $\widehat{t}: \Theta_{-1} \to \mathbb{T}$ is some feasible transfer function. This strategy is beneficial for agent 1 if

$$E_{\theta_{-1}}\left[v_1\left(\widehat{a}(\theta_{-1}),\theta_1\right)+\widehat{t}_1(\theta_{-1})\right] > E_{\theta_{-1}}\left[v_1\left(a^*(\theta_1,\theta_{-1}),\theta_1\right)+t_1(\theta_1,\theta_{-1})\right] \tag{26}$$

Agent $i \neq 1$ agrees to such renegotiation in state θ_{-1} if

$$v_i\left(\widehat{a}(\theta_{-1}),\theta_i\right)+\widehat{t}_i(\theta_{-1}) \geq v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)+t_i\left(\theta_1',\theta_{-1}\right) \tag{27}$$

Sum up (27) over $i \neq 1$ and take expectation over Θ_{-1}:

$$E_{\theta_{-1}}\left[\sum_{i\neq 1}v_i\left(\widehat{a}(\theta_{-1}),\theta_i\right)+\sum_{i\neq 1}\widehat{t}_i(\theta_{-1})\right] \geq E_{\theta_{-1}}\left[\sum_{i\neq 1}v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)+\sum_{i\neq 1}t_i\left(\theta_1',\theta_{-1}\right)\right] \tag{28}$$

Sum up (26) and (28):

$$\begin{aligned}
& E_{\theta_{-1}}\left[\sum_{i=1}^{n} v_i\left(\widehat{a}\left(\theta_{-1}\right),\theta_i\right)+\sum_{i=1}^{n}\widehat{t}_i\left(\theta_{-1}\right)\right] \\
> \; & E_{\theta_{-1}}\left[v_1\left(a^*\left(\theta_1,\theta_{-1}\right),\theta_1\right)+\sum_{i\neq 1} v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)+t_1\left(\theta_1,\theta_{-1}\right)+\sum_{i\neq 1} t_i\left(\theta_1',\theta_{-1}\right)\right] \\
= \; & E_{\theta_{-1}}\left[v_1\left(a^*\left(\theta_1,\theta_{-1}\right),\theta_1\right)+\sum_{i\neq 1} v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)+t_1\left(\theta_1,\theta_{-1}\right)-t_1\left(\theta_1',\theta_{-1}\right)\right] \\
= \; & E_{\theta_{-1}}\left[\sum_{i=1}^{n} v_i\left(a^*\left(\theta_1,\theta_{-1}\right),\theta_i\right)\right]
\end{aligned}$$

where the first equality uses budget balance ($\sum_{i\neq 1} t_i\left(\theta_1',\theta_{-1}\right)=-t_1\left(\theta_1',\theta_{-1}\right)$ for every θ_{-1}), and the second equality uses (4). But $a^*\left(\theta_1,\theta_{-1}\right)$ is ex post efficient in state $\left(\theta_1,\theta_{-1}\right)$, and $\sum_{i=1}^{n}\widehat{t}_i\left(\theta_{-1}\right)\leq 0$ for every θ_{-1}:

$$E_{\theta_{-1}}\left[\sum_{i=1}^{n} v_i\left(a^*\left(\theta_1,\theta_{-1}\right),\theta_i\right)\right]\geq E_{\theta_{-1}}\left[\sum_{i=1}^{n} v_i\left(\widehat{a}\left(\theta_{-1}\right),\theta_i\right)+\sum_{i=1}^{n}\widehat{t}_i\left(\theta_{-1}\right)\right]$$

which gives a contradiction. Hence, the truthful equilibrium of this mechanism is EPRP*.

<**Only If**> Suppose a mechanism $\langle S,m\rangle$ has an EPRP* equilibrium σ that implements the social choice function f.

First, by Lemma 4 note that $f\left(\theta\right)=\left(a\left(\theta\right),t\left(\theta\right)\right)$ must be ex post efficient: $a\left(\theta\right)=a^*\left(\theta\right)$ and $\sum_{i=1}^{n} t_i\left(\theta\right)=0$ for every θ. Next, suppose agent 1 of type θ_1 uses strategy $\sigma_1\left(\theta_1'\right)$ where $\theta_1'\in\Theta_1$. Since σ implements f, for every $\theta_{-1}\in\Theta_{-1}$ the resulting outcome must be interim payoff equivalent to $\left(a^*\left(\theta_1',\theta_{-1}\right),t\left(\theta_1',\theta_{-1}\right)\right)$. Since σ is EPRP*, agent 1 cannot improve upon his interim equilibrium payoff even if the resulting outcomes are renegotiated in the most profitable way. That is, for every θ_{-1} the resulting outcome is renegotiated to an ex post efficient outcome $\left(a^*\left(\theta\right),\widehat{t}\left(\theta_{-1}\right)\right)$ such that each agent $i\neq 1$ receives a payoff that leaves him indifferent between the two outcomes:

$$v_i\left(a^*\left(\theta\right),\theta_i\right)+\widehat{t}_i\left(\theta_{-1}\right)=v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)+t_i\left(\theta_1',\theta_{-1}\right) \tag{29}$$

Sum up (29) over $i\neq 1$ and take expectation over Θ_{-1}:

$$E_{\theta_{-1}}\left[\sum_{i\neq 1} v_i\left(a^*\left(\theta\right),\theta_i\right)+\sum_{i\neq 1}\widehat{t}_i\left(\theta_{-1}\right)\right]=E_{\theta_{-1}}\left[\sum_{i\neq 1} v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)+\sum_{i\neq 1} t_i\left(\theta_1',\theta_{-1}\right)\right] \tag{30}$$

The renegotiation-proofness constraint to prevent such deviations by agent 1 is

$$\begin{aligned} & E_{\theta_{-1}}\left[v_1\left(a^*\left(\theta\right),\theta_1\right)+t_1\left(\theta_1,\theta_{-1}\right)\right] \\ \geq\ & E_{\theta_{-1}}\left[v_1\left(a^*\left(\theta\right),\theta_1\right)-\textstyle\sum_{i\neq 1}\widehat{t}_i\left(\theta_{-1}\right)\right] \\ =\ & E_{\theta_{-1}}\left[v_1\left(a^*\left(\theta\right),\theta_1\right)+\sum_{i\neq 1}v_i\left(a^*\left(\theta\right),\theta_i\right)-\sum_{i\neq 1}v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)+t_1\left(\theta_1',\theta_{-1}\right)\right] \end{aligned} \tag{31}$$

where the equality follows from (30) and budget balance ($\widehat{t_1}\left(\theta_1',\theta_{-1}\right)=-\sum_{i\neq 1}\widehat{t_i}\left(\theta_1',\theta_{-1}\right)$ for every θ_{-1}). Rewrite (31):

$$E_{\theta_{-1}}\left[t_1\left(\theta_1,\theta_{-1}\right)\right]-E_{\theta_{-1}}\left[t_1\left(\theta_1',\theta_{-1}\right)\right]\geq E_{\theta_{-1}}\left[\sum_{i\neq 1}v_i\left(a^*\left(\theta\right),\theta_i\right)\right]-E_{\theta_{-1}}\left[\sum_{i\neq 1}v_i\left(a^*\left(\theta_1',\theta_{-1}\right),\theta_i\right)\right] \tag{32}$$

Switching the roles of θ_1 and θ_1' we find that (32) must hold with equality. Thus the transfer rule satisfies (4). The argument for other agents is identical. ■

References

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

[2] Amoros, P. (2004) "Nash implementation and uncertain renegotiation," *Games and Economic Behavior* 49, 424-434.

[3] Arrow, K. (1979) "The property rights doctrine and demand revelation under incomplete information," in *Economics and Human Welfare*, ed. M. Boskin. Academic Press.

[4] d'Aspremont, C. and L.-A. Gérard-Varet (1979) "Incentives and incomplete information." *Journal of Public Economics* 11, 25-45.

[5] Beaudry, P. and M. Poitevin (1995) "Contract renegotiation: a simple framework and implications for organization theory," *Canadian Journal of Economics* 28, 302-335.

[6] Bergemann, D. and S. Morris (2005) "Robust mechanism design," *Econometrica* 73, 1771-1813.

[7] Bergemann, D. and J. Välimäki (2002) "Information acquisition and efficient mechanism design," *Econometrica* 70, 1007-1033.

[8] Bolton, P. (1990) "Renegotiation and the dynamics of contract design," *European Economic Review* 34, 303-310.

[9] Che, Y. K. and D. B. Hausch (1999) "Cooperative investment and the value of contracting," *American Economic Review* 89, 125-147.

[10] Chung, T. (1992) "Incomplete contracts, specific investment, and risk sharing," *Review of Economic Studies* 58, 1031-1042.

[11] Chung, K. S., and J. Ely (2007) "Foundations of dominant strategy mechanisms," *Review of Economic Studies* 74, 447-476.

[12] Cramton, P. C., and T. R. Palfrey (1995) "Ratifiable mechanisms: learning from disagreement," *Games and Economic Behavior* 10, 255-283.

[13] Crawford, V. (1985) "Efficient and durable decision rules: a reformulation," *Econometrica* 53, 817-835.

[14] Crémer, J. and R. McLean (1988) "Full extraction of the surplus in bayesian and dominant strategy auctions," *Econometrica* 56, 1247-1257.

[15] Dewatripont, M. (1989) "Renegotiation and information revelation over time: the case of optimal labor contracts," *Quarterly Journal of Economics* 104, 589-619.

[16] Edlin A. and B. Hermalin (2000) "Contract renegotiation and options in agency problems," *Journal of Law, Economics and Organization* 16, 395-423.

[17] Edlin A. and S. Reichelstein (1996) "Holdups, standard breach remedies, and optimal investment," *American Economic Review* 86, 478-501.

[18] Evans, R. (2009) "Mechanism design with renegotiation and costly messages," mimeo, University of Cambridge.

[19] Forges, F. (1993) "Some thoughts on efficiency and information," in *Frontiers of Game Theory*, ed. K. Binmore, A. Kirman, and P. Tani, MIT Press.

[20] Forges, F. (1994) "Posterior efficiency," *Games and Economic Behavior* 6, 238-261.

[21] Forges, F., E. Minelli and R. Vohra (2002) "Incentives and the core of an exchange economy: a survey," *Journal of Mathematical Economics* 38, 1-41.

[22] Fudenberg, D. and J. Tirole (1990) "Moral hazard and renegotiation in agency contracts," *Econometrica* 58, 1279-1319.

[23] Fudenberg, D. and J. Tirole (1991) *Game Theory*, MIT Press.

[24] Green, J. R. and J.-J. Laffont (1979) *Incentives in Public Decision Making*, North-Holland.

[25] Green, J. R. and J.-J. Laffont (1987) "Posterior implementability in a two-person decision problem," *Econometrica* 55, 69-94.

[26] Green, J. and J.-J. Laffont (1994) "Non verifiability, costly renegotiation and efficiency," *Annales D'Economie et De Statistique* 36, 81-95.

[27] Groves, T. (1973) "Incentives in teams," *Econometrica* 41, 617-631.

[28] Holmström, B. (1979) "Groves' scheme on restricted domains," *Econometrica* 47, 1137-1144.

[29] Holmström, B. and R. Myerson (1983) "Efficient and durable decision rules with incomplete information," *Econometrica* 51, 1799-1819.

[30] Klement, A. and Z. Neeman (2005) "Against compromise: a mechanism design approach," *Journal of Law, Economics, and Organization* 21, 285-314.

[31] Kosenok, G. and S. Severinov (2008) "Individually rational, budget-balanced mechanisms and allocation of surplus," *Journal of Economic Theory* 140, 126-161.

[32] Krasa, S. (1999) "Unimprovable allocations in economies with incomplete information," *Journal of Economic Theory* 87, 144-168.

[33] Lagunoff, R. D. (1995) "Resilient allocation rules for bilateral trade," *Journal of Economic Theory* 66, 463-487.

[34] Lyon, T. and E. Rasmusen (2004) "Buyer-option contracts restored: renegotiation, inefficient threats and the hold-up problem," *Journal of Law, Economics and Organization* 20, 148-169.

[35] Ma, C.-T. A (1994) "Renegotiation and optimality in agency contracts," *Review of Economic Studies* 61, 109-129.

[36] Mailath and Postlewaite (1990) "Asymmetric information bargaining problems with many agents," *Review of Economic Studies* 57, 351-367.

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

[38] Maskin, E. and J. Tirole (1992) "The principal-agent relationship with an informed principal, II: common values," *Econometrica* 60, 1-42.

[39] Moore, J. (1992) "Implementation, contracts, and renegotiation in environments with complete information," in *Advances in Economic Theory, Sixth World Congress*, ed. by J-J. Laffont. Cambridge University Press, Cambridge.

[40] Myerson, R. B. and M. A. Satterthwaite (1983) "Efficient mechanisms for bilateral trading," *Journal of Economic Theory* 29, 265-281.

[41] Noldeke, G. and K. Schmidt (1998) "Sequential investments and options to own," *RAND Journal of Economics* 29, 633-653.

[42] Palfrey, T. R. and S. Srivastava (1993) *Bayesian Implementation*, Harwood Academic Publishers GmbH.

[43] Rogerson, W. (1992) "Contractual solutions to the hold-up problem," *Review of Economic Studies* 59, 777-793.

[44] Rubinstein A., and A. Wolinsky (1992) "Renegotiation-proof implementation and time preferences," *American Economic Review* 82, 600-614.

[45] 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.

[46] Sen, A. (2008) "Renegotiation-proof agreements under asymmetric information," mimeo, Indian Institute of Management Calcutta.

[47] Sjöström, T. (1999) "Undominated Nash implementation with collusion and renegotiation," *Games and Economic Behavior* 26, 337-352.

[48] Spier, K. E. (1994) "Pretrial bargaining and the design of fee shifting rules," *RAND Journal of Economics* 25, 197-214.

[49] Walker, M. (1980) "On the nonexistence of a dominant strategy mechanism for making optimal public decisions," *Econometrica* 48, 1521-1540.

[50] Watson, J. (2007) "Contract, mechanism design, and technological detail," *Econometrica* 75, 55-81.

[51] Watson, J. and C. Wignall (2007) "Hold-up and durable trading opportunities," mimeo, UCSD.

[52] Wickelgren, A. (2007) "The limitations of buyer-option contracts in solving the hold-up problem," *Journal of Law, Economics and Organization* 23, 127-140.

[53] Williams, S. R. (1999) "A characterization of efficient, Bayesian incentive compatible mechanisms," *Economic Theory* 14, 155-180.