Renegotiation-Proof Mechanism Design

Zvika Neeman and Gregory Pavlov

7-09

September, 2009

The Foerder Institute for Economic Research
and
The Sackler Institute of Economic Studies

Renegotiation-Proof Mechanism Design*

Zvika Neeman[†] and Gregory Pavlov[‡]

June 18, 2008

Abstract

A mechanism is said to be renegotiation-proof if it is robust against renegotation both before and after it is played. We ask the following three related questions: (1) what kind of environments or mechanism design problems admit renegotiation-proof implementation? (2) what kind of social choice rules are implementable in a way that is renegotiation-proof? and (3) what kind of mechanisms are renegotiation-proof?

We provide characterization results for environments, social choice rules, and mechanisms that facilitate renegotiation-proof implementation in complete information settings, and in incomplete information settings with independent private values. For incomplete information settings with correlated interdependent values we provide sufficient conditions for renegotiation-proof implementation. Importantly, our results imply that some common mechanism design problems do not admit the existence of any renegotiation-proof mechanism.

JEL CLASSIFICATION NUMBERS: D02, D70, D82.
KEYWORDS: ex-post renegotiation, interim renegotiation, oracle renegotiation proofness.

* Acknowledgement to be added.

† Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, and the Eitan Berglas School of Economics, Tel-Aviv University, Tel-Aviv, Israel 69978; Email `zvika@BU.edu`, http://people.bu.edu/zvika, http://www.tau.ac.il/ zvika/.

‡ Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215; Email `gpavlov@BU.edu`, http://people.bu.edu/gpavlov.

1. Introduction

In microeconomic theory, the term 'mechanism' refers to the 'rules of the game' or more simply just to 'rules,' and sometimes also to the equilibrium that is played under these rules. Mechanism design theory attempts to answer the question of when and how it is possible to design rules that induce equilibrium outcomes that are optimal with respect to some given criterion of social welfare. It is widely recognized that in order for mechanism design theory to realize its potential as a theory of institutional design it must be robust. At least five different notions of robustness have been discussed in mechanism design literature: robustness against the cost and complexity of communication and computation, robustness against collusion, robustness against uncertainty about higher order beliefs, and robustness against renegotiation.[1] This paper is devoted to the latter subject of robustness against renegotiation.

If renegotiation of a given mechanism cannot be prevented, then the players may all agree to change the mechanism they are supposed to play into a different one.[2] Renegotiation might either involve renegotiation of just the decision reached by the mechanism, or renegotiation of the equilibrium that is played under the mechanism, or renegotiation of the entire mechanism and equilibrium to be played. Renegotiation may take place either at the so called interim stage, that is *before* the mechanism is played and when each player knows only his own type, or at the so called ex-post stage, *after* the mechanism is played and when each player knows both his own type and the decision reached by the mechanism, but not necessarily the other players' types. In the former case, when renegotiation is done at the interim stage, the players might renegotiate the equilibrium they intended to play under the mechanism, or the mechanism itself. In the latter case, when renegotiation is done at the ex-post stage, the players may wish to renegotiate the decision or recommendation that is made by the mechanism.

[1] The literature on robust mechanism design has become quite voluminous. The interested reader may consult Nisan et al. (2007) and the references therein on robustness against the cost and complexity of communication and computation, see also Palfrey and Srivastava (1991) for discussion of robustness against the possibility of additional communication; Che and Kim (2006) and the references therein on robustness against collusion; and Bergemann and Morris (2005) and the references therein on robustness against uncertainty about higher order beliefs. The literature about robustness against renegotiation is surveyed in Section 2 below. Although it is not usually interpreted as such, the work on multidimensional mechanism design (see, e.g., Jehiel et al., 2006, and the references therein) may also be interpreted as part of the literature on robust mechanism design, namely, robustness against higher dimensions of the type space.

[2] The possibility of renegotiation is associated with a failure of the principal or the agents to commit to play a given mechanism. The literature distinguishes between two different versions of the failure to commit. Under the stronger version, one party has the ability to unilateraly change the mechanism in pursuit of unilateral gains. Under the more moderate version, only voluntary and hence Pareto-improving changes of the mechanism are permitted. In this paper we focus on the latter case.

A mechanism that is immune against renegotiation before it is played is said to be interim renegotiation-proof, and a mechanism that is immune against renegotiation after it is played is said to be ex-post renegotiation-proof.

Our motivation for investigating what and under what conditions mechanisms are renegotiation-proof is twofold. First, we are interested in the question of what social choice functions (or social choice rules) can be implemented when players are free to collectively renegotiate the outcome or even the mechanism itself to something else. And second, we want to know what kind of mechanisms can be expected to be 'stable' in the sense of surviving in their original form under a variety of different renegotiation scenarios. We address these two aspects in turn.

As mentioned above, mechanism design theory attempts to answer the question of when and how it is possible to design mechanisms whose equilibrium outcomes are optimal with respect to some given criterion of social welfare. If a proposed mechanism might be renegotiated, then the task of the mechanism designer becomes harder because he must either ensure that the proposed mechanism is robust against renegotiation, or, alternatively, he must be able to foresee how renegotiation will play out and make sure that the outcome under the mechanism cum renegotiation achieves the desired social goals.

The welfare criterion that is used to judge the optimality of a given mechanism is described by a social choice function (or social choice rule) that maps the states of the world into desired outcomes. In some cases, such as in the case of environments with transferable utility and complete or no asymmetric information, the possibility of renegotiation does not pose a problem for the mechanism designer provided that his sole concern is with the Pareto efficiency of the outcome in every state of the world. In such cases, even if the original mechanism is renegotiated, then the Coasian bargaining among the agents will assure the efficiency of the final outcome. However, even in such simple environments the mechanism designer may often be interested not just in ensuring the Pareto efficiency of the outcome, but also in the distribution of the monetary transfers among the players and in their payoffs. For example, in some situations some players may have opportunities to make non-contractible ex-ante investments, and creation of the appropriate incentives for such investments may require the implementation of a particular, and not just any, Pareto efficient outcome (and sometimes even an ex-post inefficient outcome).

If renegotiation happens at the interim stage, then clearly the only way the mechanism designer can ensure implementation of a given social choice function is to require the mechanism to be interim renegotiation-proof against the alternative mechanisms that may be proposed during renegotiation. Alternatively, the mechanism designer may allow interim renegotiation, but only as long as at the end of the day the same exact social choice function is realized. Obviously, in such a case the agents would have no incentive to renegotiate the original mechanism in the first place.

If renegotiation happens at the ex-post stage, then the mechanism designer has two options. Either require the mechanism to be ex-post renegotiation-proof against all the proposals to change the decision reached by the mechanism that may arise during renegotiation. Or, alternatively, not to prevent renegotiation from happening, but to ensure that the renegotiation that follows the play of the original mechanism leads to an outcome that is consistent with the social choice function he seeks to implement. Note that the second option becomes harder and harder to implement as more renegotiation scenarios become possible. If the mechanism designer is completely ignorant a priori about the way in which ex-post renegotiation will proceed, then the only way for him to follow the second option would be to try to extract the information about the anticipated renegotiation from the players themselves through the course of the play of the original mechanism, and to condition the decisions of the mechanism on this information. Such a mechanism is likely to be quite complex because it must provide the players with the incentives to reveal all their private information – their payoff relevant private information and their beliefs about likely renegotiation scenarios.[3]

In this paper we study a simpler class of mechanisms that do not attempt to extract from the players any information about anticipated renegotiation, but nonetheless implement the desired social choice function. Clearly, since we rule out the second option above, such mechanisms must be ex-post renegotiation-proof. Studying such mechanisms is a natural first step towards understanding how to do mechanism design in the presence of multiple renegotiation scenarios. 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 (or ex-post) mechanisms relate to Bayesian mechanisms. Dominant strategy and ex-post mechanisms are 'belief-free' and rely only on the payoff relevant information of the players, while Bayesian mechanisms may in addition rely on the higher-order beliefs of the players (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 players' beliefs about future renegotiation.

Our second motivation for investigating renegotiation-proof mechanisms is to learn what mechanisms are expected to be stable in the sense of surviving in their original form under a variety of renegotiation scenarios. We believe that an institution that emerges in a given environment and that remains unchanged under a variety of circumstances must be both interim and ex-post renegotiation-proof. It must be interim renegotiation-proof in order to

[3] Another complicating factor is that the players may not have any particular beliefs about future renegotiation, or may even, possibly, have conflicting beliefs, and still be able to renegotiate successfully ex-post. Consequently, it is not clear how to model the way in which players' beliefs about renegotiation constrain their renegotiation possibilities.

be immune against all possible alternative mechanisms which may be proposed to (or by) the players. And, it must be ex-post renegotiation-proof in the sense explained above in order to be robust against all the possible proposals to change the outcome reached by the mechanism ex-post. Our notion of renegotiation-proofness requires robustness against a large number of different renegotiation possibilities, and is thus stronger than what is usually assumed by the literature, where renegotiation-proofness of a mechanism is defined only with respect to one particular expected renegotiation scenario. However, we believe that if a mechanism retains its original form under a variety of different circumstances, then such a mechanism must be renegotiation-proof in our stronger sense. On the other hand, if such renegotiation-proof mechanisms can be shown to 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.

As discussed above, it is possible to think of two types of renegotiation-proof mechanisms, simple and complex. Simple mechanisms do not attempt to induce the players to reveal their beliefs about the particular kind of renegotiation scenario which is about to take place. Complex mechanisms would attempt to extract such information from the players in order to counteract expected future renegotiation. Clearly, interim-renegotiation proof mechanisms can only be simple in this sense, because renegotiation takes place before the mechanism is actually played. Ex-post renegotiation-proof mechanisms may be either simple or complex, and in this paper we restrict attention to simple mechanisms.

We provide characterization results for environments, social choice rules, and mechanisms that facilitate renegotiation-proof implementation in complete information settings, and in incomplete information settings. For complete information environments with three or more players we show that any ex-post efficient decision rule can be implemented in a way that is (both ex-post and interim) renegotiation-proof. When there are just two players ex-post renegotiation-proofness imposes more stringent requirements on the monetary transfers among the players. We show that in order for a mechanism to be ex-post renegotiation-proof in this case it must be a budget balanced Groves mechanism (Groves, 1973).[4] For incomplete information environments with independent private values, we provide a characterization of ex-post renegotiation-proof mechanisms in terms of budget balanced mechanisms that are 'Groves in expectation.'[5] For incomplete information settings with correlated, and possibly interdependent, values we provide sufficient conditions for renegotiation-proof implementation.

Groves mechanisms are well known to fail to satisfy budget balance when the play-

[4]Interestingly, Vickrey-Clarke-Groves mechanisms have also been shown by Bergemann and Välimäki (2002) to generate efficient incentives for some types of ex-ante investments (that do not involve externalities).

[5]A mechanism is 'Groves in expectation' if the expected payment to each player as a function of her type is equal to the expected payment to the player as a function of her type under some Groves mechanism.

ers' preferences depend on the state of the world in a nontrivial way (Green and Laffont, 1979; Walker, 1980), which implies that no ex-post renegotiation-proof mechanisms exist in complete information environments with two players. It is equally well known that in independent private values environments budget balanced Groves in expectation mechanisms do exist (Arrow, 1979; d'Aspremont and Gerard-Varet, 1979), but that in many economically important mechanism design problems none of these mechanisms satisfy the additional constraint of interim individual rationality (such are, for example, the problems of bilateral trade (e.g., Myerson and Satterthwaite, 1983), regulation (e.g., Baron and Myerson, 1982), and litigation and settlement (e.g., Spier 1994 and Klement and Neeman, 2005)). We believe that the implied non-existence of individually rational renegotiation-proof mechanisms is an important message of this paper. It demonstrates that doing mechanism design when renegotiation cannot be prevented is hard, and that in many environments there are no social choice functions that can be guaranteed implementation. It is of course possible that a more positive outlook for mechanism design theory with renegotiation could be provided by studying the so called 'complex mechanisms' mentioned above. But, even if this were the case, such mechanisms would probably be very complex, possibly too complex to be of practical interest.[6]

The rest of the paper proceeds as follows. In the next section, we discuss the related literature. In Section 3, we present the basic setup. Section 4 is devoted to the subject of renegotiation-proofness under complete information, and Section 5 is devoted to renegotiation-proofness under incomplete information. All proofs are relegated to the appendix.

2. Related literature

The literature on renegotiation under complete information (see especially Maskin and Moore (1999) and Segal and Whinston (2002), but also Chung (1988), Green and Laffont (1987, 1994), Hart and Moore (1988), Aghion et al. (1989), and Che and Hausch (1999)) has characterized the class of implementable social choice rules given some exogenous "renegotiation function" that for every outcome specifies another outcome to which the original outcome can be renegotiated.[7] This approach is in line with the standard approach in microeconomic theory, which is to assume that the players and the mechanism designer can foresee perfectly the outcome of any future renegotiation (see, e.g., Bolton (1990) and Dewatripont and Maskin (1990) for surveys of the early literature) and is different from the approach taken

[6] Such mechanisms would also have to overcome the conceptual difficulty described in footnote 3 above.

[7] Of related interest is the work by Bernheim et al. (1987) and Moreno and Wooders (1996) who studied coalition-proof equilibria in strategic form games. The problem of renegotiation in such environments is simpler because the players have no informational advantage vis-a-vis the social planner.

here according to which the mechanism designer is ignorant about renegotiation and the only constraint on renegotiation is that the players have to consent to it.

Rubinstein and Wolinsky (1992) also wrote about renegotiation under complete information. However, they have followed the approach that is common in the literature about renegotiation under incomplete information, which is to define a notion of renegotiation-proofness and to study the mechanisms that are renegotiation-proof given that definition. Rubinstein and Wolinsky have examined a specific buyer/seller problem, and have focused their attention on the question of how the cost of renegotiation affects the set of renegotiation-proof social choice rules.

The literature on renegotiation-proofness under incomplete information (see the seminal contribution by Holmström and Myerson (1983), as well as Crawford (1985), Palfrey and Srivastava (1991), Lagunoff (1995), and Cramton and Palfrey (1995)) has mostly confined its attention to the subject of interim renegotiation-proofness. Each paper in this literature has presented a different notion of interim renegotiation-proofness that has the property that for any mechanism design problem, there exists a mechanism that is renegotiation-proof according to that definition. The subject of ex-post renegotiation-proofness under incomplete information received considerably less attention. It was examined by Forges (1993, 1994) who concluded that the question of whether there exists a renegotiation-proof mechanism for *every* mechanism design problem remains open (1994, p. 241).[8]

3. Setup

A group of n players, indexed by $i \in N = \{1, 2, ..., n\}$, must reach a decision that involves the choice of a social alternative $a \in A$ together with the determination of monetary transfers to the players, $t = (t_1, ..., t_n) \in \mathbb{R}^n$. We require that the sum of these monetary transfers be non-positive.[9] A decision of the players (a, t), or rather an outcome (a, t) of the process of

[8] On interim renegotiation proofness, see also Maskin and Tirole (1992). On ex-post renegotiation proofness, see also Green and Laffont (1987). Beaudry and Poitevin (1995) deal with both interim and ex-post renegotiation proofness, but in a simpler model with only one privately informed player. Krasa (1999) introduced a concept of unimprovability that combines some features of interim and ex-post renegotiation proofness.

[9] Otherwise, the players may wish to renegotiate any decision just for the purpose of generating large transfer payments for themselves. Also, observe that it is possible to incorporate non budget balanced monetary transfers into the social alternative $a \in A$. For example, the social alternative a may describe whether or not a bridge is built, and how much each player is supposed to pay for it. Suppose that the cost of building the bridge is given by $c > 0$. The set of social alternatives can then include an alternative where the bridge is not built and no one pays anything, an alternative where the bridge is built and each player pays c/n, an alternative where the bridge is built, players 1 and 2 pay $c/2$ each while other players pay nothing and so on. The transfers t can then be added to a to facilitate contracting among the players.

negotiation among the players, is said to be feasible if $a \in A$ and $\sum_{i=1}^{n} t_i \leq 0$.

The players' preferences over the set $A \times \mathbb{R}^n$ as well as their beliefs about each other's preferences are determined by their types. The set of player i's types is denoted Θ_i. For simplicity, we assume that the sets Θ_i, $i \in N$, are finite, however, our results would continue to hold, with appropriate adjustments, if the players each has a continuum of types. We denote $\Theta = \Theta_1 \times \cdots \times \Theta_n$, and $\Theta_{-i} = \prod\limits_{j \neq i} \Theta_j$, with typical elements θ and θ_{-i}, respectively. A profile of types $\theta \in \Theta$ is also called a state of the world. We denote the common prior distribution over the set of states of the world Θ by P.

Each player i is assumed to be an expected utility maximizer with a quasi-linear payoff function that is given by $u_i(a, t_i, \theta) = v_i(a, \theta) + t_i$ where $v_i : A \times \Theta \rightarrow \mathbb{R}$ denotes player i's preferences over the set of social alternatives A as a function of his type and t_i denotes a possibly additional monetary transfer to player i.

In a complete information environment, the state of the world θ is commonly known among the players, although not necessarily by the social planner.[10] In an incomplete information environment, each player i knows his type θ_i, and derives his beliefs by conditioning the common prior P on his own type. A complete information mechanism design environment is thus fully described by a four-tuple $\left\langle N, A, \Theta, (u_i)_{i \in N} \right\rangle$.[11] An incomplete information mechanism design environment is described by a five-tuple $\left\langle N, A, \Theta, P, (u_i)_{i \in N} \right\rangle$.

A mechanism is a game form $\langle S, m \rangle$ that specifies a message set S_i for each player $i \in N$, and a mapping $m : S \rightarrow \Delta(A \times \mathbb{R}^n)$ from the set of message profiles $S = S_1 \times \cdots \times S_n$ into the set of lotteries over the product of the set of social alternatives A and monetary transfers $\mathbb{R}^n$. As mentioned above, we restrict our attention to feasible mechanisms. That is, we assume that the mechanism that is employed by the players is such that the sum of monetary transfers to the players is non-positive, for any profile of messages that are sent by the players. If the monetary transfers to the players sum up to zero for any profile of messages that are sent by the players then both the monetary transfers and the mechanism of which they are part are said to be budget balanced.

The combination of a mechanism $\langle S, m \rangle$ and a state of the world θ defines a complete information game $\left\langle N, S, (u_i(\cdot, \cdot, \theta) \circ m)_{i \in N} \right\rangle$. The combination of a mechanism $\langle S, m \rangle$ and a prior distribution over the states of the world P defines a Bayesian game $\left\langle N, S, \Theta, P, (u_i \circ m)_{i \in N} \right\rangle$. We denote a Nash or a Bayesian Nash equilibrium of the complete information or Bayesian game that is induced by the mechanism $\langle S, m \rangle$ by $\sigma = (\sigma_1, ..., \sigma_n)$.

A social choice rule is a mapping $f : \Theta \rightrightarrows A \times \mathbb{R}^n$ from the set of states of the world

[10] A social planner who knows the state of the world can easily implement any social choice function she likes.

[11] The fact that the state of the world is commonly known among the players in such an environment obviates the common prior.

into outcomes. A social choice rule is said to be implementable by a mechanism $\langle S, m\rangle$ in a complete or incomplete information environment, respectively, if the equilibrium outcomes that are induced by the mechanism belong to $f(\theta)$, for every $\theta \in \Theta$. We thus employ a weak notion of implementation.

Definition 1. A social alternative $a \in A$ is said to be *ex-post efficient* in a state of the world $\theta \in \Theta$ if it is such that:

$$a \in \arg\max_{a' \in A} \sum_{i=1}^{n} v_i(a', \theta).$$

An equilibrium σ is said to be ex-post efficient if it leads to the choice of ex-post efficient social alternatives in every state of the world.

We assume that for every state of the world $\theta \in \Theta$, there is a single social alternative $a \in A$ that maximizes social welfare $\sum_{i=1}^{n} v_i(a, \theta_i)$. Observe that if the players' type spaces are finite, then this assumption is generically satisfied.

Finally, to simplify the discussion, we do not discuss the subject of individual rationality. However, all of our results continue to hold if individual rationality is added as a constraint to the analysis.

4. Renegotiation-Proofness Under Complete Information

4.1. Ex-Post Renegotiation-Proofness Under Complete Information

We model the process of ex-post renegotiation in an environment with complete information in the following way: a mechanism $\langle S, m\rangle$ is chosen before the state of the world becomes known. This mechanism is played after the state of the world $\theta \in \Theta$ is realized and becomes commonly known among the players, but not known to the social planner. Consider a Nash equilibrium $\sigma = (\sigma_1, ..., \sigma_n)$ of the complete information game that is induced by the mechanism $\langle S, m\rangle$ when the state of the world is θ. Denote the Nash equilibrium outcome by $(a, t_1, ..., t_n)$.

Suppose that the process of renegotiation assumes the following form. Suppose that a different social alternative $a' \in A$, together with a profile of monetary transfers $t' = (t'_1, ..., t'_n)$ that sum up to zero (or less) is exogenously proposed to the players instead of the outcome $(a, t_1, ..., t_n)$ that was obtained under the mechanism $\langle S, m\rangle$.[12] If the players all agree to

[12] The set of social alternatives A may include lotteries. If so, ex-post renegotiation takes place before these lotteries are carried through.

switch to the renegotiated proposal, then alternative a' is implemented, and each player i receives a monetary transfer of t'_i. Otherwise, the original outcome (a,t) is implemented.[13]

For simplicity, we assume that if the outcome (a',t') Pareto dominates the outcome (a,t), which means that the former outcome is weakly preferred by all the players and strictly preferred by at least one player to the latter outcome, then the original outcome (a,t) is renegotiated to the new outcome (a',t'). Otherwise, the original outcome (a,t) is implemented.[14] This assumption leads to the following definition.

Definition 2. A Nash equilibrium σ of the complete information game $\left\langle N, S, \left(u_i\left(\cdot,\cdot,\theta\right)\circ m\right)_{i\in N}\right\rangle$ that is induced by a mechanism $\langle S,m\rangle$ when the state of the world is θ that generates an outcome (a,t) is ex-post renegotiation-proof if:

1. there does not exist an alternative feasible outcome (a',t') that Pareto dominates (a,t), and
2. there does not exist an alternative feasible outcome (a',t') that some player i prefers to (a,t), and a strategy for player i, σ'_i, such that the outcome that is generated by (σ'_i,σ_{-i}) is Pareto dominated by (a',t').[15]

The first part of the definition is straightforward. If an outcome (a,t) that is generated by an equilibrium σ is not Pareto efficient in a given state of the world $\theta\in\Theta$, then the players would agree to renegotiate it to another outcome (a',t') that Pareto dominates (a,t). This result is stated in the following lemma, which is given without proof.

Lemma 1. *In a complete information mechanism design environment, an ex-post renegotiation-proof Nash equilibrium of a feasible mechanism is ex-post efficient and budget balanced.*

The second part of the definition is more subtle. Even though there may not exist any alternative outcome that Pareto dominates (a,t), there may still exist another outcome

[13] We assume that both the original and alternative outcomes (a,t) and (a',t') are public outcomes that can be imposed directly by the social planner. We also assume that the outcomes are durable in the sense that there is no fixed date by which the outcome has to be implemented in order to yield the specified payoffs to the players. Watson (2007) (see also Noldeke and Schmidt (1995) and Lyon and Rasmusen (2004)) has shown that the effect of ex-post renegotiation may be significantly reduced in environments with inalienable actions and nondurable trade opportunities.

[14] When this is the case, the renegotiation game in which the players are asked to simultanously cast their votes in favor or against the alternative outcome (a',t') has a Nash equilibrium in which they all vote for the alternative outcome. The assumption we make allows us to focus on this equilibrium without explicitly describing the voting game.

[15] Observe that because player i may also choose not to deviate or $\sigma'_i=\sigma_i$ the first part of the definition is subsumed in the second.

(a', t') that some player i prefers to (a, t) and that player i can bring about by deviating from σ in such a way that the outcome that is produced after the deviation is Pareto dominated by (a', t'). If such a profitable deviation exists for some player then σ cannot be ex-post renegotiation-proof.

The following example illustrates the importance of the second part of the definition by demonstrating that an equilibrium may fail to be ex-post renegotiation-proof in spite of being ex-post efficient and in dominant strategies.

Example 1. Suppose that there are two players, a buyer and a seller. The seller owns an object that the buyer may want to buy. The buyer values this object at either 1 or 5. The seller's reservation value for the object is 2. The state of the world is thus determined by the buyer's valuation for the object. The set of outcomes consists of three alternatives: "no trade," "trade at the price 3," and "trade at the price 4."

Consider the following mechanism: the buyer announces whether he wants to trade or not. If he announces he wants to trade, then the buyer and seller trade at the price 4; otherwise, there is no trade. Observe that in each one of the two states of the world, the game that is induced by this mechanism has a trivial unique Nash equilibrium in dominant strategies. If the buyer's valuation for the object is 1, then in equilibrium the buyer declines to trade and the object is not traded. If the buyer's valuation for the object is 5, then in equilibrium the buyer agrees to trade and the object is traded at the price 4.

However, despite the fact that the Nash equilibrium that is played when the buyer's valuation is high is both ex-post efficient and in dominant strategies, it is not ex-post renegotiation-proof according to our definition. To see this, suppose that in the event of no trade, the buyer and seller may renegotiate the outcome to trading at the price of 3 if they so wish. A buyer who values the object at 5 and who anticipates the possibility of such renegotiation might announce that he declines to trade in the hope of renegotiating the outcome to trading at a price that is better for him. Since such renegotiation would also make the seller strictly better off compared to no trade, the seller may well agree to renegotiate the outcome. Thus, the Nash equilibrium in which the object is traded at the price 4 may be renegotiated away – the fact that the buyer's valuation for the object in this case is commonly known to be larger than 4 does not prevent this renegotiation from taking place.[16]

A mechanism may give rise to different equilibria in different states of the world, and even to different equilibria in the same state of the world. We define what it means for a mechanism to be ex-post renegotiation-proof as follows.

[16] See Forges (1993, p. 142) and (1994, p. 260) for another example in which an ex-post efficient equilibrium can be renegotiated.

Definition 3. A feasible mechanism $\langle S, m\rangle$ is *ex-post renegotiation-proof* if it has an ex-post renegotiation-proof Nash equilibrium σ^θ for every state of the world $\theta \in \Theta$.[17]

Remark 1. The difference between our notion of ex-post renegotiation-proofness and the one that is implied by Maskin and Moore (1999) (and Segal and Whinston, 2002) stems from their assumption that renegotiation is commonly known to proceed according to a given reduced form renegotiation mapping $h : A \times \mathbb{R}^n \times \Theta \to A \times \mathbb{R}^n$ that maps an outcome and a state of the world into a possibly different outcome. A mechanism is ex-post renegotiation-proof according to our definition if it is robust against renegotiation according to *any* such renegotiation procedure. Our notion of renegotiation-proofness is thus stronger, and is satisfied by fewer mechanisms.

The following proposition provides a characterization of ex-post renegotiation-proof mechanisms for the case where the number of players is larger than or equal to three.

Proposition 1. *Consider a complete information mechanism design environment with $n \geq 3$ players. In such an environment there exists a feasible ex-post renegotiation-proof mechanism that implements an outcome (a, t) if and only if the social alternative a is ex-post efficient given the state of the world, and the transfers t are budget balanced.*

The idea of the proof of Proposition 1 is the following. In order to implement a given social choice rule, a social planner needs to know the state of the world. If there are three or more players, then it is possible to use the report of player 2 about the state of the world to verify player 1's report, to use the report of player 1 about the state of the world to verify player 2's report, and to use player 3 as a budget breaker. Since this method ensures that both players 1 and 2 will reveal the true state of the world, it is possible to implement any ex-post efficient outcome given this state.

When there are only two players, it is impossible to separate the provision of incentives for reporting the truth about the state of the world from budget balance, which makes this case harder to analyze. We therefore proceed to analyze this case under two additional simplifying assumptions:

1. **Private Values:** Players' payoffs depend only on their own types, namely $v_i(a, (\theta_i, \theta_{-i})) = v_i(a, (\theta_i, \theta'_{-i}))$ for every $a \in A$, $\theta_i \in \Theta_i$, and pairs $\theta_{-i}, \theta'_{-i} \in \Theta_{-i}$. In order to simplify the notation, henceforth, until the end of this subsection, we suppress mention of other players' types in player i's payoff function and simply write $v_i(a, \theta_i)$ instead.

[17] A stronger definition would have required that every equilibrium is ex-post renegotiation proof in every state of the world.

2. **Full Support:** Every profile of states of the world $(\theta_1, ..., \theta_n) \in \Theta_1 \times \cdots \times \Theta_n$ is feasible ex-ante.

We show that when there are only two players a mechanism is ex-post renegotiation-proof if and only if it is a Groves mechanism (Groves, 1973).

Definition 4. A mechanism $\langle S, (a, t) \rangle$ is a Groves mechanism if it is such that players are asked to report their types, that is $S_i = \Theta_i$ for every player i, the decision rule $a : \Theta \to A$ is ex-post efficient, and transfers $t_i : \Theta \to \mathbb{R}$, $i \in N$, are given by

$$t_i(\theta_i, \theta_{-i}) = \sum_{j \neq i} v_j(a(\theta_i, \theta_{-i}), \theta_j) + H_i(\theta_{-i})$$

for some functions $H_i : \Theta_{-i} \to \mathbb{R}$, $i \in N$.

Proposition 2. *Consider a complete information mechanism design environment with two players. In such an environment, a feasible mechanism is ex-post renegotiation-proof if and only if it is a budget balanced Groves mechanism.*

The intuition for Proposition 2 is the following. The possibility of renegotiation implies that a player can deviate from equilibrium play in order to induce an inefficient decision and then renegotiate the outcome to one that is ex-post efficient while capturing the difference in social surplus. This implies that the possibility of ex-post renegotiation allows any player to capture the surplus or externality that he generates up to a constant. It therefore follows that a mechanism in which players already get the surplus or externality they generate is ex-post renegotiation-proof, and conversely, any mechanism that is ex-post renegotiation-proof must be a mechanism in which each player obtains a payoff that is equal to the surplus he generates up to a constant. Thus, Proposition 2 follows from the fact that the class of mechanisms in which players' payoffs are equal to the surplus they generate up to a constant coincides with the class of Groves mechanisms.[18]

Example 1 (continued). If the set of outcomes in example 1 is expanded to allow for trade at any price, then inspection of the proof of Proposition 2 reveals that any profile of ex-post renegotiation-proof Nash equilibria (one equilibrium for each state of the world) must be such that it is the buyer who determines if there is trade or not, and the price paid by the buyer when there is trade must be larger by exactly 2 than the price paid by the buyer when there is no trade (this is the Groves payment for the buyer). It follows that if we add a participation or individual rationality constraint that implies that in the event of no

[18] We do not have an intuitive explanation for the reason that the mechanism must ignore the players' reports about the other player's type.

trade the buyer does not pay anything, then in any ex-post renegotiation-proof mechanism, it is the buyer who decides if there is trade or not, and the price that is paid for the object in the event of trade is equal to 2.

It is easy to construct a budget balanced Groves mechanism if the preferences of one of the players are independent of the state of the world. However, there are several results in the literature that imply that in two player environments with sufficiently rich type spaces there does not exist any budget balanced Groves mechanism (see for example, Green and Laffont (1979), Walker (1980), and Hurwicz and Walker (1990)).

The next example demonstrates that a complete information mechanism design problems with two players may fail to admit the existence of an ex-post renegotiation-proof mechanism even in a very simple environment.

Example 2. There are two players, a buyer and a seller. The seller owns an object that the buyer may want to buy. The buyer values the object at either 1 or 5. The seller's reservation value for the object is either 2 or 6. The set of social decisions consists of a continuum of alternatives $(q, p) \in [0, 1] \times \mathbb{R}$ where $q \in [0, 1]$ is the probability of trade and $p \in \mathbb{R}$ is a balanced transfer, or the price to be paid from the buyer to the seller.

Inspection of the proof of Proposition 2 reveals that the possibility of ex-post renegotiation implies that the buyer can ensure that he does not pay more than 2 when he buys the object, and the seller can ensure that the buyer pays at least 5 when he buys the object, respectively (the two requirements are a consequence of the fact that the buyer's and seller's payments are Groves payments, respectively). Since these two requirements are inconsistent, it follows that there does not exist any ex-post renegotiation-proof mechanism for this mechanism design problem.

Finally, we remark that the "Full Support" assumption is necessary for Proposition 2. In environments where it is not satisfied, the requirements of ex-post renegotiation-proofness are weaker, and thus there may exist ex-post renegotiation-proof mechanisms that are not budget balanced Groves mechanisms.

Example 3. Suppose that there are two players, a buyer and a seller. The seller owns an object that the buyer may want to buy. There are two states of the world, L and H. In state L the buyer's valuation for the object is 2 and the seller's reservation value is 1. In state H the buyer's valuation for the object is 6 and the seller's reservation value is 5.

It is straightforward to show that no budget balanced Groves mechanism exist in this environment. Nevertheless, there exist many ex-post renegotiation-proof mechanisms. For example, a mechanism that prescribes trade with probability 1 at a constant price that is independent of the players' messages is ex-post renegotiation-proof.

4.2. Interim renegotiation-proofness Under Complete Information

Ex-post renegotiation takes place after the mechanism has been played. Interim renegotiation takes place before the mechanism is to be played. In a complete information environment, the players do not learn anything about the state of the world from playing the mechanism. It therefore follows that any equilibrium that the players would want to renegotiate in the interim stage they would also want to renegotiate ex-post. Thus any mechanism that is ex-post renegotiation-proof is also interim renegotiation-proof.[19,20]

5. Renegotiation-Proofness Under Incomplete Information

5.1. Ex-Post Renegotiation-Proofness Under Incomplete Information

Consider an equilibrium σ of a mechanism $\langle S, m \rangle$. Let $\psi : A \times \mathbb{R}^n \to A \times \mathbb{R}^n$ denote an alternative-outcome-generating-function from the set of possible outcomes into itself that for every outcome (a, t) that may be obtained under the mechanism $\langle S, m \rangle$ specifies an alternative outcome $\psi(a, t)$. A mapping ψ is feasible if it maps outcomes into feasible outcomes.

Suppose that the process of renegotiation assumes the following form. After an outcome (a, t) is produced by an equilibrium σ of a mechanism $\langle S, m \rangle$ and is communicated to the players, a different feasible outcome $\psi(a, t)$ is proposed to the players. The players vote simultaneously on whether to accept the alternative outcome $\psi(a, t)$. If all the players vote unanimously in favor of the alternative outcome $\psi(a, t)$, then it is implemented instead of the originally proposed outcome (a, t). Otherwise, the outcome (a, t) is implemented.

Under complete information or when players have private values, each player can easily tell which of any two given outcomes he prefers. But when there is incomplete information and players have interdependent valuations, whether or not a player prefers one outcome to another may depend on what types of the other players vote in favor of the alternative outcome. Furthermore, when the players observe an outcome that is consistent with play

[19] Segal and Whinston (2002) made the same observation with respect to their notions of interim and ex-post renegotiation proofness.

[20] A mechanism may be interim renegotiation-proof but not ex-post renegotiation proof. Consider the environment described in Example 1 above where the buyer can have a value of 1 or 5, and the seller has a reservation value of 2. Consider the following mechanism: the buyer announces his value, if he says "1" then there is no trade and no payment, and if he says "5" then there is trade at a price of 3. As explained above, this mechanism is not ex-post renegotiation-proof. However, this mechanism is interim renegotiation-proof. If the buyer's value is 5, then by rejecting any alternative mechanism the seller can guarantee himself a payoff of 1 and the buyer can guarantee himself a payoff of 2. Any alternative mechanism would just redistribute the surplus of 3 between the buyer and seller, and would therefore necessarily be blocked by either the buyer or the seller. The same argument applies in the case when the buyer's value is 1.

of the equilibrium σ it is reasonable to assume that they will update their beliefs about the types of the other players in a way that is consistent with σ (except, of course, for a player who deviated from σ who believes that other players played according to σ but knows he played differently); if, on the other hand, the players observe an outcome that is inconsistent with play of the equilibrium σ, then it is reasonable to assume that they will update their beliefs about the types of the other players taking into account that one of them deviated and played a different strategy.[21] These considerations lead to the following definition.

Let $P_\sigma(a,t)$ denote the probability distribution over players' types conditional on the outcome (a,t) under the assumption that the players employ the equilibrium strategies σ, and let $P_{\sigma_i',\sigma_{-i}}(a,t)$ denote the probability distribution over players' types conditional on the outcome (a,t) when player i employs the strategy σ_i' and all the other players employ the equilibrium strategies σ_{-i}.

Definition 5. An alternative (a,t) that is produced under an equilibrium σ of a mechanism $\langle S,m\rangle$ when one of the players i may have possibly played a different strategy σ_i' is said to be *renegotiated away with a positive probability* if either one of the the following conditions is satisfied:

1. If the players have all played the equilibrium strategies σ, then (a,t) is *renegotiated away with a positive probability* if there exists a set of types $T=T_1\times\cdots\times T_n$ that has a positive P_σ probability such that:

    1. All the types in T prefer the alternative outcome $\psi(a,t)$ to the outcome (a,t) conditional on the event T and for at least one type in T this preference is strict, where the beliefs of all the players are given by $P_\sigma(a,t)$,
    2. All the types outside the set T prefer the outcome (a,t) to the alternative outcome $\psi(a,t)$ conditional on the event T and their own type, where the beliefs of all the players are given by $P_\sigma(a,t)$.

2. If player i has played a strategy $\sigma_i'\neq\sigma_i$, all the other players have played the equilibrium strategies σ_{-i}, and the outcome (a,t) is consistent with play of σ, then (a,t) is *renegotiated away with a positive probability* if there exists a set of types $T=T_1\times\cdots\times T_n$ that has a positive $P_{\sigma_i',\sigma_{-i}}$ probability such that:

    1. All the types in T prefer the alternative outcome $\psi(a,t)$ to the outcome (a,t) conditional on the event T and for at least one type in T this preference is strict,

[21] Note that we do not require that the beliefs of the different types be consistent with each other. That is, two types of two different players may have different beliefs about what produced the outcome (a,t) when the latter is inconsistent with the equilibrium σ.

where the beliefs of all the players other than i are given by $P_{\sigma}(a,t)$ and player i's beliefs are given by $P_{\sigma_i',\sigma_{-i}}(a,t)$,

2. All the types outside the set T prefer the outcome (a,t) to the alternative outcome $\psi(a,t)$ conditional on the event T and their own type, where, again, the beliefs of all the players other than i are given by $P_{\sigma}(a,t)$ and player i's beliefs are given by $P_{\sigma_i',\sigma_{-i}}(a,t)$.

3. If player i has played a strategy $\sigma_i' \neq \sigma_i$, all the other players have played the equilibrium strategies σ_{-i}, and the outcome (a,t) is inconsistent with play of σ, then (a,t) is *renegotiated away with a positive probability* if there exists a set of types $T = T_1 \times \cdots \times T_n$ that has a positive $P_{\sigma_i',\sigma_{-i}}$ probability such that:

   1. All the types in T prefer the alternative outcome $\psi(a,t)$ to the outcome (a,t) conditional on the event T and for at least one type in T this preference is strict, where the beliefs of all the players other than i can be any arbitrary belief that is consistent with observation of the outcome (a,t) and player i's beliefs are given by $P_{\sigma_i',\sigma_{-i}}(a,t)$,

   2. All the types outside the set T prefer the outcome (a,t) to the alternative outcome $\psi(a,t)$ conditional on the event T and their own type, where, again, the beliefs of all the players other than i can be any arbitrary belief that is consistent with observation of the outcome (a,t) and player i's beliefs are given by $P_{\sigma_i',\sigma_{-i}}(a,t)$.

Remark 2. If players have private values, then their beliefs about the other players are irrelevant for the purpose of comparing the outcome that was obtained under the mechanism (a,t) and the alternative outcome $\psi(a,t)$. In this case, the definition of "renegotiated away with a positive probability" can be much simplified in that it is enough that there exists a profile of types that all prefer the alternative $\psi(a,t)$ to (a,t) in order to ensure that (a,t) would be renegotiated away with a positive probability. This case is analyzed in Section 5.1.1 below.

Definition 6. An equilibrium σ of a mechanism $\langle S,m\rangle$ is ex-post renegotiation-proof relative to an alternative outcome generating function ψ if none of the outcomes that can be obtained under the mechanism $\langle S,m\rangle$ when some player i adopts a strategy σ_i' while all the other players follow the strategies σ_{-i} can be renegotiated away with a positive probability in a way that benefits player i at the interim stage when i considers how to play the mechanism.

Remark 3. As in the case of renegotiation under complete information, there are two ways in which an equilibrium σ of a mechanism $\langle S,m\rangle$ can be undermined. First, following equilibrium play in the mechanism the players may renegotiate away from the mechanism's recommended decision in favor of some alternative decision, and second, the players may have an incentive to deviate from their equilibrium strategies under the mechanism in anticipation of future renegotiation, and then renegotiate as anticipated. The definition captures both of these possibilities.

Definition 7. An equilibrium σ of a mechanism $\langle S, m \rangle$ is said to be ex-post renegotiation-proof if it is ex-post renegotiation-proof against every feasible ψ.

5.1.1. The Case of Independent Private Values

The main difficulty in the analysis of the process of ex-post renegotiation is due to the fact that at the voting stage the players compare their payoff from the alternative outcome $\psi(a,t)$ with their payoff from the original outcome (a,t) *conditional on the other players voting in favor of the alternative*. The analysis becomes much simpler in the case of independent private values because the additional information revealed by the other players' voting behavior is payoff irrelevant.

The fact that an outcome that is not ex-post efficient can be renegotiated to one that is in such a way that strictly benefits all the players implies that,

Lemma 2. *In an incomplete information mechanism design environment with independent private values, an ex-post renegotiation-proof Bayesian-Nash equilibrium is ex-post efficient.*

The next example, which is the incomplete information version of Example 1, demonstrates that the converse of Lemma 2 is not true. Namely, an ex-post efficient mechanism may fail to be ex-post renegotiation-proof.

Example 1'. There are two players, a buyer and a seller. The seller owns an object that the buyer may want to buy. The buyer is equally likely to value this object at either 1 or 5. The seller's reservation value for the object is 2. The state of the world is thus determined by the buyer's valuation for the object. The set of outcomes consists of three alternatives: "no trade," "trade at the price 3," and "trade at the price 4."

Consider the following mechanism: the buyer announces whether he wants to trade or not. If he announces he wants to trade, then the buyer and seller trade at the price 4; otherwise, there is no trade. Observe that the Bayesian game that is induced by this mechanism has a trivial unique Bayesian-Nash equilibrium in dominant strategies. If the buyer's valuation for the object is 1, then the buyer declines to trade. If the buyer's valuation for the object is 5, then in equilibrium the buyer agrees to trade and the object is traded at the price 4.

However, despite the fact that the Bayesian-Nash equilibrium is both ex-post efficient and in dominant strategies, it is not ex-post renegotiation-proof according to our definition.

To see this, suppose that in the event of no trade, the buyer and seller may renegotiate the outcome to trading at the price of 3. A buyer who values the object at 5 and who anticipates such a renegotiation possibility might announce that he declines to trade hoping to renegotiate the outcome to trade at a price that is better for him. Since such renegotiation would also make the seller strictly better off compared to no trade, the seller may well agree to renegotiate the outcome. Thus, the Bayesian-Nash equilibrium outcome in which the object is traded at the price 4 may be renegotiated away.

Like in the case of example 1, if the set of outcomes is expanded to allow for trade at any price, then as shown in the appendix, ex-post renegotiation proofness implies that the buyer determines if there is trade or not, and the price paid by the buyer when there is trade is larger by exactly 2 than the price paid by the buyer when there is no trade. It follows that if we add a participation or individual rationality constraint that implies that in the event of no trade the buyer does not pay anything, then in any ex-post renegotiation-proof mechanism, the buyer decides if there is trade, and the price that is paid for the object in the event of trade is equal to 2.

The next example, which is the incomplete information version of Example 2, shows that ex-post renegotiation-proof mechanisms may also fail to exist in incomplete information environments.

Example 2'. There are two players, a buyer and a seller. The seller owns an object that the buyer may want to buy. The buyer is equally likely to value the object at either 1 or 5. The seller's reservation value for the object is equally likely to be either 2 or 6. The state of the world is thus determined both by the buyer's valuation for the object and by the seller's reservation value. The set of social decisions consists of a continuum of alternatives $(q, p) \in [0, 1] \times \mathbb{R}$ where $q \in [0, 1]$ is the probability of trade and $p \in \mathbb{R}$ is a balanced transfer, or the price to be paid from the buyer to the seller.

The proof of this is a little involved, but it can be shown, in a manner that is similar to the type of argument used in Example 1' above (see the appendix), that the possibility of ex-post renegotiation implies that the expected payment paid by a buyer of type 5 is larger by exactly 1 than the expected payment paid by a buyer of type 1. Note that 1 is the probability that the seller is of type 2, and thus trade is efficient, times the reservation value of the seller in this case. Also the expected payment received by the seller of type 2 is larger by exactly $\frac{5}{2}$ than the expected payment received by the seller of type 6. Note that $\frac{5}{2}$ is the probability that the buyer is of type 5, and thus the trade is efficient, times the valuation of the buyer in this case.

An ex-post renegotiation-proof mechanism exists in this case, but if we require interim individual rationality constraints to be satisfied then no ex-post renegotiation-proof mechanisms exist for this environment. See Example 2' in the appendix for details.

5.1.2. The Case of Correlated Interdependent Players' Types

The case of interdependent values is considerably more complicated than the case of private values. The difference is that when players have private values, they do not need to know anything about other players' types in order to decide whether an alternative outcome (a', t') Pareto dominates the mechanism's decision (a, t). In contrast, when players have interdependent valuations, whether or not it is in a player's best interest to renegotiate the outcome may depend on another player's type. And since other players willingness to renegotiate the outcome depends on their types, players have to take into account what types of other players are likely to agree to renegotiate the outcome.

The next example illustrates some of the difficulty by showing that the analog to Lemmas 1 and 2 may not hold when players have interdependent valuations. Namely, in such a case a mechanism may be ex-post renegotiation-proof in spite of not being ex-post efficient.

Example 5. There are two players. Player 1 is equally likely to be of type a or type b, player 2 has no private information. There are two decisions $\{\alpha, \beta\}$. The payoffs of the two players (u_1, u_2) are given by the following table:

	type a	type b
decision α	$0, 0$	$0, 0$
decision β	$5, -5$	$1, 1$

A mechanism that always reaches the decision α is ex-post renegotiation-proof against the alternative decision β (the fact that player 1 has a dominant strategy to vote in favor of β implies that β is unattractive for player 2). Such a mechanism is not ex-post efficient in state b.

However, when there are at least three players who have private or interdependent valuations and correlated types, then the technique of Crémer and McLean (1985, 1988), which exploits the correlation among the players' types to relax the players' incentive compatibility constraints, can be adapted to establish the existence of an ex-post renegotiation-proof mechanism that implements any ex-post efficient decision rule. The idea is that in order to implement a given social choice rule, the mechanism designer needs to know the state of the world, or the players' true types. In order to induce player i to reveal his type truthfully, it is possible to "stochastically compare" his report to the report of player j while using player k as a budget-breaker. Because in such a scheme player i's report does not affect player j's payoff, this does not influence player j's incentive to report the truth. And it is possible to "rotate" the roles of players i, j, and k, so as to provide every player with an incentive to report the truth while maintaining budget balance. Once the players are induced to report their types truthfully, the fact that the decision rule is ex-post efficient prevents them from renegotiating the outcome.

Proposition 3. *Consider an incomplete information mechanism design environment with $n \geq 3$ players. Suppose that the beliefs of every player i about all the other players are linearly independent. Let $a : \Theta \rightarrow A$ be an ex-post efficient decision rule, and let $t : \Theta \rightarrow \mathbb{R}^n$ be a budget balanced vector of transfer functions, then there exists an incentive compatible, budget balanced, and ex-post renegotiation-proof mechanism that implements (a, t).*

As shown by Neeman (2004) and Heifetz and Neeman (2006) the condition that players' beliefs about other players be linearly independent is satisfied generically in type spaces with a given finite number of types, but it fails generically when the collection of all finite type spaces is considered.[22] The question of what can be done when players' beliefs are not linearly independent and so the techniques of Crémer and McLean cannot be used remains open.

5.2. Ex-Post Oracle Renegotiation-Proofness Under Incomplete Information

The definition of ex-post renegotiation-proofness above is quite strong, since it requires a mechanism to be robust against the possibility of switching away to any feasible alternative. Nevertheless, it might be argued that it is not nearly strong enough because it does not allow the alternative proposals that are generated by ψ to depend on the private information of the players beyond what is revealed by the outcome that is produced by the mechanism. Indeed, in realistic settings renegotiation proposals result from some communication process during which the players may choose to reveal some additional private information. To capture this feature we introduce a stronger notion of renegotiation-proofness, which we call "oracle renegotiation-proofness."

Let $\widehat{\psi} : \Theta \times A \times \mathbb{R}^n \rightarrow A \times \mathbb{R}^n$ denote an alternative-outcome-generating-function that for every profile of players' types θ and every outcome (a, t) that may be obtained under the mechanism $\langle S, m \rangle$ specifies an alternative outcome $\psi(\theta, (a, t))$. A mapping $\widehat{\psi}$ is feasible if it maps profiles of types and outcomes into feasible outcomes.

Definition 8. An equilibrium σ of a mechanism $\langle S, m \rangle$ is ex-post oracle renegotiation-proof relative to a mapping $\widehat{\psi}$ if none of the outcomes that can be obtained under the mechanism

[22] Linear independence is a special case of what Heifetz and Neeman (2006) called the "belief determine preferences" or BDP property. Heifetz and Neeman established their result for private values environments, but the logic of their argument extends to environments with interdependent valuations as well.

$\langle S, m \rangle$ when some player i adopts a strategy σ_i' while all the other players follow the strategies σ_{-i} can be renegotiated away with a positive probability in a way that benefits player i.

The notion of "renegotiated away with a positive probability" is defined as above, with obvious modifications. Namely, the players may learn about other players' types from the proposed alternative $\psi(\theta, (a, t))$.

Definition 9. An equilibrium σ of a mechanism $\langle S, m \rangle$ is said to be ex-post oracle renegotiation-proof if it is ex-post renegotiation-proof against every feasible mapping $\widehat{\psi}$.

The definition of ex-post oracle renegotiation envisions an "oracle" that given the mechanism's decision and the players' types, recommends an alternative decision that the players are likely to prefer to the mechanism's original recommendation. The players treat the oracle's recommendation as exogenous.

As mentioned above, the oracle device is meant to capture the possibility that the alternative proposals may depend on the players' private information beyond what is revealed by the outcome that is produced by the mechanism. We conjecture that it is possible to show (at least for the case of private values) that if an equilibrium of a mechanism is ex-post oracle renegotiation-proof, then it is also robust against renegotiation in *any* model with an explicit renegotiation protocol. However, we postpone further investigation of this issue to future research.

Another justification for the oracle device is that in some realistic settings the state of the world may become commonly known at the ex-post stage. Thus to ensure the renegotiation-proofness of an equilibrium of a mechanism ex-post oracle renegotiation-proofness is required.

The difference between ex-post renegotiation-proofness and ex-post oracle renegotiation-proofness is illustrated in the following example that describes a mechanism that is ex-post renegotiation-proof, but not ex-post oracle renegotiation-proof.

Example 4. There are two players, a buyer and a seller. The buyer is equally likely to value an object at either 0 or 3. The seller's reservation value is equally likely to be 1 or 2. The buyer's valuation and the seller's reservation value are stochastically independent. The buyer is privately informed about his valuation and the seller is privately informed about his reservation value. The set of outcomes is given by $A = \{$"no trade," "trade at price 1," "trade at price 2"$\}$. Consider the following mechanism: the buyer announces his value. If he announces the value 0, then there is no trade; if he announces the value 3, then there is trade at the price 2. Observe that truth-telling is a dominant strategy for the buyer under this mechanism.

This mechanism is ex-post renegotiation-proof. The equilibrium payoff of the buyer whose valuation is 3 is 1. The payoff of a buyer with valuation 3 from announcing that his type is zero and then renegotiating to trade at the price 1 is $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 0 = 1$ because a seller whose reservation value is 2 would object to renegotiation. However, the mechanism is not ex-post oracle renegotiation-proof because the expected payoff to the buyer whose valuation is 3 if he announces that his valuation is zero and then renegotiates to trade at the price 1 when the seller's reservation value is 1 and to trade at the price 2 when the seller's reservation value is 2 is $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 1 = \frac{3}{2} > 1$.

5.2.1. The Case of Independent Private Values

The next proposition provides a characterization of the set of environments that admit the existence of an ex-post oracle renegotiation-proof mechanism under the assumption of independent private values. The characterization, which unlike the characterization described in Proposition 2 is independent of the number of agents, is in terms of mechanisms that are "Groves in expectation."

Definition 10. A direct revelation mechanism $\langle a, t\rangle$ is said to be *Groves in expectation* if a is an ex-post efficient decision rule and for every type $\theta_i \in \Theta_i$ of every player $i \in N$,

$$E_{\theta_{-i}}\left[t_i\left(\theta_i, \theta_{-i}\right)\right] = E_{\theta_{-i}}\left[\sum_{j \neq i} v_j\left(a\left(\theta_i, \theta_{-i}\right), \theta\right)\right] + H_i$$

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

In a mechanism that is Groves in expectation the expected payment to each player i as a function of his type is equal to the expected payment to the player as a function of his type under some Groves mechanism.[23]

Proposition 4. *Consider an incomplete information mechanism design environment with independent private values. In such a problem, a feasible mechanism is ex-post oracle renegotiation-proof if and only if it is a budget balanced Groves in expectation.*

The intuition for the proof of Proposition 3 is the following. The way we defined the process of renegotiation implies that a player can misrepresent his type when a mechanism is played and then renegotiate to an ex-post efficient outcome and capture the difference in social surplus. This implies that the possibility of ex-post renegotiation allows any player to capture the surplus or externality that he generates up to a constant. It therefore follows that a mechanism in which each player already gets the surplus or externality he generates

[23]Thus, the class of mechanisms that are Groves in expectation includes the class of Groves mechanisms, AGV mechanisms (after Arrow, 1979, and d'Aspremont and Gerard-Varet, 1979), as well as other mechanisms.

is ex-post renegotiation-proof, and conversely, any mechanism that is ex-post renegotiation-proof must be a mechanism in which each player obtains a payoff that is equal to the surplus he generates up to a constant.

Hence, Proposition 3 is a consequence of the fact that the class of mechanisms in which players' payoff are equal to the externality they generate is the class of mechanisms that are Groves in expectation. It is the class of mechanisms that are Groves in expectation and not just Groves because the players contemplate how best to misrepresent their types at the interim rather than at the ex-post stage, and this implies that the interim expected transfer to each player has to be equal to the interim expected externality, rather than the ex-post transfer equal to the ex-post externality.

Example 1' (continued). Note that the price paid by the buyer when there is trade is larger by exactly 2 than the price paid by the buyer when there is no trade, which is exactly the requirement for the payments of the buyer to be the Groves payments. Hence in this example the set of the ex-post renegotiation-proof mechanisms is the same as the set of the oracle ex-post renegotiation-proof mechanisms. This is due to the fact that the only incomplete information is on the side of the buyer, so the oracle's announcements of the state of the world are of no use here because the buyer already has this information.

Example 2' (continued). Note that the expected payment paid by a buyer of type 5 is larger by exactly 1 than the expected payment paid by a buyer of type 1, which is exactly the requirement for the payments of the buyer to be Groves in expectation. Also the expected payment received by a seller of type 2 is larger by exactly $\frac{5}{2}$ than the expected payment received by a seller of type 6, which is exactly the requirement for the payments of the seller to be Groves in expectation.

Hence in this example the set of the ex-post renegotiation-proof mechanisms is the same as the set the oracle ex-post renegotiation-proof mechanisms (which was shown to be empty if we require interim individual rationality). This is due to the fact that the oracle's announcements of the state of the world are of no use in this example because each player is anyway able to figure out the state of the world on its own. In many cases this can be done by combining the player's own private information with the information revealed by the decision suggested by the mechanism. If this is not enough, then there exist renegotiation alternatives that, if proposed, induce the other players to reveal the information that is necessary to figure out the state of the world.

Remark 4. Williams (1999) showed that if the sets of players' types are connected open subsets of $\mathbb{R}^n$ and the players' interim expected valuations are continuously differentiable in their types then any mechanism that is both ex-post efficient and Bayesian incentive compatible is payoff equivalent to a Groves mechanism at the interim stage. When this equivalence holds, Proposition 3 implies that an individually rational[24] ex-post oracle renegotiation-proof mechanism exists if and only if there exists a feasible, ex-post efficient, Bayesian incentive compatible, direct revelation mechanism that is also individually rational. The fact that for several economically important mechanism design problems, such as bilateral trade (see, e.g., Myerson and Satterthwaite, 1983), regulation (see, e.g., Baron and Myerson, 1982), and litigation and settlement (see, e.g., Spier 1994 and Klement and Neeman, 2005), no individually rational ex-post efficient mechanisms exists implies that no individually rational ex-post oracle renegotiation-proof mechanisms exist in such mechanism design problems either.

Remark 5. We do not know what is the set of mechanisms that characterizes ex-post rather than oracle ex-post renegotiation-proofness. However, Example 2' above suggests that in many cases these two notions of renegotiation-proofness would be equivalent. Thus the replacement of the notion of oracle ex-post renegotiation-proofness with the weaker notion of ex-post renegotiation-proofness would still often fail to ensure the existence of an ex-post renegotiation-proof mechanism if interim individual rationality is required.

5.2.2. The Case of Correlated Interdependent Players' Types

Inspection of the proof of Proposition 3 shows that the mechanism that is described there is also ex-post oracle renegotiation-proof.

5.3. Interim Renegotiation-Proofness Under Incomplete Information

The process of interim renegotiation is modeled in a similar way to the process of ex-post renegotiation except that renegotiation of the mechanism takes place before the mechanism is played. In this case, the beliefs of the players about how an alternative mechanism will be played and about how the original mechanism will be played after the rejection of an alternative mechanism are important. We say that an equilibrium of a mechanism is interim renegotiation-proof if it is never renegotiated for whatever rational beliefs that the players might hold.

Suppose that the process of renegotiation at the interim stage, before the mechanism is played, assumes the following form. Fix a mechanism $\langle S, m \rangle$ and a Bayesian-Nash equilibrium of this mechanism $\sigma = (\sigma_1, ..., \sigma_n)$. Suppose that an alternative mechanism $\langle S', m' \rangle$ (that has at least one equilibrium for any subset of types that choose to play it) is exogenously proposed to the players. The players vote simultaneously whether to retain the original mechanism $\langle S, m \rangle$, or to replace $\langle S, m \rangle$ by the new mechanism $\langle S', m' \rangle$. If all the players vote in favor of the alternative mechanism $\langle S', m' \rangle$, then it is played instead of $\langle S, m \rangle$.

[24] More precisely, Interim Individually Rational with zero on the right-hand-side of the individual rationality constraint.

Otherwise, the players continue to play the original mechanism $\langle S, m\rangle$ using possibly different strategies than σ that reflect what they have learned about other players' types from the rejection of the alternative mechanism $\langle S', m'\rangle$. In either case, players are only informed about the outcome of the vote, not about the votes of individual players.

A pair of mechanisms $\langle S, m\rangle$ and $\langle S', m'\rangle$ thus defines an interim renegotiation game. Player i's strategy in this interim renegotiation game is given by (i) a voting strategy $\rho_i : \Theta_i \to [0,1]$ that denotes the probability that player i votes to reject $\langle S, m\rangle$ in favor of the alternative mechanism $\langle S', m'\rangle$ as a function of player i's true type θ_i; (ii) a strategy $\sigma_i : \Theta_i \to \Delta S_i$ that is played in the mechanism $\langle S, m\rangle$ if it is retained; (iii) a strategy $\sigma'_i : \Theta_i \to \Delta S'_i$ that is played in the mechanism $\langle S', m'\rangle$ if it replaces $\langle S, m\rangle$.

Definition 11. A profile of players' strategies $(\rho_i, \sigma_i, \sigma'_i)_{i\in N}$ is a sequential equilibrium of the interim renegotiation game that is induced by the two mechanisms $\langle S, m\rangle$ and $\langle S', m'\rangle$ if

1. Every type's strategy is a best response to the other players' strategies;
2. Players update their beliefs about the other players' types using Bayes rule whenever possible, taking $(\rho_i, \sigma_i, \sigma'_i)_{i\in N}$ into account.

Definition 12. An equilibrium σ of a mechanism $\langle S, m\rangle$ is said to be *interim renegotiation-proof* if there does not exist a mechanism $\langle S', m'\rangle$ and a sequential equilibrium of the interim renegotiation game that is induced by the two mechanisms $\langle S, m\rangle$ and $\langle S', m'\rangle$ in which (i) the players vote in favor of the alternative mechanism $\langle S', m'\rangle$ with a positive probability; and (ii) at least one of the types who votes in favor of the alternative mechanism $\langle S', m'\rangle$ with a positive probability strictly prefers $\langle S', m'\rangle$ to $\langle S, m\rangle$.

Remark 6. The renegotiation game described above is similar to the one described in Holmström and Myerson (1983). The main difference between the definition presented here and Holmström and Myerson's (1983) definition of "durability" is that Holmström and Myerson defined a mechanism to be "durable" if for every alternative mechanism there is a (non trivial) voting equilibrium in which this alternative mechanism is rejected. In contrast, we define a mechanism to be interim renegotiation-proof if every alternative mechanism is rejected *in every equilibrium* in which it is preferred by at least some players' types. As shown by the next example, which is due to Holmström and Myerson (1983), our definition of interim renegotiation-proofness is strictly stronger than their definition of durability.

Example 6 (Holmström and Myerson, 1983). Suppose that there are two players with independent and equally likely types $(1a, 1b; 2a, 2b)$. There are two outcomes A and B. The players' payoffs are:

$$u_1(A, \theta) = u_2(A, \theta) = 2 \qquad \forall \theta \in \Theta$$

and

$$u_1(B,\theta) = u_2(B,\theta) = \begin{cases} 3 & \text{if} \quad \theta = (1a, 2a) \text{ or } \theta = (1b, 2b) \\ 0 & \text{if} \quad \theta = (1a, 2b) \text{ or } \theta = (1b, 2a) \end{cases}$$

The constant mechanism that selects the alternative A in every state of the world is durable because in any voting game with any alternative mechanism there is always an equilibrium rejection in which both players use uninformative voting and reporting strategies. For example, suppose that the following mechanism is suggested to the players as an alternative mechanism. The players report their types. If the types match, then alternative B is chosen; if they don't, then alternative A is chosen. This alternative mechanism has an equilibrium in which the players report their types truthfully that Pareto dominates the constant mechanism, which implies that the constant mechanism is not interim renegotiation-proof. But this alternative mechanism also has another equilibrium in which the players randomize on their type reports and where the expected payoff to each type is $7/4$. If the players believe that this is the equilibrium that will be played under the alternative mechanism rather than the Pareto efficient equilibrium then they would vote against the alternative mechanism.

The question of whether every mechanism design problem admits the existence of an interim renegotiation-proof mechanism is difficult and remains open. A partial answer to this question is given by the following proposition.

Proposition 5. *A budget balanced ex-post efficient equilibrium of a mechanism is interim renegotiation-proof. In particular, budget balanced Groves mechanisms are interim renegotiation-proof in private values environments.*

5.4. Renegotiation-Proofness Under Incomplete Information

Define a mechanism to be renegotiation-proof if it is both ex-post and interim renegotiation-proof. Our results imply that in independent private values environments any Groves mechanism gives rise to a dominant strategy equilibrium that is both ex-post and interim renegotiation-proof and hence also renegotiation-proof. In particular, in private values environments that satisfy the assumptions described in Remark 4 above, an incomplete information mechanism design problem admits the existence of a renegotiation-proof mechanism if and only if it admits the existence of a budget balanced Groves mechanism, and it admits the existence of an individually rational renegotiation-proof mechanism if and only if it admits the existence of an individually rational budget balanced Groves mechanism.

Appendix

Proof of Proposition 1.

Fix an ex-post efficient decision rule a and a budget balanced vector of transfers $t : \Theta \to \mathbb{R}^n$. Consider a mechanism (α, τ) that requires each player to report the state of the world, and that determines the outcome as a function of the players' reports $\left(\widehat{\theta}_1, ..., \widehat{\theta}_n\right)$ as follows:

$$(\alpha, \tau_1, ..., \tau_n)\left(\widehat{\theta}_1, ..., \widehat{\theta}_n\right) = \begin{cases} (a(\theta), t_1(\theta), ..., t_n(\theta)) & \text{if } \widehat{\theta}_1 = \widehat{\theta}_2 = \theta \in \Theta \\ (a_0, -M, -M, 2M, 0, ..., 0) & \text{if } \widehat{\theta}_1 \neq \widehat{\theta}_2 \end{cases}$$

where $a_0 \in A$ is some fixed social alternative, and the constant M is chosen such that $M > 2n\left(\max_{i\in N, a\in A, \theta\in\Theta} |v_i(a,\theta)| + \max_{i\in N,\theta\in\Theta} |t_i(\theta)|\right)$. The (direct revelation) mechanism (α, τ) is incentive compatible, budget balanced, and ex-post renegotiation-proof, and it implements the decision rule and the vector of transfers (a, t). ■

Proof of Proposition 2.

<**IF**> Let $\langle a, t\rangle$ be a budget balanced Groves mechanism. We show that $\langle a, t\rangle$ is ex-post renegotiation-proof.

Suppose that $\langle a, t\rangle$ is not ex-post renegotiation-proof. We show that this leads to a contradiction. The fact that a is ex-post efficient implies that if the two players report their types truthfully, then they would not be able then to renegotiate the outcome. It therefore follows that there exists a state of the world $\theta = (\theta_1, \theta_2)$, a player $i \in \{1, 2\}$, and a type of player i, $\theta_i' \neq \theta_i$ such that when the state of the world is θ, player i, whose type is commonly known between the players to be θ_i, would benefit from reporting that his type is θ_i' and then renegotiating the outcome from $(a(\theta_i', \theta_j), t(\theta_i', \theta_j))$ to $\left(a(\theta_i, \theta_j), \widehat{t}(\theta_i, \theta_j)\right)$ where $\widehat{t}$ is some ex-post budget balanced transfer function (by renegotiating the outcome to the ex-post efficient outcome $a(\theta_i, \theta_j)$, θ_i is able to capture the greatest possible surplus for himself, and so would prefer that to any other outcome $a \in A$; the transfers $\widehat{t}$ facilitate this renegotiation).

A report of θ_i' that is followed by such renegotiation is beneficial for player i if

$$v_i(a(\theta_i, \theta_j), \theta_i) + \widehat{t}_i(\theta_i, \theta_j) > v_i(a(\theta_i, \theta_j), \theta_i) + t_i(\theta_i, \theta_j) \tag{1}$$

if and only if

$$\widehat{t}_i(\theta_i, \theta_j) > t_i(\theta_i, \theta_j). \tag{2}$$

Player j agrees to the proposed renegotiation if and only if the transfer $\widehat{t}$ is such that:

$$v_j(a(\theta_i, \theta_j), \theta_j) + \widehat{t}_j(\theta_i, \theta_j) \geq v_j(a(\theta_i', \theta_j), \theta_j) + t_j(\theta_i', \theta_j),$$

or

$$\widehat{t}_j(\theta_i, \theta_j) \geq v_j(a(\theta_i', \theta_j), \theta_j) - v_j(a(\theta_i, \theta_j), \theta_j) + t_j(\theta_i', \theta_j). \tag{3}$$

The fact that both t and $\widehat{t}$ are ex-post budget balanced implies that $t_j(\theta_i', \theta_j) = -t_i(\theta_i', \theta_j)$ and $\widehat{t}_j(\theta_i, \theta_j) = -\widehat{t}_i(\theta_i, \theta_j)$. Plugging these two equations into (3) implies:

$$\widehat{t}_i(\theta_i, \theta_j) \leq v_j(a(\theta_i, \theta_j), \theta_j) - v_j(a(\theta_i', \theta_j), \theta_j) + t_i(\theta_i', \theta_j) \tag{4}$$

The fact that $\langle a, t \rangle$ is a Groves mechanism implies that

$$t_i(\theta_i, \theta_j) = v_j(a(\theta_i, \theta_j), \theta_j) + h_i(\theta_j)$$

or

$$v_j(a(\theta_i, \theta_j), \theta_j) = t_i(\theta_i, \theta_j) - h_i(\theta_j)$$

and

$$t_i(\theta_i', \theta_j) = v_j(a(\theta_i', \theta_j), \theta_j) + h_i(\theta_j)$$

or

$$v_j(a(\theta_i', \theta_j), \theta_j) - t_i(\theta_i', \theta_j) = -h_i(\theta_j)$$

for some function $h_i : \Theta_j \to \mathbb{R}$. Plugging the two equations above into (4) it follows that:

$$\begin{aligned} \widehat{t}_i(\theta_i, \theta_j) &\leq [t_i(\theta_i, \theta_j) - h_i(\theta_j)] + h_i(\theta_j) \\ &= t_i(\theta_i, \theta_j). \end{aligned}$$

A contradiction to (2).

<**Only If**> Let m be a mechanism that is ex-post budget balanced and ex-post renegotiation-proof, and let $\langle a, t \rangle$ denote its associated incentive compatible direct revelation mechanism. We show that $\langle a, t \rangle$ is a budget balanced Groves mechanism.

For every $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$ define

$$S(\theta_1, \theta_2) = \max_{a \in A} \{v_1(a, \theta_1) + v_2(a, \theta_2')\}.$$

The fact that $\langle a, t \rangle$ is ex-post renegotiation-proof implies that $a(\theta, \theta)$ must be ex-post efficient for every $\theta \in \Theta$. That is, whenever the players agree on the state of the world, the mechanism must choose efficiently given the players' reports. It therefore follows that

$$S(\theta_1, \theta_2) = v_1(a(\theta, \theta), \theta_1) + v_2(a(\theta, \theta), \theta_2) \tag{5}$$

for every $\theta \in \Theta$. But when the players fail to agree, renegotiation-proofness imposes no such obvious restriction on the mechanism $\langle a, t \rangle$, and so the definition of S implies that

$$S(\theta_1, \theta_2') \geq v_1(a(\theta, \theta'), \theta_1) + v_2(a(\theta, \theta'), \theta_2') \tag{6}$$

for any pair of states of the world $\theta, \theta' \in \Theta$.

Suppose that it is commonly known between the players that the state of the world is $\theta = (\theta_1, \theta_2)$. Player 1 can report that the state of the world is $\theta' = (\theta_1', \theta_2') \in \Theta_1 \times \Theta_2$ and then offer to renegotiate the outcome from $(a(\theta', \theta), t(\theta', \theta))$ to $\left(a(\theta, \theta), \widehat{t}(\theta, \theta)\right)$ where $\widehat{t}$ is some ex-post budget balanced transfer function.

Player 2 would agree to this renegotiation if the transfer $\widehat{t}_2$ is such that:

$$v_2(a(\theta,\theta),\theta_2)+\widehat{t}_2(\theta,\theta)\geq v_2(a(\theta',\theta),\theta_j)+t_2(\theta',\theta),$$

or

$$\widehat{t}_2(\theta,\theta)\geq v_2(a(\theta',\theta),\theta_2)-v_2(a(\theta,\theta),\theta_2)+t_2(\theta',\theta).$$

The payoff player 1 can therefore get through renegotiation is equal to

$$v_1(a(\theta,\theta),\theta_1)-\widehat{t}_2(\theta,\theta)=v_1(a(\theta,\theta),\theta_1)-v_2(a(\theta',\theta),\theta_2)+v_2(a(\theta,\theta),\theta_2)-t_2(\theta',\theta).$$

The fact that $\langle a,t\rangle$ is ex-post renegotiation-proof implies that when player 1 contemplates whether to misreport and then renegotiate, he concludes that this cannot increase his expected payoff, or:

$$v_1(a(\theta,\theta),\theta_1)+t_1(\theta,\theta)\geq v_1(a(\theta,\theta),\theta_1)-v(a(\theta',\theta),\theta_2)+v_2(a(\theta,\theta),\theta_2)-t_2(\theta',\theta)$$

or $\{t_1(\theta',\theta)+t_2(\theta',\theta)\leq 0\}$

$$t_1(\theta,\theta)+t_2(\theta',\theta)\geq -v_2(a(\theta',\theta),\theta_2)+v_2(a(\theta,\theta),\theta_2) \tag{7}$$

for every $\theta'\neq\theta$. By repeating the argument for player 2, it follows that

$$t_2(\theta,\theta)+t_1(\theta,\theta')\geq -v_1(a(\theta,\theta'),\theta_1)+v_1(a(\theta,\theta),\theta_1) \tag{8}$$

for every $\theta'\neq\theta$.

Adding (7) and (8) together and using budget balance implies that:

$$t_2(\theta',\theta)+t_1(\theta,\theta')\geq -v_2(a(\theta',\theta),\theta_2)+v_2(a(\theta,\theta),\theta_2)-v_1(a(\theta,\theta'),\theta_1)+v_1(a(\theta,\theta),\theta_1) \tag{9}$$

and by switching θ and θ' also:

$$t_2(\theta,\theta')+t_1(\theta',\theta)\geq -v_2(a(\theta,\theta'),\theta_2')+v_2(a(\theta',\theta'),\theta_2')-v_1(a(\theta',\theta),\theta_1')+v_1(a(\theta',\theta'),\theta_1') \tag{10}$$

Adding (9) and (10) together, using $t_1(\theta',\theta)+t_2(\theta',\theta)\leq 0$ and $t_1(\theta,\theta')+t_2(\theta,\theta')\leq 0$, and rearranging, implies that:

$$\begin{aligned}v_1(a(\theta,\theta'),\theta_1)+v_2(a(\theta,\theta'),\theta_2')+v_1(a(\theta',\theta),\theta_1')+v_2(a(\theta',\theta),\theta_2)&\\ \geq v_2(a(\theta,\theta),\theta_2)+v_1(a(\theta,\theta),\theta_1)+v_2(a(\theta',\theta'),\theta_2')+v_1(a(\theta',\theta'),\theta_1')&\end{aligned}$$

Thus, (6) implies that a necessary condition for ex-post renegotiation proofness is that:

$$S(\theta_1,\theta_2')+S(\theta_1',\theta_2)\geq S(\theta_1,\theta_2)+S(\theta_1',\theta_2').$$

Repeating the previous argument for the pair of states (θ_1, θ_2') and (θ_1', θ_2) instead of the pair (θ_1, θ_2) and (θ_1', θ_2') implies:

$$S(\theta_1', \theta_2') + S(\theta_1, \theta_2) \geq S(\theta_1, \theta_2') + S(\theta_1', \theta_2)$$

from which it follows that a necessary condition for ex-post renegotiation proofness is that:

$$S(\theta_1', \theta_2') + S(\theta_1, \theta_2) = S(\theta_1, \theta_2') + S(\theta_1', \theta_2). \tag{11}$$

Hence, all the possible inequalities in (6) must hold as equalities. For the decision rule $a : \Theta^N \to A$ this implies that:

$$\begin{aligned}
a((\theta_1, \theta_2), (\theta_1', \theta_2')) &\in \underset{a \in A}{\operatorname{argmax}} \{v_1(a, \theta_1) + v_2(a, \theta_2')\} \\
a((\theta_1', \theta_2'), (\theta_1, \theta_2)) &\in \underset{a \in A}{\operatorname{argmax}} \{v_1(a, \theta_1') + v_2(a, \theta_2)\} \\
a((\theta_1', \theta_2), (\theta_1, \theta_2')) &\in \underset{a \in A}{\operatorname{argmax}} \{v_1(a, \theta_1') + v_2(a, \theta_2')\} \\
a((\theta_1, \theta_2'), (\theta_1', \theta_2)) &\in \underset{a \in A}{\operatorname{argmax}} \{v_1(a, \theta_1) + v_2(a, \theta_2)\}
\end{aligned}$$

Our assumption that there is a unique decision that maximizes social welfare for any state of the world therefore implies that in a mechanism that satisfies ex-post renegotiation-proofness, players' reports about the other player's type are ignored by the mechanism, or that for any $\theta_1, \theta_1' \in \Theta_1$ and for any $\theta_2, \theta_2' \in \Theta_2$,

$$a((\theta_1, \theta_2'), (\theta_1', \theta_2)) = a((\theta_1, \theta_2), (\theta_1, \theta_2)). \tag{12}$$

Furthermore, the fact that (11) holds as an equality implies that all the inequalities that were used to generate it hold as equalities as well. In particular, $t_1(\theta', \theta) + t_2(\theta', \theta) \leq 0$ and $t_1(\theta, \theta') + t_2(\theta, \theta') \leq 0$, as well as (7) and (8) must hold as equalities, which implies

$$\begin{aligned}
& t_1((\theta_1, \theta_2), (\theta_1, \theta_2)) - t_1((\theta_1', \theta_2'), (\theta_1, \theta_2)) \qquad (13) \\
= \; & -v_2(a((\theta_1', \theta_2'), (\theta_1, \theta_2)), \theta_2) + v_2(a((\theta_1, \theta_2), (\theta_1, \theta_2)), \theta_2) \\
= \; & -v_2(a((\theta_1', \cdot), (\cdot, \theta_2)), \theta_2) + v_2(a((\theta_1, \cdot), (\cdot, \theta_2)), \theta_2)
\end{aligned}$$

for every $\theta' \neq \theta$, and

$$\begin{aligned}
& t_2((\theta_1, \theta_2), (\theta_1, \theta_2)) - t_2((\theta_1, \theta_2), (\theta_1', \theta_2')) \qquad (14) \\
= \; & -v_1(a((\theta_1, \theta_2), (\theta_1', \theta_2')), \theta_1) + v_1(a((\theta_1, \theta_2), (\theta_1, \theta_2)), \theta_1) \\
= \; & -v_1(a((\theta_1, \cdot), (\cdot, \theta_2')), \theta_1) + v_1(a((\theta_1, \cdot), (\cdot, \theta_2)), \theta_1)
\end{aligned}$$

for every $\theta' \neq \theta$, where in both cases, the second equality follows from (12). Hence, it follows that the reports of the players about the other player's type do not affect their transfer payments under the mechanism either, or that

$$t_i((\theta_1, \theta_2'), (\theta_1', \theta_2)) = t_i((\theta_1, \theta_2), (\theta_1, \theta_2)). \tag{15}$$

for $i \in \{1,2\}$, and for any $\theta_1, \theta_1' \in \Theta_1$ and any $\theta_2, \theta_2' \in \Theta_2$. Thus, (13) and (14) imply that

$$t_1((\theta_1,\cdot),(\cdot,\theta_2)) - t_1((\theta_1',\cdot),(\cdot,\theta_2)) = -v_2(a((\theta_1',\cdot),(\cdot,\theta_2)),\theta_2) + v_2(a((\theta_1,\cdot),(\cdot,\theta_2)),\theta_2)$$

and

$$t_2((\theta_1,\cdot),(\cdot,\theta_2)) - t_2((\theta_1,\cdot),(\cdot,\theta_2')) = -v_1(a((\theta_1,\cdot),(\cdot,\theta_2')),\theta_1) + v_1(a((\theta_1,\cdot),(\cdot,\theta_2)),\theta_1).$$

Another way to write the last two equations is the following:

$$\begin{aligned} t_1((\theta_1,\cdot),(\cdot,\theta_2)) &= v_2(a((\theta_1,\cdot),(\cdot,\theta_2)),\theta_2) - v_2(a((\theta_1',\cdot),(\cdot,\theta_2)),\theta_2) + t_1((\theta_1',\cdot),(\cdot,\theta_2)) \\ &= v_2(a((\theta_1,\cdot),(\cdot,\theta_2)),\theta_2) + h_1(\theta_2) \end{aligned}$$

and

$$\begin{aligned} t_2((\theta_1,\cdot),(\cdot,\theta_2)) &= v_1(a((\theta_1,\cdot),(\cdot,\theta_2)),\theta_1) - v_1(a((\theta_1,\cdot),(\cdot,\theta_2')),\theta_1) + t_2((\theta_1,\cdot),(\cdot,\theta_2')) \\ &= v_1(a((\theta_1,\cdot),(\cdot,\theta_2)),\theta_1) + h_2(\theta_1), \end{aligned}$$

which implies that the transfer payments are Groves transfer payments. ■

Example 1.

We show that the buyer is the one who determines if the trade takes place, and pays exactly 2 more than he pays if there is no trade.

By the revelation principle (for games with complete information), any mechanism can be described as a mapping from the announcements of the players into probabilities of trade q and the buyer's payments p:

B\S	1	5
1	$q_{1,1}, p_{1,1}$	$q_{1,5}, p_{1,5}$
5	$q_{5,1}, p_{5,1}$	$q_{5,5}, p_{5,5}$

Renegotiation-proof constraints require that each player prefers to report the state honestly, rather than misreport and consequently renegotiate to the efficient allocation and capture the efficient surplus ($S(1)$ or $S(5)$) less the utility of the other player prescribed by the mechanism.

$$\begin{aligned} B1 &: \; q_{1,1} - p_{1,1} \geq S(1) - (p_{5,1} - 2q_{5,1}) \\ B5 &: \; 5q_{5,5} - p_{5,5} \geq S(5) - (p_{1,5} - 2q_{1,5}) \\ S1 &: \; p_{1,1} - 2q_{1,1} \geq S(1) - (q_{1,5} - p_{1,5}) \\ S5 &: \; p_{5,5} - 2q_{5,5} \geq S(5) - (5q_{1,5} - p_{1,5}) \end{aligned}$$

Notice that the efficient surpluses are

$$\begin{aligned} S(1) &= \max_{q\in[0,1]}(1-2)\,q = 0 \\ S(5) &= \max_{q\in[0,1]}(5-2)\,q = 3 \end{aligned}$$

Also notice that by Lemma 1 the allocations must be ex-post efficient:

$$\begin{aligned} q_{1,1} &= 0 \\ q_{5,5} &= 1 \end{aligned}$$

Hence the renegotiation-proof constraints become

$$\begin{aligned} B1 &: -p_{1,1} \geq -p_{5,1} + 2q_{5,1} \\ B5 &: 5 - p_{5,5} \geq 3 - p_{1,5} + 2q_{1,5} \\ S1 &: p_{1,1} \geq -q_{1,5} + p_{1,5} \\ S5 &: p_{5,5} - 2 \geq 3 - 5q_{5,1} + p_{5,1} \end{aligned}$$

Adding up all four constraints we get

$$3 \geq 6 - 3q_{5,1} + q_{1,5}$$

or

$$3q_{5,1} - q_{1,5} \geq 3.$$

Since $0 \leq q_{1,5}, q_{5,1} \leq 1$, the only possibility is to have $q_{1,5} = 0$ and $q_{5,1} = 1$. Moreover, since the resulting equality actually holds as equality, this implies that all renegotiation-proof constraints also must hold as equalities. Rearranging we obtain

$$\begin{aligned} B1 &: p_{5,1} - p_{1,1} = 2 \\ B5 &: p_{5,5} - p_{1,5} = 2 \\ S1 &: p_{1,1} = p_{1,5} \\ S5 &: p_{5,5} = p_{5,1} \end{aligned}$$

Example 2.

We show that no renegotiation-proof mechanism exist in this environment.

By the revelation principle, any mechanism can be described as a mapping from the announcements of the players into probabilities of trade q and the buyer's payments p:

B\S	$(1,2)$	$(5,2)$	$(1,6)$	$(5,6)$
$(1,2)$	$q_{(1,2),(1,2)}, p_{(1,2),(1,2)}$	$q_{(1,2),(5,2)}, p_{(1,2),(5,2)}$	...	...
$(5,2)$	$q_{(5,2),(1,2)}, p_{(5,2),(1,2)}$	$q_{(5,2),(5,2)}, p_{(5,2),(5,2)}$	...	...
$(1,6)$	...	...	...	...
$(5,6)$	...	...	...	...

Replication of the argument from Example 1 yields:

$$p_{(5,2),(5,2)} = p_{(1,2),(1,2)} + 2$$

Repeating the argument from Example 1 for the states $(5,2)$ and $(5,6)$ yields:

$$p_{(5,2),(5,2)} = p_{(5,6),(5,6)} + 5$$

Repeating the argument from Example 1 for the states $(1,2)$ and $(1,6)$ yields:

$$p_{(1,2),(1,2)} = p_{(1,6),(1,6)}$$

Finally, repeating the argument from Example 1 for the states $(1,6)$ and $(5,6)$ yields:

$$p_{(5,6),(5,6)} = p_{(1,6),(1,6)}$$

These equalities are incompatible. Hence no renegotiation-proof mechanism exist for this environment.

Proof of Lemma 2.

Suppose that σ is an ex-post renegotiation-proof equilibrium of the mechanism $\langle S, m \rangle$. Suppose that σ is not ex-post efficient. It follows that there exists a decision $(a,t) \in A \times \mathbb{R}^n$, a profile of types $\theta = (\theta_1, ..., \theta_n)$ such that $m(\sigma(\theta)) = (a,t)$, and a feasible alternative decision $(a', t') \in A \times \mathbb{R}^n$ such that

$$v_i(a', \theta_i) + t'_i \geq v_i(a, \theta_i) + t_i \tag{16}$$

for every type θ_i, $i \in N$, with at least one strict inequality. We show that the ex-post renegotiation subgame has a sequential equilibrium in which the players all vote in favor of the alternative decision (a', t') with a positive probability. Inequality (3) implies that there exists an equilibrium in which the types θ_i, $i \in N$, all vote for the alternative (a', t') with a positive probability, and at least one of these types is made strictly better off by this vote. (Observe that since players are assumed to have private values, if other types also vote in favor of the alternative (a', t') in this equilibrium, this does not affect the payoff of the types θ_i, $i \in N$ conditional on switching to (a', t') and so does not disturb the equilibrium.) ∎

Example 1'.

By the revelation principle, any mechanism can be described as a mapping from the announcements of the players into probabilities of trade q and the buyer's payments p:

B\S	2
1	q_1, p_1
5	q_5, p_5

Renegotiation-proof constraints require the buyer to prefer to report the state honestly, rather than misreport and then renegotiate to the efficient allocation and capture the efficient surplus ($S(1)$ or $S(5)$) less the utility of the seller prescribed by the mechanism.

$$\begin{aligned} B1 &: q_1 - p_1 \geq S(1) - (p_5 - 2q_5) \\ B5 &: 5q_5 - p_5 \geq S(5) - (p_1 - 2q_1) \end{aligned}$$

As in Example 1 we have $S(1) = 0$ and $S(5) = 3$. Also notice that by Lemma 2 the allocations must be ex-post efficient, $q_1 = 0$ and $q_5 = 1$. Hence the renegotiation-proof constraints become

$$\begin{aligned} B1 &: -p_1 \geq -p_5 + 2 \\ B5 &: 5 - p_5 \geq 3 - p_1 \end{aligned}$$

This implies that

$$p_5 - p_1 = 2$$

Example 2'.

By the revelation principle, any mechanism can be described as a mapping from the announcements of the players into probabilities of trade q and the buyer's payments p:

B\S	2	6
1	$q_{1,2}, p_{1,2}$	$q_{1,6}, p_{1,6}$
5	$q_{5,2}, p_{5,2}$	$q_{5,6}, p_{5,6}$

Notice that by Lemma 2 ex-post renegotiation-proof mechanisms must be ex-post efficient: $q_{1,2} = q_{1,6} = q_{5,6} = 0$ and $q_{5,2} = 1$.

The expected payoff of the buyer of type 1 from the truthful reporting of his type is thus

$$\frac{1}{2}(-p_{1,2}) + \frac{1}{2}(-p_{1,6}).$$

If a buyer of type 1 reports that his type is 5 and the seller happens to be of type 2 then the mechanism prescribes trade with probability 1 and a payment $p_{5,2}$. Because the buyer is of type 1, no trade is efficient regardless of the type of the seller. The seller's payoff is $p_{5,2} - 2$, and he would agree to renegotiate to no trade as long as he is paid at least this much. If the buyer of type 1 reports that his type is 5 and the seller happens to be of type 6 then the mechanism prescribes no trade and a payment $p_{5,6}$. There are no potential gains from trade here and thus the prescribed decision will not be renegotiated. Hence, the renegotiation-proof constraint for a buyer of type 1 is

$$B1 : \frac{1}{2}(-p_{1,2}) + \frac{1}{2}(-p_{1,6}) \geq \frac{1}{2}(-p_{5,2} + 2) + \frac{1}{2}(-p_{5,6}).$$

The expected payoff of the buyer of type 5 from the truthful reporting of his type is

$$\frac{1}{2}(5 - p_{5,2}) + \frac{1}{2}(-p_{5,6}).$$

If a buyer of type 5 reports that his type is 1 and the seller happens to be of type 2 then the mechanism prescribes no trade and a payment $p_{1,2}$. If the seller happens to be of type 6 then the mechanism prescribes no trade and a payment $p_{1,6}$. If $p_{1,2} \neq p_{1,6}$ then the seller learns from the prescribed decision the true state of the world and thus whether there are potential gains from trade. In case the seller is of type 2 then the trade is efficient, and the seller would agree to renegotiate to trade as long as he is paid at least $p_{1,2} + 2$. In case the seller is of type 6 then there are no potential gains from trade here and thus the prescribed decision will not be renegotiated. If $p_{1,2} = p_{1,6} = p$ then the buyer does not know if there are any potential gains from trade. However, note that if offered to renegotiate to trade, the seller of type 2 would agree as long as he is paid at least $p+2$. Hence, the renegotiation-proof constraint for the buyer of type 5 in both of those cases is

$$B5 : \frac{1}{2}(5 - p_{5,2}) + \frac{1}{2}(-p_{5,6}) \geq \frac{1}{2}(5 - (p_{1,2} + 2)) + \frac{1}{2}(-p_{1,6}).$$

Similar analysis for both types of the seller gives the following renegotiation-proof constraints:

$$\begin{aligned} S2 &: \frac{1}{2}(p_{1,2}) + \frac{1}{2}(p_{5,2} - 2) \geq \frac{1}{2}(p_{1,6}) + \frac{1}{2}((p_{5,6} + 5) - 2) \\ S6 &: \frac{1}{2}(p_{1,6}) + \frac{1}{2}(p_{5,6}) \geq \frac{1}{2}(p_{1,2}) + \frac{1}{2}(p_{5,2} - 5) \end{aligned}$$

Adding up all four constraints we get

$$\frac{3}{2} \geq (=) \frac{3}{2}$$

Thus

$$\begin{aligned} B1, B5 &: \frac{1}{2}(p_{5,2} + p_{5,6}) - \frac{1}{2}(p_{1,2} + p_{1,6}) = 1 \\ S2, S6 &: \frac{1}{2}(p_{1,2} + p_{5,2}) - \frac{1}{2}(p_{1,6} + p_{5,6}) = \frac{5}{2} \end{aligned}$$

The prices that satisfy the above conditions are

$$\begin{aligned} p_{5,2} &= p_{1,6} + \frac{7}{2} \\ p_{1,2} &= p_{5,6} + \frac{3}{2} \end{aligned}$$

for any $p_{1,6}, p_{5,6} \in \mathbb{R}$. Thus the expected payoffs can be written as

$$\begin{aligned} B1 &: -\frac{3}{4} - \frac{1}{2}\left(p_{1,6} + p_{5,6}\right) \\ B5 &: \frac{3}{4} - \frac{1}{2}\left(p_{1,6} + p_{5,6}\right) \\ S2 &: \frac{3}{2} + \frac{1}{2}\left(p_{1,6} + p_{5,6}\right) \\ S6 &: \frac{1}{2}\left(p_{1,6} + p_{5,6}\right) \end{aligned}$$

Note that the ex-post renegotiation-proof mechanisms do not satisfy interim individual rationality. Interim individual rationality for the seller of type 6 requires $\frac{1}{2}\left(p_{1,6} + p_{5,6}\right)$ to be nonnegative, while interim individual rationality for the buyer of type 1 requires $\frac{1}{2}\left(p_{1,6} + p_{5,6}\right)$ to be at most $-\frac{3}{4}$.

Proof of Proposition 3.

<**IF**> Let $\langle a, t\rangle$ be a budget balanced Groves in expectation mechanism. We show that $\langle a, t\rangle$ is ex-post renegotiation-proof.

Suppose that $\langle a, t\rangle$ is not ex-post renegotiation-proof. We show that this leads to a contradiction. The fact that a is ex-post efficient implies that if all the players report their types truthfully, then they would not want to renegotiate the outcome. It therefore follows that there exists a player $i \in N$ and two types $\theta_i, \theta_i' \in \Theta_i$ such that type θ_i would benefit from reporting that his type is θ_i' and then, for every $\theta_{-i} \in \Theta_{-i}$, renegotiating the outcome from $\left(a\left(\theta_i', \theta_{-i}\right), t\left(\theta_i', \theta_{-i}\right)\right)$ to $\left(a\left(\theta_i, \theta_{-i}\right), \widehat{t}\left(\theta_i, \theta_{-i}\right)\right)$ where $\widehat{t}$ is some ex-post budget balanced transfer function (by renegotiating the outcome to the ex-post efficient outcome $a\left(\theta_i, \theta_{-i}\right)$, θ_i is able to capture the greatest possible surplus for himself, and so would prefer that to any other outcome $a \in A$; the transfers $\widehat{t}$ facilitate this renegotiation).

A report of θ_i' that is followed by renegotiation is beneficial for θ_i when he contemplates it in the interim stage if

$$E_{\theta_{-i}}\left[v_i\left(a\left(\theta_i, \theta_{-i}\right), \theta_i\right) + \widehat{t}_i\left(\theta_i, \theta_{-i}\right)\right] > E_{\theta_{-i}}\left[v_i\left(a\left(\theta_i, \theta_{-i}\right), \theta_i\right) + t_i\left(\theta_i, \theta_{-i}\right)\right]$$

if and only if

$$E_{\theta_{-i}}\left[\widehat{t}_i\left(\theta_i, \theta_{-i}\right)\right] > E_{\theta_{-i}}\left[t_i\left(\theta_i, \theta_{-i}\right)\right]. \tag{17}$$

Player j agrees to the proposed renegotiation if and only if the transfer $\widehat{t}_j$ is such that for every $\theta_{-i} \in \Theta_{-i}$:

$$v_j\left(a\left(\theta_i, \theta_{-i}\right), \theta_j\right) + \widehat{t}_j\left(\theta_i, \theta_{-i}\right) \geq v_j\left(a\left(\theta_i', \theta_{-i}\right), \theta_j\right) + t_j\left(\theta_i', \theta_{-i}\right),$$

or

$$\widehat{t}_j\left(\theta_i, \theta_{-i}\right) \geq v_j\left(a\left(\theta_i', \theta_{-i}\right), \theta_j\right) - v_j\left(a\left(\theta_i, \theta_{-i}\right), \theta_j\right) + t_j\left(\theta_i', \theta_{-i}\right).$$

Summing the previous inequalities over $j \neq i$, it follows that

$$\sum_{j\neq i} \widehat{t}_j(\theta_i,\theta_{-i}) \geq \sum_{j\neq i} v_j(a(\theta_i',\theta_{-i}),\theta_j) - \sum_{j\neq i} v_j(a(\theta_i,\theta_{-i}),\theta_j) + \sum_{j\neq i} t_j(\theta_i',\theta_{-i}). \tag{18}$$

The fact that both t and $\widehat{t}$ are ex-post budget balanced implies that $\sum_{j\neq i} t_j(\theta_i',\theta_{-i}) = -t_i(\theta_i',\theta_{-i})$ and $\sum_{j\neq i} \widehat{t}_j(\theta_i,\theta_{-i}) = -\widehat{t}_i(\theta_i,\theta_{-i})$. Plugging these two equations into (18) implies:

$$\widehat{t}_i(\theta_i,\theta_{-i}) \leq \sum_{j\neq i} v_j(a(\theta_i,\theta_{-i}),\theta_j) - \sum_{j\neq i} v_j(a(\theta_i',\theta_{-i}),\theta_j) + t_i(\theta_i',\theta_{-i})$$

for every $\theta_{-i} \in \Theta_{-i}$. Taking the expectation over $\theta_{-i} \in \Theta_{-i}$ implies

$$E_{\theta_{-i}}\left[\widehat{t}_i(\theta_i,\theta_{-i})\right] \leq E_{\theta_{-i}}\left[\sum_{j\neq i} v_j(a(\theta_i,\theta_{-i}),\theta_j)\right] + E_{\theta_{-i}}\left[t_i(\theta_i',\theta_{-i})\right] - E_{\theta_{-i}}\left[\sum_{j\neq i} v_j(a(\theta_i',\theta_{-i}),\theta_j)\right]. \tag{19}$$

The fact that $\langle a,t\rangle$ is a Groves in expectation mechanism implies that

$$E_{\theta_{-i}}\left[\sum_{j\neq i} v_j(a(\theta_i,\theta_{-i}),\theta_j)\right] = E_{\theta_{-i}}\left[t_i(\theta_i,\theta_{-i})\right] - H_i$$

and

$$E_{\theta_{-i}}\left[t_i(\theta_i',\theta_{-i})\right] - E_{\theta_{-i}}\left[\sum_{j\neq i} v_j(a(\theta_i',\theta_{-i}),\theta_j)\right] = H_i$$

for some constant H_i. Plugging the two equations above into (19) it follows that:

$$\begin{aligned} E_{\theta_{-i}}\left[\widehat{t}_i(\theta_i,\theta_{-i})\right] &\leq \left[E_{\theta_{-i}}\left[t_i(\theta_i,\theta_{-i})\right] - H_i\right] + H_i \\ &= E_{\theta_{-i}}\left[t_i(\theta_i,\theta_{-i})\right]. \end{aligned}$$

A contradiction to (17).

<**Only If**> Let $\langle a,t\rangle$ be a budget balanced incentive compatible direct revelation mechanism that is ex-post renegotiation proof. We show that $\langle a,t\rangle$ is a Groves in expectation mechanism.

Type $\theta_i \in \Theta_i$ of player i can report he is type $\theta_i' \in \Theta_i$ and then offer to renegotiate the outcome from $(a(\theta_i',\theta_{-i}), t(\theta_i',\theta_{-i}))$ to $\left(a(\theta_i,\theta_{-i}), \widehat{t}(\theta_i,\theta_{-i})\right)$ where $\widehat{t}$ is some ex-post budget balanced transfer function.

Player j would agree to this renegotiation if the transfer $\widehat{t}_j$ is such that for every $\theta_{-i} \in \Theta_{-i}$:

$$v_j(a(\theta_i,\theta_{-i}),\theta_j) + \widehat{t}_j(\theta_i,\theta_{-i}) \geq v_j(a(\theta_i',\theta_{-i}),\theta_j) + t_j(\theta_i',\theta_{-i}),$$

or

$$\widehat{t}_j(\theta_i,\theta_{-i}) \geq v_j(a(\theta_i',\theta_{-i}),\theta_j) - v_j(a(\theta_i,\theta_{-i}),\theta_j) + t_j(\theta_i',\theta_{-i}).$$

Summing the previous inequalities over $j \neq i$, it follows that renegotiation would be possible if for every $\theta_{-i} \in \Theta_{-i}$

$$\sum_{j \neq i} \widehat{t}_j (\theta_i, \theta_{-i}) \geq \sum_{j \neq i} v_j (a (\theta_i', \theta_{-i}), \theta_j) - \sum_{j \neq i} v_j (a (\theta_i, \theta_{-i}), \theta_j) + \sum_{j \neq i} t_j (\theta_i', \theta_{-i}).$$

The payoff player $\theta_i \in \Theta_i$ can therefore get through renegotiation is equal to

$$\begin{aligned} & v_i (a (\theta_i, \theta_{-i}), \theta_i) - \sum_{j \neq i} \widehat{t}_j (\theta_i, \theta_{-i}) \\ & \quad = v_i (a (\theta_i, \theta_{-i}), \theta_i) - \sum_{j \neq i} v_j (a (\theta_i', \theta_{-i}), \theta_j) + \sum_{j \neq i} v_j (a (\theta_i, \theta_{-i}), \theta_j) - \sum_{j \neq i} t_j (\theta_i', \theta_{-i}) \end{aligned}$$

The fact that $\langle a, t \rangle$ is ex-post renegotiation proof implies that, in the interim stage, when player i considers whether he should misreport and then renegotiate, he concludes that this cannot increase his expected payoff, or:

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

or

$$E_{\theta_{-i}} [t_i (\theta_i, \theta_{-i})] \geq E_{\theta_{-i}} \left[- \sum_{j \neq i} v_j (a (\theta_i', \theta_{-i}), \theta_j) + \sum_{j \neq i} v_j (a (\theta_i, \theta_{-i}), \theta_j) - \sum_{j \neq i} t_j (\theta_i', \theta_{-i}) \right]$$

for every $\theta_i, \theta_i' \in \Theta_i$. Because $t_i (\theta_i', \theta_{-i}) + \sum_{j \neq i} t_j (\theta_i', \theta_{-i}) = 0$, we have that

$$E_{\theta_{-i}} [t_i (\theta_i, \theta_{-i}) - t_i (\theta_i', \theta_{-i})] \geq E_{\theta_{-i}} \left[\sum_{j \neq i} v_j (a (\theta_i, \theta_{-i}), \theta_j) - \sum_{j \neq i} v_j (a (\theta_i', \theta_{-i}), \theta_j) \right]$$

for every $\theta_i, \theta_i' \in \Theta_i$. Because type $\theta_i' \in \Theta_i$ of player i can report that he is type $\theta_i \in \Theta_i$ and then offer to renegotiate the outcome as above, we may replace θ_i and θ_i' in the previous inequality to get:

$$E_{\theta_{-i}} [t_i (\theta_i, \theta_{-i}) - t_i (\theta_i', \theta_{-i})] \leq E_{\theta_{-i}} \left[\sum_{j \neq i} v_j (a (\theta_i, \theta_{-i}), \theta_j) - \sum_{j \neq i} v_j (a (\theta_i', \theta_{-i}), \theta_j) \right]$$

for every $\theta_i, \theta_i' \in \Theta_i$, from which it follows that

$$E_{\theta_{-i}} [t_i (\theta_i, \theta_{-i}) - t_i (\theta_i', \theta_{-i})] = E_{\theta_{-i}} \left[\sum_{j \neq i} v_j (a (\theta_i, \theta_{-i}), \theta_j) - \sum_{j \neq i} v_j (a (\theta_i', \theta_{-i}), \theta_j) \right]$$

for every $\theta_i, \theta_i' \in \Theta_i$ and $\theta_{-i} \in \Theta_{-i}$. By fixing $\theta_i' \in \Theta_i$, it therefore follows that for every $\theta_i \in \Theta_i$:

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

where

$$H_i = E_{\theta_{-i}}\left[t_i\left(\theta_i', \theta_{-i}\right) - \sum_{j \neq i} v_j\left(a\left(\theta_i', \theta_{-i}\right), \theta_j\right)\right].$$

It follows that $\langle a, t \rangle$ is a Groves in expectation mechanism. ■

Proof of Proposition 4.

Let $a : \Theta \to A$ be an ex-post efficient decision rule, and let $t : \Theta \to \mathbb{R}^n$ be a budget balanced vector of transfer functions. Denote the different types of player i by $\theta_i^1, \theta_i^2 ..., \theta_i^m$, respectively. A common prior distribution over the space of states of the world Θ induces for each type θ_i^j of each player i a belief $b_i\left(\theta_i^j\right) \in \Delta\left(\Theta_{-(i,i+1)}\right)$ about the types of all the other players except for player $i+1$ (to simplify the notation, we adopt the convention that player $n+1$ stands for player 1, and player 0 stands for player n). If these beliefs are linearly independent, then there exists a monetary transfer function $t_i^L : \Theta_{-(i+1)} \to \mathbb{R}$ that solves the following matrix equation:

$$\begin{bmatrix} b_i\left(\theta_i^1\right)(\cdot) \\ \vdots \\ b_i\left(\theta_i^m\right)(\cdot) \end{bmatrix} \cdot \left[t_i^{L\prime}\left(\theta_i^1, \cdot\right), ..., t_i^{L\prime}\left(\theta_i^m, \cdot\right)\right] = \begin{bmatrix} 0 & L & \cdots & L \\ L & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & L \\ L & \cdots & L & 0 \end{bmatrix}$$

The monetary transfers t_i^L are such that if they are applied to the players' reports about their types, and if all the players except for player i report their types truthfully, then for each type θ_i^j of player i, if θ_i^j reports his type truthfully, then his expected t_i^L payment is 0, but if θ_i^j misrepresents his type, then his expected t_i^L payment is L. Note that the monetary transfers t_i^L are such that player i's payment is independent of the report of player $i+1$. This implies that player $i+1$ has no incentive to misrepresent his type in order to receive a larger t_i^L transfer from player i.

Consider the following mechanism: each player is required to report his type. If the profile of reported types is θ then social alternative $a(\theta)$ is implemented, each player i is paid the transfer t_i, and in addition, each player $i \in \{1, ..., n-1\}$ pays $t_i^L(\theta)$ to player $i+1$, and player n pays $t_n^L(\theta)$ to player 1, where L is chosen so that it is larger than the upper bound

$$\overline{E} = \max_{\theta \in \Theta, a, b \in A} \sum_{i=1}^{n} |v_i(a, \theta) - v_i(b, \theta)| + \max_{\theta \in \Theta} \left(\sum_{i=1}^{n} |t_j(\theta)| \right)$$

on the maximum surplus that a player can obtain by renegotiation of any outcome to any other outcome.

Observe that the penalty for misreporting L is chosen to be sufficiently large so that the direct mechanism $\langle a, t \rangle$ is incentive compatible. The fact that $a(\theta)$ is an ex-post efficient decision rule and that t is budget balanced implies that if the players report their types truthfully then they cannot benefit from renegotiation of the outcome of the mechanism. And the fact that $L > \overline{E}$ implies that no player can benefit from misreporting his type and then renegotiating the outcome and capturing the implied increase in the social surplus from renegotiation. This is because misreporting has an expected cost of L to a player, while the increase in social surplus is bounded from above by $\overline{E} < L$. ■

Proof of Proposition 5.

Suppose that an ex-post efficient equilibrium σ of a mechanism $\langle S, m \rangle$ is not interim renegotiation-proof. It follows that there exists a set of profiles of players' types $\Upsilon = \Upsilon_1 \times \cdots \times \Upsilon_n$ that has a positive P probability, and an alternative equilibrium σ' of an alternative mechanism $\langle S', m' \rangle$, that the types in Υ all prefer, and some strictly prefer, conditional on their types and the set Υ to the outcome that is obtained when the players all play their equilibrium strategies σ under the mechanism $\langle S, m \rangle$. A contradiction to the ex-post efficiency of the equilibrium σ of the mechanism $\langle S, m \rangle$. ■

References

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

Arrow, K. (1979) "The Property Rights Doctrine and Demand Revelation under Incomplete Information," in *Economics and Human Welfare*, ed. M. Boskin. Academic Press.

d'Aspremont, C. and L.-A. Gérard-Varet (1979) "Incentives and Incomplete Information." *Journal of Public Economics* 11, 25-45.

Baron, D. P. and R. B. Myerson (1982) "Regulating a Monopolist with Unknown Costs," *Econometrica* 50, 911-930.

Beaudry, P. and M. Poitevin (1995) "Contract Renegotiation: A Simple Framework and Implications for Organization Theory," *Canadian Journal of Economics* 28, 302-335.

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

Bergemann, D. and J. Välimäki (2002) "Information Acquisition and Efficient Mechanism Design," *Econometrica* 70, 1007-1033.

Bernheim, D., Peleg, B. and M. Whinston (1987) "Coalition-Proof Nash Equilibria I. Concepts," *Journal of Economic Theory* 42, 1-12.

Bolton, P. (1990) "Renegotiation and the Dynamics of Contract Design," *European Economic Review* 34, 303-310.

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

Che, Y. K., and J. Kim (2006) "Robustly Collusion-Proof Implementation" *Econometrica* 74, 1063-1107.

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

Chung, K. S., and J. Ely (2007) "Foundations of Dominant Strategy Mechanisms," *Review of Economic Studies* 74, 447-476.

Clarke, E. (1971) "Multipart Pricing of Public Goods," *Public Choice* 8, 19-33.

Cramton, P. C., and T. R. Palfrey (1995) "Ratifiable Mechanisms: Learning from Disagreement," *Games and Economic Behavior* 10, 255-283.

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

Crémer, J. and R. McLean (1985) "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica* 53, 345-362.

Crémer, J. and R. McLean (1988) "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica* 56, 1247-1257.

Dasgupta, P. and E. Maskin (2000) "Efficient Auctions," *Quarterly Journal of Economics* 115, 341-388.

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

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

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

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

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

Green, J. and J.-J. Laffont (1994) "Non Verifiability, Costly Renegotiation and Efficiency," *Annales D'Economie et De Statistique* 36, 81-95.

Groves, T. (1973) "Incentives in Teams," *Econometrica* 41, 617-631.

Hart, O. and Moore, J. (1988) "Incomplete contracts and renegotiation," *Econometrica* 56, 755-785.

Heifetz, A. and Z. Neeman (2006) "On the (Im)possibility of Full Surplus Extraction in Mechanism Design," *Econometrica* 74, 213-233.

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

Hurwicz, L. and M. Walker (1990): "On the Generic Nonoptimality of Dominant-Strategy Allocation Mechanisms: A General Theorem which Includes Pure Exchange Economies," *Econometrica* 58, 683-704.

Jehiel, P., M. Meyer-ter-Vehn, B. Moldovanu, and W. R. Zame (2006) "The Limits of Ex-Post Implementation," *Econometrica* 74, 585-610.

Klement, A. and Z. Neeman (2005) "Against Compromise: A Mechanism Design Approach," *Journal of Law, Economics, and Organization* 21, 285-314.

Krasa, S. (1999) "Unimprovable Allocations in Economies with Incomplete Information," *Journal of Economic Theory* 87, 144-168.

Krishna, V. (2002) *Auction Theory*, Academic Press.

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

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.

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

Maskin, E. and J. Tirole (1992) "The Principal-Agent Relationship with an Informed Principal, II: Common Values," *Econometrica* 60, 1-42.

Moreno, D. and J. Wooders (1996) "Coalition-proof equilibrium," *Games and Economic Behavior* 17, 80-112.

Myerson, R. B. and M. A. Satterthwaite (1983) "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory* 29, 265-281.

Neeman, Z. (2004) "The irrelevance of private information in mechanism design," *Journal of Economic Theory* 117 (2004), 55-77.

Nisan, N., T. Roughgarden, E. Tardos, and V. V. Vazirani (2007) *Algorithmic Game Theory*, Cambridge University Press.

Noldeke, G. and K. Schmidt (1995) "Option Contracts and Renegotiation: A Solution to the Hold-up Problem," *RAND Journal of Economics* 26, 163-179.

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

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

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

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.

Spier, K. E. (1994) "Pretrial Bargaining and the Design of Fee Shifting Rules," *RAND Journal of Economics* 25, 197-214.

Vickrey, W. (1961) "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance* 16, 8-37.

Walker, M. (1980) "On the Nonexistence of a Dominant Strategy Mechanism for Making Optimal Public Decisions," *Econometrica* 48, 1521-1540.

Watson, J. (2007) "Contract, Mechanism Design, and Technological Detail," *Econometrica* 75, 55-81.

Williams, S. R. (1999) "A Characterization of Efficient, Bayesian Incentive Compatible Mechanisms," *Economic Theory* 14, 155-180.