

Renegotiation-Proof Mechanism Design^{*}

Zvika Neeman[†] and Gregory Pavlov[‡]

February 27, 2007

Preliminary and Incomplete

Abstract

A mechanism is said to be renegotiation-proof if it is robust against renegotiation both before and after it is played. We ask (1) what kind of environments admit the renegotiation-proof implementation of some social choice rules? (2) for a given environment, what kind of social choice rule are implementable in a way that is renegotiation-proof? and (3) for a given renegotiation-proof implementable social choice rule, how can the rule be implemented in a way that is indeed renegotiation-proof? We obtain, for environments with private values, a tight characterization of renegotiation-proof mechanisms: for complete information environments, this characterization is in terms of ex-post efficient decision rules; for incomplete information environments with independent private values, this characterization is in terms of Vickrey-Clarke-Groves (VCG) mechanisms. Importantly, we show 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.

^{*} Acknowledgements 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>.

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

1. Introduction

Mechanism design theory attempts to answer the question of when and how it is possible to design a game form (a mechanism) whose equilibrium outcomes are optimal with respect to some given criterion of social welfare. For the current mechanism design paradigm to be taken seriously as a model of institutional design, its conclusions must be robust. At least three different notions of robustness have been discussed in the literature: robustness against collusion, robustness against uncertainty about higher order beliefs, and robustness against renegotiation.¹ This paper is devoted to the subject of robustness against renegotiation.

If a proposed mechanism is renegotiated then it is impossible to ensure that it indeed achieves the goal it was designed to accomplish. Hence, if the objective of mechanism design theory is to suggest practicable (at least in principle) methods for achieving certain social goals, then renegotiation-proofness must be ensured.

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 before the mechanism is played, when each player knows only its own type, or after the mechanism is played, when each player knows both its 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 – that much is obvious. However, the possibility of ex-post renegotiation may have another, more subtle, if no less powerful, effect: namely, players may be induced to play the mechanism differently than they originally intended in anticipation of future renegotiation.

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. A mechanism that is both interim and ex-post renegotiation-proof is said to be renegotiation-proof.

The literature about renegotiation-proofness can thus be distinguished according to

¹The literature on robust mechanism design has become quite voluminous. The interested reader may consult 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 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.

whether it is about interim or ex-post renegotiation-proofness, and according to the assumptions that are imposed on the information and preferences of the players: complete information vs. independent private information vs. correlated private information; and private vs. interdependent valuations.

The literature on renegotiation-proofness under complete information (see Maskin and Moore (1999) and Segal and Whinston (2002)) has characterized the class of renegotiation-proof social choice rules relative to some exogenously given ad hoc “renegotiation function.”² This approach is in line with the standard approach in microeconomic theory, which is to assume that the players 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). In contrast, we focus on the case where the principal or designer of the mechanism is ignorant of the way renegotiation will take place, and so assumes that anything can happen.

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 focused its attention on the concept of interim renegotiation-proofness. The papers in this literature have each suggested a notion of interim renegotiation-proofness that is such that for any mechanism design problem, there exists a mechanism that is renegotiation-proof according to the notion that was proposed. The subject of ex-post renegotiation-proofness under incomplete information was examined by Forges (1993, 1994). Forges concluded that the question of whether there exists a renegotiation-proof mechanism for *every* mechanism design problem remains open (1994, p. 241).³

We present what we believe is a natural notion of renegotiation-proofness. The three main questions that are addressed in this paper are (1) what kind of environments admit the renegotiation-proof implementation of some social choice rules? (2) for a given environment, what kind of social choice rule are implementable in a way that is renegotiation-proof? and (3) for a given renegotiation-proof implementable social choice rule, how can the rule be implemented in a way that is indeed renegotiation-proof?

We obtain, for environments with private values, a tight characterization of renegotiation-proof mechanisms: for complete information environments, this characterization is in terms

²Of 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 designer of the mechanism.

³On 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.

of ex-post efficient decision rules; for incomplete information environments with independent private values, this characterization is in terms of Vickrey-Clarke-Groves (VCG) mechanisms. Importantly, we show that some common mechanism design problems *do not* admit the existence of *any* renegotiation-proof mechanism. The consideration of interdependent values and correlated types introduces considerable complications into our analysis. We provide a few results, illustrate some of the difficulties associated with this case, and point to a number of interesting open problems we have been unable to solve.

Our result about the possible inexistence of renegotiation-proof mechanisms may be interpreted as a sign that our notion of renegotiation is too permissive. Be that as it may, we believe that except for its technical contribution, our analysis also implies that more work needs to be devoted to understanding how renegotiation might be blocked and how it is actually blocked in different institutions in practice.

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

2. Set Up

A group of n players, indexed by $i \in N = \{1, 2, \dots, n\}$, 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, $t = (t_1, \dots, t_n) \in \mathbb{R}^n$. The players are allowed to arrange for monetary transfers among themselves in order to help them achieve their objectives, but we require that the sum of these monetary transfer be non-positive. Otherwise, the players may wish to renegotiate any decision just for the purpose of generating large transfer payments for themselves. A decision of the players (a, t) , or rather an outcome (a, t) of the process of 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. We denote $\Theta = \Theta_1 \times \dots \times \Theta_n$, and $\Theta_{-i} = \prod_{j \neq i} \Theta_j$, with typical elements θ and θ_{-i} , respectively. A profile of types $\theta \in \Theta$ is referred to as a state of the world.

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 its type and t_i denotes a possibly additional monetary transfer to player i .

If the environment is a complete information environment, then it is assumed that the

state of the world θ is commonly known among the players, although not necessarily by the mechanism designer.⁴ If the environment is an incomplete information environment, then it is assumed that each player i knows her type θ_i , and obtains her beliefs by conditioning the common prior, which represents the beliefs of the mechanism designer, on her own type.

A complete information *mechanism design environment* is thus fully described by a four-tuple $\langle N, A, \Theta, (u_i)_{i \in N} \rangle$. An incomplete information *mechanism design environment* is thus described by a five-tuple $\langle N, A, \Theta, P, (u_i)_{i \in N} \rangle$ where P denotes the common prior distribution over the set of states of the world Θ .

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 $m : S \rightarrow A \times \mathbb{R}^n$ from the set of message profiles $S = S_1 \times \dots \times S_n$ into the set of social alternatives A and monetary transfers \mathbb{R}^n .⁵ As mentioned above, we assume that the mechanism that is employed by the players must satisfy budget balance, or be such that the sum of monetary transfers to the players is non-positive, for any profile of messages that are sent by the players.

The combination of a mechanism $\langle S, m \rangle$ and a state of the world θ defines a complete information game $\langle N, S, (u_i(\cdot, \cdot, \theta) \circ m)_{i \in N} \rangle$. The combination of a mechanism $\langle S, m \rangle$ and a prior distribution over the states of the world P defines a Bayesian game $\langle N, S, \Theta, P, (u_i \circ m)_{i \in N} \rangle$. 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 $\sigma = (\sigma_1, \dots, \sigma_n)$.

A social choice rule is a mapping $f : \Theta \rightrightarrows A \times \mathbb{R}^n$ from the set of states of the world 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 equilibria outcomes that are induced by the mechanism belong to $f(\theta)$, for every $\theta \in \Theta$. We thus employ a weak notion of implementation.

We model the process of renegotiation in the following way: A third party proposes to the players an alternative decision, or mechanism, that they are likely to jointly prefer to the mechanism's decision, or to the original mechanism, respectively. If this third party is ever successful in inducing the players to jointly deviate from the mechanism's decision, or to reject the original mechanism in favor of the alternative mechanism, then we say that the mechanism is not ex-post or interim renegotiation-proof, respectively. If the third party can never induce the players to jointly agree to renegotiate the outcome or mechanism then we say that the mechanism is ex-post or interim renegotiation-proof, respectively.

⁴A mechanism designer who knows the state of the world can easily implement any social choice function she likes.

⁵We restrict attention to deterministic decision rules in order to simplify notation. Stochastic decision rules can be handled in a similar manner.

3. Renegotiation-Proofness Under Complete Information

3.1. 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 is realized and becomes commonly known among the players, but not known to the mechanism designer. Consider the case in which the state of the world is commonly known to be $\theta \in \Theta$. Consider a Nash equilibrium $\sigma = (\sigma_1, \dots, \sigma_n)$ 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, \dots, t_n) .

Suppose that the process of renegotiation assumes the following form: a different social alternative $a' \in A$, together with a profile of monetary transfers $t' = (t'_1, \dots, t'_n)$ that sum up to zero (or less) is exogenously proposed to the players instead of the outcome (a, t_1, \dots, t_n) that was obtained under the mechanism $\langle S, m \rangle$.⁶ If the players all agree to 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.

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.

Definition. A Nash equilibrium σ of the complete information game that is induced by a mechanism $\langle S, m \rangle$ when the state of the world is θ is said to be *ex-post renegotiation-proof* if for each player $i \in N$ there does not exist a strategy σ'_i and a feasible renegotiation proposal $(a', t') \in A \times \mathbb{R}^n$ where $\sum_{i=1}^n t'_i(\theta) \leq 0$ to renegotiate the outcome that is obtained after the players play the profile of strategies (σ'_i, σ_{-i}) that Pareto dominates the outcome that is obtained under the profile of strategies σ .

The definition implies that a Nash equilibrium σ is ex-post renegotiation-proof if and only if:

1. upon playing the Nash equilibrium strategy profile σ that generates the outcome (a, t) , there does not exist an alternative feasible outcome $(a', t') \in A \times \mathbb{R}^n$ that Pareto dominates (a, t) , and

⁶The outcome may involve a lottery. If so, ex-post renegotiation takes place before the lottery is carried through.

2. there does not exist an alternative feasible outcome (a', t') that, when anticipated by some player i , leads player i to deviate from the equilibrium σ in such a way that the outcome that is generated by the profile of strategies (σ'_i, σ_{-i}) is Pareto dominated by the alternative outcome (a', t') .

Definition. An equilibrium σ of the complete information game that is induced by a mechanism $\langle S, m \rangle$ in state $\theta \in \Theta$ is said to be *ex-post efficient* if the equilibrium outcome $(a(\theta), t(\theta))$ is such that:

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

The fact that an outcome that is not ex-post efficient can always be renegotiated to one that is in such a way that strictly benefits all the players implies the following obvious result, which is given without proof.

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

The fact that the renegotiation proposal is exogenous, and that the equilibrium must be immune to renegotiation given any alternative outcome, implies that our definition of renegotiation-proofness is strong. Just how strong is illustrated in the following example, which demonstrates 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 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 social alternatives consists of three alternatives: "no trade," "trade at the price 3," and "trade at the price 4." The mechanism, which in this context, may be thought of as a contract between the buyer and seller has to be designed before the buyer's valuation for the object becomes known, but it will be played after the buyer's valuation is realized and becomes commonly known between the buyer and seller.

Consider the following mechanism: the buyer announces whether she wants to trade or not. If she announces she 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 she declines to trade in the hope of renegotiating the outcome to trading at a price that is better for her. 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.⁷

Definition. A 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 $\theta \in \Theta$.

Remark. The difference between our notion of ex-post renegotiation-proofness and that of Maskin and Moore (1999) (and Segal and Whinston (2002)) is that Maskin and Moore define renegotiation-proofness with respect to a given specific renegotiation procedure $h : A \times \mathbb{R}^n \times \Theta \rightarrow A \times \mathbb{R}^n$ that maps an outcome and a state of the world into a possibly different outcome whereas we say that a mechanism is ex-post renegotiation-proof if it is renegotiation-proof with respect 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. 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 (i.e., such that $\sum_{i=1}^n t_i(\theta) = 0$ for every $\theta \in \Theta$), then there exists an incentive compatible, budget balanced, and ex-post renegotiation-proof mechanism that implements (a, t) .*

The idea of the proof of Proposition 1 is that when there are three or more players, then it is possible to use the report of player 2 to verify that player 1 is telling the truth, to use

⁷See Forges (1993, p. 142) and (1994, p. 260) for another example in which an ex-post efficient equilibrium can be renegotiated.

the report of player 1 to verify that player 2 is telling the truth, 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.

The fact that a decision rule that is not ex-post efficient can be easily renegotiated and that transfers that are not budget balanced are either infeasible or can also be easily renegotiated implies the following corollary:

Corollary. *Consider a complete information mechanism design environment with $n \geq 3$ players. A social choice function $(a, t) : \Theta \rightarrow A \times \mathbb{R}^n$ is implementable in a way that is ex-post renegotiation-proof if and only if a is ex-post efficient and t is budget balanced.*

When there are only two players, it is impossible to separate the provision of incentives for telling the truth from budget balance, which makes this case harder to analyze. We proceed to analyze this case under three 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 our 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.
2. **“Full Support”:** Every member of the set of states of the world $\Theta = \Theta_1 \times \dots \times \Theta_n$ is considered feasible ex-ante.
3. **“Unique Maximizer”:** For every state of the world $\theta \in \Theta$, there is a unique social alternative $a \in A$ that maximizes social welfare $\sum_{i=1}^n v_i(a, \theta)$.

We proceed to show that when there are only two players a mechanism is ex-post renegotiation-proof if and only if it is a VCG mechanism.

Definition. A mechanism $\langle S, (a, t) \rangle$ is a VCG 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 $a : \Theta \rightarrow A$ is ex-post efficient, and transfers $t_i : \Theta \rightarrow \mathbb{R}$, $i \in N$, 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 function $h_i : \Theta_{-i} \rightarrow \mathbb{R}$.

Proposition 2. *Consider a complete information mechanism design environment with two players. A budget balanced mechanism is ex-post renegotiation-proof if and only if it is a VCG mechanism.*

The intuition for Proposition 2 is the following. The possibility of renegotiation implies that a player can deviate from equilibrium in order to induce an inefficient decision and then renegotiate the outcome to one that is ex-post efficient 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 she 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 she generates up to a constant. Thus, Proposition 2 is a consequence of the fact that the class of mechanisms in which players' payoff are equal to the surplus they generate up to a constant is the class of VCG mechanisms.

Example 1 (continued). If the set of social alternatives in this example 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) under any mechanism 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 VCG payment for the buyer). It follows that if we add the constraint that in the event of no 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.

The next example demonstrates that when there are only two players, an ex-post renegotiation-proof mechanism may fail to exist.

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 alternatives consists of a continuum of alternatives: "no trade," and "trade at the price p ," where $p \in \mathbb{R}$. As in Example 1, the mechanism or contract has to be designed before the buyer's valuation and the seller's reservation value become known, but the mechanism would be played after the state of the world becomes commonly known between the buyer and seller.

Inspection of the proof of Proposition 2 reveals that the possibility of ex-post renegotiation implies that the buyer can ensure that she does not pay more than 2 when she buys the object compared to when she does not buy it, and the seller can ensure that the buyer pays at least 5 when she buys the object compared to when she does not buy it, respectively

(the two requirements are a consequence of the fact that the buyer's and seller's payments are VCG payments, respectively). Since these two requirements are inconsistent, it follows that there does not exist any ex-post renegotiation-proof mechanism for this environment.

3.2. 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 the play of 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.⁸

4. Renegotiation-Proofness Under Incomplete Information

4.1. Ex-Post Renegotiation-Proofness Under Incomplete Information

4.1.1. Definitions

There are many plausible to model the process of ex-post renegotiation. Our approach is parsimonious and at the same time sufficiently rich for our purposes. A third party that observes the decision that is made by the mechanism proposes to the players an alternative decision that they are likely to jointly prefer to the mechanism's decision. If this third party is ever successful in inducing the players to deviate from the mechanism's decision, then we say that the mechanism is not ex-post renegotiation-proof. If the third party can never induce the players to jointly agree to renegotiate the mechanism's decision, then we say that the mechanism is ex-post renegotiation-proof.

Specifically, suppose that a mechanism $\langle S, m \rangle$ is chosen and then played. After the mechanism has been played and produced a decision (a, t) , the players are informed of this decision. In addition, each player i knows its own type θ_i . Knowledge of the decision that was made by the mechanism may reveal to the players some information about other players' types, but the state of the world need not be commonly known among the players even though the decision that was made by the mechanism is.

We model the process of ex-post renegotiation by extending the game induced by the mechanism $\langle S, m \rangle$ in the following way. For every decision (a, t) that can be reached by a mechanism there is an exogenously determined alternative decision $\psi(a, t) = (a', t')$, i.e.,

$$\psi : \{(a, t) \in A \times \mathbb{R}^n \mid \exists \tilde{\sigma} \in S : m(\tilde{\sigma}) = (a, t)\} \rightarrow A \times \mathbb{R}^n.$$

⁸Segal and Whinston (2002) made the same observation with respect to their notions of interim and ex-post renegotiation proofness.

The players vote simultaneously whether to implement the original decision (a, t) or the alternative (a', t') . If all players vote in favor of the alternative outcome (a', t') , then it is implemented instead of (a, t) . Otherwise, the original decision (a, t) is implemented.⁹

Player i 's strategy in this ex-post renegotiation game is given by a strategy used in the mechanism $\sigma_i : \Theta_i \rightarrow \Delta S_i$, and a voting strategy that specifies for each reachable decision (a, t) and proposed alternative $\psi(a, t) = (a', t')$ a probability $\rho_i(a, t, \psi, \theta_i, s_i)$ that denotes the probability that player i votes to reject (a, t) in favor of the alternative $\psi(a, t)$ as a function of player i 's true type θ_i and reported message s_i .

Definition. A profile of strategies $(\sigma_i, \rho_i)_{i \in N}$ is a sequential equilibrium of the ex-post renegotiation game 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 $(\sigma_i, \rho_i)_{i \in N}$ into account.

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

Definition. An equilibrium σ of a mechanism $\langle S, m \rangle$ is said to be *ex-post renegotiation-proof against ψ* if the following two properties are satisfied:

1. Consider a collection of subgames (or voting games) that begin after the players have played σ . Then for every subgame of the ex-post renegotiation game starting with the voting stage between some reached decision (a, t) and the alternative decision $\psi(a, t) = (a', t')$ there is no equilibrium in which (i) the players change the reached decision in favor of the alternative with a positive probability; (ii) a nontrivial set of players' types strictly prefer to switch to the alternative.
2. For every equilibrium voting strategy profile $(\rho_i)_{i \in N}$ for the collection of subgames that begin after the players have played σ the following must be true. There does not exist a player j , such that there is a type θ_j of player j who strictly prefers some deviation

⁹The players are not allowed to change the decision (a, t) based on the new information that is revealed to them from the rejection of the alternative (a', t') .

strategy (σ'_j, ρ'_j) to the strategy (σ_j, ρ_j) , given that the opponents play a strategy profile $(\sigma_i, \rho_i)_{i \neq j}$.

Definition. 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 all feasible ψ .

The notion of ex-post renegotiation-proofness is quite strong, since it requires the mechanism to be robust to the possibilities of switching to all feasible sets of alternatives. Nevertheless, one may argue that it is not nearly strong enough because we do not allow the alternative proposals to depend on the private information of the players beyond what is revealed by mechanism's decision. Indeed, in realistic settings renegotiation proposals result from some communication process during which the players may choose to reveal some additional private information. We attempt to capture this feature by introducing a stronger notion of renegotiation-proofness, which we call "oracle renegotiation-proofness."

To capture this stronger notion of renegotiation-proofness we extend the game induced by the mechanism $\langle S, m \rangle$ in a different way. For every decision (a, t) that can be reached by a mechanism and for every state of the world θ there is an exogenously determined alternative decision $\widehat{\psi}(\theta, (a, t)) = (a', t')$, i.e.,

$$\widehat{\psi} : \Theta \times \{(a, t) \in A \times \mathbb{R}^n \mid \exists \tilde{\sigma} \in S : m(\tilde{\sigma}) = (a, t)\} \rightarrow A \times \mathbb{R}^n.$$

Notice that now the alternative decision may reveal some extra information about the true state of the world in addition to what is revealed by the outcome of the mechanism. As before, if all players vote in favor of the alternative outcome then it is implemented, and the original decision (a, t) is implemented otherwise.

Definition. An equilibrium σ of a mechanism $\langle S, m \rangle$ is said to be *ex-post oracle renegotiation-proof against $\widehat{\psi}$* if the following two properties are satisfied:

1. Consider a collection of subgames (or voting games) that begin after the players have played σ . Then for every subgame of the ex-post renegotiation game starting with the voting stage between some reached decision (a, t) and the alternative decision $\widehat{\psi}(\theta, (a, t)) = (a', t')$ there is no equilibrium in which (i) the players change the reached decision in favor of the alternative with a positive probability; (ii) a nontrivial set of players' types strictly prefer to switch to the alternative.
2. For every equilibrium voting strategy profile $(\rho_i)_{i \in N}$ for the collection of subgames that begin after the players have played σ the following must be true. There does not exist a player j , such that there is a type θ_j of player j who strictly prefers some deviation

strategy (σ'_j, ρ'_j) to the strategy (σ_j, ρ_j) , given that the opponents play a strategy profile $(\sigma_i, \rho_i)_{i \neq j}$.

Definition. 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 all feasible $\widehat{\psi}$.

Thus 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. Importantly, 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 private information beyond what is revealed by the outcome of the mechanism. We conjecture that it is possible to show (at least for the settings with 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, according to which the players communicate with each other when deciding on an alternative proposal.

Another justification for the oracle device is the fact that in some realistic settings the state of the world may become commonly known at the ex-post stage. Thus to assure stability of the 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 3. 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 her reservation value. The set of social alternatives consists of three alternatives: “no trade,” “trade at price 1,” and “trade at price 2”. Consider the following mechanism: the buyer announces her value. If she announces the value 0, then there is no trade; if she 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 to the buyer whose valuation is 3 is 1. The payoff to a buyer with valuation 3 from announcing that her 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 the 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 she announces that its 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 $\frac{1}{2} \cdot 2 + \frac{1}{2} \cdot 1 = \frac{3}{2} > 1$.

4.1.2. The Case of Independent Private Values

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 lemma is straightforward if we use the notion of ex-post oracle renegotiation-proofness. The point that is made by the lemma is that it is enough to require ex-post renegotiation-proofness.

The next example demonstrates that the converse of the Lemma does not hold. Namely, an ex-post efficient mechanism may fail to be ex-post renegotiation-proof.

Example 4. 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 social alternatives 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 she wants to trade or not. If she announces it 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 undominated 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 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 if they so wish. A buyer who values the object at 5 and who anticipates such a renegotiation possibility might announce that she declines to trade in the hope of renegotiating the outcome to trading at a price that is better for her. 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 – 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.

The next example goes a step further by demonstrating that ex-post renegotiation-proof mechanisms may altogether fail to exist.

Example 5. 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 alternatives consists of a continuum of alternatives: “no trade,” and “trade at the price p ,” where $p \in [2, 5]$.

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 4 above, that the possibility of ex-post renegotiation implies that the buyer can ensure that she does not pay more than 2 when she buys the object, and the seller can ensure that the buyer pays at least 5 when she buys the object, respectively. Since these two claims are inconsistent, it follows that there does not exist any ex-post renegotiation-proof mechanism for this environment.

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.

Proposition 3. *Consider an incomplete information mechanism design environment with independent private values. In such an environment, a feasible mechanism is ex-post oracle renegotiation-proof if and only if it is a VCG in expectation mechanism.*

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

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

for some constant $H_i \in \mathbb{R}$. Namely, the expected payment of each player i as a function of its type is equal to the expected payment to the player as a function of its type under some VCG mechanism.¹⁰

¹⁰The class of mechanisms that are VCG in expectation includes VCG mechanisms (after Vickrey (1961), Clarke (1971), and Groves (1973)), AGV mechanisms (after Arrow, 1979, and d’Aspremont and Gerard-Varet, 1979), as well as other mechanisms.

The intuition for the proof is the following. The way we defined the process of renegotiation implies that a player can always misrepresent her 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 it generates up to a constant. It therefore follows that the mechanisms in which players already get the surplus or externality they generate are ex-post renegotiation-proof, and conversely, any mechanisms that is ex-post renegotiation-proof must be a mechanism in which each player obtains a payoff that is equal to the surplus it 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 surplus they generate is the class of mechanisms that are VCG in expectation. It is the class of mechanisms that are VCG in expectation rather than just VCG because the players contemplate how best to misrepresent their types at the interim stage, which implies that the expected transfer to each player has to be equal to the expected externality that is generated by the player.

Remark. 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 then any mechanism that is both ex-post efficient and Bayesian incentive compatible is payoff equivalent to a VCG mechanism at the interim stage. When this equivalence holds, Proposition 3 implies that there exists a feasible ex-post oracle renegotiation-proof mechanism if and only if there exists a feasible, ex-post efficient, Bayesian incentive compatible, direct revelation mechanism. The fact that for several economically important mechanism design problems, such as bilateral trade, regulation, and litigation and settlement, no feasible ex-post efficient mechanisms exists implies that no ex-post renegotiation-proof mechanisms exist in such mechanism design problems either.

4.1.3. The Case of Correlated Private Values

When there are at least three players with correlated types, then we believe that the technique of Crémer and McLean (1985, 1988) can be adapted to establish the existence of an ex-post oracle renegotiation-proof mechanism that implements any ex-post efficient decision rule. The idea is that in order to induce player i to reveal its type truthfully, it is possible to “stochastically compare” its 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 a strong incentive to report the truth while maintaining budget balanced. 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.

We do not know yet what rules can be implemented in a way that is ex-post renegotiation-proof when there are only two players. It is a difficult problem, and it is possible that there is no “elegant” characterization for this case.

Another interesting question is what can be implemented in a way that is ex-post renegotiation-proof by a mechanism that is also robust according to another robustness criterion, such as robustness against higher order beliefs.¹¹

4.1.4. The Case of Interdependent Values

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 decision (a', t') 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 Lemma 2 that appears in Section 4.1.2 may not hold when players have interdependent valuations. Namely, in such a case a mechanism may not be ex-post efficient but still be ex-post renegotiation-proof.

Example 6. 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 undesirable for player 2). Such a mechanism is not ex-post efficient in state b .

¹¹Indeed, Heifetz and Neeman (2006) have shown that the type of arguments used by Crémer and McLean are non robust or “non-generic” according to this criterion.

Nevertheless, it is possible to show that ex-post oracle renegotiation-proofness implies ex-post efficiency, also in the case where players have interdependent valuations.

We also obtained a characterization result for the ex-post oracle renegotiation-proof mechanisms for environments with interdependent valuations and independently distributed private information analogous to Proposition 3 which appears in Section 4.1.2. The characterization in this case is less "elegant" than the one for independent private valuations, and due to space limitations we just informally describe the results. The fact that we use the notion of ex-post oracle renegotiation-proofness implies that the decisions must be ex-post efficient. The restrictions on the payments come from the requirement that no player should benefit from misrepresenting her type, and then renegotiating to an ex-post efficient outcome and capturing the difference in the social surplus.

As an illustration consider a single-good auction environment where bidders have one-dimensional signals and a single-crossing condition is satisfied. We can show that if there are two bidders then a generalized Vickrey auction (see for example Dasgupta and Maskin (2000)) is ex-post oracle renegotiation-proof. Interestingly, this is not necessarily true when there are more than two bidders. Assume that it is never efficient to allocate a good to bidder i , but the knowledge of her signal is essential for an efficient allocation of the good among the rest of the bidders. In a generalized Vickrey auction the bidder i 's payment is zero. However, bidder i may wish to misrepresent her signal to force an inefficient allocation, and later renegotiate to an efficient allocation and capture the difference in the social surplus.

4.2. Interim Renegotiation-Proofness Under Incomplete Information

To be added.

4.3. Renegotiation-Proofness Under Incomplete Information

To be added.

Appendix

Proof of Proposition 1.

Fix an ex-post efficient decision rule a and a budget balanced vector of transfers $t : \Theta \rightarrow \mathbb{R}^n$. 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 $(\hat{\theta}_1, \dots, \hat{\theta}_n)$ as follows:

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

where $a_0 \in A$ is some fixed social alternative, and the constant M is chosen such that $M > \max_{i \in N, a \in A, \theta \in \Theta} |v_i(a, \theta) + t_i(\theta)|$. 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 VCG 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 her type is θ'_i and then renegotiating the outcome from $(a(\theta'_i, \theta_j), t(\theta'_i, \theta_j))$ to $(a(\theta_i, \theta_j), \widehat{t}(\theta_i, \theta_j))$ 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 herself, 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

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

if and only if

$$\widehat{t}_i(\theta_i, \theta_j) > t_i(\theta_i, \theta_j). \quad (1)$$

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

$$\widehat{t}_j(\theta_i, \theta_j) \geq v_j(a(\theta'_i, \theta_j), \theta_j) - v_j(a(\theta_i, \theta_j), \theta_j) + t_j(\theta'_i, \theta_j). \quad (2)$$

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 (2) implies:

$$\widehat{t}_i(\theta_i, \theta_j) \leq v_j(a(\theta_i, \theta_j), \theta_j) - v_j(a(\theta'_i, \theta_j), \theta_j) + t_i(\theta'_i, \theta_j) \quad (3)$$

The fact that $\langle a, t \rangle$ is a VCG 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 $h_i : \Theta_j \rightarrow \mathbb{R}$. Plugging the two equations above into (3) 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 (1).

<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 VCG 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

$$S(\theta_1, \theta_2) = v_1(a(\theta, \theta), \theta_1) + v_2(a(\theta, \theta), \theta_2) \tag{4}$$

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

$$S(\theta_1, \theta'_2) \geq v_1(a(\theta, \theta'), \theta_1) + v_2(a(\theta, \theta'), \theta'_2) \tag{5}$$

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 $(a(\theta, \theta), \widehat{t}(\theta, \theta))$ 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:

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

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, she concludes that this cannot increase her expected payoff, or:

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

or

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

for every $\theta' \neq \theta$. Because $t_1(\theta', \theta) + t_2(\theta', \theta) = 0$, we have that

$$t_1(\theta, \theta) - t_1(\theta', \theta) \geq -v_2(a(\theta', \theta), \theta_2) + v_2(a(\theta, \theta), \theta_2) \quad (6)$$

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

$$t_2(\theta, \theta) - t_2(\theta, \theta') \geq -v_1(a(\theta, \theta'), \theta_1) + v_1(a(\theta, \theta), \theta_1) \quad (7)$$

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

Adding (6) and (7) together and using budget balance implies that:

$$-t_1(\theta', \theta) - t_2(\theta, \theta') \geq -v_2(a(\theta', \theta), \theta_2) + v_2(a(\theta, \theta), \theta_2) - v_1(a(\theta, \theta'), \theta_1) + v_1(a(\theta, \theta), \theta_1) \quad (8)$$

and by switching θ and θ' also:

$$-t_1(\theta, \theta') - t_2(\theta', \theta) \geq -v_2(a(\theta, \theta'), \theta'_2) + v_2(a(\theta', \theta'), \theta'_2) - v_1(a(\theta', \theta), \theta'_1) + v_1(a(\theta', \theta'), \theta'_1) \quad (9)$$

Adding (8) and (9) together, using budget balance, 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, (5) 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:

$$S(\theta'_1, \theta'_2) + S(\theta_1, \theta_2) = S(\theta_1, \theta'_2) + S(\theta'_1, \theta_2). \quad (10)$$

Hence, all the possible inequalities in (5) must hold as equalities. For the decision rule $a : \Theta^N \rightarrow A$ this implies that:

$$\begin{aligned} a((\theta_1, \theta_2), (\theta'_1, \theta'_2)) &\in \operatorname{argmax}_{a \in A} \{v_1(a, \theta_1) + v_2(a, \theta'_2)\} \\ a((\theta'_1, \theta'_2), (\theta_1, \theta_2)) &\in \operatorname{argmax}_{a \in A} \{v_1(a, \theta'_1) + v_2(a, \theta_2)\} \\ a((\theta'_1, \theta_2), (\theta_1, \theta'_2)) &\in \operatorname{argmax}_{a \in A} \{v_1(a, \theta'_1) + v_2(a, \theta'_2)\} \\ a((\theta_1, \theta'_2), (\theta'_1, \theta_2)) &\in \operatorname{argmax}_{a \in A} \{v_1(a, \theta_1) + v_2(a, \theta_2)\} \end{aligned}$$

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 $\theta_2, \theta'_2 \in \Theta_2$,

$$a((\theta_1, \theta'_2), (\theta'_1, \theta_2)) = a((\theta_1, \theta_2), (\theta_1, \theta_2)). \quad (11)$$

Furthermore, the fact that (10) holds as an equality implies that all the inequalities that were used to generate it hold as equalities as well. In particular, (6) and (7) must hold as equalities, or

$$\begin{aligned} &t_1((\theta_1, \theta_2), (\theta_1, \theta_2)) - t_1((\theta'_1, \theta'_2), (\theta_1, \theta_2)) \\ &= -v_2(a((\theta'_1, \theta'_2), (\theta_1, \theta_2)), \theta_2) + v_2(a((\theta_1, \theta_2), (\theta_1, \theta_2)), \theta_2) \\ &= -v_2(a((\theta'_1, \cdot), (\cdot, \theta_2)), \theta_2) + v_2(a((\theta_1, \cdot), (\cdot, \theta_2)), \theta_2) \end{aligned}$$

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

$$\begin{aligned} &t_2((\theta_1, \theta_2), (\theta_1, \theta_2)) - t_2((\theta_1, \theta_2), (\theta'_1, \theta'_2)) \\ &= -v_1(a((\theta_1, \theta_2), (\theta'_1, \theta'_2)), \theta_1) + v_1(a((\theta_1, \theta_2), (\theta_1, \theta_2)), \theta_1) \\ &= -v_1(a((\theta_1, \cdot), (\cdot, \theta'_2)), \theta_1) + v_1(a((\theta_1, \cdot), (\cdot, \theta_2)), \theta_1) \end{aligned}$$

for every $\theta' \neq \theta$, where in both cases, the second equality follows from (11). 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

$$t_i((\theta_1, \theta'_2), (\theta'_1, \theta_2)) = t_i((\theta_1, \theta_2), (\theta_1, \theta_2)). \quad (12)$$

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, (6) and (7) 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 VCG transfer payments. ■

Example 1 (continued).

We show that the buyer is the one who determines if the trade takes place, and pays exactly 2 more than she 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 :

$B \backslash S$	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 $\theta = (\theta_1, \dots, \theta_n)$ such that $m(\sigma(\theta)) = (a, t)$, and a feasible alternative decision $(a', t') \in A \times \mathbb{R}^n$ such that

$$v_i(a', \theta_i) + t'_i \geq v_i(a, \theta_i) + t_i \tag{13}$$

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 (13) 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 $\theta_i, i \in N$ conditional on switching to (a', t') and so does not disturb the equilibrium.) ■

Proof of Proposition 3.

<IF> Let $\langle a, t \rangle$ be a budget balanced VCG 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 her type is θ'_i and then, for every $\theta_{-i} \in \Theta_{-i}$, renegotiating the outcome from $(a(\theta'_i, \theta_{-i}), t(\theta'_i, \theta_{-i}))$ to $(a(\theta_i, \theta_{-i}), \hat{t}(\theta_i, \theta_{-i}))$ where \hat{t} is some ex-post budget balanced transfer function (by renegotiating the outcome to the ex-post efficient outcome $a(\theta_i, \theta_{-i})$, θ_i is able to capture the greatest possible surplus for herself, and so would prefer that to any other outcome $a \in A$; the transfers \hat{t} facilitate this renegotiation).

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

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

if and only if

$$E_{\theta_{-i}} [\hat{t}_i(\theta_i, \theta_{-i})] > E_{\theta_{-i}} [t_i(\theta_i, \theta_{-i})]. \quad (14)$$

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

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

or

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

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

$$\sum_{j \neq i} \hat{t}_j(\theta_i, \theta_{-i}) \geq \sum_{j \neq i} v_j(a(\theta'_i, \theta_{-i}), \theta_j) - \sum_{j \neq i} v_j(a(\theta_i, \theta_{-i}), \theta_j) + \sum_{j \neq i} t_j(\theta'_i, \theta_{-i}). \quad (15)$$

The fact that both t and \hat{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 $\sum_{j \neq i} \hat{t}_j(\theta_i, \theta_{-i}) = -\hat{t}_i(\theta_i, \theta_{-i})$. Plugging these two equations into (15) implies:

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

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

$$E_{\theta_{-i}} [\widehat{t}_i(\theta_i, \theta_{-i})] \leq E_{\theta_{-i}} \left[\sum_{j \neq i} v_j(a(\theta_i, \theta_{-i}), \theta_j) \right] + E_{\theta_{-i}} [t_i(\theta'_i, \theta_{-i})] - E_{\theta_{-i}} \left[\sum_{j \neq i} v_j(a(\theta'_i, \theta_{-i}), \theta_j) \right]. \quad (16)$$

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

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

and

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

for some constant H_i . Plugging the two equations above into (16) it follows that:

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

A contradiction to (14).

<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 VCG in expectation mechanism.

Type $\theta_i \in \Theta_i$ of player i can report she 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 $(a(\theta_i, \theta_{-i}), \widehat{t}(\theta_i, \theta_{-i}))$ 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

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

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

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

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 she 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 $\theta_i \in \Theta_i$:

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

where

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

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

References

- 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.
- 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.
- 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.
- Chakravorty, B., Corchon, L. and S. Wilkie (2006) "Credible Implementation," *Games and Economic Behavior* 57, 18-36.
- Che, Y. K., and J. Kim (2006) "Robustly Collusion-Proof Implementation" *Econometrica* 74, 1063-1107.
- 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.
- Cremer, J. and R. McLean (1985) "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands are Interdependent," *Econometrica* 53, 345-362.
- Cremer, 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 (1987) "Posterior Implementability in a 2-Person Decision Problem," *Econometrica* 55, 69-94.

- Groves, T. (1973) "Incentives in Teams," *Econometrica* 41, 617-631.
- 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.
- Jehiel, P., M. Meyer-ter-Vehn, B. Moldovanu, and W. R. Zame (2006) "The Limits of Ex-Post Implementation," *Econometrica* 74, 585-610.
- Krasa, S. (1999) "Unimprovable Allocations in Economies with Incomplete Information," *Journal of Economic Theory* 87, 144-168.
- Lagunoff, R. D. (1995) "Resilient Allocation Rules for Bilateral Trade," *Journal of Economic Theory* 66, 463-487.
- Maskin, E. and J. Moore (1999) "Implementation and Renegotiation," *Review of Economic Studies* 66, 39-56.
- 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.
- 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.
- 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.
- Vickrey, W. (1961) "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance* 16, 8-37.
- Williams, S. R. (1999) "A Characterization of Efficient, Bayesian Incentive Compatible Mechanisms," *Economic Theory* 14, 155-180.