

Renegotiation-Proof Mechanism Design^{*}

Zvika Neeman[†] and Gregory Pavlov[‡]

January 15, 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.

^{*} 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>.

[‡] Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215; Email gpavlov@BU.edu

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.

A standard approach in microeconomic theory 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. If a proposed mechanism is renegotiated in a way that cannot be foreseen precisely, 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 not less powerful, effect, which is that 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,

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.

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 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 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

In this paper we present what we believe is a natural and yet very strong 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 which 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.

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 indeed 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, $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ that are such that $\sum_{i=1}^n t_i \leq s(a)$. The function $s : A \rightarrow \mathbb{R}$ may be interpreted as the monetary surplus that is generated by the social alternative a that has to be shared by the players, or more naturally, when $s(a)$ is negative, as the monetary cost of implementing social alternative a . Sometimes we refer to a decision (a, t) as the outcome (a, t) .

The players' preferences over the set A 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 quasilinear payoff function that is given by $u_i(a, t_i, \theta) = v_i(a, \theta) + t_i$ where $v_i : A \times \Theta \rightarrow \mathbb{R}$. 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

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

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 five-tuple $\langle N, A, s, \Theta, (u_i)_{i \in N} \rangle$. An incomplete information *mechanism design environment* is thus described by a six-tuple $\langle N, A, s, \Theta, P, (u_i)_{i \in N} \rangle$ 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

The combination of a mechanism $\langle S, m \rangle$ and a state of the world θ defines a complete information game $\langle N, S, (u_i(\cdot, \theta, \cdot) \circ m)_{i \in N} \rangle$. 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 $\sigma = (\sigma_1, \dots, \sigma_n)$ and denote the outcome of this equilibrium, when the state of the world is θ , by $(a(\theta), t(\theta))$.

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, 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.

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. Definitions

If the state of the world θ is commonly known, then feasibility considerations imply that we may restrict our attention to mechanisms that are such that

$$\sum_{i=1}^n t_i(\theta) \leq s(a(\theta)).$$

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 $(a(\theta), t_1(\theta), \dots, t_n(\theta))$ is such that:

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

3.2. 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) .⁶

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'_1, \dots, t'_n) that sum up to $s(a')$ or less is exogenously proposed to the players instead of the Nash equilibrium outcome (a, t_1, \dots, t_n) . 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_1, \dots, t_n) is implemented.

We assume that if the outcome (a', t'_1, \dots, t'_n) Pareto dominates the outcome (a, t_1, \dots, t_n) , 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_1, \dots, t_n) is renegotiated to the new outcome (a', t'_1, \dots, t'_n) . Otherwise, the original outcome (a, t_1, \dots, t_n) is implemented.

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

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 *ex-post renegotiation proof* if for each player $i \in N$ there does not exist a strategy σ'_i and a feasible renegotiation proposal $(a', t'_1, \dots, t'_n) \in A \times \mathbb{R}^n$ that Pareto dominates the outcome that is obtained under the profile of strategies (σ'_i, σ_{-i}) .

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 decision $(a', t') \in A \times \mathbb{R}^n$ that Pareto dominates (a, t) , and
2. there does not exist an alternative feasible decision (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') .

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.

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, 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 it wants to trade or not. If it announces it 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 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 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 strategy 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 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.^{7,8}

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 $h : A \times \Theta \rightarrow A$ 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 therefore stronger, and is satisfied by fewer mechanisms.

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

⁸If the set of social alternatives in this example is expanded to allow for trade at any price, then it is possible to show by using the Revelation Principle that any profile of 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. 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. See the appendix for details.

Another difference between our framework and the one considered by Maskin and Moore (1999) is that they considered the incentives that a mechanism provides for exerting costly effort. We believe that it is possible to address the question of whether it is possible to provide incentives for exerting costly effort in our framework as well.

The following proposition provides a characterization of ex-post renegotiation proof mechanisms for the case where the number of players $n \geq 3$.

Theorem 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 $\sum_{i=1}^n t_i(\theta) = s(a(\theta))$ 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, the report of player 1 to verify that player 2 is telling the truth, and use player 3 as a budget breaker. We thus have 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 much harder to analyze. We have only some partial characterization results for this case, together with the following example, which demonstrates that an ex-post renegotiation proof mechanism may fail to exist under these circumstances.

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 $p \in [2, 5]$. 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.

The proof of this is a little involved (see the analysis of Example 1 in the appendix), but it can be shown that the possibility of ex-post renegotiation implies that the buyer can ensure

that it does not pay more than 2 when it buys the object, and the seller can ensure that the buyer pays at least 5 when it 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.

Conjecture. *Consider a complete information mechanism design environment with two players. In such an environment, a mechanism is ex-post renegotiation proof if and only if it is a VCG mechanism.*

3.3. 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 follows that any equilibrium that the players would want to renegotiate ex-ante 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.⁹

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'

⁹Segal and Whinston (2002) made the same observation with respect to their notions of interim and ex-post renegotiation proofness.

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.,

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

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.¹⁰

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:

1. There exists an equilibrium of the ex-post renegotiation game where the players follow strategies prescribed by σ in the mechanism and do not change any reached decision

¹⁰The 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') .

in favor of the proposed alternative. In this equilibrium every type θ_i , who strictly prefers the alternative decision $\psi(a, t) = (a', t')$ as opposed to the original decision (a, t) conditional on the other players voting in favor of the alternative, votes in favor of the alternative decision (a', t') with probability 1.

2. 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.

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 every 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:

1. There exists an equilibrium of the ex-post renegotiation game where the players follow strategies prescribed by σ in the mechanism and do not change any reached decision

in favor of the proposed alternative. In this equilibrium every type θ_i , who strictly prefers the alternative decision $\widehat{\psi}(\theta, (a, t)) = (a', t')$ as opposed to the original decision (a, t) conditional on the other players voting in favor of the alternative, votes in favor of the alternative decision (a', t') with probability 1.

2. 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.

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. We will investigate this issue in the further research.

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 decisions is given by $A = \{\text{“no trade,” “trade at price 1,”}$

“trade at price 2”}. Consider the following mechanism: the buyer announces its 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 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 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 main difficulty in the analysis of the ex-post renegotiation game comes from the fact that at the voting stages the players compare their payoffs evaluated at the alternative decision with their payoffs evaluated at the original decision (a, t) conditional on the other players voting in favor of the alternative. However, the analysis of the case of independent private values is considerably simpler since 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.*¹¹

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

¹¹This Lemma holds true also in the case of complete information environments, but it fails to hold when players have interdependent valuations.

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 2 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 theorem 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.

Theorem 2. *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 VCG in expectation.*

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.¹²

Space limitations prevent us from providing the proof of the Theorem. 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, the Theorem 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, the Theorem implies that there exists a feasible ex-post oracle renegotiation-proof mechanism

¹²The 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.

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.¹³

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.

¹³Indeed, 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.

The next example illustrates some of the difficulty by showing that the Lemma that appears in Section 3.2.2 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 .

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 the Theorem which appears in Section 3.2.2. 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, the 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

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 a mechanism is interim renegotiation-proof if it is never renegotiated for whatever rational beliefs that the players may 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 $\sigma = (\sigma_1, \dots, \sigma_n)$. Suppose that an alternative mechanism $\langle S', m' \rangle$ 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 $\langle S, m \rangle$. 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 $\rho_i : \Theta_i \rightarrow [0, 1]$ 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 $\sigma_i : \Theta_i \rightarrow \Delta S_i$ used in the mechanism $\langle S, m \rangle$ if it is retained; (iii) a strategy $\sigma'_i : \Theta_i \rightarrow \Delta S'_i$ used in the mechanism $\langle S', m' \rangle$ if it replaces $\langle S, m \rangle$.

Definition. 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. 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; (ii) a nontrivial set of players' types strictly prefer to switch to the alternative mechanism $\langle S', m' \rangle$.

Remark. 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 say that a mechanism is “durable” if for every alternative mechanism there is a (non trivial) voting equilibrium in which this alternative mechanism is rejected. In contrast, we say that a mechanism is interim renegotiation-proof if every alternative mechanism is rejected *in every equilibrium*. The fact that coordination problems may imply that there is a (non trivial) equilibrium rejection of every alternative implies that our definition is strictly stronger than that of Holmström and Myerson.

This is still a little tentative at this stage, but we nevertheless state the following conjecture.

Conjecture. *Consider an incomplete information mechanism design environment. In such an environment, a dominant strategy ex-post efficient equilibrium is interim renegotiation-proof.*

4.3. Renegotiation-Proofness Under Incomplete Information

Recall that a mechanism is said to be renegotiation-proof if and only if it is both ex-post and interim renegotiation-proof. The Theorem we presented in Section 3.2.2 and the conjecture presented in Section 3.3, if true, together imply that in independent private values environments any dominant strategy equilibrium that is ex-post efficient is both ex-post and interim renegotiation-proof and hence also renegotiation-proof. In particular, since a VCG mechanism has a dominant strategy equilibrium and is ex-post renegotiation-proof, it is also renegotiation-proof in independent private values environments. It therefore follows that in such environments, an incomplete information mechanism design problem admits the existence of a renegotiation-proof mechanism if and only if it admits the existence of a feasible VCG mechanism.

Appendix

Proof of Theorem 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 alternatives, 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 vector of transfers (a, t) . ■

Example 1 (continued). We show that the buyer is the one who determines if trade takes place, and pays exactly 2 more than it pays if there is no trade.

By the revelation principle (for games with complete information), any mechanism can be described as:

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

where q denotes the probability of trade and p denotes the buyer's payment. Permit lotteries but then ex-post renegotiation takes place *before* the lottery.

WTS $p_{5,5} - p_{1,1} = 2$; $q_{1,1} = q_{1,5}$; $p_{1,1} = p_{1,5}$; $q_{5,5} = q_{5,1}$; and $p_{5,5} = p_{5,1}$.

RP implies ExpEfficiency or

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

RP implies (notice that these constraints are stronger than IC)

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

In $B5$ buyer renegotiates if there's no trade and captures entire surplus; in $S5$ seller renegotiates if there's no trade and captures entire surplus. In $B1$ buyer renegotiates if there is trade but captures a surplus of only 1; in $S1$ seller renegotiates if there's trade, and captures surplus of 1)

plugging in, we get

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

combining $B5$ and $S1$ get:

$$2 - q_{1,5} \geq p_{5,5} - p_{1,1}$$

combining $B1$ and $S5$ get:

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

from which it follows that

$$2 - q_{1,5} \geq 5 - 3q_{5,1}$$

Since $0 \leq q_{1,5}, q_{5,1} \leq 1$, this can only be if $q_{1,5} = 0$ and $q_{5,1} = 1$, which, in turn implies that $p_{5,5} - p_{1,1} = 2$. I.e., the seller cannot affect the probability of trade, and the difference in the price paid by the buyer when there is trade and not is 2. We did not impose individual rationality, but if we do, and assume that the buyer pays nothing if there is no trade then the resulting mechanism is: ...

It therefore follows that in the example where buyers values are either 1 or 5, and seller's values are either 2 or 6, there exists no renegotiation proof mechanism. Analysis for the case where seller's cost is 2 yields the result above; symmetric analysis for the case where the buyer's valuation is 5 implies that it is the seller who determines probability of trade and price difference is 5 but this cannot be ... (IC implies that payment from buyer to seller should be the same when there's no trade, regardless of the types. Otherwise, traders would simply misreport their types. Suppose payment is x . Argument for buyer implies that when there is trade, it's at an expected price of $x + 2$. Argument for seller implies it's at an expected price $x + 5$. A contradiction)

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 $d \in \mathcal{D}$, a profile of types $t = (t_1, \dots, t_n)$ such that $m(\sigma(t)) = d$, and an alternative decision $a \in \mathcal{D}$ such that

$$u_i(a, t_i) \geq u_i(d, t_i) \tag{3}$$

for every type $t_i, i \in N$, with one strict inequality. We show that the ex-post renegotiation game $\langle S, m, \sigma, d, a \rangle$ has a sequential equilibrium in which the players all vote in favor of the alternative decision a with a positive probability. Inequality (3) implies that there exists an equilibrium in which the types $t_i, i \in N$, all vote for the alternative a 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 in this equilibrium, this does not affect the payoff of the types $t_i, i \in N$ conditional on switching to a and so does not disturb the equilibrium.) ■

Proof of Theorem 2. Let $\langle x, t \rangle$ be a mechanism that is VCG in expectation, individually rational, and ex-post budget balanced. We show that $\langle x, t \rangle$ is ex-post renegotiation proof.

Suppose that $\langle x, t \rangle$ is not ex-post renegotiation proof. We show that this leads to a contradiction. The fact that x 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 it is type θ'_i and then, for any $\theta_{-i} \in \Theta_{-i}$, renegotiating the outcome from $(x(\theta'_i, \theta_{-i}), t(\theta'_i, \theta_{-i}))$ to $(x(\theta_i, \theta_{-i}), \widehat{t}(\theta_i, \theta_{-i}))$ where \widehat{t} is some ex-post budget balanced transfer function (by renegotiating the outcome to the ex-post efficient outcome $x(\theta_i, \theta_{-i})$, θ_i is able to capture the greatest possible surplus for itself, and so would prefer that to any other outcome $x \in X$; the transfers \widehat{t} facilitate this renegotiation).

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

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

if and only if

$$E_{\theta_{-i}} [\widehat{t}_i(\theta_i, \theta_{-i})] > E_{\theta_{-i}} [t_i(\theta_i, \theta_{-i})]. \quad (1)$$

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

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

or

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

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

$$\sum_{j \neq i} \widehat{t}_j(\theta_i, \theta_{-i}) \geq \sum_{j \neq i} v_j(x(\theta'_i, \theta_{-i}), \theta_j) - \sum_{j \neq i} v_j(x(\theta_i, \theta_{-i}), \theta_j) + \sum_{j \neq i} t_j(\theta'_i, \theta_{-i}). \quad (2)$$

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 (2) implies:

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

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

$$E_{\theta_{-i}} [\widehat{t}_i(\theta_i, \theta_{-i})] \leq E_{\theta_{-i}} \left[\sum_{j \neq i} v_j(x(\theta_i, \theta_{-i}), \theta_j) \right] + E_{\theta_{-i}} [t_i(\theta'_i, \theta_{-i})] - E_{\theta_{-i}} \left[\sum_{j \neq i} v_j(x(\theta'_i, \theta_{-i}), \theta_j) \right]. \quad (3)$$

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

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

and

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

for some constant H_i . Plugging the two equations above into (3) 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 (1).

Let $\langle x, t \rangle$ be a direct revelation mechanism that is incentive compatible, individually rational, ex-post budget balanced, and ex-post renegotiation proof. We show that $\langle x, t \rangle$ is VCG in expectation.

Type $\theta_i \in \Theta_i$ of player i can report it is type $\theta'_i \in \Theta_i$ and then offer to renegotiate the outcome from $(x(\theta'_i, \theta_{-i}), t(\theta'_i, \theta_{-i}))$ to $(x(\theta_i, \theta_{-i}), \widehat{t}(\theta_i, \theta_{-i}))$ 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 $\theta_{-i} \in \Theta_{-i}$:

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

or

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

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

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

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

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

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

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

or

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

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

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

for every $\theta_i, \theta'_i \in \Theta_i$. Because type $\theta'_i \in \Theta_i$ of player i can report it 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:

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

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

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

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 $\theta \in \Theta$:

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

where

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

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

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.
- 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.