KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
WORKING PAPER SERIES

COMMITMENT WITHOUT REPUTATION: RENEGOTIATION-PROOF CONTRACTS UNDER ASYMMETRIC INFORMATION

Emanuele Gerratana
Levent Koçkesen

Working Paper 1323
December 2013

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
Rumelifeneri Yolu 34450 Sarıyer/Istanbul

Commitment without Reputation: Renegotiation-Proof Contracts under Asymmetric Information*

Emanuele Gerratana
SIPA, Columbia University

Levent Koçkesen†
Koç University

October 5, 2013

Abstract

This paper characterizes equilibrium outcomes of extensive form games with incomplete information in which players can sign renegotiable contracts with third-parties. Our aim is to understand the extent to which third-party contracts can be used as commitment devices when it is impossible to commit not to renegotiate them. We characterize renegotiation-proof contracts and strategies for general extensive form games with incomplete information and apply our results to two-stage games. If contracts are observable, then the second mover obtains her best possible payoff given that she plays a renegotiation-proof strategy and the first mover best responds. If contracts are unobservable, then a *"folk theorem"* type result holds: Any outcome in which the second mover best responds to the first mover's action on the equilibrium path and the first mover receives at least his "individually rational payoff", can be supported. We also apply our results to games with monotone externalities and to a model of credibility of monetary policy and show that in both cases renegotiation-proofness imposes a very simple restriction.

JEL Classification: C72, D80, L13.

Keywords: Third-Party Contracts, Commitment, Strategic Delegation, Renegotiation, Asymmetric Information, Renegotiation-Proofness, Entry-Deterrence, Monetary Policy.

*We would like to thank Alp Atakan, Emre Ozdenoren, Paolo Siconolfi, Guofu Tan, seminar participants at GAMES 2012 World Congress (2012), Econometric Society European Meeting (2012), Econometric Society North American Meeting (2012), Public Economic Theory Meeting (2013), Columbia University, and Marmara University for helpful discussions. This research has been supported by TÜBİTAK Grant No. 106K317.

†Corresponding author: Department of Economics, Koç University, Rumelifeneri Yolu, Sariyer 34450, Istanbul, Turkey. Tel: +90 212 338-1354. E-mail: lkockesen@ku.edu.tr.

1 Introduction

Could an incumbent firm deter entry by contracting with third parties, such as a bank or a labor union? Could a central bank (or a government in a union) credibly commit to monetary policy (or fiscal policy) through a contract with the government (or supranational body)? More generally can contracts with third parties change the outcome of a game to the advantage of the contracting player? When contracts are non-renegotiable, the answer to this question is in general yes.[1] In fact, there are several "folk theorem" type results for different classes of games with *observable and non-renegotiable* third-party contracts.[2] The effects of *unobservable and non-renegotiable* third-party contracts are also well-understood: Nash equilibrium outcomes of a game with and without third-party contracts are identical (Katz (1991)). In fact, all (and only) Nash equilibrium outcomes of the original game can be supported as a sequential equilibrium outcome of the game with unobservable and non-renegotiable contracts (Koçkesen and Ok (2004) and Koçkesen (2007)).[3]

In this paper we seek an answer to this question for *renegotiable* contracts. When a player in a game lacks the reputation to commit to a course of action that is not sequentially rational, writing a contract with a third-party is a natural way to achieve such a commitment. However, if the contracting parties also lack the reputation to commit not to renegotiate the contract, one has to consider renegotiable contracts. We analyze if and how renegotiable third-party contracts change the equilibrium outcomes of extensive form games with incomplete information. In the main body of the paper we consider only two-player two-stage games where the second mover (player 2) has some payoff relevant private information. In what we call the *original game*, Nature moves first and determines the state of the world θ. After that, player 1 chooses an action a_1 without observing θ. Player 2 observes both θ and a_1, chooses a_2, and the game ends. Payoff function of player $i = 1,2$ is $u_i(a_1, a_2, \theta)$. Player 1's strategy is simply a choice of action a_1 whereas player 2's strategy is a function $b_2(a_1, \theta)$.

In the *game with contracts* we let player 2 sign a contract with a neutral third-party before the original game starts. A contract specifies transfers between player 2 and the third-party as a function of the contractible outcomes, which we assume to be the action choices of the two players, (a_1, a_2). The underlying and crucial assumption is that the private information of player 2 is not observable by any other player, including the third-party, and thus non-contractible. Given a contract f, the third-party's payoff is $f(a_1, a_2)$ whereas player 2's is $u_2(a_1, a_2, \theta) - f(a_1, a_2)$.

Our main objective is to understand the outcomes of the original game that can be supported in some equilibrium of the game with contracts. The first question that we need to answer is the type of strategies $b_2(a_1, \theta)$ that can be supported by a contract, i.e., incentive compatible strategies. Since contracts cannot depend on θ, incentive compatibility imposes some restrictions on b_2. In order to get a handle on these restrictions, we assume that player 2's payoff function exhibits increasing differences in (θ, a_2). It then follows that strategy b_2 is incentive compatible if and only if it is increasing in θ (This is Proposition 2 on page 9).

[1] See, among many others, Vickers (1985), Fershtman and Judd (1987), Sklivas (1987), Koçkesen et.al. (2000), Brander and Lewis (1986), Bolton and Scharfstein (1990), Snyder (1996), Spencer and Brander (1983), Brander and Spencer (1985), Eaton and Grossman (1986), Walsh (1995).

[2] See Fershtman, Judd, and Kalai (1991), Polo and Tedeschi (2000), and Katz (2006).

[3] Prat and Rustichini (2003) and Jackson and Wilkie (2005) analyze related models in which players can write action contingent contracts before the game is played. Unlike the current paper, in these papers contractual relationships are not exclusive and the focus is on the efficiency properties of the equilibrium set. Also related is Bhaskar (2009), in which players need to pay a price to a supplier in order to play certain actions that are controlled by the supplier.

The second important step is to characterize the restrictions imposed by renegotiation. We model renegotiation as a game form: After player 1 moves, player 2 can make a renegotiation offer to the third-party, who knows a_1, but not θ, and can either accept the offer or reject it. We define renegotiation-proof equilibrium as a perfect Bayesian equilibrium in which the equilibrium contract is not renegotiated after any (θ, a_1) and characterize the renegotiation-proof contracts and strategies (This is Theorem 1 on page 11). We also provide necessary and sufficient conditions for a strategy to be renegotiation-proof (Propositions 3 and 4).[4] These results generalize quite readily to arbitrary extensive form games with incomplete information where players are free to use mixed strategies, to environments in which the third-party is not neutral, and to stronger notions of renegotiation-proofness (See Section 4).

In Section 5 we present the implications of the above results in terms of the outcomes of the original game. We allow contracts to be observable or unobservable (by player 1) and renegotiable or non-renegotiable. We show that if contracts are observable, then player 2 can commit credibly to his Stackelberg payoff, i.e., the best payoff that he can achieve given that player 1 plays a best response. If the contracts are non-renegotiable, then the only restriction on the Stackelberg payoff is that player 2 uses an increasing strategy (Proposition 6). If contracts are renegotiable, then they also have to be renegotiation-proof (Proposition 8). We also show that these are the only outcomes that can be supported (Propositions 7 and 9). In other words, as long as one respects the restrictions imposed by incentive compatibility and renegotiation-proofness, contracts indeed serve as credible commitment devices.

We next consider unobservable contracts. We show that if contracts are non-renegotiable, then any Bayesian Nash equilibrium of the original game in which player 2's strategy is increasing can be supported (Proposition 10). In fact, we prove a folk theorem type result: any outcome $(a_1^*, a_2^*(\theta))$ of the original game in which $a_2^*(\theta)$ is a best response to a_1^* for each θ and player 1's payoff is at least as large as his "individually rational" payoff, can be supported (Corollary 1). Definition of individually rational payoff is different from the standard one in that player 2, in minimizing player 1's payoff, is restricted to using increasing strategies. Similar results hold for renegotiable contracts except that in the definition of the individually rational payoff player 2's strategy is restricted to be increasing and renegotiation-proof (Proposition 11 and Corollary 2).

Note that renegotiation affects the games with observable and with unobservable contracts in the same way: In both cases the only additional restriction is that the strategies used by the contracting party must be renegotiation-proof. This means that the difference between what can be supported with observable and unobservable contracts does not depend on renegotiation. In particular, if the Stackelberg payoff obtained with observable contracts is a Bayesian Nash equilibrium, then unobservable and renegotiation-proof contracts may still serve as commitment devices, but the Stackelberg payoff is not the only equilibrium outcome anymore. On the other hand, if the Stackelberg payoff obtained with observable contracts is not a Bayesian Nash equilibrium, then unobservable contracts might not serve as commitment devices. Using a distinction introduced by Schelling (1960), with unobservable contracts (renegotiable or not) we can only *commit to a threat*, while with observable

[4] Our assumption that the third-party cannot observe θ during renegotiation is crucial. Otherwise, the result is trivial: One can only support the perfect Bayesian equilibria of the original game. This is because, if both a_1 and θ are common knowledge, then player 2 and the third-party would renegotiate away any strategy of player 2 that does not maximize the joint surplus, i.e., player 2's payoff in the original game.

contracts (renegotiable or not) we can also *commit to a promise*.[5] We illustrate this point with two applications: entry-deterrence and credibility of monetary policy. We show that it is possible to deter entry with renegotiation-proof contracts even when they are unobservable, whereas, in the monetary policy game, commitment to a low inflation rate is possible only with observable contracts.

Entry-deterrence game is an example of a more general class of games, which we call games with monotone externalities and analyze in Section 6.1. In games with monotone externalities, player 1's payoff is increasing or decreasing in Player 2's action for a given action a_1. Suppose, for concreteness, that player 1's payoff is increasing in a_2. If contracts are non-renegotiable, then player 2 can obtain a favorable outcome by punishing player 1 by playing the smallest a_2 whenever he plays an unfavorable action. Since a constant strategy is increasing, incentive compatibility does not bring any further restrictions on the outcomes that can be supported with non-renegotiable contracts. On the other hand, renegotiation-proofness imposes a very specific type of constraint on the kind of punishment player 2 can inflict upon player 1: The highest type of player 2 must play a best response while the other types could keep playing the smallest action (Corollary 3).[6] In other words, the additional restriction renegotiation-proofness brings about depends on the probability of the highest type: The lower this probability, the less severe is the effect of renegotiation.

The class of games with monotone externalities is large and contains many economic models. The canonical example, of course, is the Stackelberg competition. We show that in this game, the follower firm indeed benefits from renegotiation-proof third-party contracts. This game can also be construed as an entry deterrence game, in which case we show that entry can always be deterred with non-renegotiable contracts but only under certain conditions with renegotiation-proof contracts, i.e., renegotiation has real bite in these games.

In Section 6.2, we apply our results to the well-known problem of credibility of monetary policy: Once the public forms expectations of inflation, the government has an incentive to create surprise inflation in order to decrease unemployment. This implies that inflation is inefficiently high in the rational expectations equilibrium (Kydland and Prescott (1977), Barro and Gordon (1983)). Contracts with central banks have been suggested as a possible solution to this problem (Walsh (1995), Persson and Tabellini (1993)). In fact, in many countries central banks have mandates in the form of inflationary targets, which may be interpreted as contracts. With observable and non-renegotiable contracts we obtain a result that is well-known in the literature: The equilibrium outcome of the game with contracts has no inflationary bias and induces the socially optimal policy (Proposition 14). If the contracts are observable and renegotiable, the inflationary bias cannot be eliminated completely but is lower than the one resulting from discretionary monetary policy (Proposition 15). In other words, the possibility to renegotiate the contract reduces the commitment role played by contracts but does not eliminate it. Also in this application, renegotiation-proofness imposes a very specific type of constraint on the equilibrium strategies of the contracting party (monetary authority): The inflation rate chosen by the monetary authority in the "worst" state of the world (e.g., a financial crisis) is the same as the inflation rate chosen without contracts (i.e., higher), while the inflation rate chosen in all other states is the same as the inflation rate chosen in the game with non-renegotiable contracts. There-

[5] In other words, with unobservable contracts we can support non-sequentially rational strategies only off the equilibrium path while with observable contracts, we can support non-sequentially rational strategies also on the equilibrium path (See Koçkesen and Ok (2004) and Koçkesen (2007)).

[6] This is true when player 1's payoff is increasing in player 2's action. If his payoff is decreasing, then the harshest punishment player 2's can impose is to play the highest action for all types other than the lowest, who must play a best response.

fore, in this case too, the higher the probability of the worst state, the more severe is the effect of renegotiation.

These two applications have also an interesting implication about optimal credible commitments. A commitment (or a contract) that is optimal to carry out in every state of the world will never be renegotiated and therefore is very credible. However, such contracts are rarely optimal ex ante because in many situations it is beneficial to commit to a different strategy. Monetary policy is a case in point. Another way to achieve high credibility is to make renegotiation very difficult. For example, in many countries central banks have a mandate to hit a particular inflation target, which cannot be altered. Such contracts, however, may turn out to be suboptimal ex post, for example when a small probability event such as a financial crisis occurs. This is exactly what happened in many countries including England, where the government has an incentive to create some inflation in order to escape the liquidity trap but cannot do so because of the aforementioned mandate of the central bank.[7] Our results suggest that in order to achieve optimal commitments that are also credible, one needs to allow flexibility in some states of the world. For example, optimal monetary rule that is also credible would need to allow for a higher inflation rate in case of a serious crisis.

Relationship to the Literature

The general message of this paper is that while renegotiation does limit the commitment value of third-party contracts (in general, renegotiation has bite), it does not completely destroy it. This message complements and clarifies the results in the existing literature. Dewatripont (1988) analyzes an entry-deterrence game in which the incumbent signs a contract with a labor union before the game begins. A potential entrant observes the contract and then decides whether to enter or not. Renegotiation takes place after the entry decision is made, during which the union offers a new contract to the incumbent. The crucial assumption is that the incumbent has some payoff relevant private information during the renegotiation process. Dewatripont (1988) shows that commitment effects exist in such a model and may deter entry when contracts are publicly observable. One contribution of our paper is to show that commitment effects exist in arbitrary two-stage games: For example oligopoly models with price competition and games in which credible commitment to monetary and fiscal policy is valuable. Furthermore, we show that commitment effects exist even if the contracts are unobservable and hence they exist also if we allow the contracts to be renegotiated immediately after they are signed. In fact, allowing for secret renegotiation right after the contracts are signed has no effect on the results if the game is with unobservable contracts and reduces the case of observable contracts to unobservable ones.[8]

Caillaud et al. (1995) applies Dewatripont's (1988) idea to a game between two principal-agent hierarchies. In the first stage of their game each principal decides whether to publicly offer a contract to the agent; in the second stage each principal offers a secret contract to the agent, which, if accepted, overwrites the public contract that might have been offered in stage 1; in the third stage each agent receives payoff relevant information, decides whether to quit, and if he does not quit, he plays a normal form game with the other agent. Their main question is whether there exist equilibria of this game in which the principals choose not to offer a public contract in stage 1. If the answer

[7] See Tim Harford's article in *Financial Times* on March 1, 2013 for more on this issue.

[8] There is another difference between our model and Dewatripont (1988): In our renegotiation protocol, the informed party makes the new contract offer, whereas in Dewatripont's, it is the uninformed party who makes the offer. This turns out to make a difference as we discuss in Section 6.1 (Proposition 12)

to this question is no, then the interpretation is that contracts have commitment value. They show that contracts have commitment value if the market game stage is of Cournot type, but not if it is of Bertrand type. Moreover, when contracts have commitment value, they reduce the payoff of the contracting parties. The received message of Caillaud et al. (1995), in comparison with Dewatripont (1988), is that by allowing secret renegotiation, Caillaud et al. (1995) enhanced the realism of the model and clarified the role of strategic contracting.[9] The crucial difference between our model and Caillaud et al. is that they assume that agents play a simultaneous move game (and principals offer contracts to the agents simultaneously) whereas we focus on sequential move games. Therefore, one contribution of our paper is to show that the differences between the results of Caillaud et al. (1995) and Dewatripont (1988) do not depend on allowing secret renegotiation right after the contracts are signed, but instead depend on the fact that Dewatripont (1988) studies a sequential move game while Caillaud et al. (1995) a simultaneous move game.

In a related paper, Gerratana and Koçkesen (2012) also study the effects of renegotiation-proof third-party contracts in two-stage games. However, that paper assumes that the original game is with perfect information whereas the current one assumes it is a game with incomplete information. Some aspects of the analyses of these two models are similar and use similar tools, namely theorems of the alternative. Indeed, results on renegotiation-proof contracts and strategies in Section 3 (Theorem 1 and Propositions 3 and 4) are exact analogs of their counterparts in Gerratana and Koçkesen (2012). However, the games to which these are applied are completely different, and so are the effects of renegotiation. For the class of games for which Gerratana and Koçkesen (2012) obtain sharp results, renegotiation has no bite, that is, the outcomes of games with unobservable and non-renegotiable contracts are robust to the introduction of renegotiation.[10] This is not true in the current paper, where renegotiation, in general, has a bite. Furthermore, in the current paper we extend our results to (1) arbitrary extensive form games and mixed strategies; (2) to non-neutral third-parties; and (3) to the case of observable contracts.

In addressing the question of how renegotiation affects the commitment value of third-party contracts we assumed that in the renegotiation stage the contracting party and the third-party have asymmetric information.[11] This choice seems natural in settings where the original game has incomplete information. Therefore, we hope that the analysis developed in this paper could be used for other applications not considered in this paper.

2 The Model

Our aim is to understand the effects of renegotiation-proof third-party contracts in extensive form games. In this section, we will do this in a particularly simple environment: two-stage games with private information, which we call the *original game*. The main reason we present our results for two-stage games is ease of exposition. Still, we should note that many models in economics such as the entry game, the Stackelberg game, and monopolistic screening belong to this class of games. Furthermore, we show in Section 4.1 that our main characterization results extend to arbitrary extensive

[9] See Bolton and Dewatripont (2005) pages 631-636.

[10] See Gerratana and Koçkesen (2012) Section 6.

[11] This modeling choice follows Dewatripont (1988) and Caillaud et al. (1995). However, other researchers analyzed different kind of frictions in the renegotiation stage (see for example Bensaid and Gary-Bobo (1993), who find that renegotiable contract have commitment effects in a model with non-transferable utility).

form games with incomplete information as long as they satisfy an increasing differences property (see Definition 8).

We allow one of the players to sign a contract with a third-party before the original game begins and call this new game the *game with third-party contracts*. The contract specifies a transfer between the player and the third-party as a function of the contractible outcomes of the original game. The crucial aspect of our model is the presence of asymmetric information between this player and the third-party during the renegotiation phase.

More precisely, we define the *original game*, denoted G, as follows: Nature chooses $\theta \in \Theta$ according to probability distribution $p \in \Delta(\Theta)$. After the move of Nature, player 1, without observing θ, chooses $a_1 \in A_1$. Lastly, player 2 observes (θ, a_1) and chooses $a_2 \in A_2$. We assume that A_1, A_2, and Θ are finite and let $p(\theta)$ denote the probability of Nature choosing θ. Payoff function of player $i \in \{1,2\}$ is given by $u_i : A \times \Theta \to \mathbb{R}$, where $A = A_1 \times A_2$.

The *game with third-party contracts* is a three player extensive form game described by the following sequence of events: Player 2 offers a contract $f : A \to \mathbb{R}$ to a third-party. The third-party accepts (denoted y) or rejects (denoted n) the contract. In case of rejection the game ends and the third-party receives a fixed payoff of $\delta \in \mathbb{R}$ while player 2 receives $-\infty$.[12] In case of acceptance, player 1 and 2 play the original game. We assume that throughout the entire game θ remains the private information of player 2.

Since offering a contract that is rejected yields player 2 a very small payoff, the contract offer will be accepted in all equilibria. Therefore, for simplicity, we omit the third-party's acceptance decision from histories and represent an outcome of the game with third-party contracts as (f, θ, a_1, a_2). The payoff functions in the game with contracts are given by $v_1\left(f, a_1, a_2, \theta\right) = u_1\left(a_1, a_2, \theta\right)$, $v_2\left(f, a_1, a_2, \theta\right) = u_2\left(a_1, a_2, \theta\right) - f\left(a_1, a_2\right)$, $v_3\left(f, a_1, a_2, \theta\right) = f\left(a_1, a_2\right)$, where v_3 is the payoff function of the third-party. Note that the payoff function of the third-party assumes that he is neutral towards the outcome of the game, i.e., he cares only about the transfer. In Section 4.2 we relax this assumption and allow the third-party also to have intrinsic preferences over the outcomes of the original game.

The game is with renegotiable contracts if the contracting parties can renegotiate the contract after player 1 plays a_1 and before player 2 chooses a_2. We assume that player 2, who is the informed party, initiates the renegotiation process by offering a new contract, which the third-party may accept or reject. If the third-party rejects the renegotiation offer g, then player 2 chooses $a_2 \in A_2$ and the outcome is payoff equivalent to $\left(f, \theta, a_1, a_2\right)$. If he accepts, then player 2 chooses $a_2 \in A_2$ and the outcome is payoff equivalent to $\left(g, \theta, a_1, a_2\right)$.

We say that the game is with observable contracts if the initial contract is observed by player 1. Otherwise, we say that the game is with unobservable contracts. In other words, there are four possible games with third-party contracts depending upon whether the contract is renegotiable or non-renegotiable and observable or unobservable. Given an original game G, we will denote the game with non-renegotiable and observable contracts with $\Gamma_{NO}(G)$, non-renegotiable and unobservable contracts with $\Gamma_{NU}(G)$, renegotiable and observable contracts with $\Gamma_{RO}(G)$, and renegotiable and unobservable contracts with $\Gamma_{RU}(G)$.

A *behavior strategy* for player $i \in \{1,2,3\}$ is defined as a set of probability measures $\beta_i \equiv \{\beta_i(I) : I \in \mathscr{I}_i\}$, where $\mathscr{I}_i$ is the set of information sets of player i and $\beta_i(I)$ is defined on the set of actions available at information set I. One may write $\beta_i(h)$ for $\beta_i(I)$ for any history $h \in I$. By a *system of*

[12] This assumption is made only to eliminate equilibria in which no contract has been signed and can easily be relaxed.

beliefs, we mean a set $\mu \equiv \{\mu(I) : I \in \mathscr{I}_i \text{ for some } i\}$, where $\mu(I)$ is a probability measure on I. A pair (β, μ) is called an *assessment*. An assessment (β, μ) is said to be a *perfect Bayesian equilibrium* (PBE) if (1) each player's strategy is optimal at every information set given her beliefs and the other players' strategies; and (2) beliefs at every information set are consistent with observed histories and strategies.[13]

We will limit our analysis to pure behavior strategies, and hence a strategy profile of the original game G is given by $(b_1, b_2) \in A_1 \times A_2^{A_1 \times \Theta}$.[14] For any behavioral strategy profile (b_1, b_2) of G, define the expected payoff of player $i = 1,2$ as $U_i(b_1, b_2) = \sum_{\theta \in \Theta} p(\theta) u_i(b_1, b_2(b_1, \theta), \theta)$ and the best response correspondences as $BR_1(b_2) = \text{argmax}_{a_1 \in A_1} U_1(a_1, b_2)$ for all $b_2 \in A_2^{A_1 \times \Theta}$ and $BR_2(a_1, \theta) = \text{argmax}_{a_2 \in A_2} u_2(a_1, a_2, \theta)$ for all $(a_1, \theta) \in A_1 \times \Theta$. We say that a strategy profile (b_1^*, b_2^*) is a *Bayesian Nash equilibrium* of G if $b_1^* \in BR_1(b_2^*)$ and $b_2^*(b_1^*, \theta) \in BR_2(b_1^*, \theta)$ for all θ. The difference between a perfect Bayesian equilibrium and a Bayesian Nash equilibrium, of course, is that the former requires player 2 to best respond to every action of player 1, whereas the latter requires best response to only the equilibrium action. Therefore, every perfect Bayesian equilibrium is a Bayesian Nash equilibrium but not conversely.

For any behavior strategy profile (b_1, b_2) in G, we say that an assessment (β, μ) in $\Gamma_k(G)$, $k = NO, NU, RO, RU$, *induces* (b_1, b_2) if in $\Gamma_k(G)$ player 1 plays according to b_1 and, after the equilibrium contract, player 2 plays according to b_2.[15]

Our ultimate aim is to characterize renegotiation-proof equilibria, in which the equilibrium contract is not renegotiated after any history.[16] More precisely,

Definition 1 (Renegotiation-Proof Equilibrium)**.** A perfect Bayesian equilibrium (β^*, μ^*) of $\Gamma_{RO}(G)$ and $\Gamma_{RU}(G)$ is *renegotiation-proof* if the equilibrium contract is not renegotiated after any $a_1 \in A_1$ and $\theta \in \Theta$.

We say that a strategy profile (b_1, b_2) of the original game G can be *supported with observable and non-renegotiable contracts* if there exists a perfect Bayesian equilibrium of $\Gamma_{NO}(G)$ that induces (b_1, b_2). Similarly, a strategy profile (b_1, b_2) of the original game G can be *supported with observable renegotiation-proof contracts* if there exists a renegotiation-proof perfect Bayesian equilibrium of $\Gamma_{RO}(G)$ that induces (b_1, b_2). Similarly for unobservable and non-renegotiable and unobservable renegotiation-proof contracts.

One important question is whether we are "missing" equilibria by restricting the analysis to renegotiation-proof equilibria. The following result, known as *renegotiation-proofness principle*, shows that the answer is no.

Proposition 1 (Renegotiation-Proofness Principle)**.** *If there is a perfect Bayesian equilibrium of the game $\Gamma_{RO}(G)$ (resp. $\Gamma_{RU}(G)$) that induces a strategy profile (b_1, b_2) of the original game G, then there exist a renegotiation-proof perfect Bayesian equilibrium of $\Gamma_{RO}(G)$ (resp. $\Gamma_{RU}(G)$) that induces the same strategy profile (b_1, b_2).*

[13] See Fudenberg and Tirole (1991) for a precise definition of perfect Bayesian equilibrium.

[14] In Section 4.1 we relax this and allow also mixed strategies. This introduces some technical difficulties but our main results go through.

[15] Note that in $\Gamma_{RO}(G)$ and $\Gamma_{RU}(G)$, player 2 may choose an action $a_2 \in A_2$ either without renegotiating the initial contract or after attempting renegotiation.

[16] We follow the previous literature in our definition of renegotiation-proof equilibrium. See, for example, Maskin and Tirole (1992) and Beaudry and Poitevin (1995).

Proof. In Section 8. □

AN EXAMPLE: ENTRY DETERRENCE

In order to illustrate our main query as well as some of our results later on, we introduce a very simple entry game in this section (See Figure 1). Player 1 is a potential entrant, who may enter (E) or stay out (O) and player 2, who is the incumbent, may fight (F) or accommodate (A) entry.

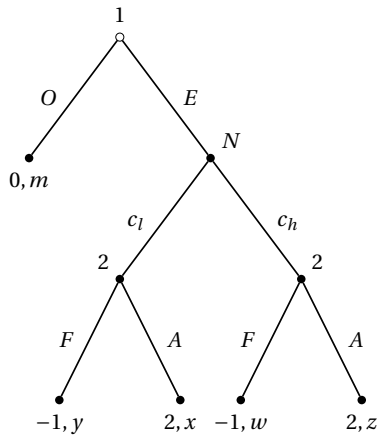

Figure 1: Entry Game

We assume that fighting is costly, and it is costlier for the high cost incumbent (type c_h) than for the low cost (type c_l): $z-w > x-y > 0$. The entrant believes that the incumbent's type is low cost with probability $p \in (0,1)$.

The unique perfect Bayesian equilibrium (PBE) of this game is (E, AA), i.e., the entrant enters and both types of the incumbent accommodate. We assume that the monopoly profit is larger than the highest possible profit following entry, i.e., $m > x$. In other words, the incumbent would benefit from deterring entry, and one way of achieving this would be to sign a contract with a third-party that makes fighting optimal. For example, the following contract makes playing FF optimal: $f(F) = \delta$, $f(A) = \delta + (z-w)$. Is such a contract renegotiation-proof? If not, can entry still be deterred with renegotiation-proof contracts? In what follows we will answer these questions and also characterize the equilibrium outcomes that can be supported with third-party contracts under different assumptions regarding their observability and renegotiation-proofness.

3 Renegotiation-Proof Contracts

In this section we will provide results that help identify the set of outcomes of any original game G that can be supported by renegotiation-proof perfect Bayesian equilibria of the game with observable (or unobservable) and renegotiable contracts.

In order to decide whether to accept a new contract offer in the renegotiation phase of the game with renegotiable contracts, the third-party forms beliefs regarding player 2's strategy under the new contract and compares his payoffs from the old and the new contracts. In equilibrium, these beliefs

must be such that player 2's strategies are sequentially rational, i.e., incentive compatible, under the new contract. Let the contract space be $\mathscr{C} = \mathbb{R}^{A_1 \times A_2}$ and define incentive compatibility as a property of any contract-strategy pair $(f, b_2) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$.

Definition 2 (Incentive Compatibility)**.** $(f, b_2) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is *incentive compatible* if

$$u_2(a_1, b_2(a_1,\theta),\theta) - f(a_1, b_2(a_1,\theta)) \geq u_2(a_1, b_2(a_1,\theta'),\theta) - f(a_1, b_2(a_1,\theta')) \text{ for all } a_1 \in A_1 \text{ and } \theta,\theta' \in \Theta.$$

We say that a strategy b_2 is incentive compatible if there is a contract f such that (f, b_2) is incentive compatible. We can obtain a sharp characterization of incentive compatible strategies if we impose more structure on the original game. To this end, let $\succsim_\theta$ be a linear order on Θ and $\succsim_2$ a linear order on A_2, and denote their asymmetric parts by $\succ_\theta$ and $\succ_2$, respectively.

Definition 3 (Increasing Differences)**.** $u_2 : A_1 \times A_2 \times \Theta \to \mathbb{R}$ is said to have *increasing differences* in $(\succsim_\theta, \succsim_2)$ if for all $a_1 \in A_1$, $\theta \succsim_\theta \theta'$ and $a_2 \succsim_2 a_2'$ imply that $u_2(a_1, a_2, \theta) - u_2(a_1, a_2, \theta') \geq u_2(a_1, a_2', \theta) - u_2(a_1, a_2', \theta')$. It is said to have *strictly increasing differences* if $\theta \succ_\theta \theta'$ and $a_2 \succ_2 a_2'$ imply that $u_2(a_1, a_2, \theta) - u_2(a_1, a_2, \theta') > u_2(a_1, a_2', \theta) - u_2(a_1, a_2', \theta')$.

Definition 4 (Increasing Strategies)**.** $b_2 : A_1 \times \Theta \to A_2$ is called *increasing* in $(\succsim_\theta, \succsim_2)$ if for all $a_1 \in A_1$, $\theta \succsim_\theta \theta'$ implies that $b_2(a_1,\theta) \succsim_2 b_2(a_1,\theta')$. Denote the set of all increasing b_2 by B_2^+.

For the rest of the paper, we restrict attention to games G in which there exist a linear order on Θ and a linear order on A_2 such that u_2 has strictly increasing differences in $(\succsim_\theta, \succsim_2)$. Standard arguments show that under increasing differences, incentive compatibility implies that b_2 is increasing. The following proposition states this result and shows that its converse also holds.

Proposition 2. *If $u_2 : A_1 \times A_2 \times \Theta \to \mathbb{R}$ has strictly increasing differences, then a strategy $b_2 : A_1 \times \Theta \to A_2$ is incentive compatible if and only if it is increasing.*

Proof. Omitted. □

The only if part follows from a standard argument in contract theory. In order to prove the if part fix an arbitrary $a_1 \in A_1$, let the number of elements of Θ be n, and order its elements so that $\theta^n \succsim_\theta \theta^{n-1} \succsim_\theta \cdots \theta^2 \succsim_\theta \theta^1$. For any contract-strategy pair (f, b_2), define $f(a_1)_j = f(a_1, b_2(a_1,\theta^j))$, $j = 1,\ldots,n$, and let, with an abuse of notation, $f(a_1) \in \mathbb{R}^n$ be the vector whose jth component is given by $f(a_1)_j$. When u_2 has increasing differences, incentive compatibility of (f, b_2) is equivalent to the local upward and downward constraints:[17]

$$\begin{aligned} f(a_1)_j - f(a_1)_{j+1} &\leq u_2(a_1, b_2(a_1,\theta^j),\theta^j) - u_2(a_1, b_2(a_1,\theta^{j+1}),\theta^j), \quad j = 1,\ldots,n-1 \\ -f(a_1)_{j-1} + f(a_1)_j &\leq u_2(a_1, b_2(a_1,\theta^j),\theta^j) - u_2(a_1, b_2(a_1,\theta^{j-1}),\theta^j), \quad j = 2,\ldots,n \end{aligned}$$

For any $a_1 \in A_1$, we can write these inequalities in matrix form as $Df(a_1) \leq \vec{U}_2(a_1, b_2)$, where D is a matrix of coefficients and $\vec{U}_2(a_1, b_2)$ a column vector with $2(n-1)$ components, whose component $2j-1$ is given by

$$\vec{U}_2(a_1, b_2)_{2j-1} = u_2(a_1, b_2(a_1,\theta^j),\theta^j) - u_2(a_1, b_2(a_1,\theta^{j+1}),\theta^j)$$

[17] See, for example, Bolton and Dewatripont (2005), p. 78.

and component $2j$ is given by

$$\vec{U}_2(a_1, b_2)_{2j} = u_2(a_1, b_2(a_1, \theta^{j+1}), \theta^{j+1}) - u_2(a_1, b_2(a_1, \theta^j), \theta^{j+1})$$

Therefore, the proof will be completed if we can show that if u_2 has strictly increasing differences and b_2 increasing, then there exists $f(a_1) \in \mathbb{R}^n$ such that $Df(a_1) \leq \vec{U}_2(a_1, b_2)$. This follows easily from Gale's theorem for linear inequalities (Mangasarian (1994), p. 33).

We next define our renegotiation-proofness concept, which follows from the definition of renegotiation-proof perfect Bayesian equilibrium (Definition 1).

Definition 5 (Renegotiation-Proofness)**.** We say that $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is *renegotiation-proof* if for all $a_1 \in A_1$ and $\theta \in \Theta$ for which there exists an incentive compatible (g, b_2) such that

$$u_2(a_1, b_2(a_1, \theta), \theta) - g(a_1, b_2(a_1, \theta)) > u_2(a_1, b_2^*(a_1, \theta), \theta) - f(a_1, b_2^*(a_1, \theta)) \tag{1}$$

there exists a $\theta' \in \Theta$ such that

$$f(a_1, b_2^*(a_1, \theta')) \geq g(a_1, b_2(a_1, \theta')) \tag{2}$$

In words, if, for some (θ, a_1), there is a contract g and an incentive compatible continuation play b_2 such that player 2 prefers g over f (i.e., (1) holds), there must exist a belief of the third-party (over θ) under which it is optimal to reject g, which is implied by (2).[18]

Finally, we define a renegotiation-proof strategy as,

Definition 6 (Renegotiation-Proof Strategy)**.** A strategy $b_2 \in A_2^{A_1 \times \Theta}$ is *renegotiation-proof* if there exists an $f \in \mathscr{C}$ such that (f, b_2) is incentive compatible and renegotiation-proof. Denote the set of all renegotiation-proof strategies by B_2^R.

Definitions 5 and 6 are indeed the correct definitions to work with, in the sense that they identify the conditions that any contract f and strategy b_2 must satisfy to be part of a renegotiation-proof perfect Bayesian equilibrium of $\Gamma_{RO}(G)$ or $\Gamma_{RU}(G)$. Indeed, if a strategy b_2 of the original game is not renegotiation-proof, then there is no perfect Bayesian equilibrium (of the game with renegotiable contracts) in which a contract f is offered and b_2 is played without renegotiating f. This simply follows from the fact that if (f, b_2) is not renegotiation-proof, then there is (a_1, θ) and a contract g that would be accepted for any belief of the third-party at the renegotiation stage and increase player 2's payoff. In other words, f will be renegotiated after (a_1, θ) and therefore the equilibrium is not renegotiation-proof. In fact, the converse of that statement also holds: If b_2 is renegotiation-proof, we can construct a perfect Bayesian equilibrium of the game with renegotiable contracts in which the equilibrium contract is not renegotiated after any a_1 and θ. Of course, the equilibrium contract and b_2 will also have to satisfy other conditions for them to be part of an equilibrium, but these would depend on whether the contracts are observable or unobservable, an issue which we will address in Section 5.

In this section we present a result that characterizes renegotiation-proof contracts and strategies. In order to understand this result one should first realize that condition (2) in Definition 5 is satisfied

[18] This definition allows beliefs to be arbitrary following an off-the-equilibrium renegotiation offer. An alternative definition would be to require the beliefs to satisfy intuitive criterion. In Section 4.3 we show that our results go through with minor modifications when we adopt this stronger version.

trivially if the strategy b_2 does not lead to a higher surplus for the contracting parties after (a_1,θ). In other words, for each a_1 and $i=1,\dots n$, we need to check renegotiation-proofness of (f,b_2^*) only against strategies that belong to the following set:

$$\mathfrak{B}(a_1,i,b_2^*)=\{b_2\in A_2^{A_1\times\Theta}: b_2 \text{ is increasing and } u_2(a_1,b_2(a_1,\theta^i),\theta^i)>u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)\}. \tag{3}$$

By Definition 5, (f,b_2^*) is *not* renegotiation-proof if and only if there exist $a_1\in A_1$, $i=1,\dots n$, and incentive compatible (g,b_2) such that $u_2(a_1,b_2(a_1,\theta^i),\theta^i)-g(a_1)_i>u_2(a_1,b_2^*(a_1,\theta_i),\theta_i)-f(a_1)_i$ and $g(a_1)_j>f(a_1)_j$ for all $j=1,\dots,n$. As we have discussed after Proposition 2, when u_2 has increasing differences, incentive compatibility of (g,b_2) is equivalent to $Dg(a_1)\le\vec{U}_2(a_1,b_2)$. Therefore, (f,b_2^*) is not renegotiation-proof if and only if there exist a_1, i, b_2 and $\varepsilon\in\mathbb{R}^n$ such that

$$D(f(a_1)+\varepsilon)\le\vec{U}_2(a_1,b_2),\quad \varepsilon_i<u_2(a_1,b_2(a_1,\theta^i),\theta^i)-u_2(a_1,b_2^*(a_1,\theta^i),\theta^i),\quad \varepsilon\gg 0.$$

These conditions can be written as $[Ax\gg 0, Cx\ge 0 \text{ has a solution } x]$, once the vector x and matrices A and C are appropriately defined. Motzkin's theorem of the alternative then implies that the necessary and sufficient condition for being renegotiation-proof is $[A'y_1+C'y_2=0, y_1>0, y_2\ge 0 \text{ has a solution } y_1,y_2]$. The fact that u_2 has increasing differences can then be used to prove the equivalence of this condition to the one stated in the following theorem.

Theorem 1. *$(f,b_2^*)\in\mathscr{C}\times A_2^{A_1\times\Theta}$ is renegotiation-proof if and only if for any $a_1\in A_1$, $i\in\{1,2,\dots,n\}$, and $b_2\in\mathfrak{B}(a_1,i,b_2^*)$ there exists a $k\in\{1,2,\dots,i-1\}$ such that*

$$u_2(a_1,b_2(a_1,\theta^i),\theta^i)-u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)+\sum_{j=k}^{i-1}\vec{U}_2(a_1,b_2)_{2j-1}\le f(a_1)_k-f(a_1)_i \tag{4}$$

or there exists an $l\in\{i+1,i+2,\dots,n\}$ such that

$$u_2(a_1,b_2(a_1,\theta^i),\theta^i)-u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)+\sum_{j=i+1}^{l}\vec{U}_2(a_1,b_2)_{2(j-1)}\le f(a_1)_l-f(a_1)_i \tag{5}$$

Proof. Omitted. □

Theorem 1 characterizes the conditions for which $\left(f,b_2^*\right)$ is renegotiation-proof. Our next step is to find conditions for a strategy b_2^* to be supported with renegotiation-proof contracts. The following definition facilitates the exposition.

Definition 7. For any a_1, $i=1,\dots,n$ and $b_2\in\mathfrak{B}(a_1,i,b_2^*)$ we say that $m(b_2)\in\{1,2,\dots,n\}$ is a *blocking type* if

$$u_2(a_1,b_2(a_1,\theta^i),\theta^i)-u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)\le\sum_{j=m(b_2)}^{i-1}\left[\vec{U}_2(a_1,b_2^*)_{2j-1}-\vec{U}_2(a_1,b_2)_{2j-1}\right] \tag{6}$$

or

$$u_2(a_1,b_2(a_1,\theta^i),\theta^i)-u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)\le\sum_{j=i+1}^{m(b_2)}\left[\vec{U}_2(a_1,b_2^*)_{2(j-1)}-\vec{U}_2(a_1,b_2)_{2(j-1)}\right] \tag{7}$$

We obtain the following necessary conditions for a strategy b_2^* to be renegotiation-proof.

Proposition 3. *A strategy $b_2^* \in A_2^{A_1 \times \Theta}$ is renegotiation-proof only if for any $a_1 \in A_1$, $i \in \{1,2,\ldots,n\}$, and $b_2 \in \mathfrak{B}(a_1, i, b_2^*)$ there is a blocking type.*

Proof. Omitted. □

The above condition becomes also sufficient for renegotiation-proofness with an additional requirement about the relation of blocking types for different renegotiation opportunities.

Proposition 4. *A strategy $b_2^* \in A_2^{A_1 \times \Theta}$ is renegotiation-proof if for any $a_1 \in A_1$, $i \in \{1,2,\ldots,n\}$, and $b_2 \in \mathfrak{B}(a_1, i, b_2^*)$ there is a blocking type $m(b_2^j)$ such that $k < l$, $m(b_2^k) > k$, and $m(b_2^l) < l$ imply $m(b_2^k) \leq m(b_2^l)$.*

Proof. Omitted. □

The conditions given in Proposition (3) and (4) coincide when player 2 has only two types. Therefore, Proposition (3) is a full characterization result for such games. Although, they fall short of providing a full characterization in games with more than two types, they help us do so in environments with more structure as we demonstrate in Section 6.

EXAMPLE: ENTRY DETERRENCE

Let $c_h \succ_\theta c_l$ and $A \succ_2 F$ and observe that $z - w > x - y$ implies that u_2 has strictly increasing differences. Proposition 2 therefore implies that the set of incentive compatible strategies are $\{FF, FA, AA\}$. Are these strategies renegotiation-proof? AA is clearly renegotiation-proof because both types are best responding and hence $\mathfrak{B}(E, c_l, AA) = \mathfrak{B}(E, c_h, AA) = \emptyset$. How about FF? For both types playing A is a better response and hence $\mathfrak{B}(E, c_l, FF) = \{AA\}$ and $\mathfrak{B}(E, c_h, FF) = \{FA, AA\}$. Is there a blocking type for c_l, i.e., does (7) hold for $m(b_2) = c_h$? Since

$$u_2(E, A, c_l) - u_2(E, F, c_l) = x - y > u_2(E, F, c_h) - u_2(E, F, c_h) - (u_2(E, A, c_h) - u_2(E, A, c_h)) = 0$$

the answer is no, i.e., FF is not renegotiation-proof. Is FA renegotiation-proof? In this case $\mathfrak{B}(E, c_l, FA) = \{AA\}$ and $\mathfrak{B}(E, c_h, FA) = \emptyset$. Is there a blocking type for c_l? Since

$$u_2(E, A, c_l) - u_2(E, F, c_l) = x - y \leq u_2(E, A, c_h) - u_2(E, F, c_h) - (u_2(E, A, c_h) - u_2(E, A, c_h)) = z - w$$

the answer is yes. Therefore, the set of renegotiation-proof strategies is $\{FA, AA\}$. In other words, renegotiation-proofness in this example is satisfied whenever the high cost type best responds. Also note that for the high cost type, not best responding is costlier, i.e., $z - w > x - y$. Credible commitment, in this example, requires best responding when it is very costly no to do so. Finally, an example of a contract that supports FA is $f(F) = \delta$, $f(A) = \delta + (x - y)$.

4 Extensions

So far we have assumed that the original game has only two stages and conducted the equilibrium analysis in pure strategies. Furthermore, we have assumed that the third-party is neutral. In this section we show that all our main results can be generalized to a much more general class of extensive form games with incomplete information and they are true in mixed strategy equilibria as well. We

also discuss how our results are modified when third-parties are not neutral. Finally, we present the implications of a stronger definition of renegotiation-proofness.

4.1 General Extensive Form Games

Although we have stated our results for two stage games in which only player 2 has private information and has the right to sign a contract with a third-party, we can generalize them to arbitrary finite extensive form games with incomplete information and perfect recall. The only restriction we impose is that players' payoff functions in the original game exhibit increasing differences in a sense that we will make precise.

Define the original game G as an extensive form game with incomplete information in which player $i \in \{1,\dots,n\}$ privately learns his type $\theta_i \in \Theta_i$ at the beginning of the game. Assume that Nature chooses types independently and let $|\Theta_i| = n_i$. After the types are determined players start taking actions. We denote the set of histories (excluding the moves of Nature at the beginning) by H and denote a typical history by $h = (a_0, a_1, a_2, \dots, a_k)$, where a_0 is the initial node (or empty history) and a_j denotes the jth action taken in this history. Payoff function of player i is given by $u_i : Z \times \Theta \to \mathbb{R}$, where Z is the set of (finite) terminal histories and Θ is the set of all type profiles $\theta = (\theta_1, \dots, \theta_n)$.

The set of pure strategies of player i is given by S_i and a mixed strategy for player i is a probability distribution over S_i for each θ_i, i.e., a mapping $\sigma_i : \Theta_i \to \Delta(S_i)$. A pure strategy profile is denoted by s and mixed strategy profile by σ. Denote the set of all information sets at which player i moves by $\mathscr{I}_i$ and the set of all information sets in the game by $\mathscr{I}$. At any information set $I \in \mathscr{I}_i$, player i has a set of pure strategies available for the rest of the game, denoted $S_i|_I$ and defined as the restriction of S_i to information sets of player i that follows (and includes) I. A belief system is a collection of probability measures $m = \{m(I) : I \in \mathscr{I}\}$, where $m(I)$ for $I \in \mathscr{I}_i$ is defined over $I \times \Theta_{-i}$. A pair (σ, m) is called an assessment.[19] We say that an assessment is consistent if beliefs at every information set are derived from prior beliefs, observed histories, and strategies using Bayes' Law whenever possible.

We assume that for every player i and information set $I \in \mathscr{I}_i$, $S_i|_I$ is a chain. In other words, there is a binary relation $\succsim_i$ on Θ_i and $\succsim_{si}$ on S_i that is reflexive, antisymmetric, transitive, and complete.[20] We denote the asymmetric parts of $\succsim_i$ and $\succsim_{si}$ by $\succ_i$ and $\succ_{si}$, respectively.

Fix a player i and an information set $I \in \mathscr{I}_i$. Given an history $h \in I$, if the type profile is θ, player i plays $s_i \in S_i|_I$, and other players play $s_{-i} \in S_{-i}|_I$, payoff of player i can be written as $u_i(h, s_i, s_{-i}, \theta_i, \theta_{-i})$.[21]

Definition 8 (Increasing Differences)**.** We say that an original game G has increasing differences if for any $i \in N$, $I \in \mathscr{I}_i$, $\theta_i \succsim_i \theta_i'$ and $s_i \succsim_{si} s_i'$ imply that

$$u_i(h, s_i, s_{-i}, \theta_i, \theta_{-i}) - u_i(h, s_i', s_{-i}, \theta_i, \theta_{-i}) \geq u_i(h, s_i, s_{-i}, \theta_i', \theta_{-i}) - u_i(h, s_i', s_{-i}, \theta_i', \theta_{-i})$$

for all $h \in I$, $s_{-i} \in S_{-i}|_I$, and $\theta_{-i} \in \Theta_{-i}$. It is said to have *strictly increasing differences* if $\theta_i \succ_i \theta_i'$ and

[19] The original game as well as the games with contracts that we will define shortly have perfect recall and therefore for every behavior strategy there is an outcome equivalent mixed strategy and vice versa. Therefore, we are free to work with either the behavior or mixed strategies, whichever more convenient.

[20] Note that any chain is a lattice. Also, we omit the dependence of $\succsim_{si}$ on I to avoid notational clutter.

[21] We can do that by finding the terminal history $z(h, s_i, s_{-i})$ that is reached when players play according to (s_i, s_{-i}) after h and defining $u_i(h, s_i, s_{-i}, \theta_i, \theta_{-i}) = u_i(z(h, s_i, s_{-i}), \theta_i, \theta_{-i})$.

$s_i \succ_{si} s_i'$ imply that

$$u_i(h, s_i, s_{-i}, \theta_i, \theta_{-i}) - u_i(h, s_i', s_{-i}, \theta_i, \theta_{-i}) > u_i(h, s_i, s_{-i}, \theta_i', \theta_{-i}) - u_i(h, s_i', s_{-i}, \theta_i', \theta_{-i})$$

for all $h \in I$, $s_{-i} \in S_{-i}|_I$, and $\theta_{-i} \in \Theta_{-i}$.

In other words, an original game G has increasing differences if the payoff functions have increasing differences in $(s_i, \theta_i) \in S_i|_I \times \Theta_i$ at every information set I irrespective for how the other players play and what the types of the other players are. We assume that the original game G has strictly increasing differences. Examples of games with increasing differences include repeated ultimatum bargaining and chain store games.

The induced game with non-renegotiable and renegotiable contracts are straightforward generalizations of their counterparts for two stage games. Each player i independently offers a contract $f_i : Z \to \mathbb{R}$ to a distinct third-party t_i, who accepts or rejects it. In case of rejection the game ends, t_i receives a fixed payoff of $\delta_i \in \mathbb{R}$, and player i receives $-\infty$. In case of acceptance Nature chooses θ and players in N play G.

We assume that only player i observes his type θ_i, and that this is the only source of asymmetric information between i and the third-party t_i. In other words, at any point in the game both the main player and his third-party observe the same histories. The payoff functions are given by $v_i(f, z, \theta) = u_i(z, \theta) - f_i(z)$, $v_{t_i}(f, z, \theta) = f_i(z)$.

One of the conditions that strategies must satisfy in any perfect Bayesian equilibrium of the game with contracts, is that player i's strategy must be sequentially rational, or incentive compatible, under the contract. Increasing differences imply that this is equivalent to strategies being increasing. Define a mixed strategy as increasing if any pure strategy in its support is increasing in the type. We have the following counterpart to Proposition 2.

Proposition 5. *If the original game G has strictly increasing differences, then a mixed strategy is incentive compatible if and only if it is increasing.*

Proof. Omitted. □

The game is with renegotiable contracts if the contracting parties can renegotiate the contract at any point throughout the game. At any information set $I \in \mathscr{I}_i$ player i either offers a new contract $g_i : Z \to \mathbb{R}$ to the third-party t_i or chooses an action. If player i offers a new contract, the third-party either accepts or rejects it.

We retain the same definition of renegotiation-proofness. Fix a consistent assessment (σ, m). We say that a contract strategy pair (f_i, σ_i) is renegotiation-proof at (σ_{-i}, m), if whenever there is a contract g_i and an incentive compatible continuation play σ_i' such that player i of type θ_i strictly prefers (g_i, σ_i') over (f_i, σ_i) at information set I, there must exist a type θ_i' for which expected transfers under (f_i, σ_i) at least as high as the transfers under (g_i, σ_i').

In order to state the counterpart to our main result for renegotiation-proof contracts we need a few more definitions. Fix a consistent assessment (σ, m) and let $U_i^{\sigma,m}(\sigma_i', \theta_i|I)$ be the expected payoff of player i of type θ_i to playing mixed strategy $\sigma_i' \in \Delta(S_i|_I)$ conditional on reaching information set I. Similarly, let $F_i^{\sigma,m}(\sigma_i'|I)$ be the expected transfers.

For any $i \in N$, $I \in \mathscr{I}_i$, consistent assessment (σ, m), and $\sigma_i' : \Theta_i \to \Delta(S_i|_I)$ define $\vec{U}_i^{\sigma,m}(\sigma_i'|I)$ as a column vector with $2(n_i - 1)$ components, where component $2j-1$ is given by $U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta^j|I) -$

$U_i^{\sigma,m}(\sigma_i'(\theta_i^{j+1}),\theta^j|I)$ and component $2j$ is given by $U_i^{\sigma,m}(\sigma_i'(\theta_i^{j+1}),\theta^{j+1},|I) - U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta^{j+1}|I)$, $j = 1,2,\ldots,n_i-1$. Similarly, define $\vec{F}_i^{\sigma,m}(\sigma_i'|I)$ as the n_i vector whose jth component is given by $F_i^{\sigma,m}(\sigma_i'(\theta_i^j)|I)$, $j = 1,2,\ldots,n_i-1$.[22] Let

$$\widehat{\Sigma}_i(I,j,\sigma_i) = \{\sigma_i' : \Theta_i \to \Sigma_i|_I : \sigma_i' \text{ is increasing and } U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) > U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I)\}$$

be the set of all mixed strategies at information set I that is increasing and increases the payoff of player i type j over his payoff under σ_i.

Theorem 1 generalizes in a quite straightforward way:

Theorem 2. *Fix a consistent assessment (σ,m) and $i \in N$. $(f_i,\sigma_i) \in \mathscr{C} \times \Sigma_i$ is* renegotiation-proof *at (σ_{-i},m) if and only if for any $I \in \mathscr{I}_i$, $j = 1,\ldots,n_i$, and $\sigma_i' \in \widehat{\Sigma}_i(I,j,\sigma_i)$ there exists a $k \in \{1,2,\ldots,j-1\}$ such that*

$$U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) + \sum_{t=k}^{j-1} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t-1} \leq \vec{F}_i^{\sigma,m}(\sigma_i|I)_k - \vec{F}_i^{\sigma,m}(\sigma_i|I)_j \tag{8}$$

or there exists an $l \in \{j+1, j+2,\ldots,n_i\}$ such that

$$U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) + \sum_{t=j+1}^{l} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2(t-1)} \leq \vec{F}_i^{\sigma,m}(\sigma_i|I)_l - \vec{F}_i^{\sigma,m}(\sigma_i|I)_j \tag{9}$$

Proof. Omitted. □

Once the definition of a blocking type is appropriately modified (see Definition 7), Propositions 3 and 4 also generalize in a straightforward manner.

4.2 Interested Third-Party

In our model we assumed that the third-party has no interest in the outcome of the original game other than the transfer from (or to) player 2. This is not always the case with third-party contracts. For example, the government in its contractual relationship with a central bank is likely to be interested in the outcome of the game between the central bank and the public. Similarly, the European Union in its contractual relationships with Airbus is interested in the entry game played by Airbus and Boeing. We can easily think of many other instances of games with third-party contracts in which the third-party itself is interested in the outcome of the game. How do our results change if this is the case? The answer turns out to be straightforward and intuitive.

Let $u_3(a_1,a_2,\theta)$ be the third-party's payoff function so that under contract f his payoff would be $u_3(a_1,a_2,\theta) + f(a_1,a_2)$. We say that (f,b_2^*) is renegotiation-proof if for all $a_1 \in A_1$ and $\theta \in \Theta$ for which there exists an incentive compatible (g,b_2) such that

$$u_2(a_1,b_2(a_1,\theta),\theta) - g(a_1,b_2(a_1,\theta)) > u_2(a_1,b_2^*(a_1,\theta),\theta) - f(a_1,b_2^*(a_1,\theta))$$

[22] Note that independence of types across players implies that the beliefs of player i over h and θ_{-i}, i.e., $m(h,\theta_{-i}|I)$, do not depend on θ_i. For the same reason expected transfers to pure strategy s_i do not depend on θ_i.

there exists a $\theta' \in \Theta$ such that

$$u_3(a_1, b_2^*(a_1,\theta'),\theta') + f(a_1, b_2^*(a_1,\theta')) \geq u_3(a_1, b_2(a_1,\theta'),\theta') + g(a_1, b_2(a_1,\theta'))$$

In the model with neutral third-party a renegotiation opportunity arises whenever there is an increasing strategy that increases player 2's payoff $u_2(a_1,a_2,\theta)$, which is the total surplus available to player 2 and the third-party in that model. When the third-party is no longer neutral, total surplus available becomes $u_2(a_1,a_2,\theta) + u_3(a_1,a_2,\theta)$. Accordingly, a renegotiation opportunity arises whenever there is an increasing strategy that increases total surplus $u_2(a_1,a_2,\theta) + u_3(a_1,a_2,\theta)$. Therefore, we modify the definition of $\mathfrak{B}(a_1,i,b_2^*)$ as the set of strategies b_2 that are increasing and satisfy

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) + u_3(a_1, b_2(a_1,\theta^i),\theta^i) > u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) + u_3(a_1, b_2^*(a_1,\theta^i),\theta^i)\}.$$

We can now state the modified version of Theorem 1:

Theorem 3. *(f, b_2^*) is renegotiation-proof if and only if for any $a_1 \in A_1$, $i \in \{1,2,\ldots,n\}$, and $b_2 \in \mathfrak{B}(a_1,i,b_2^*)$ there exists a $k \in \{1,2,\ldots,i-1\}$ such that*

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) + u_3(a_1, b_2(a_1,\theta^k),\theta^k) - u_3(a_1, b_2^*(a_1,\theta^k),\theta^k) + \sum_{j=k}^{i-1} \vec{U}_2(a_1,b_2)_{2j-1} \leq f(a_1)_k - f(a_1)_i \quad (10)$$

or there exists an $l \in \{i+1, i+2,\ldots,n\}$ such that

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) + u_3(a_1, b_2(a_1,\theta^l),\theta^l) - u_3(a_1, b_2^*(a_1,\theta^l),\theta^l) \sum_{j=i+1}^{l} \vec{U}_2(a_1,b_2)_{2(j-1)} \leq f(a_1)_l - f(a_1)_i \quad (11)$$

Proof. Omitted. □

Note that an interested third-party introduces two changes into the result: First, a renegotiation opportunity arises only if it increases the total surplus rather than just player 2's payoff. This might in fact help a contract become renegotiation-proof, if, for example, the third-party and player 2 have completely opposite preferences. Second, compared with (4) and (5), inequalities (10) and (11) have extra terms on the left hand side, which might help or hurt a contract become renegotiation-proof depending upon the sign of those terms.

Again, once the definition of a blocking type is appropriately modified, Propositions 3 and 4 can be easily generalized to the case of non-neutral third-party.

4.3 Strong Renegotiation-Proofness

Our definition of renegotiation-proofness follows directly from the assumed game form for the renegotiation procedure, i.e., player 2, who is the the informed party, makes a new contract offer and the third-party, who is uninformed, accepts or rejects. In a renegotiation-proof equilibrium, the contract is never renegotiated, and therefore any renegotiation offer is an out-of-equilibrium event. This allows us to specify the beliefs of the third-party freely after a new contract offer. This may be found

unreasonable and a more plausible alternative could be to require beliefs satisfy the conditions specified in the intuitive criterion introduced by Cho and Kreps (1987).

In our setting, intuitive criterion requires that, given an equilibrium contract strategy pair (f, b_2^*) and following a renegotiation offer (g, b_2), beliefs put positive probability only on types for which (g, b_2) is not equilibrium-dominated, i.e., only on those types θ' for which

$$u_2(a_1, b_2(a_1,\theta'),\theta') - g(a_1, b_2(a_1,\theta')) \geq u_2(a_1, b_2^*(a_1,\theta'),\theta') - f(a_1, b_2^*(a_1,\theta'))$$

This leads to the following definition.

Definition 9 (Strong Renegotiation-Proofness)**.** We say that $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is *strongly renegotiation-proof* if for all $a_1 \in A_1$ and $\theta \in \Theta$ for which there exists an incentive compatible (g, b_2) such that

$$u_2(a_1, b_2(a_1,\theta),\theta) - g(a_1, b_2(a_1,\theta)) > u_2(a_1, b_2^*(a_1,\theta),\theta) - f(a_1, b_2^*(a_1,\theta)) \tag{12}$$

there exists a $\theta' \in \Theta$ such that

$$f(a_1, b_2^*(a_1,\theta')) \geq g(a_1, b_2(a_1,\theta')) \tag{13}$$

and

$$u_2(a_1, b_2(a_1,\theta'),\theta') - g(a_1, b_2(a_1,\theta')) \geq u_2(a_1, b_2^*(a_1,\theta'),\theta') - f(a_1, b_2^*(a_1,\theta')) \tag{14}$$

This is exactly the same as renegotiation-proofness except that it adds condition (14), which allows us to construct beliefs that satisfy intuitive criterion after any renegotiation offer. It can be shown that when we work with this definition, Theorem 1 needs to be modified as follows.

Theorem 4. *(f, b_2^*) is strongly renegotiation-proof if and only if for any $a_1 \in A_1$, $i \in \{1, 2, \ldots, n\}$, and $b_2 \in \mathfrak{B}(a_1, i, b_2^*)$ there exists a $k \in \{1, 2, \ldots, i-1\}$ such that*

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) - \min\{0, u_2(a_1, b_2(a_1,\theta^k),\theta^k) - u_2(a_1, b_2^*(a_1,\theta^k),\theta^k)\} + \sum_{j=k}^{i-1} \vec{U}_2(a_1, b_2)_{2j-1} \leq f(a_1)_k - f(a_1)_i \tag{15}$$

or there exists an $l \in \{i+1, i+2, \ldots, n\}$ such that

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - u_2(a_1 b_2^*(a_1,\theta^i),\theta^i) - \min\{0, u_2(a_1, b_2(a_1,\theta^l),\theta^l) - u_2(a_1, b_2^*(a_1,\theta^l),\theta^l)\} + \sum_{j=i+1}^{l} \vec{U}_2(a_1, b_2)_{2(j-1)} \leq f(a_1)_l - f(a_1)_i \tag{16}$$

Proof. Omitted. □

It is also easy to show Propositions 3 and 4 go through with a minor modification similar to the one made in Theorem 4.

5 Equilibrium Outcomes of Games with Contracts

There may be legal or technological constraints that might render contracts non-renegotiable and therefore outcomes that can be supported by non-renegotiable contracts are of interest on their own.

Furthermore, understanding non-renegotiable contracts will help place our results within the literature and allow us to isolate the effects of renegotiation. Similarly, and irrespective of whether a contract is renegotiable, there may be valid reasons why a contract maybe observable or unobservable. Legal contracts between a firm and a bank, or a government and an international body, and many compensation contracts are observable yet subject to renegotiation if the parties find it in their benefit to do so. Other contracts can be either secret or subject to renegotiation before the game begins, i.e., they can be unobservable. In this section we will present results regarding the outcomes that can be supported under different assumptions about the contracts.

5.1 Observable Contracts

Let us assume that the contract signed between player 2 and the third-party before the game begins is observable to player 1 but may or may not be renegotiated after player 1 moves in the game.

5.1.1 Non-renegotiable Contracts

If the contracts are observable but not renegotiable, then we can show that player 2 can obtain the best payoff possible given that player 2 plays an increasing strategy and player 1 best responds. More precisely, define the best Stackelberg payoff of player 2 as $\bar{U}_2^B = \max_{b_2 \in B_2^+} \max_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$ and the worst Stackelberg payoff as $\bar{U}_2^W = \max_{b_2 \in B_2^+} \min_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$.

Proposition 6. *If contracts are observable, then $\bar{U}_2^B - \delta$ can be supported with non-renegotiable contracts.*

Proof. In Section 8. □

The proof of this result is quite easy. In the definition of the best Stackelberg payoff, player 2 is playing the best increasing strategy, say b_2^*, given that player 1 is playing a best response that is most favorable for player 2. Proposition 2 implies that b_2^* is incentive compatible, i.e., there is a contract, say f^*, that makes it optimal to play. It is easy to show that there is a perfect Bayesian equilibrium of the game with observable and non-renegotiable contracts in which player 2 offers f^* with expected value δ, player 1 plays the most favorable best response to that, say b_1^*, and player 2 plays $b_2^*(b_1^*, \theta)$ after (f^*, b_1^*, θ). Expected payoff of player 2 in such an equilibrium is $\bar{U}_2^B - \delta$.

We can also show that player 2 cannot get a payoff that is smaller than his worst Stackelberg payoff.

Proposition 7. *If contracts are observable, then $\bar{U}_2^W - \delta$ is the smallest payoff that can be supported with non-renegotiable contracts.*

Proof. In Section 8. □

In order to see why let $\hat{b}_{2,a_1} \operatorname{argmin}_{b_2 \in B_2^+} U(a_1, b_2)$ for any $a_1 \in A_1$. In other words, for any a_1, $\hat{b}_{2,a_1}$ is the worst increasing strategy for player 1 that player 2 can play. Since $\hat{b}_{2,a_1}$ is increasing, it can be shown that there is a contract that makes it uniquely optimal to play. Now let $b_1^*(b_2) \in \operatorname{argmin}_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$, $b_2^* \in \operatorname{argmax}_{b_2 \in B_2^+} U_2(b_1^*(b_2), b_2)$, and $a_1^* = b_1^*(b_2^*)$. Note that $U_2(a_1^*, b_1^*) = \bar{U}_2^W$ and suppose, for contradiction, that player 2 gets a payoff that is strictly smaller than $\bar{U}_2^W - \delta$. We show that there exists a contract that makes it uniquely optimal to play $b_2^*(a_1^*, \theta)$ after a_1^* and $\hat{b}_{2,a_1}(a_1, \theta)$ after any other a_1. If Player 2 offers this contract, player 1 must play a best response to

b_2^*. This is because for any $a_1 \notin BR_1(b_1^*)$, we have $U_1(br_1(b_2^*), b_2^*) > U_1(a_1, b_2^*) \geq U_1(a_1, \hat{b}_{2,a_1})$. Therefore, deviation to such a contract yields a gross payoff of at least $U_2(a_1^*, b_1^*) = \bar{U}_2^W$ and a net payoff arbitrarily close to $\bar{U}_2^W - \delta$, a contradiction.

Of course, we have full characterization if player 1's best response correspondence is single-valued, i.e., $BR_1(b_2)$ is a singleton for any $b_2 \in B_2^+$: The unique equilibrium payoff of player 2 that can be supported with observable and non-renegotiable contracts is $\bar{U}_2^B - \delta$.

5.1.2 Renegotiable Contracts

If the contracts are observable and renegotiable, then player 2 can again achieve his Stackelberg payoff, except that the definition of this payoff must reflect the fact that player 2 plays a renegotiation-proof strategy. Define the best and worst renegotiation-proof Stackelberg payoffs of player 2 as $\bar{U}_2^{BR} = \max_{b_2 \in B_2^R} \max_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$ and the worst Stackelberg payoff as $\bar{U}_2^{WR} = \max_{b_2 \in B_2^R} \min_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$ and note that the difference in the definitions comes from the fact that player 2 has to play a renegotiation-proof strategy.

Proposition 8. *If contracts are observable, then $\bar{U}_2^{BR} - \delta$ can be supported with renegotiation-proof contracts.*

Proof. In Section 8. □

The proof of this result also constructs an equilibrium in which player 2 receives the best Stackelberg payoff that he can get by playing a renegotiation-proof strategy. There is however a complication in the proof compared with the proof of Proposition 6. When contracts are non-renegotiable any deviation from the contract that induces the best Stackelberg outcome under increasing strategies must still induce an increasing strategy. This implies that no deviation can yield a higher payoff. When contracts are renegotiable a deviation may or may not induce a renegotiation-proof strategy and hence we cannot tell whether such a deviation can yield a payoff that is strictly higher than the best Stackelberg payoff that can be obtained by a renegotiation-proof strategy. However, in the proof of Proposition 8 we construct an equilibrium in which any deviation obtained via renegotiation can also be obtained via a renegotiation-proof strategy, and this gives us the desired result.

We can again show that player 2 cannot get a payoff that is smaller than his worst Stackelberg payoff that can be obtained with renegotiation-proof strategies.

Proposition 9. *If contracts are observable, then $\bar{U}_2^{WR} - \delta$ is the smallest payoff that can be supported with non-renegotiable contracts.*

Proof. Proof of this result is similar to that of Proposition 7 and omitted. □

The above results provide sharp predictions for equilibrium outcomes of the games with observable contracts. In particular, they show that third-party contracts play the role of a commitment device to the extent that player 2's strategy respects the constraints brought about by incentive compatibility, in the case of non-renegotiable contracts, and renegotiation-proofness, in the case of renegotiable contracts. The implications of these results in terms of the equilibrium outcomes depend on the specifics of the original game. We present some of these implications in our running example and further in Section 6.

EXAMPLE: ENTRY DETERRENCE

Assume that $p \neq 2/3$ so that player 1's best response correspondence is single-valued:

$$br_1(FF) = O, br_1(AA) = E, br_1(FA) = \begin{cases} O, & p > 2/3 \\ E, & p < 2/3 \end{cases}$$

Remember that the set of incentive compatible strategies is $B_2^+ = \{FF, FA, AA\}$ and the set of renegotiation-proof strategies is $B_2^R = \{FA, AA\}$. Therefore, the Stackelberg payoff of player 2 given that he plays an incentive compatible strategy is m, which he achieves by playing FF. Proposition 6 and 7 imply that this is the unique payoff that can be supported with observable and non-renegotiable contracts. In other words, entry-deterrence is the unique equilibrium outcome. How about with RP contracts? If $p > 2/3$, then the Stackelberg payoff is m, obtained by playing FA, whereas if $p < 2/3$, FA does not deter entry and the best player 2 can do in this case is to play AA, with payoff $px + (1-p)z$. In other words, if $p > 2/3$ unique equilibrium outcome is entry-deterrence and if $p < 2/3$ unique equilibrium outcome is entry and accommodate.

5.2 Unobservable Contracts

We now assume that the initial contract between player 2 and the third party is not observable to player 1. Again there are two possibilities: the contract could be renegotiable or non-renegotiable.

5.2.1 Non-Renegotiable Contracts

If contracts are non-renegotiable, we have the following characterization.

Proposition 10. *A strategy profile (b_1^*, b_2^*) of the original game G can be supported with unobservable and non-renegotiable contracts if and only if (b_1^*, b_2^*) is a Bayesian Nash equilibrium of G and b_2^* is increasing in $(\succsim_\theta, \succsim_2)$.*

Proof. In Section 8 □

This result shows that unobservable third-party contracts potentially enlarges the set of outcomes that can arise in equilibrium. Furthermore, while earlier papers showed that, when there is no asymmetric information, any Nash equilibrium of the original game can be supported with unobservable contracts, this result shows that only the subset of Bayesian Nash equilibria in which the second player plays an increasing strategy can be supported if, instead, there is asymmetric information.

This result also has an immediate corollary in terms of the outcomes that can be supported. For any strategy profile $(b_1, b_2) \in A_1 \times A_2^{A_1 \times \Theta}$, we define an *outcome* $(a_1, a_2) \in A_1 \times A_2^{\Theta}$ of G as $a_1 = b_1$ and $a_2(\theta) = b_2(b_1, \theta)$. Define the individually rational payoff of player 1 as

$$\underline{U}_1^+ = \max_{a_1 \in A_1} \min_{b_2 \in B_2^+} U_1(a_1, b_2). \tag{17}$$

This is the best payoff player 1 can guarantee for herself in game G, given that player 2 plays an increasing strategy.[23] The following easily follows from Proposition 10.

[23] We should also note that this is different from the definition of individually rational payoff used in the repeated games literature, which is the minmax payoff rather than the maxmin payoff. The maxmin payoff is at most equal to the minmax payoff.

Corollary 1. *An outcome (a_1^*, a_2^*) of the original game G can be supported with unobservable and non-renegotiable contracts if and only if (1) $a_2^*(\theta) \in BR_2(a_1^*, \theta)$ for all θ and (2) $U_1(a_1^*, a_2^*) \geq \underline{U}_1^+$.*

Again, note that, in general, outcomes that are not perfect Bayesian equilibrium outcomes of the original game can also be supported. This can be achieved by writing a contract that leads player 2 to punish player 1 when he deviates from his equilibrium action. Since contracts cannot be conditioned on θ and u_2 has increasing differences, player 2 can only use punishment strategies that are increasing in θ. The best that player 1 can do by deviating is therefore given by $\underline{U}_1$, and his equilibrium payoff cannot be smaller than this payoff. This is condition (2). Condition (1), on the other hand, simply follows from the requirement that only Bayesian Nash equilibrium outcomes can be supported, and hence, player 2 must be best responding along the equilibrium path.

Note that if θ were contractible as well, we would not need to limit the punishment strategies to be increasing. In this case, condition (2) would have the individually rational payoff defined as $\max_{a_1 \in A_1} \min_{b_2 \in A_2^{A_1 \times \Theta}} U_1(a_1, b_2)$. In that case, the result would be the exact analog of those in models without asymmetric information, i.e., Koçkesen and Ok (2004) and Koçkesen (2007).

We should also note that there are interesting environments in which non-contractibility of θ does not restrict the set of outcomes that can be supported with non-renegotiable contracts. For example if player 1's payoff does not depend on θ, then the punishment does not have to depend on θ either. Therefore, one can simply use a constant punishment after each deviation, which would be increasing by construction. A second environment is *games with monotone externalities*, in which u_1 is increasing (or decreasing) in a_2. In this case, after any a_1, the harshest punishment is the lowest (or highest) a_2, which is constant and hence increasing.

5.2.2 Renegotiable Contracts

Suppose now that the contracts are unobservable and renegotiable. The counterpart to Proposition 10 is the following:

Proposition 11. *A strategy profile (b_1^*, b_2^*) of the original game G can be supported with unobservable and renegotiation-proof contracts if and only if (b_1^*, b_2^*) is a Bayesian Nash equilibrium of G and b_2^* is increasing and renegotiation-proof.*

Proof. In Section 8 □

This result too has an immediate corollary. Define the best payoff player 1 can guarantee for herself in game G, given that player 2 plays a renegotiation-proof strategy as $\underline{U}_1^R = \max_{a_1 \in A_1} \min_{b_2 \in B_2^R} U_1(a_1, b_2)$. We have the following corollary.

Corollary 2. *An outcome (a_1^*, a_2^*) of the original game G can be supported with unobservable and renegotiation-proof contracts if and only if (1) $a_2^*(\theta) \in BR_2(a_1^*, \theta)$ for all θ and (2) $U_1(a_1^*, a_2^*) \geq \underline{U}_1^R$.*

In order to apply the results on unobservable contracts the crucial piece of information is the individually rational payoff of player 1 given that player 2 plays an increasing or a renegotiation-proof strategy. We illustrate how this can be done in Section 6.1 for a large class of games that we call games with monotone externalities. We show that in those games the only thing that distinguishes the case of renegotiation-proof contracts from non-renegotiable contracts is that the highest type of player 2

must play a best response to any a_1 under renegotiation-proof contracts, whereas the only restriction is that he plays an increasing strategy in the case of non-renegotiable contracts.

EXAMPLE: ENTRY DETERRENCE

Individually rational payoffs of player 1 are given by

$$U_1^+ = \max_{a_1 \in A_1} \min_{b_2 \in B_2^+} U_1(a_1, b_2) = U_1(O, FF) = 0$$

$$U_1^R = \max_{a_1 \in A_1} \min_{b_2 \in B_2^R} U_1(a_1, b_2) = \begin{cases} U_1(O, FA) = 0, & p > 2/3 \\ U_1(E, AA) = 2 - 3p, & p < 2/3 \end{cases}$$

Corollary 1 implies that (O, FF) and (E, AA) can both be supported with unobservable and non-renegotiable contracts. Corollary 2 implies that if $p > 2/3$ both (O, FA) and (E, AA) can be supported with unobservable and RP contracts, whereas if $p < 2/3$ only (E, AA) can be supported.

6 Applications

6.1 Games with Monotone Externalities

We say that an original game G is a game with monotone externalities if player 1's payoff is monotonically increasing or decreasing in player 2's action, i.e., for any a_1 and θ, $a_2' \succsim_2 a_2$ implies either $u_1(a_1, a_2', \theta) \geq u_1(a_1, a_2, \theta)$ or $u_1(a_1, a_2', \theta) \leq u_1(a_1, a_2, \theta)$.[24] Such positive or negative externalities are very common in economic models. Indeed, the class of games that satisfy these conditions includes Stackelberg and entry games, sequential Bertrand games with differentiated products, monopolistic screening, and ultimatum bargaining, among others.

Assume that contracts are unobservable. Fix $a_1 \in A_1$, let $\underline{a}_2$ ($\overline{a}_2$) be the smallest (largest) element of A_2, and define

$$b_2^+(a_1, \theta) = \begin{cases} \underline{a}_2, & \forall \theta \text{ if } u_1(a_1, a_2, \theta) \text{ increasing in } a_2 \\ \overline{a}_2, & \forall \theta \text{ if } u_1(a_1, a_2, \theta) \text{ decreasing in } a_2 \end{cases}$$

Note that this strategy is increasing in θ and it is the harshest punishment player 2 can inflict upon player 1, i.e., $b_2^+ \in \operatorname{argmin}_{b_2 \in B_2^+} U_1(a_1, b_2)$ for all a_1. In other words, the individually rational payoff of player 1 given that player 2 plays an increasing strategy is given by $\underline{U}_1^+ = \max_{a_1} U_1(a_1, b_2^+)$. We can directly apply Corollary 1 and conclude that an outcome (a_1^*, a_2^*) of the original game G with monotone externalities can be supported with non-renegotiable contracts if and only if $a_2^*(\theta) \in BR_2(a_1^*, \theta)$ for all θ and $U_1(a_1^*, a_2^*) \geq \underline{U}_1^+$.

Which outcomes can be supported with unobservable and renegotiation-proof contracts? Proposition 3 implies that the harshest punishment strategy b_2^+ is not renegotiation proof.[25] Using Proposition 4 we can show that, if u_1 is increasing in a_2, the harshest renegotiation-proof punishment is to make the highest type of player 2 play a best response, while the other types play the smallest a_2 (see

[24] Note that player 1's payoff may be increasing in a_2 for some a_1 and decreasing for others.

[25] See Lemma 3 in Section 8, which shows that renegotiation-proofness of a strategy $b_2 \in A_2^{A_1 \times \Theta}$ implies that the highest (lowest, resp.) type does not have a profitable deviation to a higher (lower, resp.) action.

Lemma 2 in Section 8). Similarly, if u_1 is decreasing in a_2, the harshest renegotiation-proof punishment is to make the smallest type best respond and the other types play the largest a_2.

More precisely, let $b_2^n(a_1) \in \operatorname{argmin}_{a_2 \in BR_2(a_1,\theta^n)} u_1(a_1,a_2,\theta^n)$ and $b_2^1(a_1) \in \operatorname{argmin}_{a_2 \in BR_2(a_1,\theta^1)} u_1(a_1,a_2,\theta^1)$. Define the punishment strategy as

$$b_2^R(a_1,\theta) = \begin{cases} \underline{a}_2, & \text{if } u_1(a_1,a_2,\theta) \text{ increasing in } a_2 \text{ and } \theta \prec_\theta \theta^n \\ b_2^n(a_1), & \text{if } u_1(a_1,a_2,\theta) \text{ increasing in } a_2 \text{ and } \theta = \theta^n \\ \overline{a}_2, & \text{if } u_1(a_1,a_2,\theta) \text{ decreasing in } a_2 \text{ and } \theta \succ \theta^1 \\ b_2^1(a_1), & \text{if } u_1(a_1,a_2,\theta) \text{ decreasing in } a_2 \text{ and } \theta = \theta^1 \end{cases} \tag{18}$$

The best payoff that player 1 can achieve against this strategy is $\underline{U}_1^R = \max_{a_1} U_1(a_1, b_2^R)$. We can then apply Corollary 2 to games with monotone externalities.

Corollary 3. *An outcome (a_1^*, a_2^*) of an original game with monotone externalities can be supported with unobservable and renegotiation-proof contracts if and only if (1) $a_2^*(\theta) \in BR_2(a_1^*,\theta)$ for all θ and (2) $U_1(a_1^*, a_2^*) \geq \underline{U}_1^R$.*

Therefore, the effect of renegotiation in this environment is very clear. If the contracts are unobservable and non-renegotiable, then any outcome (a_1, a_2) in which player 2 best responds on the equilibrium path and punishes player 1 in the harshest possible way off-the-equilibrium can be supported. With unobservable and renegotiation-proof contracts player 2 cannot punish player 1 in the harshest possible way anymore: the highest (or the lowest) type must play a best response even off-the-equilibrium path.

We next apply these results to a game that has been a canonical example for issues related to commitment: Stackelberg and entry-deterrence games.

Quantity Competition and Entry-Deterrence

Consider a Stackelberg game in which firm 1 moves first by choosing an output level $q_1 \in Q_1$ and firm 2, after observing q_1, chooses its own output level $q_2 \in Q_2$. Inverse demand function is given by $P(q_1,q_2) = \max\{0, \alpha - q_1 - q_2\}$, where $\alpha > 0$, and we assume Q_i is a rich enough finite subset of $\mathbb{R}_+$ whose largest element is α.[26] Cost function of firm 1 is $C_1(q_1) = cq_1$, where c is common knowledge, whereas the cost function of firm 2 is $C_2(q_2) = \theta q_2$. We assume that $\theta \in \{\theta^1, \theta^2, \ldots, \theta^n\}$, where $n \geq 2$, is private information of firm 2 and $\theta^1 < \theta^2 < \cdots < \theta^n$. Firm 1 believes that the probability of θ^i is given by $p(\theta^i)$ and for ease of exposition we assume that expected value of θ is equal to c. The profit function of firm i is given by $\pi_i(q_1,q_2,\theta) = P(q_1,q_2)q_i - C_i(q_i)$ and we assume that both firms are profit maximizers.

To ensure positive output levels in equilibrium we assume that $\alpha > 2\theta^n - c$, in which case the (Stackelberg) equilibrium outcome of this game is given by

$$\left(q_1^s, q_2^s(\theta)\right) = \left(\frac{\alpha - c}{2}, \frac{\alpha - 2\theta + c}{4}\right)$$

Define the game G as follows: Let $A_1 = Q_1$ and $A_2 = \{-q_2 : q_2 \in Q_2\}$ and define $\succsim_i$ on A_i as $a_i \succsim_i a_i' \Leftrightarrow a_i \geq a_i'$ and $\succsim_\theta$ as $\theta \succsim_\theta \theta' \Leftrightarrow \theta \geq \theta'$. Let the payoff function of player i be given by $u_i(a_1,a_2,\theta) =$

[26] We introduce this assumption so that player 2 can choose a high enough output level to drive the price to zero.

$\pi_i(a_1, -a_2, \theta)$, for any $(a_1, a_2) \in A_1 \times A_2$. The game G is strategically equivalent to the Stackelberg game defined in the previous paragraph. It is also easy to show that u_2 has strictly increasing differences in (a_2, θ) and u_1 is increasing in a_2.

Let us first assume that contracts are unobservable. Since u_2 has strictly increasing differences and u_1 is increasing in a_2, we can apply Corollary 1 and Corollary 2 to characterize all the outcomes that can be supported with non-renegotiable as well as renegotiation-proof third-party contracts. In order to apply Corollary 1, we need to calculate the individually rational payoff of player 1, i.e., $\underline{U}_1^+$ as defined in equation (17). The harshest punishment firm 2 can inflict is to drive the price down to zero by producing α for any type θ. Since this is a constant (and hence an increasing) strategy, it follows that $\underline{U}_1 = 0$. In other words, any outcome $(a_1^*, a_2^*(\theta))$ such that firm 2 best responds to a_1^* and firm 1 gets at least zero profit can be supported with non-renegotiable contracts. In particular, entry can be deterred with non-renegotiable contracts.

Can entry be deterred with renegotiation-proof contracts? In order to apply Corollary 2, we need to first calculate player 1's individually rational payoff given that player 2 plays a renegotiation-proof strategy. The discussion above implies that the harshest punishment is obtained when the highest type of player 2 best responds while the other types choose the lowest a_2, i.e., $a_2 = -\alpha$. Player 1's expected payoff when player 2 plays this strategy is given by $\frac{1}{2} p(\theta^n)\,(\alpha + \theta^n - a_1)\,a_1 - c a_1$. Its maximum, i.e., player 1's individually rational payoff, is therefore equal to

$$\underline{U}_1^R = \begin{cases} 0, & p(\theta^n)(\alpha+\theta^n) - 2c \leq 0 \\ \frac{(p(\theta^n)(\alpha+\theta^n)-2c)^2}{8p(\theta^n)}, & p(\theta^n)(\alpha+\theta^n) - 2c > 0 \end{cases}$$

Condition (1) of Corollary 2 requires that $a_2^*(\theta) = \frac{a_1^*+\theta-\alpha}{2}$ for all θ, and hence $U_1(a_1^*, a_2^*) = \frac{1}{2}(\alpha - c - a_1^*)a_1^*$. Therefore, by condition (2), any outcome such that $\frac{1}{2}(\alpha - c - a_1^*)a_1^* \geq \underline{U}_1^R$ can be supported.

Also note that if $p(\theta^n)(\alpha+\theta^n) - 2c > 0$, then $\underline{U}_1^R$ is strictly positive. This implies that entry cannot be deterred if $p(\theta^n)(\alpha+\theta^n) - 2c > 0$. Therefore, we have the following result:

Corollary 4. *Entry can be deterred with unobservable and non-renegotiable contracts. It can be deterred with unobservable and renegotiation-proof contracts if and only if* $p(\theta^n)(\alpha+\theta^n) - 2c \leq 0$.

Now let us assume that contracts are observable. The best payoff that player 2 can obtain in the original game is the monopoly outcome, i.e., $a_1^* = 0$ and $a_2^*(\theta) = \frac{a_1^*+\theta-\alpha}{2}$. If contracts are non-renegotiable, then Player 2 can obtain this outcome exactly the same way as with unobservable contracts: If player 1 plays any $a_1 > 0$, punish him by flooding the market, i.e., choose $a_2 = -\alpha$. In other words, with observable and non-renegotiable contracts the unique outcome is the monopoly (entry-deterrence) outcome.

Could player 2 achieve the monopoly outcome with renegotiation-proof contracts? The above analysis implies that the answer is yes as long as $\underline{U}_1^R = 0$, i.e., $p(\theta^n)(\alpha+\theta^n) - 2c \leq 0$. It is easy to see that if this condition holds, then the unique equilibrium outcome that can be achieved with observable and renegotiation-proof contracts is the monopoly outcome. If, on the other hand, $p(\theta^n)(\alpha+\theta^n) - 2c > 0$, then monopoly outcome can no longer be supported with renegotiation-proof contracts. However, player 2 can obtain the following outcome: Player 1 plays a_1^*, where a_1^* is the smallest a_1

such that

$$\frac{1}{2}(\alpha - c - a_1^*)a_1^* \geq \frac{\left(p(\theta^n)(\alpha+\theta^n) - 2c\right)^2}{8p(\theta^n)}$$

and player 2 plays $a_2^*(\theta) = \frac{a_1^*+\theta-\alpha}{2}$ for all θ. In this outcome, player 1 produces the smallest amount consistent with player 2 punishing with the harshest possible renegotiation-proof strategy off-the-equilibrium and best responding on the equilibrium path.

Dewatripont (1988) has analyzed a similar entry game and showed that entry can be deterred with renegotiation-proof contracts under certain conditions. His conditions are different from ours because he uses a different renegotiation-proofness concept, namely durability, first introduced by Holmstrom and Myerson (1983). A decision rule is durable if and only if the parties involved would never unanimously approve a change from this decision rule to any other decision rule. Holmstrom and Myerson showed that this is equivalent to interim incentive efficiency when there is only one player with private information. In our context, only player 2 has private information and hence a contract-strategy pair (f, b_2^*) is interim incentive efficient (and therefore durable) if and only if there is no $a_1 \in A_1$ and an incentive compatible (g, b_2) such that after a_1 every type of player 2 and the third-party do better under (g, b_2), with at least one doing strictly better.

We have a characterization of durable strategies for the two-type case, i.e., when $\Theta = \{\theta^1, \theta^2\}$, and even in that case, the relationship between our concept of renegotiation-proofness and durability turns out to be quite subtle. It is not difficult to show that neither concept implies the other one in general. However, in games with monotone externalities it can be shown that durability implies renegotiation-proofness. The entry-deterrence game is a game with monotone externalities, and therefore, if entry can be deterred with durable contracts, it can also be deterred with renegotiation-proof contracts. In fact, in the entry-deterrence game player 2's payoff function is single-peaked and for such environments we have a complete characterization of durable outcomes that is particularly easy to apply. Using this characterization, we can show that the relationship between durability and renegotiation-proofness is strict.

Proposition 12. *In the entry-deterrence game with two types, if $p_1(\theta^2 + \alpha) > (\theta^2 - \theta^1)$, then entry can be deterred with renegotiation-proof contracts but not with durable contracts.*

Proof. Available upon request. □

Remember that the harshest renegotiation-proof punishment strategy of the incumbent is to flood the market if entry occurs, except for the highest type (type θ^2), who has to best respond. Durability still requires that the highest type best responds. The difference is that flooding the market for type θ^1 is not a durable strategy: There is a restriction on how much the incumbent can produce in response to entry, which is condition (d) of Proposition 1 in Dewatripont (1988).

6.2 Commitment to monetary policy

In this section we apply our results to the problem of credibility of monetary policy (Kydland and Prescott (1977), Barro and Gordon (1983)). Consider the following model. Public (player 1) forms expectations of inflation rate π^e after which the economy receives a zero mean shock θ and the policy maker (player 2) sets the inflation rate π. Unemployment rate is given by an expectation augmented

Phillips curve with a stochastic shock, $U = U^* - \alpha(\pi - \pi^e) + \theta$, where $\alpha > 0$ and U^* is the natural rate of unemployment. The policy maker's objective function is $-U^2 - \beta\pi^2$ whereas the public aims to accurately forecast the inflation rate.

We can study this model in our framework by defining the following game G. Let $a_1 = \pi^e$, $a_2 = \pi$, $A_1 = A_2 = \mathbb{R}$, $\theta \in \{\theta^1, \theta^2, \ldots, \theta^n\}$, and $\theta^1 < \theta^2 < \cdots < \theta^n$.[27] Let probability of θ^i be $p(\theta^i)$ and $\sum_i p(\theta^i)\theta^i = 0$. Define $\succsim_2$ on A_2 as $a_2 \succsim_2 a_2' \Leftrightarrow a_2 \geq_2 a_2'$ and $\succsim_\theta$ as $\theta \succsim_\theta \theta' \Leftrightarrow \theta \geq \theta'$. Let the payoff functions be $u_1(\pi^e, \pi, \theta) = -(\pi^e - \pi)^2$, $u_2(\pi^e, \pi, \theta) = -(U^* - \alpha(\pi - \pi^e) + \theta)^2 - \beta\pi^2$.[28] It is easy to show that u_2 has strictly increasing differences in (π, θ).

With an abuse of notation denote the public strategy as $\pi^e \in \mathbb{R}$ and the policy maker's strategy as $\pi(\pi^e, \theta)$. The literature looks at a particular class of equilibria known as *rational expectations equilibria* in which the public always predicts the inflation rate accurately: $\pi^e = \sum_i p(\theta^i)\pi(\pi^e, \theta^i)$. In order to allow comparison with the extant literature, we restrict attention to strategies of policy maker that are separable in θ and π^e, i.e., $\pi(\pi^e, \theta) = \tilde{\pi}(\pi^e) + \hat{\pi}(\theta)$, and such that $\frac{d\tilde{\pi}(\pi^e)}{d\pi^e} \neq 1$.[29] The equilibrium outcome of the original game G, then, is the same as what is known as the rational expectations equilibrium in the literature:

Proposition 13. *The unique perfect Bayesian equilibrium of the original game G is given by $\pi^e = \frac{\alpha}{\beta}U^*$ and $\pi(\pi^e, \theta) = \frac{\alpha^2}{\alpha^2+\beta}\pi^e + \frac{\alpha}{\alpha^2+\beta}(U^* + \theta)$, with outcome $\pi^e = \frac{\alpha}{\beta}U^*$ and $\pi(\theta) = \frac{\alpha}{\beta}U^* + \frac{\alpha}{\alpha^2+\beta}\theta$*

Proof. Omitted. □

As it is well known this outcome is inefficient and the policy maker would rather commit to the policy $\pi(\theta) = \frac{\alpha}{\alpha^2+\beta}\theta$, which results in zero expected inflation. Contracts have been shown to achieve this commitment (Walsh (1995) and Persson and Tabellini (1993)) and our next result reproduces this result.

Proposition 14. *The unique outcome supported by observable and non-renegotiable third party contracts is $\pi^e = 0$ and $\pi(\theta) = \frac{\alpha}{\alpha^2+\beta}\theta$.*

Proof. In Section 8. □

However, the interpretation of this result has been criticized by McCallum (1995, 1997), whose "second fallacy" argument can be rephrased in terms of renegotiation: How credible can be a commitment through a contract when both parties have an incentive not to enforce it? The next result states what can be obtained with renegotiable contracts and therefore addresses the criticism by McCallum.

Proposition 15. *The unique outcome supported by observable and renegotiable third party contracts is*

$$\pi^e = \frac{\alpha p_n}{\alpha^2(1-p_n)+\beta}U^* \tag{19}$$

$$\pi(\theta) = \begin{cases} \frac{\alpha}{\alpha^2+\beta}\theta, & \theta < \theta^n \\ \frac{\alpha}{\alpha^2+\beta}\theta + \frac{\alpha U^*}{\alpha^2(1-p_n)+\beta}, & \theta = \theta^n \end{cases} \tag{20}$$

[27] In our model we assumed that A_1 and A_2 are finite. In this application, we are assuming that they have the cardinality of the continuum. In order for the results of the paper to go through, we need to assume that player 2 is restricted to offer contracts f after which player 1 and 2's optimal continuation plays are well defined.

[28] We assume that the third-party is neutral. If not, we can modify our results as suggested in Section 4.2.

[29] See Lemma 4 for the reasons behind these restrictions.

Proof. In Section 8. □

Proposition 15 tells us that if contracts can be renegotiated, the inflationary bias cannot be completed eliminated. However, the average inflation rate is less than the one produced by the discretionary monetary policy (i.e., the one without commitment). In other words, the possibility to renegotiate the contracts reduces the commitment role of contracts, but does not completely eliminate it.

An interesting feature of the equilibrium outcome with renegotiable contracts is that the realized inflation differ from the non-renegotiable case only in the "worst state" of the world θ^n. This implies that the average inflation rate is increasing in the probability p_n. As p_n approaches one, π^e approaches $\frac{\alpha}{\beta}U^*$ (the inflation rate without commitment), while as p_n approaches zero, the average inflation rate approaches zero (the inflation rate under full commitment).

Another observation is that in this model, unlike in the entry-deterrence model, unobservable contracts (whether they are renegotiable or not) are completely ineffectual in solving the problem of credibility of monetary policy. Indeed, if the contracts are unobservable, then the only outcome of the original game that can be supported is the equilibrium outcome of the original game given by the rational expectations equilibrium without commitment.[30]

7 Conclusion

In this paper we studied the extent to which renegotiation reduces the commitment value of third-party contracts in extensive form games with incomplete information. In order to isolate the effect of renegotiation, we compared the outcomes that can be supported with non-renegotiable contracts with those that can be supported with renegotiable contracts, and showed that when renegotiation cannot be excluded the committing party is restricted to choose renegotiation-proof strategies. More precisely we showed that when the contracts are observable to the first mover, then the second mover obtains her Stackelberg payoff that can be achieved with renegotiation-proof strategies. When the contracts are not observable, then some kind of a "folk theorem" is true: Any outcome in which the second mover best responds to the first mover's action and the first mover obtains his individually rational payoff can be supported. In the definition of the individually rational payoff, player 2 is restricted to using increasing and renegotiation-proof strategies.

The restriction imposed by renegotiation-proofness is particularly transparent in some interesting economic applications. In particular, we apply our results to games with monotone externalities, i.e., games in which the first mover's payoff is monotonically increasing (or decreasing) in the second mover's action, and to a model of credibility of monetary policy (which is not a game with monotone externalities). The general message is that, while renegotiation does limit the commitment value of third-party contracts and hence, in general, has bite, renegotiable contracts still have an important strategic value. In games with monotone externalities, player 2 can induce player 1 to play player 2's favorite action by punishing him if he plays some other action. If player 1's payoff is increasing in player 2's action, then the worst punishment is to play the lowest possible action for every type of player 2. However, this is not a renegotiation-proof strategy. The worst renegotiation-proof punishment is to best respond for the highest type while the others play the smallest action.

[30] This follows immediately from Propositions 10 and 11 and the restriction in our strategy space.

In the problem of credibility of monetary policy, it is well-known that with non-renegotiable contracts it is possible to eliminate the inflationary bias and to achieve the socially optimal policy. We show that if contracts are renegotiable, the rate of inflation in the worst state of the world (e.g., a financial crisis) must be the discretionary one, while the rate of inflation in all the other states of the world are the same as the rate of inflation chosen in the case of non-renegotiable contracts.

The class of games considered in this paper is quite large and we believe that there are many other interesting models or economic problems that can benefit from the analysis and the methods developed in this paper.

8 Proofs

In the game with non-renegotiable contracts, player 2 has an information set at the beginning of the game, which we identify with the null history $\emptyset$, and an information set for each $(f,\theta,a_1) \in \mathscr{C}\times\Theta\times A_1$, where $\mathscr{C} = \mathbb{R}^{A_1\times A_2}$. Player 3 has an information set for each $f \in \mathscr{C}$. If contracts are unobservable, then player 1 has only one information set, given by $\mathscr{C}$. If contracts are observable, then player 1 has an information set for each $f \in \mathscr{C}$. In the game with renegotiable contracts, player 2 has additional information sets corresponding to each history (f,θ,a_1,g,y) and (f,θ,a_1,g,n) and player 3 has an additional information set of each (f,a_1,g), which we denote by $I_3(f,a_1,g)$.

Proof of Proposition 1. Take a perfect Bayesian equilibrium (PBE) (β,μ) of the game $\Gamma_{RO}(G)$ that induces the strategy profile (b_1^*,b_2^*) of G and let f^* be the contract offered by player 2 to the third party in the first stage, that is $\beta_2(\emptyset) = f^*$. For each a_1, partition the set Θ in two sets, $\Theta_{a_1}^R$ and $\Theta_{a_1}^{NR} = \Theta - \Theta_{a_1}^R$, where $\theta \in \Theta_{a_1}^R$ if the contract f^* is renegotiated after (a_1,θ); that is $\theta \in \Theta_{a_1}^R$ if $\beta_2(f^*,\theta,a_1) = g$ for some $g \in \mathscr{C}$ and $\beta_3(f^*,\theta,a_1,g) = y$. If $\Theta_{a_1} = \emptyset$ for all a_1, then (β,μ) is renegotiation-proof and we are done. For each a_1 and $\theta \in \Theta_{a_1}^R$ denote with $g_{(\theta,a_1)}$ the contract that is offered and accepted after (f^*,θ,a_1), that is $\beta_2(f^*,\theta,a_1) = g_{(\theta,a_1)}$ and $\beta_3(f^*,\theta,a_1,g_{(\theta,a_1)}) = y$. Also, for each a_1 and $\theta \in \Theta_{a_1}^R$ denote with $b_{g_{(\theta,a_1)}}$ the strategy that player 2 chooses after the contract $g_{(\theta,a_1)}$ is offered and accepted, i.e., $\beta_2(f^*,\theta,a_1,g_{(\theta,a_1)},y) = b_{g_{(\theta,a_1)}}$. For each a_1 and $\theta \in \Theta_{a_1}^{NR}$, denote with $b_{f^*}(a_1,\theta)$ the strategy that player 2 chooses after (f^*,a_1,θ), that is $\beta_2(f^*,\theta,a_1) = b_{f^*}(a_1,\theta)$. Define

$$h^*(a_1,a_2) = \begin{cases} f^*(a_1,a_2), & \text{if } \exists\theta\in\Theta_{a_1}^{NR} \text{ such that } a_2 = b_{f^*}(a_1,\theta) \\ g_{(\theta,a_1)}(a_1,a_2), & \text{if } \exists\theta\in\Theta_{a_1}^{R} \text{ such that } a_2 = b_{g_{(\theta,a_1)}}(a_1,\theta) \\ \infty & \text{otherwise} \end{cases}$$

$$b_{h^*}(a_1,\theta) = \begin{cases} b_{f^*}(a_1,\theta), & \text{if } \theta\in\Theta_{a_1}^{NR} \\ b_{g_{(\theta,a_1)}}(a_1,\theta), & \text{if } \theta\in\Theta_{a_1}^{R} \end{cases}$$

Consider the following assessment, denoted (β',μ'), in which player 2 offers h^* and plays according to b_{h^*} without attempting to renegotiate. Player 1 has the same beliefs as in (β,μ), $\beta_1'(h^*) = \beta_1(f^*)$, and $\beta_1'(f) = \beta_1(f)$ for $f \neq h^*$; Player 3 has the same beliefs as in (β,μ), $\beta_3'(h^*) = y$, and $\beta_3'(I) = \beta_3(I)$ at any other information set I; $\beta_2'(\emptyset) = h^*$, $\beta_2'(h^*,\theta,a_1) = b_{h^*}(a_1,\theta)$, and $\beta_2'(I) = \beta_2(I)$ at any the other information set I.

We will show that (β',μ') is a renegotiation-proof PBE of $\Gamma_{RO}(G)$ and induces (b_1^*,b_2^*). Since

$\beta_1'(h^*) = b_1^*$ and $\beta_2'(h^*,\theta,a_1) = b_{h^*}(a_1,\theta) = b_2^*(a_1,\theta)$, the assessment (β',μ') induces (b_1^*,b_2^*), and since h^* is not renegotiated after any (a_1,θ), it is renegotiation-proof.

In order to prove that (β',μ') is a PBE, we first show that $h^*(a_1,a_2)$ is well defined. Indeed, if there exist $a_1 \in A_1$, $\theta \in \Theta_{a_1}^R$ and a $\theta' \in \Theta_{a_1}^{NR}$ such that $b_{f^*}(a_1,\theta') = b_{g_{(\theta,a_1)}}(a_1,\theta) = a_2$, then $f^*(a_1,a_2) = g_{(\theta,a_1)}(a_1,a_2)$. Suppose, for contradiction, that $f^*(a_1,a_2) \neq g_{(\theta,a_1)}(a_1,a_2)$. Assume first that $f^*(a_1,a_2) > g_{(\theta,a_1)}(a_1,a_2)$. In the PBE (β,μ), after (f^*,a_1,θ'), player 2 chooses $b_{f^*}(a_1,\theta') = a_2$ and gets a payoff of $u_2(a_1,a_2) - f^*(a_1,a_2)$. However by choosing $g_{(\theta,a_1)}$ and then playing a_2, player 2 would get $u_2(a_1,a_2) - g_{(\theta,a_1)}(a_1,a_2) > u_2(a_1,a_2) - f^*(a_1,a_2)$.[31] This implies that player 2 has a profitable deviation, contradicting that (β,μ) is a PBE.

Assume now that $f^*(a_1,a_2) < g_{(\theta,a_1)}(a_1,a_2)$. In (β,μ), after (f^*,a_1,θ), player 2 renegotiates by offering $g_{(\theta,a_1)}$, which is accepted, and 2 chooses a_2. He receives a payoff of $u_2(a_1,a_2,\theta) - g_{(\theta,a_1)}(a_1,a_2)$, which is smaller than the payoff he could receive by playing a_2. Therefore player 2 has a profitable deviation contradicting that (β,μ) is a PBE.

Next, we prove that (h^*,b_{h^*}) is incentive compatible, that is, for all a_1 and θ:

$$u_2(a_1,b_{h^*}(a_1,\theta),\theta) - h^*(a_1,b_{h^*}(a_1,\theta)) \geq u_2(a_1,a_2,\theta) - h^*(a_1,a_2) \text{ for any } a_2 \in A_2 \tag{21}$$

First, assume $\theta \in \Theta_{a_1}^{NR}$ so that $b_{h^*}(a_1,\theta) = b_{f^*}(a_1,\theta)$. If a_2 is such that $\exists \theta' \in \Theta_{a_1}^{NR}$, with $a_2 = b_{h^*}(a_1,\theta') = b_{f^*}(a_1,\theta')$, then (21) becomes

$$u_2(a_1,b_{f^*}(a_1,\theta),\theta) - f^*(a_1,b_{f^*}(a_1,\theta)) \geq u_2(a_1,b_{f^*}(a_1,\theta'),\theta) - f^*(a_1,b_{f*}(a_1,\theta'))$$

which holds by optimality of $b_{f^*}(a_1,\theta)$. If a_2 is such that there exist $\theta' \in \Theta_{a_1}^R$ such that $a_2 = b_{h^*}(a_1,\theta') = b_{g_{(\theta',a_1)}}(a_1,\theta')$, then (21) holds because otherwise after (a_1,θ) player 2 could offer $g_{(\theta',a_1)}$ and once accepted, play $a_2 = b_{g_{(\theta',a_1)}}(a_1,\theta')$. This yields a payoff of $u_2(a_1,a_2,\theta) - g_{(\theta',a_1)}(a_1,a_2) > u_2(a_1,b_{f^*}(a_1,\theta),\theta) - f^*(a_1,b_{f^*}(a_1,\theta))$, implying that player 2 has a profitable deviation after (f^*,θ,a_1) and contradicting that (β,μ) is a perfect Bayesian equilibrium.[32] Finally, if there is no $\theta' \in \Theta$, with $a_2 = b_{h^*}(a_1,\theta')$, then (21) holds trivially.

Next, assume $\theta \in \Theta_{a_1}^R$ so that $b_{h^*}(a_1,\theta) = b_{g_{(\theta,a_1)}}(a_1,\theta)$. If a_2 is such that $\exists \theta' \in \Theta_{a_1}^{NR}$, with $a_2 = b_{h^*}(a_1,\theta') = b_{f^*}(a_1,\theta')$, then (21) holds, because, otherwise after (a_1,θ) player 2 could choose $a_2 = b_{f^*}(a_1,\theta')$ (without renegotiating f^*) and could get a payoff of

$$u_2(a_1,b_{f^*}(a_1,\theta'),\theta) - f^*(a_1,b_{f^*}(a_1,\theta')) > u_2(a_1,b_{g_{(\theta,a_1)}}(a_1,\theta),\theta) - g_{(\theta,a_1)}(a_1,b_{g_{(\theta,a_1)}}(a_1,\theta)).$$

This implies that player 2 has a profitable deviation after history (f,θ,a_1), contradicting that (β,μ) is a PBE. If a_2 is such that there exists a $\theta' \in \Theta_{a_1}^R$ with $a_2 = b_{h^*}(a_1,\theta') = b_{g_{(\theta',a_1)}}(a_1,\theta')$, then (21) holds, because, otherwise player 2 could offer $g_{(\theta',a_1)}$ after (a_1,θ) and once accepted play $a_2 = b_{g_{(\theta',a_1)}}(a_1,\theta')$. This yields

$$u_2(a_1,a_2,\theta) - g_{(\theta',a_1)}(a_1,a_2) > u_2(a_1,b_{g_{(\theta,a_1)}}(a_1,\theta),\theta) - g_{(\theta,a_1)}(a_1,b_{g_{(\theta,a_1)}}(a_1,\theta))$$

and shows that player 2 has a profitable deviation after history (f,θ,a_1), contradicting that (β,μ) is a

[31] Note that $g_{(\theta,a_1)}(a_1,a_2)$ is accepted after a_1 in (β,μ) since $\theta \in \Theta_{a_1}^R$ by hypothesis.

[32] Note that $g_{(\theta',a_1)}$ is accepted by construction.

PBE.[33] Finally, if there is no $\theta' \in \Theta$, with $a_2 = b_{h^*}(a_1, \theta')$, then (21) holds trivially.

We now verify that (β', μ') is indeed a perfect Bayesian equilibrium of $\Gamma_{RO}(G)$. For player 1, $\beta'_1(h^*) = \beta_1(f^*)$ is optimal because for any a_1, h^* and f^* induce the same outcome. Similarly, for any $f \neq h^*$, β'_{-1} and β_{-1} induce the same continuation play, which implies that $\beta'_1(f) = \beta(f)$ is optimal. For player 3, $\beta_3(h^*) = y$ is optimal because $\beta_3(f^*) = y$ is optimal and h^* and f^* induce the same continuation play; $\beta'_3(f, a_1, g) = \beta_3(f, a_1, g)$ is optimal because $\beta'_2(f, \theta, a_1, g, x) = \beta_2(f, \theta, a_1, g, x)$, for $x = y, n$.

For player 2, $\beta'_2(f, \theta, a_1, g, x) = \beta_2(f, \theta, a_1, g, x)$, for $x = y, n$, is optimal by construction. Similarly, for any $f \neq h^*$, $\beta'_2(f, \theta, a_1) = \beta_2(f, \theta, a_1)$ is optimal since continuation plays after (f, θ, a_1) are the same under β and β'. Now consider optimality of $\beta'_2(h^*, \theta, a_1) = b_{h^*}(a_1, \theta)$. Suppose first that $\theta \in \Theta^{NR}_{a_1}$. Incentive compatibility of (h^*, b_{h^*}) implies that there is no profitable deviation to a different action a_2. There cannot be a profitable deviation to offering a contract either, because continuation play after any such contract is the same under β and β' and hence if there was such a contract, then playing according to $b_{f^*}(a_1, \theta)$ would not be optimal under β. Now suppose $\theta \in \Theta^{R}_{a_1}$. Again, incentive compatibility of (h^*, b_{h^*}) implies that there is no profitable deviation to a different action a_2. Suppose, for contradiction, that there is a profitable deviation to offering a contract g'. If g' is rejected, then incentive compatibility of (h^*, b_{h^*}) is contradicted. If g' is accepted, then it is also accepted in under β. If this were to bring a higher payoff, then renegotiating to $g_{(\theta, a_1)}$ would not be optimal in β. Finally, $\beta'_2(\emptyset) = h^*$ is optimal because h^* and f^* yield the same expected payoffs and any $f \neq h^*$ induces the same continuation play under β' and β. Consistency of beliefs follows easily.

The proof for a perfect Bayesian equilibrium (β, μ) of the game $\Gamma_{RU}(G)$ is virtually identical and omitted. □

We first introduce some notation. Let the number of elements in Θ be equal to n and order its elements so that $\theta^n \succsim_\theta \theta^{n-1} \succsim_\theta \cdots \theta^2 \succsim_\theta \theta^1$. Let e_i be the i^{th} standard basis row vector for $\mathbb{R}^n$ and define the row vector $d_i = e_i - e_{i+1}, i = 1, 2, \ldots, n-1$. Let D be the $2(n-1) \times n$ matrix whose row $2i-1$ is d_i and row $2i$ is $-d_i$, $i = 1, \ldots, n-1$. For any $a_1 \in A_1$ and $b_2 \in A_2^{A_1 \times \Theta}$ define $\vec{U}_2(a_1, b_2)$ as a column vector with $2(n-1)$ components, where component $2i-1$ is given by $u_2(a_1, b_2(a_1, \theta^i), \theta^i) - u_2(a_1, b_2(a_1, \theta^{i+1}), \theta^i)$ and component $2i$ is given by $u_2(a_1, b_2(a_1, \theta^{i+1}), \theta^{i+1}) - u_2(a_1, b_2(a_1, \theta^i), \theta^{i+1})$, $i = 1, 2, \ldots, n-1$.

Notation 1. Given two vectors $x, y \in R^n$

1. $x \geq y$ if and only if $x_i \geq y_i$, for all $i = 1, 2, \ldots, n$;
2. $x > y$ if and only if $x_i \geq y_i$, for all $i = 1, 2, \ldots, n$ and $x \neq y$;
3. $x \gg y$ if and only if $x_i > y_i$, for all $i = 1, 2, \ldots, n$.

Similarly for $\leq$, $<$, and $\ll$.

For any $a_1 \in A_1$, $b_2 \in A_2^{A_1 \times \Theta}$ and $f \in \mathscr{C}$, let $f(a_1, b_2)$ be the column vector with n components, where i^{th} component is given by $f(a_1, b_2(a_1, \theta^i)), i = 1, 2, \ldots, n$. For any strategy profile (b_1, b_2) of G define the expected transfer from player 2 to the third-party as $F(b_1, b_2) = \sum_{\theta \in \Theta} p(\theta) f(b_1, b_2(b_1, \theta))$.

Proof of Proposition 6. Let $b_2^* \in \operatorname{argmax}_{b_2 \in B_2^+} \max_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$ and $b_1^* \in \operatorname{argmax}_{b_1 \in BR_1(b_2^*)} U_2(b_1, b_2^*)$. Note that $\bar{U}_2^B = U_2(b_1^*, b_2^*)$. Since b_2^* is increasing by construction, there exists a contract f^* such

[33] Again note that $g_{(\theta', a_1)}$ is accepted under (β, μ) by construction.

that (f^*, b_2^*) is incentive compatible and $F(b_1^*, b_2^*) = \delta$. For any $f \in \mathscr{C}, a_1 \in A_1, \theta \in \Theta$ choose $b_{2,f} \in \text{argmax}_{a_2 \in A_2} u_2(a_1, a_2, \theta) - f(a_1, a_2)$ and $b_{1,f} \in \text{argmax}_{a_1' \in A_1} U_1(a_1', b_{2,f})$.

Consider the following assessment (β, μ) of $\Gamma(G)$: $\beta_2(\emptyset) = f^*$, $\beta_3(f^*) = y$, $\beta_3(f) = y$ iff $F(b_{1,f}, b_{2,f}) \geq \delta$, $\beta_1(f^*) = b_1^*$, $\beta_1(f) = b_{1,f}$, for $f \neq f^*$, $\beta_2(f^*, \theta, a_1) = b_2^*(a_1, \theta)$, $\beta_2(f, \theta, a_1) = b_{2,f}(a_1, \theta)$ for all $f \neq f^*$, $a_1 \in A_1$, and $\theta \in \Theta$.

If player 2 offers any contract $f \neq f^*$, the continuation play will be $(b_{1,f}, b_{2,f})$. If $F(b_{1,f}, b_{2,f}) < \delta$ it will be rejected and hence it cannot be a profitable deviation. If $F(b_{1,f}, b_{2,f}) \geq \delta$, then

$$U_2(b_1^*, b_2^*) - F(b_1^*, b_2^*) = U_2(b_1^*, b_2^*) - \delta \geq U_2(b_{1,f}, b_{2,f}) - F(b_{1,f}, b_{2,f})$$

by construction. Therefore, it is optimal for player 2 to offer f^*. Sequential rationality at other information sets are easily checked and we conclude that this assessment is a perfect Bayesian equilibrium of the game with observable contracts. □

Proof Proposition 7. Let $b_1^*(b_2) \in \text{argmin}_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$, $b_2^* \in \text{argmax}_{b_2 \in B_2^+} U_2(b_1^*(b_2), b_2)$, and $a_1^* = b_1^*(b_2^*)$. Note that $U_2(a_1^*, b_1^*) = \bar{U}_2^W$ and suppose, for contradiction, that player 2 gets a payoff $\tilde{U}_2 < \bar{U}_2^W - \delta$. We will show that player 2 can offer a contract that supports (a_1^*, b_2^*) and yields a higher payoff.

For any a_1 choose $\hat{b}_{2,a_1} \in \text{argmin}_{b_2 \in B_2^+} U_1(a_1, b_2)$. By construction $\hat{b}_{2,a_1}$ is increasing and hence there exists a contract that makes it optimal to play. We will further show that there exists a contract that makes it the unique optimal strategy after a_1. Assume without loss of generality that $\hat{b}_{2,a_1}(a_1, \theta) \neq \hat{b}_{2,a_1}(a_1, \theta')$ whenever $\theta \neq \theta'$ and hence $\hat{b}_{2,a_1}(a_1, \theta^i) \succ_\theta \hat{b}_{2,a_1}(a_1, \theta^{i-1})$ for all $i = 1, \ldots, n$.[34] Define $\vec{U}_2(a_1, \hat{b}_{2,a_1})$ as usual and note that strictly increasing differences and $\hat{b}_{2,a_1}(a_1, \theta^i) \succ_\theta \hat{b}_{2,a_1}(a_1, \theta^{i-1})$ imply that

$$\vec{U}_2(a_1, \hat{b}_{2,a_1})_{2i-1} + \vec{U}_2(a_1, \hat{b}_{2,a_1})_{2i} > 0, \quad \forall i = 1, \ldots, n-1.$$

We will show that there exists f^{a_1} such that $Df^{a_1} \ll \vec{U}_2(a_1, \hat{b}_{2,a_1})$. Define

$$A = \begin{pmatrix} \vec{U}_2(a_1, \hat{b}_2) & -D \\ 1 & 0 \end{pmatrix}$$

and note that there exists f^{a_1} such that $Df^{a_1} \ll \vec{U}_2(a_1, \hat{b}_{2,a_1})$ iff there exists x such that $Ax \gg 0$.[35] By Gordan's Theorem, this is true iff $A'y = 0$ implies $y \leq 0$. It is easy to show that $A'y = 0$ implies $y_1 = y_2, y_3 = y_4, \cdots, y_{2(n-1)-1} = y_{2(n-1)}$. Therefore,

$$A'y = y_{2(n-1)+1} + \sum_{i=1}^{n-1} (\vec{U}_2(a_1, \hat{b}_{2,a_1})_{2i-1} + \vec{U}_2(a_1, \hat{b}_{2,a_1})_{2i}) y_{2i-1}$$

$\vec{U}_2(a_1, \hat{b}_{2,a_1})_{2i-1} + \vec{U}_2(a_1, \hat{b}_{2,a_1})_{2i} > 0, \forall i = 1, \ldots, n-1$, and $A'y = 0$ imply $y \leq 0$.

[34] If there exist i such that $\hat{b}_{2,a_1}(a_1, \theta^i) = \hat{b}_{2,a_1}(a_1, \theta^{i-1})$ simply eliminate the incentive compatibility constraint between them and set $\hat{f}_{a_1}(a_1, \hat{b}_{2,a_1}(a_1, \theta^i)) = \hat{f}_{a_1}(a_1, \hat{b}_{2,a_1}(a_1, \theta^{i-1}))$.

[35] To see this let $x = \begin{pmatrix} \zeta \\ f^{a_1} \end{pmatrix}$

Let $\varepsilon > 0$ be small and define $f(b_1^*, a_2) = \delta + \varepsilon$ for all a_2. For any $a_1 \neq b_1^*$ define

$$f(a_1, a_2) = \begin{cases} f_i^{a_1}, & a_2 = \hat{b}_{2,a_1}(a_1, \theta^i) \\ \infty, & \text{otherwise} \end{cases}$$

Under this contract, player 2 plays a best response to a_1^* and according to $\hat{b}_{2,a_1}$ after any $a_1 \neq a_1^*$. Player 1, on the other hand, must play a best response to b_2^*. This is because for any $a_1 \notin BR_1(b_1^*)$, we have $U_1(br_1(b_2^*), b_2^*) > U_1(a_1, b_2^*) \geq U_1(a_1, \hat{b}_{2,a_1})$. Therefore, deviation to such a contract yields a payoff of $U_2(b_1^*, b_2^*) - \delta - \varepsilon > \tilde{U}_2$, for small enough ε. In other words, player 2 has a profitable deviation, contradicting that $\tilde{U}_2$ is an equilibrium payoff. □

Before we proceed to the proof of Proposition 8 we need to introduce a definition and prove a supplementary lemma.

Definition 10. We say that a perfect Bayesian equilibrium (β, μ) of the game with renegotiable contracts has conservative beliefs if

$$\beta_2(f, a_1, \theta) = g \in \mathscr{C}, \beta_2(f, a_1, \theta, g, y) = b_g(a_1, \theta), \beta_2(f, a_1, \theta, g, n) = b_f(a_1, \theta), \beta_3(I_3(f, a_1, \theta)) = y$$

imply $g(b_g(a_1, \theta)) \geq f(b_f(a_1, \theta))$.

In other words, whenever, in equilibrium, type θ renegotiates the contract from f to g, the third-party should not expect a decrease in the transfer from that type.

Lemma 1. *Take any a perfect Bayesian equilibrium of the game with renegotiable contracts and assume that it has conservative beliefs. Suppose that a contract f is renegotiated after some a_1 and θ. Then, there exists a contract strategy pair that is incentive compatible, renegotiation-proof, and induces the same outcome as f after a_1.*

This lemma tell us that in any equilibrium with conservative beliefs, one can achieve any outcome that is achieved via renegotiation after a_1 by using a renegotiation-proof contract.

Proof of Lemma 1. Fix a a perfect Bayesian equilibrium with conservative beliefs and suppose that contract f is renegotiated after some a_1 and θ. Let the set of types after which f is renegotiated be Θ^R and $\Theta^{NR} = \Theta \setminus \Theta^R$. For any $\theta \in \Theta^R$, let g_θ be the new contract and $b_{g_\theta}(a_1, \theta)$ be the new strategy of player 2 after a_1 and θ. Also let $b_f(a_1, \theta)$ be the equilibrium strategy of player 2 after a_1 and θ if he does not renegotiate f. In other words, we have $\beta_2(f, \theta, a_1) = b_f(a_1, \theta)$, $\forall \theta \in \Theta^{NR}, \beta_2(f, \theta, a_1) = g_\theta$, $\forall \theta \in \Theta^R$, $\beta_2(f, \theta, a_1, g_\theta, y) = b_{g_\theta}(a_1, \theta)$, $\beta_2(f, \theta, a_1, g_\theta, n) = b_f(a_1, \theta)$, and $\beta_3(I_3(f, a_1, g_\theta)) = y$. For ease of exposition we will omit the reference to a_1 in the following. Consider the following mixture menu:

$$\{(g_\theta(b_{g_\theta}(\theta)), b_{g_\theta}(\theta))_{\theta \in \Theta^R}\} \cup \{(f(b_f(\theta)), b_f(\theta))_{\theta \in \Theta^{NR}}\}$$

It is clear that this menu replicates the outcome induced by f after a_1. We also claim that this menu is incentive compatible and renegotiation proof after a_1.

Incentive compatibility of (f, b_f) implies that no two types in Θ^{NR} has an incentive to mimic each other. Consider $\theta, \theta' \in \Theta^R$ and suppose, for contradiction, that

$$u_2(b_{g_\theta}(\theta), \theta) - g_\theta(b_{g_\theta}(\theta)) < u_2(b_{g_{\theta'}}(\theta'), \theta) - g_{\theta'}(b_{g_{\theta'}}(\theta'))$$

But then type θ could increase her payoff after (f, a_1) by offering $g_{\theta'}$ and playing $b_{g_{\theta'}}(\theta')$ rather than offering g_θ and playing $b_{g_\theta}(\theta)$.[36]

Now let $\theta' \in \Theta^{NR}$ and $\theta \in \Theta^R$, and suppose for contradiction that

$$u_2(b_f(\theta'), \theta') - f(b_f(\theta')) < u_2(b_{g_\theta}(\theta), \theta') - g_\theta(b_{g_\theta}(\theta))$$

This implies that after (f, a_1) offering g_θ, which is accepted in equilibrium, and playing $b_{g_\theta}(\theta)$ is a profitable deviation for type θ'.

Finally, let $\theta' \in \Theta^{NR}$ and $\theta \in \Theta^R$ and suppose, for contradiction, that

$$u_2(b_{g_\theta}(\theta), \theta) - g_\theta(b_{g_\theta}(\theta)) < u_2(b_f(\theta'), \theta) - f(b_f(\theta'))$$

But then type θ could play $b_f(\theta')$ after (f, a_1) and receive a higher payoff rather than offering g_θ, which is accepted, and playing $b_{g_\theta}(\theta)$. This proves that the mixture menu is incentive compatible.

Suppose now, for contradiction, that the mixture menu is not renegotiation-proof after a_1. Then, there exists θ and an incentive compatible contract strategy pair (h, b_h) such that if $\theta \in \Theta^{NR}$, then,

$$u_2(b_h(\theta), \theta) - h(b_h(\theta)) > u_2(b_f(\theta), \theta) - f(b_f(\theta)) \tag{22}$$

if $\theta \in \Theta^R$, then

$$u_2(b_h(\theta), \theta) - h(b_h(\theta)) > u_2(b_{g_\theta}(\theta), \theta) - g_\theta(b_{g_\theta}(\theta)) \tag{23}$$

and

$$h(b_h(\hat{\theta})) > f(b_f(\hat{\theta})), \forall \hat{\theta} \in \Theta^{NR} \tag{24}$$

$$h(b_h(\hat{\theta})) > g_{\hat{\theta}}(b_{g_{\hat{\theta}}}(\hat{\theta})), \forall \hat{\theta} \in \Theta^R \tag{25}$$

Since $g_{\hat{\theta}}$ is accepted for all $\hat{\theta} \in \Theta^R$ and the equilibrium has conservative beliefs,

$$g_{\hat{\theta}}(b_{g_\theta}(\hat{\theta})) \geq f(b_f(\hat{\theta})), \ \forall \hat{\theta} \in \Theta^R \tag{26}$$

which, together with (24) and (25), implies that

$$h(b_h(\hat{\theta})) > f(b_f(\hat{\theta})), \forall \hat{\theta} \in \Theta \tag{27}$$

Suppose first that $\theta \in \Theta^{NR}$. Inequalities (22) and (27) imply that after (f, a_1) type θ could offer h, which would be accepted, and increase her payoff, a contradiction that in equilibrium she plays $b_f(\theta)$ after (f, a_1).

Similarly, if $\theta \in \Theta^R$, then (23) and (27) imply that after (f, a_1) type θ could offer h, which would

[36] Note that $g_{\theta'}$ is accepted after (f, a_1) in equilibrium by assumption.

be accepted, and increase her payoff, rather than offering g_θ, a contradiction. Therefore, the mixture menu is renegotiation-proof.

Since the mixture is incentive compatible we can easily extend it to a contract whose domain is the entire A_2 rather than just the range of b_f and b_{g_θ}. Define the new contract as

$$h(a_2) = \begin{cases} f(a_2), & \exists \theta : a_2 = b_f(\theta) \\ g_\theta(a_2), & \exists \theta : a_2 = b_{g_\theta}(\theta) \\ \infty, & \text{otherwise} \end{cases}$$

and note that h is well-defined since incentive compatibility of the mixture menu implies that whenever $b_f(\theta') = b_{g_\theta}(\theta) = a_2$ for some $\theta \in \Theta^R$ and $\theta' \in \Theta^{NR}$ we must also have $f(a_2) = g_\theta(a_2)$. □

Proof of Proposition 8. Let $b_2^* \in \text{argmax}_{b_2 \in B_2^R} \max_{b_1 \in BR_1(b_2)} U_2(b_1, b_2)$ and $b_1^* = \text{argmax}_{b_1 \in BR_1(b_2^*)} U_2(b_1, b_2^*)$. Note that $\bar{U}_2^{BR} = U_2(b_1^*, b_2^*)$. Since b_2^* is increasing and renegotiation-proof, there exists $f^* \in \mathscr{C}$ such that (f^*, b_2^*) is incentive compatible and renegotiation-proof with $F^*(b_1^*, b_2^*) = \delta$. For any $f \in \mathscr{C}$, a_1, and θ, let $b_{2,f}(a_1, \theta) \in \text{argmax}_{a_2} u_2(a_1, a_2, \theta) - f(a_1, a_2)$ and $g_{(f,\theta,a_1)} \in \text{argmax}_g u_2(a_1, b_{2,g}(a_1, \theta), \theta) - g(a_1, b_{2,g}(a_1, \theta))$ subject to $g(a_1, b_{2,g}(a_1, \theta')) \geq f(a_1, b_{2,f}(a_1, \theta'))$ for all θ'.

Consider the following assessment (β, μ): $\beta_2(\emptyset) = f^*$; $\beta_1(f^*) = b_1^*$, $\beta_3(f^*) = y$, $\beta_2(f^*, \theta, a_1) = b_2^*(a_1, \theta)$ for all (a_1, θ);

$$\beta_2(f, \theta, a_1) = \begin{cases} g_{(f,\theta,a_1)}, \text{ if } & u_2(a_1, b_{2,g_{(f,\theta,a_1)}}(a_1, \theta), \theta) - g_{(f,\theta,a_1)}(a_1, b_{2,g_{(f,\theta,a_1)}}(a_1, \theta)) \\ & \quad > u_2(a_1, b_{2,f}(a_1, \theta), \theta) - f(a_1, b_{2,f}(a_1, \theta)) \\ b_{2,f}(a_1, \theta), & \text{otherwise} \end{cases}$$

for any $f \neq f^*$ and (θ, a_1); $\beta_2(f, \theta, a_1, g, y) = b_{2,g}(a_1, \theta)$ and $\beta_2(f, \theta, a_1, g, n) = b_{2,f}(a_1, \theta)$ for all $f \neq f^*$ and (a_1, θ, g); $\beta_2(f^*, \theta, a_1, g, n) = b_2^*(a_1, \theta)$ for all (a_1, θ, g);

$$\beta_3(I_3(f^*, a_1, g)) = \begin{cases} y, & g(a_1, b_{2,g}(a_1, \theta)) > f^*(a_1, b_2^*(a_1, \theta)) \quad \forall \theta \\ n, & \text{otherwise} \end{cases}$$

and

$$\beta_3(I_3(f, a_1, g)) = \begin{cases} y, & \text{if } g(a_1, b_{2,g}(a_1, \theta)) \geq f(a_1, b_{2,f}(a_1, \theta)) \quad \forall \theta \\ n, & \text{otherwise} \end{cases}$$

for any a_1, g and $f \neq f^*$. Obviously, any $f \neq f^*$ induces a continuation strategy b_2^f for player 2, which may involve renegotiation after some θ. Let player 1 play the same best response to the continuation play irrespective of the contract that induces it. Let the third-party accept f iff continuation play yields expected transfers at least equal to δ. Specify beliefs as follows: $\mu(I_3(f^*, a_1, g))(\theta) = p(\theta)$ if $g(a_1, b_{2,g}(a_1, \theta)) > f^*(a_1, b_2^*(a_1, \theta))$ for all θ and $\mu(I_3(f^*, a_1, g))(\theta') = 1$ if there exists θ' such that $f^*(a_1, b_2^*(a_1, \theta')) \geq g(a_1, b_{2,g}(a_1, \theta'))$; For any $f \neq f^*$ and (a_1, g), $\mu(I_3(f, a_1, g))(\theta) = p(\theta)$ if $g(a_1, b_{2,g}(a_1, \theta)) \geq f(a_1, b_{2,f}(a_1, \theta))$ for all θ and $\mu^*(I_3(f, a_1, g))(\theta') = 1$ if there exists θ' such that $f(a_1, b_{2,f}(a_1, \theta')) > g(a_1, b_{2,g}(a_1, \theta'))$.

Now consider any contract $f \neq f^*$. If $(f, b_{2,f})$ is renegotiation-proof, then $b_{2,f} \in B_2^R$ and hence f cannot yield a higher payoff than f^*. Therefore, suppose that $(f, b_{2,f})$ is not renegotiation-proof and

let b_2^f be the induced strategy, which includes renegotiation after some a_1 and θ. Since $\beta_3(I_3(f,a_1,g)) = y$ iff $g(a_1,b_{2,g}(a_1,\theta)) \geq f(a_1,b_{2,f}(a_1,\theta))$ for all $\theta \in \Theta$, the equilibrium constructed above has conservative beliefs. Lemma 1 therefore implies that there exists $(h,b_{2,h})$ which is incentive compatible and renegotiation-proof and induces the same outcome as (f,b_2^f). But no renegotiation-proof strategy can yield a payoff that is higher than $\bar{U}_2^{BR}$ and hence deviation to f cannot be profitable.

Sequential rationality at other information sets and consistency of beliefs can be checked easily to show that the above assessment is a perfect Bayesian equilibrium. □

Proof. (Proof of Proposition 10) **(Only if)** Suppose that (b_1^*,b_2^*) can be supported. Then, there exists a perfect Bayesian equilibrium (β^*,μ^*) that induces (b_1^*,b_2^*), i.e., $\beta_2^*(\emptyset) = f^*$, $\beta_3(f^*) = y$, $\beta_1^*(\mathscr{C}) = b_1^*$, $\beta_2^*(f^*,\theta,a_1) = b_2^*(a_1,\theta)$ for all $a_1 \in A_1$ and $\theta \in \Theta$. The fact that (b_1^*,b_2^*) is a Bayesian Nash equilibrium of G is a direct consequence of sequential rationality of players 1 and 2. It must also be the case that it is optimal to play according to b_2^* under f^*. Increasing differences and Proposition 2 implies that b_2^* is increasing.

[If] Let (b_1^*,b_2^*) be a Bayesian Nash equilibrium of G such that b_2^* is increasing. Proposition 2 implies that there exists a contract f' such that (f',b_2^*) is incentive compatible. It is not difficult to show that we can find such a contract whose expected value under (b_1^*,b_2^*) is equal to δ. So assume $F'(b_1^*,b_2^*) = \delta$. For any $b_2 \in A_2^{A_1\times\Theta}$ and $a_1 \in A_1$, let $b_2(a_1,\Theta)$ be the image of Θ under $b_2(a_1,.)$. Define

$$f^*(a_1,a_2) = \begin{cases} f'(a_1,a_2), & \text{if } a_2 \in b_2^*(a_1,\Theta) \\ \infty, & \text{otherwise} \end{cases}$$

for any $(a_1,a_2) \in A_1 \times A_2$, and

$$b_{2,f}^*(a_1,\theta) = \begin{cases} b_2^*(a_1,\theta), & f = f^* \\ \in \operatorname{argmax}_{a_2} u_2(a_1,a_2,\theta) - f(a_1,a_2), & f \neq f^* \end{cases}$$

for any $f \in \mathscr{C}$, $a_1 \in A_1$, and $\theta \in \Theta$. Consider the assessment (β^*,μ^*) of $\Gamma(G)$, where $\beta_2^*[\emptyset] = f^*$, $\beta_3[f] = y$ iff $F(b_1^*,b_{2,f}^*) \geq \delta$, $\beta_1^*[\mathscr{C}] = b_1^*$, $\beta_2^*[f,\theta,a_1] = b_{2,f}^*(a_1,\theta)$ for all $f \in \mathscr{C}$, $a_1 \in A_1$, and $\theta \in \Theta$, and $\mu^*[\mathscr{C}](f^*) = 1$. It is easy to check that this assessment induces (b_1^*,b_2^*) and is a perfect Bayesian equilibrium of $\Gamma(G)$. □

Proof of Proposition 11. **[If]** Let (b_1^*,b_2^*) be a Bayesian Nash equilibrium of G such that b_2^* is increasing and renegotiation-proof. This implies that there exists $f' \in \mathscr{C}$ such that (f',b_2^*) is incentive compatible and renegotiation-proof. Let $f^*(a_1,a_2) = f'(a_1,a_2) - F'(b_1^*,b_2^*) + \delta$ for all (a_1,a_2) and note that $F^*(b_1^*,b_2^*) = \delta$. Furthermore, using Theorem 1, it can be easily checked that (f^*,b_2^*) is incentive compatible and renegotiation-proof. For any $f \in \mathscr{C}$, a_1, and θ, let $b_{2,f}(a_1,\theta) \in \operatorname{argmax}_{a_2} u_2(a_1,a_2,\theta) - f(a_1,a_2)$ and $g_{(f,\theta,a_1)} \in \operatorname{argmax}_g u_2(a_1,b_{2,g}(a_1,\theta),\theta) - g(a_1,b_{2,g}(a_1,\theta))$ subject to $g(a_1,b_{2,g}(a_1,\theta')) \geq f(a_1,b_{2,f}(a_1,\theta'))$ for all θ'.

Consider the following assessment (β^*,μ^*) of $\Gamma_R(G)$: $\beta_2^*(\emptyset) = f^*$; $\beta_3(f) = y$ iff continuation play

yields an expected transfer of at least δ, $\beta_1^*(\mathscr{C}) = b_1^*$, $\beta_2^*(f^*,\theta,a_1) = b_2^*(a_1,\theta)$ for all (a_1,θ);

$$\beta_2^*(f,\theta,a_1) = \begin{cases} g_{(f,\theta,a_1)}, \text{ if } & u_2(a_1, b_{2,g_{(f,\theta,a_1)}}(a_1,\theta),\theta) - g_{(f,\theta,a_1)}(a_1, b_{2,g_{(f,\theta,a_1)}}(a_1,\theta)) \\ & \quad > u_2(a_1, b_{2,f}(a_1,\theta),\theta) - f(a_1, b_{2,f}(a_1,\theta)) \\ b_{2,f}(a_1,\theta), & \text{otherwise} \end{cases}$$

for any $f \neq f^*$ and (θ,a_1); $\beta_2^*(f,\theta,a_1,g,y) = b_{2,g}(a_1,\theta)$ and $\beta_2(f,\theta,a_1,g,n) = b_{2,f}(a_1,\theta)$ for all $f \neq f^*$ and (a_1,θ,g); $\beta_2(f^*,\theta,a_1,g,n) = b_2^*(a_1,\theta)$ for all (a_1,θ,g);

$$\beta_3^*(I_3(f^*,a_1,g)) = \begin{cases} y, & g(a_1, b_{2,g}(a_1,\theta)) > f^*(a_1, b_2^*(a_1,\theta)) \quad \forall\theta \\ n, & \text{otherwise} \end{cases}$$

and

$$\beta_3^*(I_3(f,a_1,g)) = \begin{cases} y, & \text{if } g(a_1, b_{2,g}(a_1,\theta)) \geq f(a_1, b_{2,f}(a_1,\theta)) \quad \forall\theta \\ n, & \text{otherwise} \end{cases}$$

for any a_1, g and $f \neq f^*$; $\mu^*(\mathscr{C})(f^*) = 1$; $\mu^*(I_3(f^*,a_1,g))(\theta) = p(\theta)$ if $g(a_1, b_{2,g}(a_1,\theta)) > f^*(a_1, b_2^*(a_1,\theta))$ for all θ and $\mu^*(I_3(f^*,a_1,g))(\theta') = 1$ if there exists θ' such that $f^*(a_1, b_2^*(a_1,\theta')) \geq g(a_1, b_{2,g}(a_1,\theta'))$; For any $f \neq f^*$ and (a_1,g), $\mu^*(I_3(f,a_1,g))(\theta) = p(\theta)$ if $g(a_1, b_{2,g}(a_1,\theta)) \geq f(a_1, b_{2,f}(a_1,\theta))$ for all θ and $\mu^*(I_3(f,a_1,g))(\theta') = 1$ if there exists θ' such that $f(a_1, b_{2,f}(a_1,\theta')) > g(a_1, b_{2,g}(a_1,\theta'))$. This assessment induces (b_1^*, b_2^*) and is a renegotiation-proof perfect Bayesian equilibrium.

[Only if] Suppose that $\Gamma_R(G)$ has a renegotiation-proof perfect Bayesian equilibrium (β^*,μ^*) that induces (b_1^*, b_2^*). Letting $\beta_2^*(\emptyset) = f^*$, we have $\beta_1^*(\mathscr{C}) = b_1^*$, $\beta_2(f^*,\theta,a_1) = b_2^*(a_1,\theta)$ for all (a_1,θ), and $\mu^*(\mathscr{C})(f^*) = 1$. Sequential rationality of player 1 implies that

$$b_1^* \in \operatorname*{argmax}_{a_1} U_1(a_1, b_2^*) \tag{28}$$

whereas that of player 2 implies $u_2(a_1, b_2^*(a_1,\theta),\theta) - f^*(a_1, b_2^*(a_1,\theta)) \geq u_2(a_1, b_2^*(a_1,\theta'),\theta) - f^*(a_1, b_2^*(a_1,\theta'))$ for all a_1 and θ,θ', which, together with increasing differences, implies that b_2^* is increasing.

We also claim that

$$b_2^*(b_1^*,\theta) \in \operatorname*{argmax}_{a_2} u_2(b_1^*, a_2,\theta) \quad \forall\theta. \tag{29}$$

Suppose, for contradiction, that this is not the case for θ' and let $\hat{a}_2 \in \operatorname{argmax}_{a_2} u_2(b_1^*, a_2,\theta')$ and define $\varepsilon = u_2(b_1^*,\hat{a}_2,\theta') - u_2(b_1^*, b_2^*(b_1^*,\theta'),\theta') > 0$. Define $f'(a_1,a_2) = F^*(b_1^*, b_2^*) + \varepsilon/2$ and note that the third-party accepts f'. Assume first that f' is not renegotiated after b_1^* and note that sequential rationality of player 2 implies that $\beta_2^*(f',\theta,b_1^*) \in \operatorname{argmax}_{a_2} u_2(b_1^*, a_2,\theta)$. Let $b_{2,f'}(a_1,\theta) = \beta_2^*(f',\theta,a_1)$. Player 2's expected payoff under f' is

$$U_2(b_1^*, b_{2,f'}) - F^*(b_1^*, b_2^*) - \varepsilon/2 > U_2(b_1^*, b_2^*) - F^*(b_1^*, b_2^*)$$

contradicting that (β^*,μ^*) is a PBE. A similar argument goes through if f' is renegotiated after b_1^*.

Therefore, by (28) and (29), (b_1^*, b_2^*) is a Bayesian Nash equilibrium of G and b_2^* is increasing. Finally, suppose that b_2^* is not renegotiation-proof. This implies that for any contract f such that (f, b_2^*) is incentive compatible, there exist a_1', θ', and an incentive compatible (g, b_2) such that $u_2(a_1', b_2(a_1',\theta'),\theta') -$

$g(a_1', b_2(a_1', \theta')) > u_2(a_1', b_2^*(a_1', \theta'), \theta') - f(a_1', b_2^*(a_1', \theta'))$ and $g(a_1', b_2(a_1', \theta)) > f(a_1', b_2^*(a_1', \theta))$ for all θ. This implies that, in any perfect Bayesian equilibrium, after history (f, θ', a_1') player 2 strictly prefers to renegotiate and offer g and the third-party accepts it. In other words, there exists no renegotiation-proof perfect Bayesian equilibrium which induces (b_1^*, b_2^*), completing the proof. □

Proof of Corollary 3. Given Corollary 2, we only need to prove that $\underline{U}_1^R = \max_{a_1} U_1(a_1, b_2^R)$. We first need the following definition

Definition 11. For any $b_2 \in A_2^{A_1 \times \Theta}$ we say that (a_1, i), $i \in \{1, 2, \ldots, n\}$ has *right deviation* (*left deviation*) at b_2 if there exists an $a_2 \in A_2$ such that $a_2 \succsim_2 b_2(a_1, \theta^i)$ ($b_2(a_1, \theta^i) \succsim_2 a_2$) and $u_2(a_1, a_2, \theta^i) > u_2(a_1, b_2(a_1, \theta^i)\theta^i)$. Otherwise, we say that i has *no right deviation* (*no left deviation*)at b_2.

For any $b_2 \in A_2^{A_1 \times \Theta}$ and (a_1, i), $i \in \{1, \cdots, n\}$, that has right deviation at b_2, define

$$R(a_1, i) = \{k > i : b_2(a_1, \theta^k) \in BR_2(a_1, \theta^k) \text{ and } i < j < k \text{ implies that } (a_1, j) \text{ has no left deviation at } b_2\}$$

Similarly, for any (a_1, i) with $i \in \{1, \cdots, n\}$, that has a left deviation at b_2, define

$$L(a_1, i) = \{k < i : b_2(a_1, \theta^k) \in BR_2(a_1, \theta^k) \text{ and } k < j < i \text{ implies that } (a_1, j) \text{ has no right deviation at } b_2\},$$

We need the following lemma:

Lemma 2. *b_2^* is renegotiation-proof if for any (a_1, i_1) that has right deviation and any (a_1, i_2) that has left deviation at b_2^*, $R(a_1, i_1) \neq \emptyset$, $L(a_1, i_2) \neq \emptyset$, and $i_1 < i_2$ implies $R(a_1, i_1) \cap L(a_1, i_2) \neq \emptyset$.*

Proof of Lemma 2. Similar to the proof of Lemma 6 in Gerratana and Koçkesen (2012) and omitted. □

We can now proceed to the proof of Corollary 3. We first prove that b_2^R is renegotiation proof. Fix a_1 and assume u_1 is increasing in a_2. Then, there is no (a_1, i) with left deviation by construction of b_2^R. For any (a_1, i) with right deviation, we have $n \in R(a_1, i)$. Similarly, if u_1 is decreasing in a_2. Lemma 2, therefore, implies that b_2^R is renegotiation-proof.

We next prove that for any a_1 and renegotiation-proof strategy $b_2 \in B_2^R$, we have $U_1(a_1, b_2) \geq U_1(a_1, b_2^R)$. We will use the following lemma

Lemma 3. *If $b_2 \in A_2^{A_1 \times \Theta}$ is renegotiation-proof, then (a_1, θ^n) has no right deviation at b_2 for any $a_1 \in A_1$*

Proof of Lemma 3. Similar to the proof of Lemma 7 in Gerratana and Koçkesen (2012) and omitted. □

Fix a_1 and assume that u_1 is increasing in a_2. Let $b_2 \in B_2^R$. Lemma 3 implies that $b_2(a_1, \theta^n) \succsim_2 b_2^n(a_1)$ and hence $u_1(a_1, b_2(a_1, \theta^n), \theta^n) \geq u_1(a_1, b_2^n(a_1), \theta^n) = u_1(a_1, b_2^R(a_1, \theta^n), \theta^n)$. Also, $u_1(a_1, b_2(a_1, \theta), \theta) \geq u_1(a_1, \underline{a}_2, \theta)$ for all θ, which implies that $U_1(a_1, b_2) \geq U_1(a_1, b_2^R)$. Therefore, $\underline{U}_1^R = \max_{a_1} U_1(a_1, b_2^R)$. □

Proof of Proposition 14. Denote with B_2^+ the set of strategies of player 2, $\pi(\pi^e, \theta)$, that are increasing in θ. To avoid notational clutter, we denote such strategies simply with π. Recall that we denote with

$\bar{U}_2^B$ the best Stackelberg payoff, and with $\bar{U}_2^W$ the worst Stackelberg payoff, which are given in the monetary policy example by

$$\bar{U}_2^B = \max_{\pi \in B_2^+} \max_{\pi^e \in BR_1(\pi)} \{-\sum_{\theta \in \Theta} [U^* + \theta - \alpha(\pi - \pi^e)]^2 p(\theta) - \beta \sum_{\theta \in \Theta} \pi^2\}$$

$$\bar{U}_2^W = \max_{\pi \in B_2^+} \min_{\pi^e \in BR_1(\pi)} \{-\sum_{\theta \in \Theta} [U^* + \theta - \alpha(\pi - \pi^e)]^2 p(\theta) - \beta \sum_{\theta \in \Theta} \pi^2\}$$

First, we prove that for any $\pi^e \in BR_1(\pi)$

$$\bar{U}_2 = \max_{\pi \in B_2^+} \{-\sum_{\theta \in \Theta} [U^* + \theta - \alpha(\pi - \pi^e)]^2 p(\theta) - \beta \sum_{\theta \in \Theta} \pi^2\}$$

is unique and therefore $\bar{U}_2^B = \bar{U}_2^W$. In order to do this we need the lemmata (4)-(5)

Lemma 4. *If* $\pi(\pi^e, \theta) = \tilde{\pi}(\pi^e) + \hat{\pi}(\theta)$ *with* $\pi(\pi^e, \theta)$ *differentiable and* $\frac{d\tilde{\pi}(\pi^e)}{d\pi^e} \neq 1$, *then* $BR_1(\pi) \subset \{\pi^e : \pi^e = \tilde{\pi}(\pi^e) + \sum_{\theta \in \Theta} \hat{\pi}(\theta) p(\theta)\}$

Proof of Lemma 4. Since we assumed $\pi(\pi^e, \theta) = \tilde{\pi}(\pi^e) + \hat{\pi}(\theta)$, the set $BR_1(\pi)$ is

$$BR_1(\pi) = \underset{\pi^e}{\operatorname{argmax}} - \sum_{\theta \in \Theta} [\pi^e - \tilde{\pi}(\pi^e) - \hat{\pi}(\theta)]^2 p(\theta)$$

The first order condition of the maximization problem for player 1 is

$$-2 \sum_{\theta \in \Theta} [\pi^e - \tilde{\pi}(\pi^e) - \hat{\pi}(\theta)] p(\theta) (1 - \frac{d\tilde{\pi}(\pi^e)}{d\pi^e}) = 0$$

Since we assumed $\frac{d\tilde{\pi}(\pi^e)}{d\pi^e} \neq 1$, the first order condition is satisfied if $\pi^e - \tilde{\pi}(\pi^e) = \sum_{\theta \in \Theta} \hat{\pi}(\theta) p(\theta)$. It, then, follows that

$$BR_1(\pi) \subset \{\pi^e \mid \pi^e - \tilde{\pi}(\pi^e) = \sum_{\theta \in \Theta} \hat{\pi}(\theta) p(\theta)\}$$

□

Lemma 5. *The following problem:*

$$\bar{V} = \max_{\{\tilde{\pi}(\pi^e), \hat{\pi}(\theta)\}} V(\tilde{\pi}(\pi^e), \hat{\pi}(\theta)) = -\sum_{\theta \in \Theta} \{U^* + \theta - \alpha[\tilde{\pi}(\pi^e) - \pi^e + \hat{\pi}(\theta)]\}^2 p(\theta) - \beta \sum_{\theta \in \Theta} \{\tilde{\pi}(\pi^e) + \hat{\pi}(\theta)\}^2 p(\theta) \quad (30)$$

subject to

$$\hat{\pi}(\theta) \text{ increasing} \quad (31)$$

$$\pi^e - \tilde{\pi}(\pi^e) = \sum_{\theta \in \Theta} \hat{\pi}(\theta) p(\theta) \quad (32)$$

has solution $\tilde{\pi}(\pi^e) + \hat{\pi}(\theta) = \frac{\alpha}{\alpha^2 + \beta} \theta$

Proof of Lemma 5. We first solve the relaxed problem, obtained by dropping the constraint $\hat{\pi}(\theta)$ increasing, and then verify that the solution to the relaxed problem is indeed increasing in θ.

By substituting (32) in (30) the relaxed problem is equivalent to:

$$\max_{\{\pi^e, \hat{\pi}(\theta)\}} \{-\sum_{\theta \in \Theta} \{U^* + \theta - \alpha[\hat{\pi}(\theta) - \sum_{\theta \in \Theta} \hat{\pi}(\theta) p(\theta)]\}^2 p(\theta) - \beta \sum_{\theta \in \Theta} \{\pi^e + \hat{\pi}(\theta) - \sum_{\theta \in \Theta} \hat{\pi}(\theta) p(\theta)\}^2 p(\theta)\} \quad (33)$$

which has the solution

$$\hat{\pi}(\theta) = k + \frac{\alpha}{\alpha^2+\beta}\theta \tag{34}$$

$$\pi^e = 0 \tag{35}$$

In order to obtain the solution to the original problem we substitute (34) and (35) in (32), obtaining $\tilde{\pi}(\pi^e) = -k$, which implies

$$\tilde{\pi}(\pi^e) + \hat{\pi}(\theta) = \frac{\alpha}{\alpha^2+\beta}\theta$$

Since $\hat{\pi}(\theta)$ defined by this equation is increasing in θ, it solves the problem (30). □

By Lemma 4, we have $\bar{V} \geq \bar{U}_2(\pi^e)$. Moreover when $\tilde{\pi}(\pi^e) + \hat{\pi}(\theta) = \frac{\alpha}{\alpha^2+\beta}\theta$,

$$BR_1(\pi) = \{\pi^e \mid \pi^e - \tilde{\pi}(\pi^e) = \sum_{\theta\in\Theta} \hat{\pi}(\theta)p(\theta)\} = \{\pi^e = 0\}$$

This implies that $\bar{U}_2(\pi^e) = \bar{V}$, hence $\bar{U}_2(\pi^e)$ is unique and therefore $\bar{U}_2^B = \bar{U}_2^W$.

Since we established $\bar{U}_2^B = \bar{U}_2^W$, Propositions 6 and 7 imply that the only equilibrium payoff of player 2 that can be supported with observable and non-renegotiable contracts is $\bar{U}_2^B - \delta$. Moreover by Lemma 5 we know that $\bar{U}_2^B - \delta$ is achieved by player 2 committing to the strategy $\pi(\pi^e,\theta) = \frac{\alpha}{\alpha^2+\beta}\theta$. Therefore the unique outcome supported by observable and non-renegotiable third party contracts is $\pi^e = 0$ and $\pi(\theta) = \frac{\alpha}{\alpha^2+\beta}\theta$.

□

Proof of Proposition 15. Denote with B_2^R the set of strategies of player 2 that are increasing in θ and renegotiation-proof. Recall that we denote with $\bar{U}_2^{BR}$ the best renegotiation-proof Stackelberg payoff, and with $\bar{U}_2^{WR}$ the worst renegotiation-proof Stackelberg payoff, that in the monetary policy example are respectively

$$\bar{U}_2^{BR} = \max_{\pi\in B_2^R} \max_{\pi^e\in BR_1(\pi)} \{-\sum_{\theta\in\Theta} [U^* + \theta - \alpha(\pi - \pi^e)]^2 p(\theta) - \beta \sum_{\theta\in\Theta} \pi^2\}$$

$$\bar{U}_2^{WR} = \max_{\pi\in B_2^R} \min_{\pi^e\in BR_1(\pi)} \{-\sum_{\theta\in\Theta} [U^* + \theta - \alpha(\pi - \pi^e)]^2 p(\theta) - \beta \sum_{\theta\in\Theta} \pi^2\}$$

First, we prove that for any $\pi^e \in BR_1(\pi)$

$$\bar{U}_2^R = \max_{\pi\in B_2^R} \{-\sum_{\theta\in\Theta} [U^* + \theta - \alpha(\pi - \pi^e)]^2 p(\theta) - \beta \sum_{\theta\in\Theta} \pi^2\}$$

is unique and therefore $\bar{U}_2^{BR} = \bar{U}_2^{WR}$.

Lemma 6. *The following problem: choose $\tilde{\pi}(\pi^e)$ and $\hat{\pi}(\theta)$ to maximize*

$$f(\tilde{\pi}(\pi^e), \hat{\pi}(\theta)) = -\sum_{\theta\in\Theta} \{U^* + \theta - \alpha[\tilde{\pi}(\pi^e) - \pi^e + \hat{\pi}(\theta)]\}^2 p(\theta) - \beta \sum_{\theta\in\Theta} \{\tilde{\pi}(\pi^e) + \hat{\pi}(\theta)\}^2 p(\theta) \tag{36}$$

subject to

$$\hat{\pi}(\theta) \text{ increasing} \tag{37}$$

$$\pi^e - \tilde{\pi}(\pi^e) = \sum_{\theta\in\Theta} \hat{\pi}(\theta)p(\theta) \tag{38}$$

and

$$\tilde{\pi}(\pi^e) + \hat{\pi}(\theta_n) \geq \frac{\alpha}{\alpha^2+\beta} U^* + \frac{\alpha}{\alpha^2+\beta}\theta_n + \frac{\alpha^2}{\alpha^2+\beta}\pi^e \tag{39}$$

has a unique solution

$$\tilde{\pi}(\pi^e) + \hat{\pi}(\theta) = \begin{cases} \frac{\alpha}{\alpha^2+\beta}\theta, & \theta < \theta^n \\ \frac{\alpha}{\alpha^2+\beta}\theta + \frac{\alpha}{\alpha^2(1-p_n)+\beta}U^*, & \theta = \theta^n \end{cases}$$

Proof. We first solve the relaxed problem obtained by ignoring the constraint $\hat{\pi}(\theta)$ increasing and then we verify that the solution of the relaxed problem $\hat{\pi}(\theta)$ is indeed increasing in θ.

By substituting (38) in (36) and (39), the relaxed problem is equivalent to the following: choose $\hat{\pi}(\theta)$ and π^e to maximize

$$g(\hat{\pi}(\theta), \pi^e) = -\sum_{\theta\in\Theta} [U^* + \theta - \alpha(\hat{\pi}(\theta) - \sum_{\theta} \hat{\pi}(\theta)p(\theta))]^2 p(\theta) - \beta \sum_{\theta\in\Theta} [\hat{\pi}(\theta) - \sum_{\theta} \hat{\pi}(\theta)p(\theta) + \pi^e]^2 p(\theta) \tag{40}$$

subject to

$$\pi^e \frac{\beta}{\alpha^2+\beta} - \sum_{\theta} \hat{\pi}(\theta)p(\theta) + \hat{\pi}(\theta_n) - \frac{\alpha}{\alpha^2+\beta}(U^* + \theta_n) \geq 0$$

Define

$$h(\hat{\pi}(\theta), \pi^e) = \pi^e \frac{\beta}{\alpha^2+\beta} - \sum_{\theta} \hat{\pi}(\theta)p(\theta) + \hat{\pi}(\theta_n) - \frac{\alpha}{\alpha^2+\beta}(U^* + \theta_n). \tag{41}$$

The Lagrangean of the problem (40) is

$$L(\hat{\pi}(\theta), \pi^e, \lambda) = -\sum_{\theta\in\Theta} [U^* + \theta - \alpha(\hat{\pi}(\theta) - \sum_{\theta} \hat{\pi}(\theta)p(\theta))]^2 p(\theta) - \beta \sum_{\theta\in\Theta} [\hat{\pi}(\theta) - \sum_{\theta} \hat{\pi}(\theta)p(\theta) + \pi^e]^2 p(\theta) + \lambda h(\hat{\pi}(\theta), \pi^e)$$

The critical points of the Lagrangean solve the following system of equations:

$$\pi^e = \frac{\lambda}{2}\frac{1}{\alpha^2+\beta} \tag{42}$$

For $\theta < \theta_n$

$$\hat{\pi}(\theta) = \frac{\alpha}{\alpha^2+\beta}\theta + \sum_{\theta} \hat{\pi}(\theta)p(\theta) - \frac{\lambda}{2(\alpha^2+\beta)} \tag{43}$$

For $\theta = \theta_n$

$$\hat{\pi}(\theta_n) = \frac{\alpha}{\alpha^2+\beta}\theta_n + \sum_{\theta} \hat{\pi}(\theta)p(\theta) + \frac{\lambda}{2(\alpha^2+\beta)}(\frac{1}{p(\theta_n)} - 1) \tag{44}$$

$$\lambda \geq 0,\ h(\hat{\pi}(\theta), \pi^e) \geq 0,\ \lambda h(\hat{\pi}(\theta), \pi^e) = 0 \tag{45}$$

First, we show that $h(\hat{\pi}(\theta)) = 0$. For contradiction assume that $h(\hat{\pi}(\theta), \pi^e) > 0$. Then $\lambda = 0$ by (45). This implies $\pi^e = 0$ (by 42) and $\hat{\pi}(\theta_n) = \frac{\alpha}{\alpha^2+\beta}\theta_n + \sum_{\theta} \hat{\pi}(\theta)p(\theta)$ by (44), which, once substituted in (41), implies $h(\hat{\pi}(\theta), \pi^e) < 0$, a contradiction to (45).

Hence, we established that $h(\hat{\pi}(\theta), \pi^e) = 0$, i.e.,

$$h(\hat{\pi}(\theta), \pi^e) = \pi^e \frac{\beta}{\alpha^2+\beta} - \sum_{\theta} \hat{\pi}(\theta)p(\theta) + \hat{\pi}(\theta_n) - \frac{\alpha}{\alpha^2+\beta}(U^* + \theta_n) = 0 \tag{46}$$

The system of equations (42)-(44) and (46) has a unique solution (up to the constant $k = \sum_\theta \hat{\pi}(\theta)p(\theta)$):

$$\lambda = \frac{2(\alpha^2+\beta)\alpha p_n}{\alpha^2(1-p_n)+\beta}U^*$$

$$\pi^e = \frac{\alpha p_n}{\alpha^2(1-p_n)+\beta}U^* \tag{47}$$

$$\hat{\pi}(\theta) = \begin{cases} k + \frac{\alpha}{\alpha^2+\beta}\theta - \frac{\alpha p_n}{\alpha^2(1-p_n)+\beta}U^*, & \theta < \theta^n \\ k + \frac{\alpha}{\alpha^2+\beta}\theta + \frac{\alpha p_n}{\alpha^2(1-p_n)+\beta}\frac{1-p_n}{p_n}U^*, & \theta = \theta^n \end{cases} \tag{48}$$

which is also the unique solution of the problem (40) since $\lambda \geq 0$. To obtain the solution to the original problem (36), we substitute (47) and (48) in (38), obtaining

$$\tilde{\pi}(\pi^e) = \frac{\alpha p_n}{\alpha^2(1-p_n)+\beta}U^* - k \tag{49}$$

and hence

$$\tilde{\pi}(\pi^e) + \hat{\pi}(\theta) = \begin{cases} \frac{\alpha}{\alpha^2+\beta}\theta, & \theta < \theta^n \\ \frac{\alpha}{\alpha^2+\beta}\theta + \frac{\alpha U^*}{\alpha^2(1-p_n)+\beta}, & \theta = \theta^n \end{cases} \tag{50}$$

Since $\hat{\pi}(\theta)$ is increasing in θ, then (50) is also a solution to the original problem (36). □

Now, notice that (39) is a necessary condition for $\tilde{\pi}(\pi^e) + \hat{\pi}(\theta)$ to be a renegotiation-proof strategy (this follows from Lemma 3), therefore the set of strategies $\tilde{\pi}(\pi^e) + \hat{\pi}(\theta)$ that satisfy (39) contains the set of renegotiation-proof strategies. Moreover, from Lemma 2 on page 37 it follows that the strategy (50) is renegotiation-proof. To see this, observe that for (50), there is no (π^e, θ), that has left deviation and for each (π^e, θ), we have $\theta^n \in R(\pi^e, \theta)$. Finally for $\pi = \tilde{\pi}(\pi^e) + \hat{\pi}(\theta)$, the best response of player 1 is equal to the set defined by (38) that is

$$BR_1(\pi) = \{\pi^e \mid \pi^e - \tilde{\pi}(\pi^e) = \sum_{\theta\in\Theta} \hat{\pi}(\theta)p(\theta)\} = \{\pi^e = \frac{\alpha p_n}{\alpha^2(1-p_n)+\beta}U^*\}$$

Therefore $\bar{U}^R$ is equal to the value of maximization problem (36), hence it is unique and therefore $U^{BR} = U^{WR}$. Since we established $\bar{U}_2^{BR} = \bar{U}_2^{WR}$, Propositions 8 and 9 imply that the only equilibrium payoff of player 2 that can be supported with observable and renegotiation-proof contracts is $\bar{U}_2^{BR} - \delta$. Moreover because of lemma 6, we know that $\bar{U}_2^{BR} - \delta$ is achieved by player 2 committing to the strategy (50). Therefore, the unique outcome supported by observable and renegotiable third party contracts is

$$\pi^e = \frac{\alpha p_n}{\alpha^2(1-p_n)+\beta}U^*$$

$$\pi(\theta) = \begin{cases} \frac{\alpha}{\alpha^2+\beta}\theta, & \theta < \theta^n \\ \frac{\alpha}{\alpha^2+\beta}\theta + \frac{\alpha U^*}{\alpha^2(1-p_n)+\beta}, & \theta = \theta^n \end{cases}$$

□

References

[1] Barro, R. J. and D. B. Gordon (1983), "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy*, 91, 589-610.

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

[3] Bensaid, B. and R. J. Gary-Bobo (1993), "Commitment Value of Contracts under Renegotiation Constraints," *Econometrica*, 61, 1423-29.

[4] Bhaskar, V. (2009), "Games Played in a Contracting Environment," *Games and Economic Behavior*, 67, 389-398.

[5] Bolton, P. and M. Dewtripont (2005), *Contract Theory*, Cambridge: The MIT Press.

[6] Bolton, P. and D. S. Scharfstein (1990), "A Theory of Predation Based on Agency Problems in Financial Contracting," *American Economic Review*, 80, 93-106.

[7] Brander, J. A. and T. R. Lewis (1986), "Oligopoly and Financial Structure: The Limited Liability Effect," *American Economic Review*, 76, 956-970.

[8] Brander, J. A. and B. J. Spencer (1985), "Export Subsidies and International Market Share Rivalry," *Journal of International Economics*, 18, 83-100.

[9] Caillaud, B., B. Jullien, and P. Picard (1995), "Competing Vertical Structures: Precommitment and Renegotiation," *Econometrica*, 63, 621-646.

[10] Cho, I. K. and D. Kreps (1987), "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179-221.

[11] Dewatripont, M. (1988), "Commitment through Renegotiation-Proof Contracts with Third Parties," *Review of Economic Studies*, 55, 377-389.

[12] Eaton, J. and G. M. Grossman (1986), "Optimal Trade and Industrial Policy under Oligopoly," *Quarterly Journal of Economics*, 101, 383-406.

[13] Fershtman, C. and K. Judd (1987), "Equilibrium Incentives in Oligopoly," *American Economic Review*, 77, 927-940.

[14] Fershtman, C., K. Judd, and E. Kalai (1991), "Observable Contracts: Strategic Delegation and Cooperation," *International Economic Review*, 32, 551-559.

[15] Fudenberg, D. and J. Tirole (1991), "Perfect Bayesian and Sequential Equilibrium," *Journal of Economic Theory*, 53, 236-260.

[16] Gerratana E. and L. Koçkesen (2012), "Strategic Effects of Renegotiation-Proof Contracts," *The B.E. Journal of Theoretical Economics (Advances)*, 12:1.

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

[18] Jackson, M. O. and S. Wilkie (2005), "Endogenous Games and Mechanisms: Side Payments Among Players," *Review of Economic Studies*, 72, 543-566.

[19] Katz, M. (1991), "Game-Playing Agents: Unobservable Contracts as Precommitments," *Rand Journal of Economics*, 22, 307-328.

[20] Katz, M. (2006), "Observable Contracts as Commitments: Interdependent Contracts and Moral Hazard," *Journal of Economics & Management Strategy*, 15, 685-706.

[21] Kydland, F. E. and E. C. Prescott (1977), "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, 85, 473-492.

[22] Koçkesen, L. (2007), "Unobservable Contracts as Precommitments," *Economic Theory*, 31, 539-552.

[23] Koçkesen, L. and E. A. Ok (2004), "Strategic Delegation by Unobservable Incentive Contracts," *Review of Economic Studies*, 71, 397-424.

[24] Koçkesen, L., E. A. Ok, and R. Sethi (2000), "The Strategic Advantage of Negatively Interdependent Preferences," *Journal of Economic Theory*, 92, 274-299.

[25] Mangasarian, O. L. (1994) *Nonlinear Programming*, New York: McGraw-Hill.

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

[27] McCallum, B.T. (1995), "Two Fallacies Concerning Central Bank Independence," *American Economic Review Papers and Proceedings*, 85, 207-211.

[28] McCallum, B.T. (1997), "Crucial Issues Concerning Central Bank Independence," *Journal of Monetary Economics*, 39, 99-112.

[29] Persson, T. and G. Tabellini (1993), "Designing Institutions for Monetary Stability," *Carnegie-rochester Conference Series on Public Policy*, 39, 53-84.

[30] Polo, M. and P. Tedeschi (2000), "Delegation Games and Side-Contracting," *Research in Economics*, 54, 101-116.

[31] Prat, A. and A. Rustichini (2003), "Games Played through Agents," *Econometrica*, 71, 989-1026.

[32] Sklivas, S. D. (1987), "The Strategic Choice of Managerial Incentives," *Rand Journal of Economics*, 18, 452-458.

[33] Snyder, C. M. (1996), "Negotiation and Renegotiation of Optimal Financial Contracts Under the Threat of Predation," *Journal of Industrial Economics*, 44, 325-343.

[34] Spencer, B. J. and J. A. Brander (1983), "International R & D Rivalry and Industrial Strategy," *Review of Economic Studies*, 50, 707-722.

[35] Vickers, J. (1985), "Delegation and the Theory of the Firm," *Economic Journal, Supplement*, 95, 138-147.

[36] Walsh, C. E. (1995), "Optimal Contracts for Central Bankers," *American Economic Review*, 85, 150-167.

Supplement to "Commitment without Reputation: Renegotiation-Proof Contracts under Asymmetric Information"

(not for publication)

Emanuele Gerratana
SIPA, Columbia University

Levent Koçkesen
Koç University

July 20, 2013

1 Introduction

In what follows we provide the omitted proofs of the statements made in our paper "Commitment without Reputation: Renegotiation-Proof Contracts under Asymmetric Information." In order to distinguish statements made in that paper from the ones made in this document we will add a note "(of the main paper)" after those from the main paper.

2 Proofs for Section 3

It is well-known that if b_2 is increasing, then, under increasing differences, incentive compatibility reduces to local incentive compatibility. We state it as a claim for future reference.

Claim 1. *If u_2 has increasing differences in $(\succsim_\theta, \succsim_2)$ and $b_2 \in A_2^{A_1\times\Theta}$ is increasing in $(\succsim_\theta, \succsim_2)$, then for any $f \in \mathscr{C}$*

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - f(a_1, b_2(a_1,\theta^i)) \geq u_2(a_1, b_2(a_1,\theta^j),\theta^i) - f(a_1, b_2(a_1,\theta^j)), \text{ for all } i,j = 1,2,\ldots,n$$

holds if and only if

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - f(a_1, b_2(a_1,\theta^i)) \geq u_2(a_1, b_2(a_1,\theta^{i-1}),\theta^i) - f(a_1, b_2(a_1,\theta^{i-1})), \text{ for all } i = 2,\ldots,n,$$

and

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - f(a_1, b_2(a_1,\theta^i)) \geq u_2(a_1, b_2(a_1,\theta^{i+1}),\theta^i) - f(a_1, b_2(a_1,\theta^{i+1})), \text{ for all } i = 1,2,\ldots,n-1.$$

Proof of Proposition 2 (of the main paper). **(Only if)** Suppose that b_2 is incentive compatible, i.e., there exists a contract f such that (f, b_2) is incentive compatible. Fix orders $(\succsim_\theta, \succsim_2)$ in which u_2 has strictly increasing differences. Take any $a_1 \in A_1$ and $\theta, \theta' \in \Theta$ and assume without loss of generality, that $\theta \succ_\theta \theta'$. Suppose, for contradiction, that $b_2(a_1, \theta') \succ_2 b_2(a_1, \theta)$. Sequential rationality of player 2 implies

$$u_2(a_1, b_2(a_1,\theta),\theta) - f(a_1, b_2(a_1,\theta)) \geq u_2(a_1, b_2(a_1,\theta'),\theta) - f(a_1, b_2(a_1,\theta'))$$
$$u_2(a_1, b_2(a_1,\theta'),\theta') - f(a_1, b_2(a_1,\theta')) \geq u_2(a_1, b_2(a_1,\theta),\theta') - f(a_1, b_2(a_1,\theta))$$

and hence

$$u_2(a_1, b_2(a_1,\theta'),\theta) - u_2(a_1, b_2(a_1,\theta),\theta) \leq u_2(a_1, b_2(a_1,\theta'),\theta') - u_2(a_1, b_2(a_1,\theta),\theta'),$$

contradicting that u_2 has strictly increasing differences in $(\succsim_\theta, \succsim_2)$. Therefore, b_2 must be increasing in $(\succsim_\theta, \succsim_2)$.
[If] Suppose u_2 has strictly increasing differences and b_2 is increasing. We need to prove the existence of a contract $f \in \mathscr{C}$ such that

$$u_2(a_1, b_2(a_1,\theta^i),\theta^i) - f(a_1, b_2(a_1,\theta^i)) \geq u_2(a_1, b_2(a_1,\theta^j),\theta^i) - f(a_1, b_2(a_1,\theta^j)), \text{ for all } i, j = 1,2,...,n. \tag{1}$$

By Claim 1, (1) holds if and only if $Df(a_1, b_2) \leq \vec{U}_2(a_1, b_2)$. Therefore, we need to show that for any $a_1 \in A_1$ there exists $f(a_1, b_2) \in \mathbb{R}^n$ such that $Df(a_1, b_2) \leq \vec{U}_2(a_1, b_2^*)$. By Gale's theorem for linear inequalities (Mangasarian (1994), p. 33), there exists such an $f(a_1, b_2) \in \mathbb{R}^n$ if and only if for any $y \in \mathbb{R}_+^{2(n-1)}$, $D'y = 0$ implies $y'\vec{U}_2(a_1, b_2^*) \geq 0$. It is easy to show that $D'y = 0$ if and only if $y_1 = y_2, y_3 = y_4, \cdots, y_{2(n-1)-1} = y_{2(n-1)}$. Let $\vec{U}_2(a_1, b_2)_i$ denote the i^{th} row of $\vec{U}_2(a_1, b_2)$ and note that since b_2 is increasing and u_2 has strictly increasing differences, $\vec{U}_2(a_1, b_2)_{2i-1} + \vec{U}_2(a_1, b_2)_{2i} \geq 0$, for any $i = 1, 2, \ldots, n-1$. Therefore,

$$y'\vec{U}_2(a_1, b_2^*) = \sum_{i=1}^{n-1} (\vec{U}_2(a_1, b_2)_{2i-1} + \vec{U}_2(a_1, b_2)_{2i}) y_{2i-1} \geq 0$$

This proves the existence of a $f(a_1, b_2) \in \mathbb{R}^n$ such that (1) is satisfied for all $a_1 \in A_1$. We can complete the proof by defining $\tilde{f} \in \mathscr{C}$ as

$$\tilde{f}(a_1, a_2) = \begin{cases} f(a_1, a_2), & \exists \theta : a_2 = b_2(a_1, \theta) \\ \infty, & \text{otherwise} \end{cases}$$

□

Proof of Theorem 1 (of the main paper). By definition $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is not renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1, 2, \ldots, n$ and an incentive compatible $(g, b_2) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ such that $u_2(a_1, b_2(a_1,\theta^i),\theta^i) - g(a_1, b_2(a_1,\theta^i)) > u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) - f(a_1, b_2^*(a_1,\theta^i))$ and $g(a_1, b_2(a_1,\theta^j)) > f(a_1, b_2^*(a_1,\theta^j))$ for all $j = 1, 2, \ldots, n$. For any $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$, let

$f(a_1, b_2^*) \in \mathbb{R}^n$ be a vector whose j-th component, $j = 1,2,\ldots,n$, is given by $f(a_1, b_2^*(a_1, \theta^j))$. Note that incentive compatibility of $(g, b_2) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is equivalent to $Dg(a_1, b_2) \leq \vec{U}_2(a_1, b_2)$ for all $a_1 \in A_1$. Therefore, $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is not renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1,2,\ldots,n$ and $(g(a_1, b_2), b_2) \in \mathbb{R}^n \times A_2^{A_1 \times \Theta}$ such that $Dg(a_1, b_2) \leq \vec{U}_2(a_1, b_2)$, $u_2(a_1, b_2(a_1, \theta^i), \theta^i) - g(a_1, b_2(a_1, \theta^i)) > u_2(a_1, b_2^*(a_1, \theta^i), \theta^i) - f(a_1, b_2^*(a_1, \theta^i))$, and $g(a_1, b_2) \gg f(a_1, b_2^*)$. Also note that $g(a_1, b_2) \gg f(a_1, b_2^*)$ if and only if there exists an $\varepsilon \gg 0$ such that $g(a_1, b_2) = f(a_1, b_2^*) + \varepsilon$. Therefore, we have the following

Lemma 1. *$(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is not renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1,2,\ldots,n$, $b_2 \in A_2^{A_1 \times \Theta}$, and $\varepsilon \in \mathbb{R}^n$ such that $D(f(a_1, b_2^*) + \varepsilon) \leq \vec{U}_2(a_1, b_2)$, $\varepsilon_i < u_2(a_1, b_2(a_1, \theta^i), \theta^i) - u_2(a_1, b_2^*(a_1^i, \theta^i), \theta^i)$, and $\varepsilon \gg 0$.*

We first state a theorem of the alternative, which we will use in the sequel.

Lemma 2 (Motzkin's Theorem)**.** *Let A and C be given matrices, with A being non-vacuous. Then either*

1. *$Ax \gg 0$ and $Cx \geq 0$ has a solution x*

   or

2. *$A'y_1 + C'y_2 = 0$, $y_1 > 0$, $y_2 \geq 0$ has a solution y_1, y_2*

   but not both.

Proof of Lemma 2. See Mangasarian (1994), p. 28. □

For any $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$, $a_1 \in A_1$, $b_2 \in A_2^{A_1, \times \Theta}$, and $i = 1,2,\ldots,n$, define $V = \vec{U}_2(a_1, b_2) - Df(a_1, b_2^*)$, $C = \begin{pmatrix} V & -D \end{pmatrix}$, and

$$A = \begin{pmatrix} I_{n+1} \\ l_i \end{pmatrix}$$

where $l_i = (u_2(a_1, b_2(a_1, \theta^i), \theta^i) - u_2(a_1, b_2^*(a_1, \theta^i), \theta^i))e_1 - e_{i+1}$. Note that C and A depend on and are uniquely defined by (f, b_2^*), a_1 and (i, b_2) but we suppress this dependency for notational convenience. The following lemma uses Motzkin's Theorem to express renegotiation-proofness as an alternative.

Lemma 3. *$(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is renegotiation-proof if and only if for any $a_1 \in A_1$, $i = 1,2,\ldots,n$ and $b_2 \in A_2^{A_1 \times \Theta}$ there exist $y \in \mathbb{R}^{n+2}$ and $z \in \mathbb{R}^{2(n-1)}$ such that $A'y + C'z = 0$, $y > 0$, $z \geq 0$.*

Proof of Lemma 3. By Lemma 1, (f, b_2^*) is not renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1,2,\ldots,n$, $b_2 \in A_2^{A_1 \times \Theta}$, and $\varepsilon \in \mathbb{R}^n$ such that $D(f(a_1, b_2^*) + \varepsilon) \leq \vec{U}_2(a_1, b_2)$, $\varepsilon_i < u_2(a_1, b_2(a_1, \theta^i), \theta^i) - u_2(a_1, b_2^*(a_1, \theta^i), \theta^i)$, and $\varepsilon \gg 0$. This is true if and only if for some a_1, i and b_2 there exists an $x \in \mathbb{R}^{n+1}$ such that $Ax \gg 0$ and $Cx \geq 0$. To see this let $\xi > 0$ and define

$$x = \begin{pmatrix} \xi \\ \xi\varepsilon \end{pmatrix}$$

Then $D(f(a_1,b_2^*)+\varepsilon) \leq \vec{U}_2(a_1,b_2)$ if and only if $Cx \geq 0$. Also, $\varepsilon \gg 0$ and $\varepsilon_i < u_2(a_1,b_2(a_1,\theta^i),\theta^i) - u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)$ if and only if $Ax \gg 0$. The lemma then follows from Motzkin's Theorem. □

For any $(f,b_2^*) \in \mathscr{C} \times A_2^{A_1\times\Theta}$, $b_2 \in A_2^{A_1\times\Theta}$, $a_1 \in A_1$, and $i = 1,2,\ldots,n$, let $\vec{U}_2(a_1,b_2)_j$ denote the j-th component of vector $\vec{U}_2(a_1,b_2)$ and define $\alpha_1 = 1$, $\alpha_{i+1} = u_2(a_1,b_2(a_1,\theta^i),\theta^i) - u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)$, and

$$\alpha_{k+1} = \sum_{j=k}^{i-1} \vec{U}_2(a_1,b_2)_{2j-1} + \alpha_{i+1} - f(a_1,b_2^*(a_1,\theta^k)) + f(a_1,b_2^*(a_1,\theta^i)), \qquad \text{for } k = 1,2,\ldots,i-1,$$

$$\alpha_{l+1} = \sum_{j=i+1}^{l} \vec{U}_2(a_1,b_2)_{2(j-1)} + \alpha_{i+1} - f(a_1,b_2^*(a_1,\theta^l)) + f(a_1,b_2^*(a_1,\theta^i)), \quad \text{for } l = i+1,i+2,\ldots,n,$$

$$\beta_j = \vec{U}_2(a_1,b_2)_{2j} + \vec{U}_2(a_1,b_2)_{2j-1}, \qquad \text{for } j = 1,2,\ldots,n-1.$$

Again, note that α_j and β_j depend on and are uniquely defined by (f,b_2^*), a_1 and (i,b_2) but we suppress this dependency in the notation. We have the following lemma.

Lemma 4. *For any $(f,b_2^*) \in \mathscr{C} \times A_2^{A_1\times\Theta}$, $b_2 \in A_2^{A_1\times\Theta}$, $a_1 \in A_1$ and $i = 1,2,\ldots,n$, there exist $y \in \mathbb{R}^{n+2}$ and $z \in \mathbb{R}^{2(n-1)}$ such that $A'y + C'z = 0$, $y > 0$, and $z \geq 0$ if and only if there exist $\hat{y} \in \mathbb{R}^{n+1}$ and $\hat{z} \in \mathbb{R}^{(n-1)}$ such that $\hat{y} > 0$, $\hat{z} \geq 0$, and*

$$\sum_{j=1}^{n+1} \alpha_j \hat{y}_j + \sum_{j=1}^{n-1} \beta_j \hat{z}_j = 0 \tag{2}$$

Proof of Lemma 4. Fix $(f,b_2^*) \in \mathscr{C} \times A_2^{A_1\times\Theta}$, $b_2 \in A_2^{A_1\times\Theta}$, $a_1 \in A_1$ and $i = 1,2,\ldots,n$. First note that for any y and z, $A'y + C'z = 0$ if and only if

$$y_1 + (u_2(a_1,b_2(a_1,\theta^i),\theta^i) - u_2(a_1,b_2^*(a_1,\theta^i),\theta^i))\,y_{n+2} + V'z = 0 \tag{3}$$

$$D'z = \left[A'y\right]_{-1} \tag{4}$$

where $\left[A'y\right]_{-1}$ is the n-dimensional vector obtained from $A'y$ by eliminating the first row. Recursively adding row 1 to row 2, row 2 to row 3, and so on, we can reduce $\begin{pmatrix} D' & \left[A'y\right]_{-1} \end{pmatrix}$ to a row echelon form and show that (4) holds if and only if

$$z_{2j-1} = z_{2j} + \sum_{k=1}^{j} y_{k+1}, \quad j = 1,2,\ldots,i-1 \tag{5}$$

$$z_{2j} = z_{2j-1} + \sum_{k=j+1}^{n} y_{k+1}, \quad j = i,i+1,\ldots,n-1 \tag{6}$$

$$y_{n+2} = \sum_{k=1}^{n} y_{k+1} \tag{7}$$

Substituting (5)-(7) into (3) we get

$$y_1+\alpha_{i+1}\sum_{k=1}^{n}y_{k+1}+\sum_{j=1}^{i-1}\vec{U}_2(a_1,b_2)_{2j-1}\sum_{k=1}^{j}y_{k+1}+\sum_{j=i}^{n-1}\vec{U}_2(a_1,b_2)_{2j}\sum_{k=j+1}^{n}y_{k+1}+\sum_{j=1}^{i-1}(\vec{U}_2(a_1,b_2)_{2j-1}$$
$$+\vec{U}_2(a_1,b_2)_{2j})z_{2j}+\sum_{j=i}^{n-1}(\vec{U}_2(a_1,b_2)_{2j-1}+\vec{U}_2(a_1,b_2)_{2j})z_{2j-1}-\sum_{k=1}^{n}(f(a_1,b_2^*(a_1,\theta^k))-f(a_1,b_2^*(a_1,\theta^i)))y_{k+1}=0 \tag{8}$$

Therefore, $A'y+C'z=0$ if and only if equations (5) through (8) hold. Now suppose that there exist $y\in\mathbb{R}^{n+2}$ and $z\in\mathbb{R}^{2(n-1)}$ such that $y>0$, $z\geq 0$, and (5) through (8) hold. Define $\hat{y}_j=y_j$, for $j=1,\ldots,n+1$ and

$$\hat{z}_j=\begin{cases}z_{2j}, & j=1,\ldots,i-1\\ z_{2j-1}, & j=i,\ldots,n-1\end{cases}$$

It is easy to verify that $\hat{y}>0$, $\hat{z}\geq 0$, and $\sum_{j=1}^{n+1}\alpha_j\hat{y}_j+\sum_{j=1}^{n-1}\beta_j\hat{z}_j=0$.

Conversely, suppose that there exist $\hat{y}\in\mathbb{R}^{n+1}$ and $\hat{z}\in\mathbb{R}^{(n-1)}$ such that $\hat{y}>0$, $\hat{z}\geq 0$, and (2) holds. Define $y_j=\hat{y}_j$ for $j=1,\ldots,n+1$ and $y_{n+2}=\sum_{i=1}^{n+1}\hat{y}_j$. For any $j=1,\ldots,i-1$, let $z_{2j-1}=\hat{z}_j+\sum_{k=1}^{j}\hat{y}_{k+1}$ and $z_{2j}=\hat{z}_j$, and for any $j=i,\ldots,n-1$, let $z_{2j-1}=\hat{z}_j$ and $z_{2j}=\hat{z}_j+\sum_{k=j+1}^{n}\hat{y}_{k+1}$. It is straightforward to show that $y>0$, $z\geq 0$, and (5) through (8) hold. This completes the proof of Lemma 4. □

Lemma 3 and 4 imply that $(f,b_2^*)\in\mathscr{C}\times A_2^{A_1\times\Theta}$ is renegotiation-proof if and only if for any $a_1\in A_1$, $i\in\{1,2,\ldots,n\}$ and $b_2\in A_2^{A_1\times\Theta}$, there exist $\hat{y}\in\mathbb{R}^{n+1}$ and $\hat{z}\in\mathbb{R}^{(n-1)}$ such that $\hat{y}>0$, $\hat{z}\geq 0$, and equation (2) holds. We can now complete the proof of Theorem 1 (of the main paper).

[Only if] Suppose, for contradiction, that there exist $a_1\in A_1$, $i=1,2,\ldots,n$ and an increasing $b_2\in A_2^{A_1\times\Theta}$ such that $u_2(a_1,b_2(a_1,\theta^i),\theta^i)>u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)$, but there is no $k=1,2,\ldots,i-1$ such that (4) holds and no $l=i+1,\ldots,n$ such that (5) holds. This implies that $\alpha_j>0$ for all $j=1,\ldots,n+1$. Since u_2 has increasing differences, $\beta_j\geq 0$ for all $j=1,\ldots,n-1$. Therefore, $\hat{y}>0$ and $\hat{z}\geq 0$ imply that $\sum_{j=1}^{n+1}\alpha_j\hat{y}_j+\sum_{j=1}^{n-1}\beta_j\hat{z}_j>0$, which, by Lemma 4, contradicts that (f,b_2^*) is renegotiation-proof.

[If] Fix arbitrary $a_1\in A_1$, $i=1,2,\ldots,n$ and increasing $b_2\in A_2^{A_1\times\Theta}$ such that $u_2(a_1,b_2(a_1,\theta^i),\theta^i)>u_2(a_1,b_2^*(a_1,\theta^i),\theta^i)$. Suppose first that there exists a $k\in\{1,\ldots,i-1\}$ such that (4) holds. This implies that $\alpha_{i+1}>0$ and $\alpha_{k+1}\leq 0$. Let $\hat{y}_{k+1}=1$, $\hat{y}_{i+1}=\frac{-\alpha_{k+1}}{\alpha_{i+1}}\geq 0$, and all the other $\hat{y}_j=0$ and $\hat{z}_j=0$. This implies that equation (2) holds and, by Lemma 3 and 4, that (f,b_2^*) is renegotiation-proof. Suppose now that there exists an $l\in\{i+1,\ldots,n\}$ such that (5) holds. Then, $\alpha_{i+1}>0$ and $\alpha_{l+1}\leq 0$. Let $\hat{y}_{l+1}=1$, $\hat{y}_{i+1}=\frac{-\alpha_{l+1}}{\alpha_{i+1}}\geq 0$ and all the other $\hat{y}_j=0$ and $\hat{z}_j=0$. This, again, implies that (2) holds and that (f,b_2^*) is renegotiation-proof. □

Proof of Proposition 3 (of the main paper). Suppose that b_2^* is renegotiation-proof and fix a_1, $i=1,\ldots,n$ and a $b_2(a_1,\theta^i)\in\mathfrak{B}(a_1,i,b_2^*)$. For any $j=1,\ldots,n$, let $c_j=e_i-e_j$, where e_j is the

j^{th} standard basis row vector for $\mathbb{R}^n$, and define

$$E_j = \begin{pmatrix} D \\ c_j \end{pmatrix}$$

Also let

$$w_k = u_2(a_1, b_2(a_1,\theta^i),\theta^i) - u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) + \sum_{j=k}^{i-1} \vec{U}_2(a_1,b_2)_{2j-1}$$

$$w_l = u_2(a_1, b_2(a_1,\theta^i),\theta^i) - u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) + \sum_{j=i+1}^{l} \vec{U}_2(a_1,b_2)_{2(j-1)}$$

for any $k \in \{1,\dots,i-1\}$ and $l \in \{i+1,\dots,n\}$ and define

$$V_j = \begin{pmatrix} \vec{U}_2(a_1,b_2^*) \\ -w_j \end{pmatrix}$$

Incentive compatibility of (f, b_2^*) implies that $Df(a_1,b_2^*) \le \vec{U}_2(a_1,b_2^*)$. Renegotiation proofness, by Theorem 1 (of the main paper), implies that $c_k f(a_1,b_2^*) \le -w_k$ for some $k \in \{1,\dots,i-1\}$ or $c_l f(a_1,b_2^*) \le -w_l$ for some $l \in \{i+1,\dots,n\}$. Suppose first that there exists a $k \in \{1,\dots,i-1\}$ such that $c_k f(a_1,b_2^*) \le -w_k$. Then we must have $E_k f(a_1,b_2^*) \le V_k$. By Gale's theorem of linear inequalities, this implies that $x \ge 0$ and $E_k' x = 0$ implies $x' V_k \ge 0$. Denote the first $2(n-1)$ elements of x by y and the last element by z. It is easy to show that $E_k' x = 0$ implies that $y_{2j-1} = y_{2j} + z$ for $j \in \{k, k+1,\dots,i-1\}$ and $y_{2j-1} = y_{2j}$ for $j \notin \{k, k+1,\dots,i-1\}$. Therefore,

$$\begin{aligned} x' V_k &= \sum_{j=1}^{n-1} \vec{U}_2(a_1,b_2^*)_{2j} y_{2j} + \sum_{j=1}^{n-1} \vec{U}_2(a_1,b_2^*)_{2j-1} y_{2j-1} - z w_k \\ &= \sum_{j=1}^{n-1} (\vec{U}_2(a_1,b_2^*)_{2j} + \vec{U}_2(a_1,b_2^*)_{2j-1}) y_{2j} + z(-w_k + \sum_{j=k}^{i-1} \vec{U}_2(a_1,b_2^*)_{2j-1}) \\ &\ge 0 \end{aligned}$$

This implies that $-w_k + \sum_{j=k}^{i-1} \vec{U}_2(a_1,b_2^*)_{2j-1} \ge 0$ and hence k is a blocking type.

Similarly, we can show that, if there exists an $l \in \{i+1,\dots,n\}$ such that $c_l f(a_1,b_2^*) \le -w_l$, then l is a blocking type, and this completes the proof. □

Proof of Proposition 4 (of the main paper). Let $b_2^* \in A_2^{A_1 \times \Theta}$ be an increasing strategy satisfying the conditions of the proposition. We will show that there exist an $f \in \mathscr{C}$ such that (f, b_2^*) is incentive-compatible and renegotiation-proof. Fix an $a_1 \in A_1$ and for each $i = 1,\cdots,n$ and $b_2^i \in \mathscr{B}(a_1,i,b_2^*)$ pick a blocking type $m(b_2^i) = 1,\cdots,n$ that satisfies the conditions given in the proposition. For each $i = 1$ and $b_2^i \in \mathscr{B}(a_1,i,b_2^*)$ define the n-dimensional row vector

$c_{b_2^i} = e_i - e_{m(b_2^i)}$, where e_j is the j^{th} standard basis row vector for $\mathbb{R}^n$, and the scalar $w_{b_2^i}$ as

$$w_{b_2^i} = u_2(a_1, b_2^i(a_1,\theta^i),\theta^i) - u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) + \mathbf{1}_{\{m(b_2^i)\le i-1\}} \sum_{j=m(b_2^i)}^{i-1} \vec{U}_2(a_1, b_2^i)_{2j-1} + \mathbf{1}_{\{i\le m(b_2^i-1\}} \sum_{j=i+1}^{m(b_2^i)} \vec{U}_2(a_1, b_2^i)_{2(j-1)}. \quad (9)$$

Note that for a given $a_1 \in A_1$ and $i = 1,\cdots,n$, $\mathscr{B}(a_1,i,b_2^*)$ is finite and let $\sum_{i=1}^n |\mathscr{B}(a_1,i,b_2^*)| = p$. Denote with $C(a_1)$, the $p \times n$ matrix composed of all the rows $c_{b_2^i}$ and with $W(a_1)$the p dimensional vector with component $w_{b_2^i}$ corresponding to each b_2^i. Let $E(a_1)$ be the matrix

$$E(a_1) = \begin{pmatrix} D \\ C(a_1) \end{pmatrix}$$

and $V(a_1)$ the column vector

$$V(a_1) = \begin{pmatrix} \vec{U}_2(a_1, b_2^*) \\ -W(a_1) \end{pmatrix}$$

Now, if for each $a_1 \in A_1$, we can find an $f(a_1,b_2^*)$ such that $E(a_1)f(a_1,b_2^*) \le V(a_1)$ the proof would be completed. In fact, if $E(a_1)f(a_1,b_2^*) \le V(a_1)$, then $Df(a_1,b_2^*) \le \vec{U}_2(a_1,b_2^*)$, which implies that (f,b_2^*) incentive compatible. Furthermore, $E(a_1)f(a_1,b_2^*) \le V(a_1)$ implies $W(a_1) \le -C(a_1)f(a_1,b_2^*)$ and, by Theorem 1 (of the main paper), that (f,b_2^*) is renegotiation-proof. Gale's theorem of linear inequalities implies that there exist $f(a_1,b_2^*) \in \mathbb{R}^n$ such that $E(a_1)f(a_1,b_2^*) \le V(a_1)$ if and only if $x \in \mathbb{R}^{p+2(n-1)}$, $x \ge 0$ and $E(a_1)'x = 0$ implies $x'V(a_1) \ge 0$. Decompose x into two vectors so that the first $2(n-1)$ elements constitute y and the remaining p components constitute z. Notice that for any $i = 1,\ldots,n$ and $b_2^i \in \mathfrak{B}(a_1,i,b_2^*)$ there is a corresponding element of z, which we will denote $z_{b_2^i}$.

Recursively adding row 1 to row 2, row 2 to row 3, and so on, we can reduce $E(a_1)'$ to a row echelon form and show that $E(a_1)'x = 0$ if and only if

$$y_{2j-1} = y_{2j} + \sum_{b_2^i} z_{b_2^i}[\mathbf{1}_{\{m(b_2^i)\le j\le i-1\}} - \mathbf{1}_{\{i\le j\le m(b_2^i)-1\}}] \quad (10)$$

for $j = 1,\ldots,n-1$.

Let $J_- = \{j \in \{1,\ldots,n-1\} : \exists b_2^i$ such that $i \le j \le m(b_2^i)-1\}$ and $J_+ = \{j \in \{1,\ldots,n-1\} : \exists b_2^i$ such that $m(b_2^i) \le j \le i-1\}$ and note that $J_- \cap J_+ = \emptyset$. To see this, suppose, for contradiction, that there exists a $j \in J_- \cap J_+$. Therefore, there exists a b_2^i such that $i \le j \le m(b_2^i)-1$ and $b_2^{i'}$ such that $m(b_2^{i'}) \le j \le i'-1$. This implies that $i < i'$, $m(b_2^i) > i$, $m(b_2^{i'}) < i'$, but $m(b_2^i) > m(b_2^{i'})$, contradicting the conditions of the proposition. We can therefore write (10) as

$$y_{2j} = y_{2j-1} + \sum_{b_2^i} z_{b_2^i} \mathbf{1}_{\{i\le j\le m(b_2^i)-1\}} \quad (11)$$

for $j \in J_-$ and

$$y_{2j-1} = y_{2j} + \sum_{b_2^i} z_{b_2^i} \mathbf{1}_{\{m(b_2^i) \le j \le i-1\}} \tag{12}$$

for $j \in J_+$.

Finally note that

$$x'V(a_1) = \sum_{j=1}^{n-1} \vec{U}_2(a_1, b_2^*)_{2j} y_{2j} + \sum_{j=1}^{n-1} \vec{U}_2(a_1, b_2^*)_{2j-1} y_{2j-1} - \sum_{b_2^i} z_{b_2^i} w_{b_2^i}$$

Substituting from (11) and (12) we obtain

$$\begin{aligned} x'V(a_1) = &\sum_{j \in J_-} \left[\vec{U}_2(a_1, b_2^*)_{2j} + \vec{U}_2(a_1, b_2^*)_{2j-1}\right] y_{2j-1} + \sum_{j \in J_+} \left[\vec{U}_2(a_1, b_2^*)_{2j} + \vec{U}_2(a_1, b_2^*)_{2j-1}\right] y_{2j} \\ &+ \sum_{b_2^i} z_{b_2^i} \left[-w_{b_2^i} + \mathbf{1}_{\{m(b_2^i) \le i-1\}} \sum_{j=m(b_2^i)}^{i-1} \vec{U}_2(a_1, b_2^*)_{2j-1} + \mathbf{1}_{\{i \le m(b_2^i)-1\}} \sum_{j=i}^{m(b_2^i)-1} \vec{U}_2(a_1, b_2^*)_{2j} \right] \end{aligned}$$

Increasing differences, the definition of $m(b_2^i)$, and $y, z \ge 0$ imply that $x'V \ge 0$, and the proof is completed. □

For any strategy profile (b_1, b_2) of G define the expected transfer from player 2 to the third-party as $F(b_1, b_2) = \sum_{\theta \in \Theta} p(\theta) f(b_1, b_2(b_1, \theta))$.

3 Proofs for section 4.1

In the game with non-renegotiable contracts, player i has an information set at the beginning of the game, which we denote $\emptyset$, and an information set for each $(f_i, \theta_i, I) \in \mathscr{C} \times \Theta_i \times \mathscr{I}_i$. In the game with renegotiable contracts $\Gamma_R(G)$, player i has additional information sets corresponding to accepted or rejected renegotiation offers, (f_i, θ_i, I, g, y) and (f_i, θ_i, I, g, n), and player t_i has an additional information set of each (f_i, I, g).

The following is well-known but we state it without proof for easy reference:

Lemma 5. *Let (σ, m) be a consistent assessment and $f_i \in \mathbb{R}^Z$. For any i, θ_i, $I \in \mathscr{I}_i$*

$$\sigma_i(I, \theta_i) \in \underset{\sigma_i' \in \Delta_i|_I}{\operatorname{argmax}} \left\{ U_i(\sigma_i', \sigma_{-i}, m|I, \theta_i) - F_i(\sigma_i', \sigma_{-i}, m|I) \right\} \tag{13}$$

if and only if every pure strategy $s_i \in \operatorname{supp}(\sigma_i(I, \theta_i))$ satisfies

$$s_i \in \underset{s_i \in S_i|_I}{\operatorname{argmax}} U_i^{\sigma,m}(s_i, \theta_i|I) - F_i^{\sigma,m}(s_i|I) \tag{14}$$

Lemma 6. *Let (σ, m) be a consistent assessment and $f_i \in \mathbb{R}^Z$. Then for any i, θ_i, $I \in \mathscr{I}_i$, (13) implies that σ_i is increasing at I.*

Proof of Lemma 6. Fix a consistent assessment (σ, m). Our assumption that $u_i(h, s_i, s_{-i}, \theta_i, \theta_{-i})$ has strictly increasing differences in (s_i, θ_i) implies that

$$v_i(h, s_i, s_{-i}, \theta_i, \theta_{-i}) = u_i(h, s_i, s_{-i}, \theta_i, \theta_{-i}) - f_i(h, s_i, s_{-i})$$

has strictly increasing differences. Furthermore, since S_i is a chain, $v_i(h, s_i, s_{-i}, \theta_i, \theta_{-i})$ is supermodular on S_i. Finally, note that $U_i^{\sigma,m}(s_i, \theta_i|I) - F_i^{\sigma,m}(s_i|I)$ is nothing but $v_i(h, s_i, s_{-i}, \theta_i, \theta_{-i})$ integrated over (s_{-i}, h, θ_{-i}) and, since supermodularity and strictly increasing differences are preserved under integration, we conclude that $U_i^{\sigma,m}(s_i, \theta_i|I) - F_i^{\sigma,m}(s_i|I)$ is supermodular on S_i and has strictly increasing differences in (s_i, θ_i).

Now let $\sigma_i(I, \theta_i)$ satisfy (13). Then, by Lemma 5, every pure strategy $s_i \in \text{supp}(\sigma_i(I, \theta_i))$ satisfies (14). Well known monotone comparative statics results (see for example Theorem 2.8.4 in Topkis (1998)) then imply that

$$\theta_i \succ_i \theta_i', s_i \in \text{supp}(\sigma_i(I, \theta_i)), s_i' \in \text{supp}(\sigma_i(I, \theta_i')) \Rightarrow s_i \succsim_{si} s_i'$$

which proves that σ_i is increasing at I. □

Let the number of elements in Θ_i be equal to n_i and order its elements so that $\theta_i^{n_i} \succsim_i \theta_i^{n_i-1} \succsim_i \cdots \theta_i^2 \succsim_i \theta_i^1$. Lemma 5 has shown that sequential rationality of i at information set I is equivalent to

$$s_i \in \underset{s_i \in S_i|_I}{\text{argmax}} U_i^{\sigma,m}(s_i, \theta_i|I) - F_i^{\sigma,m}(s_i|I)$$

for every $s_i \in \text{supp}(\sigma_i(I, \theta_i))$. The following lemma shows that this is equivalent to an incentive compatibility condition.

Lemma 7. *Fix a consistent assessment (σ, m), $i \in N$, $I \in \mathscr{I}_i$, and $\sigma_i : \Theta_i \to \Delta(S_i|_I)$. There exists a contract $\hat{f}_i$ such that*

$$\sigma_i(\theta_i^j) \in \underset{\sigma_i' \in \Sigma_i|_I}{\text{argmax}} U_i^{\sigma,m}(\sigma_i', \theta_i^j|I) - \hat{F}_i^{\sigma,m}(\sigma_i'|I), \forall\, j = 1, \ldots, n_i. \tag{15}$$

iff there exists a contract f_i such that

$$U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j|I) - F_i^{\sigma,m}(\sigma_i(\theta_i^j)|I) \geq U_i^{\sigma,m}(\sigma_i(\theta_i^k), \theta_i^j|I) - F_i^{\sigma,m}(\sigma_i(\theta_i^k)|I), \forall\, j, k = 1, \ldots, n_i. \tag{16}$$

Proof of Lemma 7. Necessity is immediate. For sufficiency, suppose that there exists a contract f_i such that $\sigma_i : \Theta_i \to \Delta(S_i|_I)$ satisfies (16). Define $\hat{f}_i$ as follows

$$\hat{f}_i(h, s_i, s_{-i}) = \begin{cases} f_i(h, s_i, s_{-i}), & \text{if } \exists j : s_i \in \text{supp}(\sigma_i(\theta_i^j)) \\ \max_{\theta_i}\{U_i^{\sigma,m}(s_i, \theta_i) - U_i^{\sigma,m}(\sigma_i(\theta_i), \theta_i) + F_i^{\sigma,m}(\sigma_i(\theta_i)|I)\}, & \text{otherwise} \end{cases}$$

It is easy to show that $s_i : \Theta_i \to S_i|_I$ and $\hat{f}_i$ satisfy (15). □

For any $\sigma_i, \sigma_i' \in \Sigma_i$ define $\succsim_{\sigma i}$ by $\sigma_i \succsim_{\sigma i} \sigma_i'$ iff $s_i \succsim_{si} s_i'$ for all $s_i \in \text{supp}(\sigma_i)$ and $s_i' \in \text{supp}(\sigma_i')$. The following lemma shows that increasing differences in pure strategies implies increasing differences in mixed strategies.

Lemma 8. *Fix an assessment (σ, m), $i \in N$, and $I \in \mathscr{I}_i$. If $U_i^{\sigma,m}(s_i, \theta_i|I)$ has (strictly) increasing differences in $(s_i, \theta_i) \in S_i|_I \times \Theta_i$, then $U_i^{\sigma,m}(\sigma_i, \theta_i|I)$ has (strictly) increasing differences in $(\sigma_i, \theta_i) \in \Sigma_i|_I \times \Theta_i$.*

Proof of Lemma 8. Let $\sigma_i \succsim_{\sigma i} \sigma_i'$ and $\theta_i \succsim_i \theta_{i'}$. This implies that $s_i \succsim_{si} s_i'$ for all $s_i \in \text{supp}(\sigma_i)$ and $s_i' \in \text{supp}(\sigma_i')$. Therefore, increasing differences in (s_i, θ_i) implies

$$U_i^{\sigma,m}(s_i, \theta_i|I) - U_i^{\sigma,m}(s_i, \theta_i'|I) \geq U_i^{\sigma,m}(s_i', \theta_i|I) - U_i^{\sigma,m}(s_i', \theta_i'|I)$$

for all $s_i \in \text{supp}(\sigma_i)$ and $s_i' \in \text{supp}(\sigma_i')$. This implies

$$\sum_{s_i \in \text{supp}(\sigma_i)} \sigma_i(s_i)[U_i^{\sigma,m}(s_i, \theta_i|I) - U_i^{\sigma,m}(s_i, \theta_i'|I)] \geq \sum_{s_i' \in \text{supp}(\sigma_i')} \sigma_i'(s_i')[U_i^{\sigma,m}(s_i', \theta_i|I) - U_i^{\sigma,m}(s_i', \theta_i'|I)]$$

and hence

$$U_i^{\sigma,m}(\sigma_i, \theta_i|I) - U_i^{\sigma,m}(\sigma_i, \theta_i'|I) \geq U_i^{\sigma,m}(\sigma_i', \theta_i|I) - U_i^{\sigma,m}(\sigma_i', \theta_i'|I).$$

□

This lemma has a useful corrolary:

Corollary 1. *Fix an assessment (σ, m), $i \in N$, and $I \in \mathscr{I}_i$. If $U_i^{\sigma,m}(s_i, \theta_i|I)$ has (strictly) increasing differences in $(s_i, \theta_i) \in S_i|_I \times \Theta_i$, then $U_i^{\sigma,m}(\sigma_i, \theta_i|I) - F_i^{\sigma,m}(\sigma_i|I)$ has (strictly) increasing differences in $(\sigma_i, \theta_i) \in \Sigma_i|_I \times \Theta_i$.*

Let e_j be the j^{th} standard basis row vector for $\mathbb{R}^{n_i}$ and define the row vector $d_j = e_j - e_{j+1}$, $j = 1, 2, \ldots, n_i - 1$. Let D be the $2(n_i - 1) \times n_i$ matrix whose row $2j-1$ is d_j and row $2j$ is $-d_j$, $j = 1, \ldots, n-1$. For any $i \in N$, $I \in \mathscr{I}_i$, consistent assessment (σ, m), and $\sigma_i' : \Theta_i \to \Delta(S_i|_I)$ define $\vec{U}_i^{\sigma,m}(\sigma_i'|I)$ as a column vector with $2(n_i - 1)$ components, where component $2j-1$ is given by $U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta^j|I) - U_i^{\sigma,m}(\sigma_i'(\theta_i^{j+1}), \theta^j|I)$ and component $2j$ is given by $U_i^{\sigma,m}(\sigma_i'(\theta_i^{j+1}), \theta^{j+1}, |I) - U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta^{j+1}|I)$, $j = 1, 2, \ldots, n_i - 1$. Similarly, define $\vec{F}_i^{\sigma,m}(\sigma_i'|I)$ as the n_i vector whose jth component is given by $F_i^{\sigma,m}(\sigma_i'(\theta_i^j)|I)$, $j = 1, 2, \ldots, n_i - 1$.

The following lemmas show that if $s_i : \Theta_i \to S_i|_I$ is increasing then it satisfies (16) iff it satisfies local conditions and express it in a convenient matrix notation.

Lemma 9. *Fix $i \in N$ and $I \in \mathscr{I}_i$ and let $\sigma_i : \Theta_i \to \Delta(S_i)$ be increasing and $V_i : \Sigma_i \times \Theta_i \to \mathbb{R}$ has increasing differences in (σ_i, θ_i). Then,*

$$V_i(\sigma_i(\theta_i^j), \theta_i^j) \geq V_i(\sigma_i(\theta_i^k), \theta_i^j), \forall\, j, k = 1, \ldots, n_i \tag{17}$$

iff

$$V_i(\sigma_i(\theta_i^j),\theta_i^j) \geq V_i(\sigma_i(\theta_i^{j-1}),\theta_i^j), \forall\, j = 2,\ldots,n_i \tag{18}$$

and

$$V_i(\sigma_i(\theta_i^j),\theta_i^j) \geq V_i(\sigma_i(\theta_i^{j+1}),\theta_i^j), \forall\, j = 1,\ldots,n_i - 1. \tag{19}$$

Proof of Lemma 9. Necessity is immediate. For sufficiency, suppose that (18) holds. This implies $V_i(\sigma_i(\theta_i^j),\theta_i^j) \geq V_i(\sigma_i(\theta_i^{j-1}),\theta_i^j)$ and $V_i(\sigma_i(\theta_i^{j-1}),\theta_i^{j-1}) \geq V_i(\sigma_i(\theta_i^{j-2}),\theta_i^{j-1})$, which, in turn, implies

$$V_i(\sigma_i(\theta_i^j),\theta_i^j) - V_i(\sigma_i(\theta_i^{j-1}),\theta_i^j) + V_i(\sigma_i(\theta_i^{j-1}),\theta_i^{j-1}) - V_i(\sigma_i(\theta_i^{j-2}),\theta_i^{j-1}) - V_i(\sigma_i(\theta_i^{j-2}),\theta_i^j) + V_i(\sigma_i(\theta_i^{j-2}),\theta_i^j) \geq 0$$

Therefore,

$$V_i(\sigma_i(\theta_i^j),\theta_i^j) - V_i(\sigma_i(\theta_i^{j-2}),\theta_i^j) \geq V_i(\sigma_i(\theta_i^{j-1}),\theta_i^j) - V_i(\sigma_i(\theta_i^{j-2}),\theta_i^j) - [V_i(\sigma_i(\theta_i^{j-1}),\theta_i^{j-1}) - V_i(\sigma_i(\theta_i^{j-2}),\theta_i^{j-1})] \geq 0$$

where the second inequality follows from increasing differences. We have proved that (18) implies all downward constraints. It can be similarly shown that (19) implies all upward conditions. □

Lemma 10. *Fix a consistent assessment (σ, m), $i \in N$, a contract f_i, $I \in \mathscr{I}_i$, and let $\sigma_i : \Theta_i \to \Sigma_i|_I$ be increasing. Then* (16) *holds iff $D\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq \vec{U}_i^{\sigma,m}(\sigma_i|I)$.*

Proof of Lemma 10. Follows from Corollary 1 and lemma 9. □

Lemma 11. *Fix $i \in N$, $\sigma_i : \Theta_i \to \Delta(S_i)$, $I \in \mathscr{I}_i$, and let $|S_i| = k$. If there exists n_i vector F, then there also exists a k vector f such that*

$$F_j = \sum_{l=1}^{k} \sigma_i(s_i^l|\theta_i^j) f_l$$

whenever $\sigma_i(\theta_i) \neq \sigma_i(\theta_i')$ for any $\theta_i, \theta_i' \in \Theta_i$.

Proof of Lemma 11. The proof is done if we can show that the following system of equations has a solution for f:

$$\Upsilon f = F$$

where jth row of Υ is given by $\sigma_i * \theta_i^j$. Since $\sigma_i(\theta_i) \neq \sigma_i(\theta_i')$ for any $\theta_i, \theta_i' \in \Theta_i$, the rows of Υ are independent. Therefore, $\text{rank}(\Upsilon) = \text{rank}\begin{pmatrix}\Upsilon & F\end{pmatrix}$, and the proof is completed. □

Proof of Proposition 5 (of the main paper). [Only If] Follows from Lemma 6.

[If] We need to prove the existence of a contract f_i such that for any $I \in \mathscr{I}_i$

$$U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) - F_i^{\sigma,m}(\sigma_i(\theta_i^j)|I) \geq U_i^{\sigma,m}(\sigma_i(\theta_i^k),\theta_i^j|I) - F_i^{\sigma,m}(\sigma_i(\theta_i^k)|I), \forall\, j,k = 1,\ldots,n_i. \tag{20}$$

By Lemma 10, (20) holds if and only if $D\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq \vec{U}_i^{\sigma,m}(\sigma_i|I)$. By Gale's theorem for linear inequalities (Mangasarian (1994), p. 33), there exists a $\vec{F}_i^{\sigma,m}(\sigma_i|I) \in \mathbb{R}^{n_i}$ such that $D\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq \vec{U}_i^{\sigma,m}(\sigma_i|I)$ if and only if for any $y \in \mathbb{R}_+^{2(n_i-1)}$, $D'y = 0$ implies $y'\vec{U}_i^{\sigma,m}(\sigma_i|I) \geq 0$. It is easy to show that $D'y = 0$ if and only if $y_1 = y_2, y_3 = y_4, \cdots, y_{2(n_i-1)-1} = y_{2(n_i-1)}$. Let $\vec{U}_i^{\sigma,m}(\sigma_i|I)_j$ denote the j^{th} row of $\vec{U}_i^{\sigma,m}(\sigma_i|I)$ and note that since σ_i is increasing and $U_i^{\sigma,m}(\sigma_i,\theta_i|I)$ has strictly increasing differences, $\vec{U}_i^{\sigma,m}(\sigma_i|I)_{2j-1} + \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2j} \geq 0$, for any $j = 1,2,\ldots,n_i-1$. Therefore,

$$y'\vec{U}_i^{\sigma,m}(\sigma_i|I) = \sum_{j=1}^{n_i-1} (\vec{U}_i^{\sigma,m}(\sigma_i|I)_{2j-1} + \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2j}) y_{2j-1} \geq 0$$

Lemma 11 implies that as long as different types play different mixed strategies we can use this $\vec{F}_i^{\sigma,m}(\sigma_i|I) \in \mathbb{R}^{n_i}$ to construct $F(s_i)$ for each $s_i \in S_i|_I$.[1]. We can then define

$$f_i(h,s_i,s_{-i}) = F(s_i) \forall\, h,s_{-i}.$$

and use this f_i to construct a contract that satisfies incentive whose expectation satisfies (16). This completes the proof. □

Proof of Theorem 2 (of the main paper). Fix $i \in N$ and a consistent assessment (σ,m). By definition (f_i,σ_i) is not renegotiation-proof at (σ_{-i},m) if and only if there exist $I \in \mathscr{I}_i$, $j = 1,\ldots,n_i$, $g_i \in \mathscr{C}$, and $\sigma_i' : \Theta_i \to \Sigma_i|_I$ such that

$$U_i^{\sigma,m}(\sigma_i'(\theta_i^l),\theta_i^l|I) - G_i^{\sigma,m}(\sigma_i'(\theta_i^l)|I) \geq U_i^{\sigma,m}(\sigma_i'(\theta_i^k),\theta_i^l|I) - G_i^{\sigma,m}(\sigma_i'(\theta_i^k)|I), \forall\, l,k = 1,\ldots,n_i \tag{21}$$

$$U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) - G_i^{\sigma,m}(\sigma_i'(\theta_i^j)|I) > U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) - F_i^{\sigma,m}(\sigma_i(\theta_i^j)|I) \tag{22}$$

$$G_i^{\sigma,m}(\sigma_i'(\theta_i^k)|I) > F_i^{\sigma,m}(\sigma_i(\theta_i^k)|I), \forall\, k = 1,\ldots,n_i. \tag{23}$$

Define $\vec{G}_i^{\sigma,m}(\sigma_i'|I)$ as n_i vector whose kth component is given by $G_i^{\sigma,m}(\sigma_i'(\theta_i^k)|I)$. By Lemma 10, (21) is equivalent to $D\vec{G}_i^{\sigma,m}(\sigma_i'|I) \leq \vec{U}_i^{\sigma,m}(\sigma_i'|I)$. Therefore, we have the following:

[1] If different types use the same strategies, we can simply eliminate one of them from the incentive compatibility constraints.

Lemma 12. *Fix $i \in N$ and a consistent assessment (σ, m). (f_i, σ_i) is not renegotiation-proof at (σ_{-i}, m) if and only if there exist $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, $g_i \in \mathscr{C}$, $\sigma_i' : \Theta_i \to \Sigma_i|_I$, and $\varepsilon \in \mathbb{R}^{n_i}$ such that $D(\vec{F}_i^{\sigma,m}(\sigma_i|I) + \varepsilon) \leq \vec{U}_i^{\sigma,m}(\sigma_i'|I)$, $\varepsilon_j < U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j|I)$, and $\varepsilon \gg 0$.*

Fix $i \in N$, a consistent assessment (σ, m), (f_i, σ_i), $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, $\sigma_i' : \Theta_i \to \Sigma_i|_I$. Define $V = \vec{U}_i^{\sigma,m}(\sigma_i'|I) - D\vec{F}_i^{\sigma,m}(\sigma_i|I)$, $C = \begin{pmatrix} V & -D \end{pmatrix}$, and

$$A = \begin{pmatrix} I_{n_i+1} \\ l_j \end{pmatrix}$$

where $l_j = (U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j|I))e_1 - e_{j+1}$. The following lemma uses Motzkin's Theorem to express renegotiation-proofness as an alternative.

Lemma 13. *Fix $i \in N$ and a consistent assessment (σ, m). (f, σ_i) is renegotiation-proof if and only if for any $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, $\sigma_i' : \Theta_i \to \Sigma_i|_I$, there exist $y \in \mathbb{R}^{n_i+2}$ and $z \in \mathbb{R}^{2(n_i-1)}$ such that $A'y + C'z = 0$, $y > 0$, $z \geq 0$.*

Proof of Lemma 13. Let $\xi > 0, \varepsilon \in \mathbb{R}^{n_i}$, and define

$$x = \begin{pmatrix} \xi \\ \xi\varepsilon \end{pmatrix}$$

Then $D(\vec{F}_i^{\sigma,m}(\sigma_i|I) + \varepsilon) \leq \vec{U}_i^{\sigma,m}(\sigma_i'|I)$ if and only if $Cx \geq 0$. Also, $\varepsilon \gg 0$ and $\varepsilon_j < U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j|I)$ if and only if $Ax \gg 0$. The lemma then follows from Motzkin's Theorem. □

Fix $i \in N$, a belief system m, $\sigma \in \Sigma_{-i}$, (f, σ_i), $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, and $\sigma_i' \in \Sigma_i|_I$. Define $\alpha_1 = 1$, $\alpha_{j+1} = U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j|I)$, and

$$\alpha_{k+1} = \sum_{t=k}^{j-1} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t-1} + \alpha_{j+1} - F_i^{\sigma,m}(\sigma_i(\theta_i^k)|I) + F_i^{\sigma,m}(\sigma_i(\theta_i^j)|I), \qquad \text{for } k = 1, 2, \ldots, j-1,$$

$$\alpha_{l+1} = \sum_{t=j+1}^{l} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2(t-1)} + \alpha_{j+1} - F_i^{\sigma,m}(\sigma_i(\theta_i^l)|I) + F_i^{\sigma,m}(\sigma_i(\theta_i^j)|I), \quad \text{for } l = j+1, j+2, \ldots, n_i,$$

$$\beta_t = \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t} + \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t-1}, \qquad \text{for } t = 1, 2, \ldots, n_i - 1.$$

We have the following lemma.

Lemma 14. *Fix $i \in N$, a consistent assessment (σ, m), (f_i, σ_i), $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, $\sigma_i' : \Theta_i \to \Sigma_i|_I$. There exist $y \in \mathbb{R}^{n_i+2}$ and $z \in \mathbb{R}^{2(n_i-1)}$ such that $A'y + C'z = 0$, $y > 0$, $z \geq 0$ if and only if there exist $\hat{y} \in \mathbb{R}^{n_i+1}$ and $\hat{z} \in \mathbb{R}^{(n_i-1)}$ such that $\hat{y} > 0$, $\hat{z} \geq 0$, and*

$$\sum_{t=1}^{n_i+1} \alpha_t \hat{y}_t + \sum_{t=1}^{n_i-1} \beta_t \hat{z}_t = 0 \tag{24}$$

Proof of Lemma 14. First note that for any y and z, $A'y + C'z = 0$ if and only if

$$y_1 + (U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I))\, y_{n_i+2} + V'z = 0 \tag{25}$$

$$D'z = \left[A'y\right]_{-1} \tag{26}$$

where $\left[A'y\right]_{-1}$ is the n_i-dimensional vector obtained from $A'y$ by eliminating the first row. Recursively adding row 1 to row 2, row 2 to row 3, and so on, we can reduce $\begin{pmatrix} D' & \left[A'y\right]_{-1} \end{pmatrix}$ to a row echelon form and show that (26) holds if and only if

$$z_{2t-1} = z_{2t} + \sum_{k=1}^{t} y_{k+1}, \quad t = 1,2,\ldots,j-1 \tag{27}$$

$$z_{2t} = z_{2t-1} + \sum_{k=t+1}^{n_i} y_{k+1}, \quad t = j, j+1,\ldots,n_i-1 \tag{28}$$

$$y_{n_i+2} = \sum_{k=1}^{n_i} y_{k+1} \tag{29}$$

Substituting (27)-(29) into (25) we get

$$\begin{aligned} &y_1 + \alpha_{j+1} \sum_{k=1}^{n_i} y_{k+1} + \sum_{t=1}^{j-1} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t-1} \sum_{k=1}^{t} y_{k+1} + \sum_{t=j}^{n_i-1} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t} \sum_{k=t+1}^{n_i} y_{k+1} + \sum_{t=1}^{j-1} (\vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t-1} \\ &+ \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t}) z_{2t} + \sum_{t=j}^{n_i-1} (\vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t-1} + \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t}) z_{2t-1} - \sum_{k=1}^{n_i} (F_i^{\sigma,m}(\sigma_i(\theta_i^k)|I) - F_i^{\sigma,m}(\sigma_i(\theta_i^j)|I))\, y_{k+1} = 0 \end{aligned} \tag{30}$$

Therefore, $A'y + C'z = 0$ if and only if equations (27) through (30) hold. Now suppose that there exist $y \in \mathbb{R}^{n_i+2}$ and $z \in \mathbb{R}^{2(n_i-1)}$ such that $y > 0$, $z \geq 0$, and (27) through (30) hold. Define $\hat{y}_t = y_t$, for $t = 1,\ldots,n+1$ and

$$\hat{z}_t = \begin{cases} z_{2t}, & t = 1,\ldots,j-1 \\ z_{2t-1}, & t = j,\ldots,n_i-1 \end{cases}$$

It is easy to verify that $\hat{y} > 0$, $\hat{z} \geq 0$, and $\sum_{t=1}^{n_i+1} \alpha_t \hat{y}_t + \sum_{t=1}^{n_i-1} \beta_t \hat{z}_t = 0$.

Conversely, suppose that there exist $\hat{y} \in \mathbb{R}^{n_i+1}$ and $\hat{z} \in \mathbb{R}^{(n_i-1)}$ such that $\hat{y} > 0$, $\hat{z} \geq 0$, and (24) holds. Define $y_t = \hat{y}_t$ for $t = 1,\ldots,n_i+1$ and $y_{n_i+2} = \sum_{t=1}^{n_i+1} \hat{y}_t$. For any $t = 1,\ldots,j-1$, let $z_{2t-1} = \hat{z}_t + \sum_{k=1}^{t} \hat{y}_{k+1}$ and $z_{2t} = \hat{z}_t$, and for any $t = j,\ldots,n_i-1$, let $z_{2t-1} = \hat{z}_t$ and $z_{2t} = \hat{z}_t + \sum_{k=t+1}^{n_i} \hat{y}_{k+1}$. It is straightforward to show that $y > 0$, $z \geq 0$, and (27) through (30) hold. This completes the proof of Lemma 14. □

Fix $i \in N$ and a consistent assessment (σ, m). (f, σ_i) is renegotiation-proof at (σ_{-i}, m) if and only if for any $I \in \mathscr{I}_i$, $j = 1,\ldots,n_i$, $\sigma_i' : \Theta_i \to \Sigma_i|_I$, there exist $\hat{y} \in \mathbb{R}^{n+1}$ and $\hat{z} \in \mathbb{R}^{(n-1)}$ such that $\hat{y} > 0$, $\hat{z} \geq 0$, and equation (24) holds. We can now complete the proof of Theorem 2 (of the main paper).

[Only if] Fix a consistent assessment (σ, m), $i \in N$, and (f_i, σ_i). Suppose, for contradiction, that there exist $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, and increasing $\sigma_i' : \Theta_i \to \Sigma_i|_I$ such that $U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j | I) > U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j | I)$, but there is no $k = 1, 2, \ldots, j-1$ such that (4) holds and no $l = j+1, \ldots, n_i$ such that (5) holds. This implies that $\alpha_t > 0$ for all $t = 1, \ldots, n_i + 1$. Since $U_i^{\sigma,m}(\sigma_i', \theta_i | I)$ has increasing differences, $\beta_t \geq 0$ for all $t = 1, \ldots, n_i - 1$. Therefore, $\hat{y} > 0$ and $\hat{z} \geq 0$ imply that $\sum_{t=1}^{n_i+1} \alpha_t \hat{y}_t + \sum_{t=1}^{n_i-1} \beta_t \hat{z}_t > 0$, which, by Lemma 14, contradicts that (f_i, σ_i) is renegotiation-proof.

[If] Fix a consistent assessment (σ, m), $i \in N$, and (f_i, σ_i). Let $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, and $\sigma_i' : \Theta_i \to \Sigma_i|_I$ be increasing and $U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j | I) > U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j | I)$. Suppose first that there exists a $k \in \{1, \ldots, j-1\}$ such that (4) holds. This implies that $\alpha_{j+1} > 0$ and $\alpha_{k+1} \leq 0$. Let $\hat{y}_{k+1} = 1$, $\hat{y}_{j+1} = \frac{-\alpha_{k+1}}{\alpha_{j+1}} \geq 0$, and all the other $\hat{y}_t = 0$ and $\hat{z}_t = 0$. This implies that equation (24) holds and, by Lemma 13 and 14, that (f_i, σ_i) is renegotiation-proof. Suppose now that there exists an $l \in \{j+1, \ldots, n\}$ such that (5) holds. Then, $\alpha_{j+1} > 0$ and $\alpha_{l+1} \leq 0$. Let $\hat{y}_{l+1} = 1$, $\hat{y}_{j+1} = \frac{-\alpha_{l+1}}{\alpha_{j+1}} \geq 0$ and all the other $\hat{y}_t = 0$ and $\hat{z}_t = 0$. This, again, implies that (24) holds and that (f_i, σ_i) is renegotiation-proof. □

Definition 1. Fix a consistent assessment (σ, m) and $i \in N$. For any $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, and $\sigma_i' \in \widehat{\Sigma}_i(I, j, \sigma_i)$ we say that $b_j \in \{1, 2, \ldots, n_i\}$ is a *blocking type* if there exists a $k \in \{1, 2, \ldots, j-1\}$ such that

$$U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j | I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j | I) \leq \sum_{t=b_j}^{j-1} [\vec{U}_i^{\sigma,m}(\sigma_i | I)_{2t-1} - \vec{U}_i^{\sigma,m}(\sigma_i' | I)_{2t-1}] \tag{31}$$

or there exists an $l \in \{j+1, j+2, \ldots, n_i\}$ such that

$$U_i^{\sigma,m}(\sigma_i'(\theta_i^j), \theta_i^j | I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j), \theta_i^j | I) \leq \sum_{t=b_j}^{j-1} [\vec{U}_i^{\sigma,m}(\sigma_i | I)_{2t-1} - \vec{U}_i^{\sigma,m}(\sigma_i' | I)_{2(t-1)}] \tag{32}$$

We obtain the following necessary conditions for a strategy σ_i to be renegotiation-proof.

Proposition 1. *Fix a consistent assessment (σ, m) and $i \in N$. A mixed strategy σ_i is renegotiation-proof only if for any $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, and $\sigma_i' \in \widehat{\Sigma}_i(I, j, \sigma_i)$ there is a blocking type.*

The above condition becomes also sufficient for renegotiation-proofness with an additional requirement about the relation of blocking types for different renegotiation opportunities.

Proposition 2. *Fix a consistent assessment (σ, m) and $i \in N$. An increasing mixed strategy σ_i is renegotiation-proof if for any $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, and $\sigma_i' \in \widehat{\Sigma}_i(I, j, \sigma_i)$ there is a blocking type b_j such that $k < l$, $b_k > k$, and $b_l < l$ imply $b_k \leq b_l$.*

Proof of Proposition 1. Fix a consistent assessment (σ, m) and $i \in N$ and suppose that the mixed strategy σ_i is renegotiation-proof. Fix $I \in \mathscr{I}_i$, $j = 1, \ldots, n_i$, and an increasing $\sigma_i' : \Theta_i \to$

$\Sigma_i|_I$ such that $U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) > U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I)$. For any $t = 1,\ldots,n_i$, let $c_t = e_j - e_t$, where e_t is the t^{th} standard basis row vector for $\mathbb{R}^{n_i}$, and define

$$E_t = \begin{pmatrix} D \\ c_t \end{pmatrix}$$

Also let

$$w_k = U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) + \sum_{t=k}^{j-1} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2t-1}$$

$$w_l = U_i^{\sigma,m}(\sigma_i'(\theta_i^j),\theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) + \sum_{t=j+1}^{l} \vec{U}_i^{\sigma,m}(\sigma_i'|I)_{2(t-1)}$$

for any $k \in \{1,\ldots,j-1\}$ and $l \in \{j+1,\ldots,n_i\}$ and define

$$V_j = \begin{pmatrix} \vec{U}_i^{\sigma,m}(\sigma_i|I) \\ -w_j \end{pmatrix}$$

Incentive compatibility of (f_i,σ_i) implies that $D\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq \vec{U}_i^{\sigma,m}(\sigma_i|I)$. Renegotiation proofness, by Theorem 2 (of the main paper), implies that $c_k\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq -w_k$ for some $k \in \{1,\ldots,j-1\}$ or $c_l\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq -w_l$ for some $l \in \{j+1,\ldots,n\}$. Suppose first that there exists a $k \in \{1,\ldots,j-1\}$ such that $c_k\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq -w_k$. Then we must have $E_k\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq V_k$. By Gale's theorem of linear inequalities, this implies that $x \geq 0$ and $E_k'x = 0$ implies $x'V_k \geq 0$. Denote the first $2(n_i-1)$ elements of x by y and the last element by z. It is easy to show that $E_k'x = 0$ implies that $y_{2t-1} = y_{2t} + z$ for $t \in \{k,k+1,\ldots,j-1\}$ and $y_{2t-1} = y_{2j}$ for $t \notin \{k,k+1,\ldots,j-1\}$. Therefore,

$$\begin{aligned} x'V_k &= \sum_{t=1}^{n_i-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t}y_{2t} + \sum_{t=1}^{n_i-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1}y_{2t-1} - zw_k \\ &= \sum_{t=1}^{n_i-1} (\vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t} + \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1})y_{2t} + z(-w_k + \sum_{t=k}^{j-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1}) \\ &\geq 0 \end{aligned}$$

This implies that $-w_k + \sum_{t=k}^{j-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1}) \geq 0$ and hence k is a blocking type.

Similarly, we can show that, if there exists an $l \in \{i+1,\ldots,n_i\}$ such that $c_l\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq -w_l$, then l is a blocking type, and this completes the proof. □

Before we proceed to the proof of Proposition 2 we need the following lemma:

Lemma 15. *Fix a consistent assessment (σ,m) and $i \in N$. $(f_i,\sigma_i) \in \mathscr{C} \times \Sigma_i$ is* renegotiation-proof *at (σ_{-i},m) if for all $I \in \mathscr{I}_i, \theta_i \in \Theta_i$ such that there exists $(g_i,s_i') \in \mathscr{C} \times S_i|_I$ that is incentive compatible at (σ_{-i},m) and*

$$U_i(s_i'(I,\theta_i),\sigma_{-i},m|I,\theta_i) - G_i(s_i'(I,\theta_i),\sigma_{-i},m|I,\theta_i) > U_i(\sigma_i(I,\theta_i),\sigma_{-i},m|I,\theta_i) - F_i(\sigma_i(I,\theta_i),\sigma_{-i},m|I,\theta_i)$$

there exists a $\theta_i' \in \Theta_i$ *such that*

$$F_i(\sigma_i(I,\theta_i'),\sigma_{-i},m|I,\theta_i') \geq G_i(s_i'(I,\theta_i'),\sigma_{-i},m|I,\theta_i')$$

Proof of Lemma 15. By definition (f_i,σ_i) is not renegotiation-proof at (σ_{-i},m) if and only if there exist $I \in \mathscr{I}_i$, $j=1,\ldots,n_i$, $g_i \in \mathscr{C}$, and $\sigma_i' : \Theta_i \to \Delta(S_i|_I)$ such that (21)-(23) hold. This is true iff there exists $s_i' : \Theta_i \to S_i|_I$ such that

$$U_i^{\sigma,m}(s_i'(\theta_i^l),\theta_i^l|I) - G_i^{\sigma,m}(s_i'(\theta_i^l)|I) \geq U_i^{\sigma,m}(s_i'(\theta_i^k),\theta_i^l|I) - G_i^{\sigma,m}(s_i'(\theta_i^k)|I), \forall\, l,k=1,\ldots,n_i \quad (33)$$

$$U_i^{\sigma,m}(s_i'(\theta_i^j),\theta_i^j|I) - G_i^{\sigma,m}(s_i'(\theta_i^j)|I) > U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) - F_i^{\sigma,m}(\sigma_i(\theta_i^j)|I) \quad (34)$$

$$G_i^{\sigma,m}(s_i'(\theta_i^k)|I) > F_i^{\sigma,m}(\sigma_i(\theta_i^k)|I), \forall\, k=1,\ldots,n_i. \quad (35)$$

This is because if there exists an s_i', that satisfies (33)-(35) then there trivially exists σ_i' that satisfies (21)-(23). Conversely, let there be a σ_i' that satisfies (21)-(23). This implies that there is an $s_i' \in \text{supp}(\sigma_i')$ that satisfies (35) for all $k=1,\ldots,n_i$. Then this s_i' also satisfies (33) and (34) because $\sigma_i'(\theta_i^l) \in \text{argmax}_{\sigma_i''} U_i^{\sigma,m}(\sigma_i'',\theta_i^l|I) - G_i^{\sigma,m}(\sigma_i''|I)$ iff $s_i' \in \text{argmax}_{s_i''} U_i^{\sigma,m}(s_i'',\theta_i^l|I) - G_i^{\sigma,m}(s_i''|I)$ for any $s_i' \in \text{supp}(\sigma_i')$. □

Proof of Proposition 2. Fix a consistent assessment (σ,m) and $i \in N$ and suppose that the mixed strategy σ_i is increasing and satisfies the conditions of the proposition. We will show that there exist an $f_i \in \mathscr{C}$ such that (f_i,σ_i) is incentive-compatible and renegotiation-proof at (σ_{-i},m). Fix $I \in \mathscr{I}_i$ and for each $j=1,\ldots,n_i$ define

$$\widehat{S}_i(I,j,\sigma_i) = \{s_i' : \Theta_i \to S_i|_I : s_i' \text{ is increasing and } U_i^{\sigma,m}(s_i'(\theta_i^j),\theta_i^j|I) > U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I)\}. \quad (36)$$

For each $j=1,\ldots,n_i$ and $s_i^j \in \widehat{S}_i(I,j,\sigma_i)$ pick a blocking type $b(s_i^j)$ such that $k<l$, $b(s_i^k)>k$, and $b(s_i^l)<l$ imply $b(s_i^k) \leq b(s_i^l)$. By Lemma 15, if (f_i,σ_i) is renegotiation-proof against any $s_i' \in \widehat{S}_i(I,j,\sigma_i)$ it is also renegotiation-proof against any $\sigma_i' \in \widehat{\Sigma}_i(I,j,\sigma_i)$.

For each $j=1,\ldots,n_i$ and $s_i^j \in \widehat{S}_i(I,j,\sigma_i)$ define the n_i-dimensional row vector $c_{s_i^j} = e_j - e_{b(s_i^j)}$, where e_t is the tth standard basis row vector for $\mathbb{R}^{n_i}$, and the scalar $w_{s_i^j}$ as

$$w_{s_i^j} = U_i^{\sigma,m}(s_i^j(\theta_i^j),\theta_i^j|I) - U_i^{\sigma,m}(\sigma_i(\theta_i^j),\theta_i^j|I) + \mathbf{1}_{\{b(s_i^j)\leq j-1\}} \sum_{t=b(s_i^j)}^{j-1} \vec{U}_i^{\sigma,m}(s_i^j|I)_{2t-1}$$
$$+ \mathbf{1}_{\{j\leq b(s_i^j)-1\}} \sum_{t=j+1}^{b(s_i^j)} \vec{U}_i^{\sigma,m}(s_i^j|I)_{2(t-1)}.$$

Note that $\widehat{S}_i(I,j,\sigma_i)$ is finite and let $\sum_{j=1}^{n_i} |\widehat{S}_i(I,j,\sigma_i)| = q$. Denote with $C(I)$, the $q \times n$ matrix

composed of all the rows $c_{s_i^j}$ and with $W(I)$ the q dimensional vector with component $w_{s_i^j}$ corresponding to each s_i^j. Let $E(I)$ be the matrix

$$E(I) = \begin{pmatrix} D \\ C(I) \end{pmatrix}$$

and $V(I)$ the column vector

$$V(I) = \begin{pmatrix} \vec{U}_i^{\sigma,m}(\sigma_i|I) \\ -W(I) \end{pmatrix}$$

Now, if for each $I \in \mathscr{I}_i$, we can find an $\vec{F}_i^{\sigma,m}(\sigma_i|I)$ such that $E(I)\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq V(I)$ the proof would be completed. Indeed, if $E(I)\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq V(I)$, then $D\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq \vec{U}_i^{\sigma,m}(\sigma_i|I)$, which implies that there exists f_i such that (f_i, σ_i) is incentive compatible (see Lemma 11). Furthermore, $E(I)\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq V(I)$ implies $W(I) \leq -C(I)\vec{F}_i^{\sigma,m}(\sigma_i|I)$ and, by Theorem 2, that (f_i, σ_i) is renegotiation-proof.

Gale's theorem of linear inequalities implies that there exist $\vec{F}_i^{\sigma,m}(\sigma_i|I)$ such that $E(I)\vec{F}_i^{\sigma,m}(\sigma_i|I) \leq V(I)$ if and only if $x \in \mathbb{R}^{q+2(n_i-1)}$, $x \geq 0$ and $E(I)'x = 0$ implies $x'V(I) \geq 0$. Decompose x into two vectors so that the first $2(n_i-1)$ elements constitute y and the remaining q components constitute z. Notice that for any $j = 1, \ldots, n_i$ and $s_i^j \in \widehat{S}_i(I, j, \sigma_i)$ there is a corresponding element of z, which we will denote $z_{s_i^j}$.

Recursively adding row 1 to row 2, row 2 to row 3, and so on, we can reduce $E(I)'$ to a row echelon form and show that $E(I)'x = 0$ if and only if

$$y_{2t-1} = y_{2t} + \sum_{s_i^j} z_{s_i^j} [\mathbf{1}_{\{b(s_i^j) \leq t \leq j-1\}} - \mathbf{1}_{\{j \leq t \leq b(s_i^j)-1\}}] \tag{37}$$

for $t = 1, \ldots, n_i - 1$.

Let $T_- = \{t \in \{1, \ldots, n_i - 1\} : \exists s_i^j$ such that $j \leq t \leq b(s_i^j) - 1\}$ and $T_+ = \{t \in \{1, \ldots, n_i - 1\} : \exists s_i^j$ such that $b(s_i^j) \leq t \leq j - 1\}$ and note that $T_- \cap T_+ = \emptyset$. To see this, suppose, for contradiction, that there exists a $t \in T_- \cap T_+$. Therefore, there exists a s_i^j such that $j \leq t \leq b(s_i^j) - 1$ and $s_i^{j'}$ such that $b(s_i^{j'}) \leq t \leq j' - 1$. This implies that $j < j'$, $b(s_i^j) > j$, $b(s_i^{j'}) < j'$, but $b(s_i^j) > b(s_i^{j'})$, contradicting the conditions of the proposition. We can therefore write (37) as

$$y_{2t} = y_{2t-1} + \sum_{s_i^j} z_{s_i^j} \mathbf{1}_{\{j \leq t \leq b(s_i^j)-1\}} \tag{38}$$

for $t \in T_-$ and

$$y_{2t-1} = y_{2t} + \sum_{s_i^j} z_{s_i^j} \mathbf{1}_{\{b(s_i^j) \leq t \leq j-1\}} \tag{39}$$

for $t \in T_+$.

Finally note that

$$x'V(I) = \sum_{t=1}^{n_i-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t}\, y_{2t} + \sum_{t=1}^{n_i-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1}\, y_{2t-1} - \sum_{s_i^j} z_{s_i^j} w_{s_i^j}$$

Substituting from (38) and (39) we obtain

$$x'V(I) = \sum_{t\in T_-} \left[\vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t} + \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1}\right] y_{2t-1} + \sum_{t\in T_+} \left[\vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t} + \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1}\right] y_{2t}$$
$$+ \sum_{s_i^j} z_{s_i^j} \left[-w_{s_i^j} + \mathbf{1}_{\{b(s_i^j)\le j-1\}} \sum_{t=b(s_i^j)}^{j-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t-1} + \mathbf{1}_{\{j\le b(s_i^j)-1\}} \sum_{t=i}^{b(s_i^j)-1} \vec{U}_i^{\sigma,m}(\sigma_i|I)_{2t} \right]$$

Increasing differences, the definition of $b(s_i^j)$, and $y, z \ge 0$ imply that $x'V \ge 0$, and the proof is completed. □

4 Proofs for section 4.2

Proof of Theorem 3 (of the main paper). By definition $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1\times\Theta}$ is not renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1,2,\ldots,n$ and an incentive compatible $(g, b_2) \in \mathscr{C} \times A_2^{A_1\times\Theta}$ such that $u_2(a_1, b_2(a_1,\theta^i),\theta^i) - g(a_1, b_2(a_1,\theta^i)) > u_2(a_1, b_2^*(a_1,\theta^i),\theta^i) - f(a_1, b_2^*(a_1,\theta^i))$ and $u_3(a_1, b_2(a_1,\theta^j),\theta^j) + g(a_1, b_2(a_1,\theta^j)) > u_3(a_1, b_2^*(a_1,\theta^j),\theta^j) + f(a_1, b_2^*(a_1,\theta^j))$ for all $j = 1,2,\ldots,n$. Therefore, we have the following

Lemma 16. *$(f, b_2^*) \in \mathscr{C} \times A_2^{A_1\times\Theta}$ is not renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1,2,\ldots,n$, $b_2 \in A_2^{A_1\times\Theta}$, and $\varepsilon \in \mathbb{R}^n$ such that $D(f(a_1, b_2^*) + \varepsilon) \le \vec{U}_2(a_1, b_2)$, $\varepsilon_i < u_2(a_1, b_2(a_1,\theta^i),\theta^i) - u_2(a_1, b_2^*(a_1^i,\theta^i),\theta^i)$, and $u_3(a_1, b_2(a_1,\theta^j),\theta^j) - u_3(a_1, b_2^*(a_1,\theta^j),\theta^j) + \varepsilon_j > 0$ for all $j = 1,\ldots,n$.*

Define the matrices V and C as in the proof of Theorem 1 (of the main paper), and define the matrix A as follows: its row 1 is e_1, row $n+2$ is l_i, and row $j+1$, for $j = 1,\ldots,n$, is given by $[u_3(a_1, b_2(a_1,\theta^j),\theta^j) - u_3(a_1, b_2^*(a_1,\theta^j),\theta^j)]e_1 + e_{j+1}$. We have the following lemma, whose proof is similar to that of Lemma 3.

Lemma 17. *$(f, b_2^*) \in \mathscr{C} \times A_2^{A_1\times\Theta}$ is strongly renegotiation-proof if and only if for any $a_1 \in A_1$, $i = 1,2,\ldots,n$ and $b_2 \in A_2^{A_1\times\Theta}$ there exist $y \in \mathbb{R}^{n+2}$ and $z \in \mathbb{R}^{2(n-1)}$ such that $A'y + C'z = 0$, $y > 0$, $z \ge 0$.*

The rest of the proof is almost identical to that of Theorem 1 (of the main paper), and therefore is omitted.

□

5 Proofs for section 4.3

Proof of Theorem 4 (of the main paper). By definition $(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is not strongly renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1, 2, \ldots, n$ and an incentive compatible $(g, b_2) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ such that $u_2(a_1, b_2(a_1, \theta^i), \theta^i) - g(a_1, b_2(a_1, \theta^i)) > u_2(a_1, b_2^*(a_1, \theta^i), \theta^i) - f(a_1, b_2^*(a_1, \theta^i))$, $g(a_1, b_2(a_1, \theta^i)) > f(a_1, b_2^*(a_1, \theta^i))$, and $g(a_1, b_2(a_1, \theta^i)) - f(a_1, b_2^*(a_1, \theta^i)) > \min\{0, u_2(a_1, b_2(a_1, \theta^j), \theta^j) - u_2(a_1, b_2^*(a_1, \theta^j), \theta^j)\}$ for all $j = 1, 2, \ldots, n$. The following lemma easily follows.

Lemma 18. *$(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is not strongly renegotiation-proof if and only if there exist $a_1 \in A_1$, $i = 1, 2, \ldots, n$, $b_2 \in A_2^{A_1 \times \Theta}$, and $\varepsilon \in \mathbb{R}^n$ such that $D(f(a_1, b_2^*) + \varepsilon) \leq \vec{U}_2(a_1, b_2)$, $0 < \varepsilon_i < u_2(a_1, b_2(a_1, \theta^i), \theta^i) - u_2(a_1, b_2^*(a_1^i, \theta^i), \theta^i)$, and $\varepsilon_j > \min\{0, u_2(a_1, b_2(a_1, \theta^j), \theta^j) - u_2(a_1, b_2^*(a_1, \theta^j), \theta^j)\}$ for all $j = 1, 2, \ldots, n$.*

Define the matrices V and C as in the proof of Theorem 1 (of the main paper), and define the matrix A as follows: its row 1 is e_1, row $n+2$ is l_i, and row $j+1$, for $j = 1, \ldots, n$, is given by $-\min\{0, u_2(a_1, b_2(a_1, \theta^j), \theta^j) - u_2(a_1, b_2^*(a_1, \theta^j), \theta^j)\} e_1 + e_{j+1}$. We have the following lemma, whose proof is similar to that of Lemma 3.

Lemma 19. *$(f, b_2^*) \in \mathscr{C} \times A_2^{A_1 \times \Theta}$ is strongly renegotiation-proof if and only if for any $a_1 \in A_1$, $i = 1, 2, \ldots, n$ and $b_2 \in A_2^{A_1 \times \Theta}$ there exist $y \in \mathbb{R}^{n+2}$ and $z \in \mathbb{R}^{2(n-1)}$ such that $A'y + C'z = 0$, $y > 0$, $z \geq 0$.*

The rest of the proof is almost identical to that of Theorem 1 (of the main paper), and therefore is omitted. □

6 Proofs for Section 5.1.2

Proof Proposition 9 (of the main paper). Let

$$b_1^*(b_2) \in \underset{b_1 \in BR_1(b_2)}{\operatorname{argmin}} \; U_2(b_1, b_2)$$

$$b_2^* \in \underset{b_2 \in B_2^R}{\operatorname{argmax}} U_2(b_1^*(b_2), b_2)$$

and $a_1^* = b_1^*(b_2^*)$. Note that $U_2(a_1^*, b_1^*) = \bar{U}_2^{WR}$ and suppose, for contradiction, that player 2 gets a payoff $\tilde{U}_2 < \bar{U}_2^{WR} - \delta$. We will show that player 2 can offer a contract that supports (a_1^*, b_2^*) and yields a higher payoff.

For any a_1 choose $\hat{b}_{2,a_1} \in \operatorname{argmin}_{b_2 \in B_2^R} U_1(a_1, b_2)$. By construction $\hat{b}_{2,a_1}$ is increasing and hence there exists a contract that makes it optimal to play. As in the proof of Proposition 7, we can show that there exists a contract f^{a_1} such that $\hat{b}_{2,a_1}$ is the unique optimal strategy and $(f^{a_1}, \hat{b}_{2,a_1})$ is renegotiation-proof after a_1. Let $\varepsilon > 0$ be small and define $f(b_1^*, a_2) = \delta + \varepsilon$

for all a_2. For any $a_1 \neq b_1^*$ define

$$f(a_1,a_2) = \begin{cases} f_i^{a_1}, & a_2 = \hat{b}_{2,a_1}(a_1,\theta^i) \\ \infty, & \text{otherwise} \end{cases}$$

Under this contract, player 2 plays a best response to a_1^* and according to $\hat{b}_{2,a_1}$ after any $a_1 \neq a_1^*$. Player 1, on the other hand, must play a best response to b_2^*. This is because for any $a_1 \notin BR_1(b_1^*)$, we have $U_1(br_1(b_2^*),b_2^*) > U_1(a_1,b_2^*) \geq U_1(a_1,\hat{b}_{2,a_1})$. Therefore, deviation to such a contract yields a payoff of $U_2(b_1^*,b_2^*)-\delta-\varepsilon > \tilde{U}_2$, for small enough ε. In other words, player 2 has a profitable deviation, contradicting that $\tilde{U}_2$ is an equilibrium payoff. □

7 Proofs for section 6.2

Proof of Proposition 13 (of the main paper). In the original game G, player 2 chooses π after observing π^e and the realization of θ. For each π^e and θ, player 2 chooses π to maximize $-(U^*+\theta-\alpha(\pi-\pi^e))^2-\beta\pi^2$. The first order conditions yield the optimal strategy for player 2 as

$$\pi(\pi^e,\theta) = \frac{\alpha^2}{\alpha^2+\beta}\pi^e + \frac{\alpha}{\alpha^2+\beta}(U^*+\theta). \tag{40}$$

Given the optimal strategy of player 2, player 1 chooses π^e to maximize $-\sum_{\theta\in\Theta}(\pi^e-\pi(\pi^e,\theta))^2 p(\theta)$. The first order condition is

$$-2(1-\frac{\alpha^2}{\alpha^2+\beta})\sum_{\theta\in\Theta}(\pi^e-\pi(\pi^e,\theta))p(\theta) = 0$$

which is solved as

$$\pi^e = \sum_{\theta\in\Theta}\pi(\pi^e,\theta)p(\theta)$$

Substituting it in to (40), we obtain

$$\pi^e = \frac{\alpha}{\beta}U^*$$

and

$$\pi(\pi^e,\theta) = \frac{\alpha}{\beta}U^* + \frac{\alpha}{\alpha^2+\beta}\theta$$

completing the proof. □

8 Proofs for section 8

Proof of Lemma 2 (of the main paper). Fix $a_1 \in A_1$, $i \in \{1,\cdots,n\}$, and $b_2^i \in \mathfrak{B}(a_1,i,b_2^*)$. Since A_2 is linearly ordered, we have $b_2^i(a_1,\theta_i) \succsim_2 b_2^*(a_1,\theta^i)$ or $b_2^*(a_1,\theta_i) \succsim_2 b_2^i(a_1,\theta^i)$. First, assume that $b_2^i(a_1,\theta_i) \succsim_2 b_2^*(a_1,\theta^i)$, i.e., (a_1,i) has right deviation at b_2^*, and note that $R(a_1,i) \neq \emptyset$ by assumption. Let $J = \{j \in \mathbb{N} : i+1 \leq j \leq \min R(a_1,i) - 1 \text{ and } b_2^*(a_1,\theta^j) \succ_2 b_2^i(a_1,\theta^j)\}$. If $J = \emptyset$, let $m(b_2^i) = \min R(a_1,i)$ and if $J \neq \emptyset$, let $m(b_2^i) = \min J$. It is simple to show that

$$\begin{aligned}\sum_{j=i+1}^{m(b_2^i)} \Big(& u_2(a_1, b_2^i(a_1,\theta^{j-1}),\theta^j) - u_2(a_1, b_2^*(a_1,\theta^{j-1}),\theta^j) \\ & -[u_2(a_1, b_2^i(a_1,\theta^{j-1}),\theta^{j-1}) - u_2(a_1, b_2^*(a_1,\theta^{j-1}),\theta^{j-1})]\Big) \\ & + u_2(a_1, b_2^*(a_1,\theta^{m(b_2^i)}),\theta^{m(b_2^i)}) - u_2(a_1, b_2^i(a_1,\theta^{m(b_2^i)}),\theta^{m(b_2^i)}) \geq 0 \quad (41)\end{aligned}$$

Inequality (41) implies that $m(b_2^i)$ is a blocking type.

Now assume that $b_2^*(a_1,\theta^i) \succsim_2 b_2^i(a_1,\theta^i)$, i.e., (a_1,i) has left deviation at b_2^*, and note that $L(a_1,i) \neq \emptyset$. Let $J = \{j \in \mathbb{N} : \max L(i) + 1 \leq j \leq i-1 \text{ and } b_2^i(a_1,\theta^j) \succ_2 b_2^*(a_1,\theta^j)\}$. If $J = \emptyset$, let $m(b_2^i) = \max L(i)$ and if $J \neq \emptyset$, let $m(b_2^i) = \max J$ and note that

$$\begin{aligned}\sum_{j=m(b_2^i)}^{i-1} \Big(& u_2(a_1, b_2^*(a_1,\theta^{j+1}),\theta^{j+1}) - u_2(a_1, b_2^i(a_1,\theta^{j+1}),\theta^{j+1}) \\ & - [u_2(a_1, b_2^*(a_1,\theta^{j+1}),\theta^j) - u_2(a_1, b_2^i(a_1,\theta^{j+1}),\theta^j)]\Big) \\ & + u_2(a_1, b_2^*(a_1,\theta^{m(b_2^i)}),\theta^{m(b_2^i)}) - u_2(a_1, b_2^i(a_1,\theta^{m(b_2^i)}),\theta^{m(b_2^i)}) \geq 0 \quad (42)\end{aligned}$$

Inequality (42) implies that $m(b_2^i)$ is a blocking type.

Finally assume that there exist (a_1,i_1) and (a_1,i_2) with $i_1 < i_2$ such that $m(b_2^{i_1}) > i_1$ and $m(b_2^{i_2}) < i_2$. This implies that (a_1,i_1) has right deviation and (a_1,i_2) has left deviation at b_2^*, which imply that $R(a_1,i_1) \neq \emptyset$, $L(a_1,i_2) \neq \emptyset$ and $R(a_1,i_1) \cap L(a_1,i_2) \neq \emptyset$. But this implies that $m(b_2^{i_1}) \leq m(b_2^{i_2})$ and the proof is completed by applying Proposition 4. □

Proof of Lemma 3 (of the main paper). Suppose, for contradiction, that there exists an $a_1' \in A_1$ such that (a_1',θ^n) has right deviation at b_2, i.e., there exists an $a_2' \in A_2$ such that $a_2' \succsim_2 b_2(a_1',\theta^n)$ and $u_2(a_1',a_2',\theta^n) > u_2(a_1',b_2(a_1',\theta^n),\theta^n)$. Define

$$b_2'(a_1',\theta) = \begin{cases} a_2', & \theta = \theta^n \\ b_2(a_1',\theta), & \theta \prec_\theta \theta^n \end{cases}$$

Note that b_2' is increasing and therefore $b_2' \in \mathfrak{B}(a_1',n,b_2)$. It is easy to show that for (a_1',n,b_2') there is no blocking type and therefore, by Proposition 3 (of the main paper), b_2 is not renegotiation proof. □