WORKING PAPERS

N° 1522

March 2024

"Mediated Renegotiation"

Andrea Attar, Lorenzo Bozzoli, Roland Strausz

Toulouse School of Economics

Mediated Renegotiation

Andrea Attar, Lorenzo Bozzoli, Roland Strausz*

March 27, 2024

Abstract

We develop a new approach to contract renegotiation under informational frictions. Specifically, we consider mediated mechanisms which *cannot* be contingent on any subsequent offer, but *can* generate a new source of asymmetric information between the contracting parties. Taking as a reference the canonical framework of Fudenberg and Tirole (1990), we show that, if mediated mechanisms are allowed, the corresponding renegotiation game admits only one equilibrium allocation, which coincides with the second-best one. Thus, the inefficiencies typically associated to the threat of renegotiation may be completely offset by the design of more sophisticated trading mechanisms. (*JEL* D43, D82, D86)

1 Introduction

We identify a class of mediated mechanisms that mitigate all inefficiencies due to renegotiation in the canonical moral hazard problem as studied in Fudenberg and Tirole (1990). In so doing, we show that *any* equilibrium of the corresponding mediated renegotiation game implements the second-best allocation, which obtains with moral hazard but *without* any possibility of renegotiation (as in Holmström (1979) or Shavell (1979)).

With this result we contribute to a growing literature that studies the extent to which the design of trading mechanisms alleviates the inefficiencies arising under limited commitment (Bester and Strausz (2007), Doval and Skreta (2022), Georgiadis-Harris et al. (2023), and Lomys and Yamashita (2022)). While these contributions investigate this issue mainly in the context of the Coase conjecture, our study is the first to analyze the role of more sophisticated mechanisms in problems of contract renegotiation.

*Attar: CNRS, Toulouse School of Economics University of Toulouse Capitole, and Università degli Studi di Roma "Tor Vergata" (email: andrea.attar@tse-fr.eu); Bozzoli: Università degli Studi di Roma "Tor Vergata" (email: lorenzo.bozzoli@uniroma2.it); Strausz: School of Business and Economics, Humboldt-Universität zu Berlin (email: roland.strausz@hu-berlin.de). We thank Eloisa Campioni, Dino Gerardi, Soenje Reiche and Francois Salanié, for very thoughtful comments. We also thank seminar audiences at Toulouse School of Economics and Università degli Studi di Roma "Tor Vergata" for many useful discussions. This research has benefited from financial support from the Agence Nationale de la Recherche (ANR) (project ANR-23-CE26-0006-01).

Our mechanisms are mediated in the sense of Rahman and Obara (2010) in that the final allocation which the mechanism selects, depends on signals generated by the mechanism itself.[1] The required sophistication of mediation is however very minor; the optimal mechanism only communicates to the agent the outcome of a 50-50 toss coin. Specifically, the agent receives this signal privately before she has to accept or reject the principal's renegotiation offer. Hence, our mediated mechanism extends on the non-mediated ones in Fudenberg and Tirole (1990) only by the private communication of a toss coin. Although the extension seems insignificant, we show that it enables the principal to fully negate the threat of renegotiation.

To understand how such a minor extension leads to such an economically significant effect, it is helpful to point out that while Fudenberg and Tirole (1990) also allow the agent to send a binary message, its use differs from ours. In Fudenberg and Tirole (1990), the agent uses the message to indicate whether she chose a high or low effort. By contrast, our mechanism uses the binary message to delegate to the agent the punishment of the principal's attempts to renegotiate. By sending a message in the original mechanism after observing a renegotiated offer, the agent is able to correlate her (privately known) outside options to the offers she receives at the renegotiation stage. This provides a flexible device to deter renegotiation. In particular, the renegotiation offer which Fudenberg and Tirole (1990) identify as undermining the second best is not profitable anymore: observing this offer, the agent's optimal report in the original mechanism yields her a better outside option with positive probability, in which case she is lead to reject the new offer, thereby hindering renegotiation.

Because one can frame the renegotiation problem also as one of a contract designer competing with its myopic future self over incentivizing a common agent, our results naturally connect to recent literature on competing mechanisms that points out strong strategic effects of agents who are privately and asymmetrically informed about some characteristics of the posted mechanisms (Attar et al. (2023)). More generally, our results hinge on the effect that with multiple contracting parties endogenously generated asymmetric information can alleviate contractual frictions. This effect has also been studied in contracting environments with the threat of contractual collusion (Ortner and Chassang (2018), Asseyer (2020) and von Negenborn and Pollrich (2020)).

[1] See Strausz (2012) for how mediated mechanisms provide an alternative representation for direct mechanisms in the light of the revelation principle derived in Myerson (1982).

2 The Model

We consider the canonical framework of Fudenberg and Tirole (1990) (FT, henceforth), in which a risk-neutral principal (P, he) incentivizes a risk-averse agent (A, she), who takes an unobservable effort. There are two outputs (states), a good one g and a bad one b, where $g > b > 0$, and the probability distribution over outputs depends on the binary effort $e \in E \doteq \{L, H\}$. Let $p_e \doteq prob(g|e)$, and $p_H > p_L$ so that $\Delta p \doteq p_H - p_L > 0$. The effort e yields the expected output $Y_e \doteq p_e g + (1 - p_e)b$.

Payoffs and Allocations. A's utility is additively separable in income $w \in \mathbb{R}$ and effort $e \in E$, so that we can express it as $U(w) - D(e)$. The utility function U exhibits $U'(w) > 0$ and $U''(w) < 0$ for each $w \in \mathbb{R}$, and U is unbounded over its domain $\mathbb{R}$, i.e., $\lim_{w \to -\infty} U(w) = -\infty$ and $\lim_{w \to \infty} U(w) = \infty$. These assumptions imply that the inverse Φ of U is well–defined for any $u \in \mathbb{R}$ with $\Phi'(u) > 0$, and $\Phi''(u) > 0$ for each $u \in \mathbb{R}$. We normalize A's low effort cost to $D(e = L) = 0$ and denote A's high effort cost by $D(e = H) = d > 0$.

For any $e \in E$, final payoffs are determined by the state-contingent transfers that P makes to A. We let a contract be a pair $(w_g, w_b) \in \mathbb{R}^2$ of such transfers. Because it is often more convenient to represent a contract in terms of the induced utilities it provides to A, we also write (with slight abuse of notation) a contract as $c = (U_g, U_b)$, with $U_g = U(w_g)$ and $U_b = U(w_b)$.

A (deterministic) allocation is a pair $(e, c) \in E \times \mathbb{R}^2$ of payoff-relevant decisions.

The agent's expected payoff from (e, c) is

$$U_e(c) = p_e U_g + (1 - p_e)U_b - D(e),$$

where U^0 is her reservation payoff.[2]

The principal's expected payoff from (e, c) is

$$V_e(c) = Y_e - p_e\Phi(U_g) - (1 - p_e)\Phi(U_b).$$

Efficient and Incentive Compatible Allocations. Because A is risk-averse, while P is risk-neutral, any Pareto-efficient allocation exhibits full insurance. For any $e \in E$, let $c_e^{FI}(U) = (U + D(e), U + D(e))$ denote the full-insurance contract that yields A the expected payoff $U \in \mathbb{R}$. We also define, for each $e \in E$, the function $V_e^{FI} : \mathbb{R} \to \mathbb{R}$ where

$$V_e^{FI}(U) \doteq V_e(c_e^{FI}(U)) = Y_e - \Phi(U + D(e))$$

[2]In FT, it holds $U^0 = 0$. Writing the outside option as U^0 is more insightful for interpreting results.

identifies the P's payoff associated with the full-insurance contract leaving the expected payoff U to A. Since $\Phi' > 0$, V_e^{FI} is strictly decreasing in U for any $e \in E$.

Because P has all the bargaining power, the optimal contract with observable effort implements the efficient allocation that yields P his maximal payoff while still guaranteeing A her outside option U^0. We refer to this outcome as the first-best. Thus the first-best contract is $c^{FB} \doteq c_H^{FI}(U^0)$, which yields $V^{FB} \doteq V_H^{FI}(U^0)$ to P, and $U = U^0$ to A.[3]

If, instead, effort is unobservable, any feasible allocation must be incentive-compatible. In this case, P's optimal contract induces $e = H$ and gives at least U^0 to A.[4] We refer to this contract as the second-best and it obtains from solving:

$$\arg\max_{c\in\mathbb{R}^2} \quad V_H(c) = p_H(g - \Phi(U_g)) + (1 - p_H)(b - \Phi(U_b)) \tag{1}$$

$$\text{s.t.} \quad p_H U_g + (1 - p_H)U_b - d \geq p_L U_g + (1 - p_L)U_b. \tag{2}$$

$$p_H U_g + (1 - p_H)U_b - d \geq U^0. \tag{3}$$

In the solution, the incentive constraint (2) binds. Accordingly, let $c^{IC}(U) = (U_g^{IC}(U), U_b^{IC}(U))$ denote the contract on the incentive-compatibility frontier that leaves an expected payoff $U \in \mathbb{R}$ to A. That is:

$$U_g^{IC}(U) \doteq U + \frac{1 - p_L}{\Delta p}d \quad \text{and} \quad U_b^{IC}(U) \doteq U - \frac{p_L}{\Delta p}d.$$

It is convenient to define, for each $e \in E$, the function $V_e^{IC} : \mathbb{R} \to \mathbb{R}$ denoting P's payoff when A takes $e \in E$ and $c^{IC}(U)$ is implemented. That is:

$$V_e^{IC}(U) \doteq V_e(c^{IC}(U)) = Y_e - p_e\Phi\left(U + \frac{1 - p_L}{\Delta p}d\right) - (1 - p_e)\Phi\left(U - \frac{p_L}{\Delta p}d\right).$$

Since V_H^{IC} is decreasing in U, the participation constraint (3) binds at the solution, which implies that the second-best contract is $c^{SB} \doteq c^{IC}(U^0))$. It yields $V^{SB} \doteq V_H^{IC}(U^0)$ to P, and $U_H(c^{IC}(U^0)) = U^0$ to A.

The Standard Contracting Game. The second-best allocation (H, c^{SB}) is the only one to be supported in a subgame-perfect equilibrium of the standard contracting game G:

1. P offers a contract $c \in \mathbb{R}^2$.

2. A accepts or rejects c. If A rejects, the game ends and outside options accrue. If A accepts, the game continues as follows:

[3] As we follow FT in focusing on the non-trivial case that $e = H$ is optimal in the second best, we have that $e = H$ is also optimal in the first-best.

[4] As in FT, we assume that $e = H$ is optimal in the second best.

3. A chooses $e = H$ with probability $x \in [0, 1]$ and $e = L$ with probability $1 - x$.

4. The output realizes and transfers flow in accordance with the contract c.

Remark 1 *While the contract c^{SB} makes A indifferent between $e = L$ and $e = H$, the game G has a unique equilibrium allocation, which coincides with (H, c^{SB}). This is so even though in the subgame that starts after A accepts c^{SB}, any mixed strategy $x \in (0, 1)$ is (part of) a Nash equilibrium. Yet, for any $x < 1$, P can offer a perturbed contract $c_\varepsilon^{SB} = (U_g^{IC}(U^0) + \varepsilon, U_b^{IC}(U^0))$, which induces a subgame where, for $\varepsilon > 0$ but small, accepting the contract and picking $e = H$ is uniquely optimal for A and yields P strictly more than what he obtains under c^{SB}. Thus, the equilibrium allocation is unique because there is a sequence of contracts, each yielding a subgame in which it is uniquely optimal for A to accept and pick $e = H$, and this sequence converges to c^{SB}.*

The Renegotiation Game. FT point out that the second-best allocation (H, c^{SB}) is *interim* inefficient, i.e., after the effort is chosen but before the output is realized. This leads them to analyze an extension G^r of the game G, in which P can renegotiate away any such inefficiency by offering a new contract at the *interim* stage, after stage 3 but before stage 4.

Because, at this stage, A is privately informed about her effort choice, a proper analysis of the renegotiation game G^r requires the introduction of more complex mechanisms than the simple take-it or leave-it pairs (U_g, U_b) of the game G. For this reason, they let P design a revelation mechanism, which specifies a contract for each effort announced by A. Specifically, a (deterministic) revelation mechanism is a mapping $\gamma_c : E \to \mathbb{R}^2$, and we let C be the set of all such mechanisms. The extensive form of G^r is as follows:

1. P offers a mechanism $\gamma_c \in C$.

2. A accepts or rejects γ_c. If A rejects, the game ends and outside options accrue. If A accepts, the game unfolds as follows:

3. A chooses $e = H$ with probability $x \in [0, 1]$ and $e = L$ with probability $1 - x$.

4. Without observing e, P makes a renegotiation offer $\gamma_c^r \in C \cup \{\emptyset\}$, where $\{\emptyset\}$ represents P's action not to renegotiate.

5. A accepts or rejects γ_c^r by declaring $\rho \in \{y, n\}$. She then sends a message $m \in E$ in the mechanism she participates in.

6. If $\rho = n$, transfers occur according to $\gamma_c(m)$. If $\rho = y$, transfers occur according to $\gamma_c^r(m)$.

Any mechanism γ_c that A accepts in stage 2 yields a subgame $G^r(\gamma_c)$ that starts in stage 3. FT note that, in any such subgame, choosing $x = 1$ is not part of an equilibrium. To see this, suppose A takes $e = H$ with probability one. Then, the best reply of P is to offer a full insurance mechanism in stage 4 that is accepted by A. But against this renegotiation offer, A would be strictly better off choosing $e = L$. FT then show that the overall game G^r admits only one (perfect Bayesian) equilibrium allocation. At equilibrium, renegotiation is successfully prevented. Yet, A takes $e = H$ with probability $x^{FT} < 1$, and obtains an expected payoff $U^{FT} \geq U^0$.

Thus, the threat of renegotiation fundamentally constraints the provision of insurance, and the agent's incentives to undertake the efficient level of effort.

In their analysis, FT restrict attention to revelation mechanisms,[5] which, by construction, do not incorporate any *private* communication to the agent. We argue that this restriction is critical. In the remaining of the paper, we show the key role of mechanisms in which the wage may also depend on a random signal s about which only the agent gets informed. Because any such mechanism conditions the transfers on the signal s that it sends to the agent, we refer to it as a "mediated mechanism" in line with Rahman and Obara (2010) connotation of a "mediated contract".

Specifically, we show that, if P can use mediated mechanisms, then the unique equilibrium allocation is exactly (H, c^{SB}). This result demonstrates that the frictions usually associated to contract renegotiation can be fully overcome by more sophisticated contract design.

3 Mediated Mechanisms

A *mediated* mechanism $\gamma = \{\mathcal{M}, \mathcal{S}, \sigma, \tau\}$ features a finite set of messages $\mathcal{M}$ sent from A, a finite set of signals $\mathcal{S}$ to be privately received by A according to the distribution $\sigma : \mathcal{M} \to \Delta(\mathcal{S})$, and a decision rule $\tau : \mathcal{M} \times \mathcal{S} \to \mathbb{R}^2$ that associates a profile of transfers $\tau(m, s)$ to each profile of messages and signals.[6] We let Γ denote the set of mediated mechanisms.

In the *mediated renegotiation game* G_Γ, P posts mechanisms in Γ, and, at the renegotiation stage, he may offer any other mechanism in Γ to A as an alternative.

[5] They write: "there is no loss of generality in restricting the contract space to be C", since "the revelation principle implies that, at the interim stage, the principal can implement any allocation obtained through a complex contract" (Fudenberg and Tirole, 1990, p. 1283)

[6] To circumvent potential measure theoretical complications, we focus on finite sets.

For $|\mathcal{M}| = 2$ and $|\mathcal{S}| = 1$, mechanisms in Γ reduce to mechanisms in C as defined in FT.[7] Hence, the mechanisms that we consider extend the ones in FT in the message dimension $\mathcal{M}$ as well as in the signal dimension $\mathcal{S}$.

In the following subsection, we consider two classes of mediated mechanisms, $\Gamma_0 \subset \Gamma$ and $\Gamma_1 \subset \Gamma$, that differ according to the *extensive form* of the game they induce, that is, the sequence in which the reports and the disclosures are made. As the rest of this section will clarify, different implementation results can be achieved by allowing the principal to propose mediated mechanisms from these different classes.

3.1 The Mediated Renegotiation Game

The mediated renegotiation game G_Γ takes place as follows:

1. P offers a mediated mechanism $\gamma \in \Gamma$.
2. A accepts or rejects the offer. If she rejects the outside options accrue, otherwise the subgame $G_\Gamma(\gamma)$ takes place.

The details of $G_\Gamma(\gamma)$ depend on the timing of the reports and the disclosures featured by γ. In this subsection, the extensive form (sub)games induced by the acceptance of mechanisms belonging to the classes $\Gamma_0 \subset \Gamma$ and $\Gamma_1 \subset \Gamma$ are described. For a given subgame $G_\Gamma(\gamma)$, we let $\lambda(\gamma)$ be a (continuation) strategy for the agent in $G_\Gamma(\gamma)$, while a principal's pure strategy is a renegotiated mechanism $\gamma^r \in \Gamma$.

Consider first the class of mechanisms Γ_0, in which the agent observes a signal at the ex ante stage, before taking the effort decision, and sends a report after rejecting a renegotiated offer.

The extensive form of the subgame $G_\Gamma(\gamma)$ that starts after A accepts a mechanism $\gamma \in \Gamma_0$, is as follows:

1. A observes a private signal $s \in \mathcal{S}$ extracted from the distribution $\sigma \in \Delta(\mathcal{S})$.[8]
2. Having observed $s \in \mathcal{S}$, A chooses $e = H$ with probability $x \in [0, 1]$ and $e = L$ with probability $1 - x$.
3. Without observing neither e nor s, P makes a renegotiation offer $\gamma^r = \{\mathcal{M}^r, \mathcal{S}^r, \sigma^r, \tau^r\} \in \Gamma \cup \{\emptyset\}$.

[7] In particular, concerning the extensive form of the game induced by a mechanism, FT let the agent report into γ only after γ^r has been rejected; however, as they argue in Fudenberg and Tirole (1990, p. 1283), their results still hold when other extensive forms and timing of the reports are considered.

[8] Since s is extracted before m is sent, $\sigma(s|m) = \sigma(s)$ for each $(s, m) \in \mathcal{S} \times \mathcal{M}$.

4. A accepts or rejects γ^r by declaring $\rho \in \{y, n\}$.

5. If $\rho = n$ or $\gamma^r = \{\emptyset\}$, the agent sends $m \in \mathcal{M}$. If $\rho = y$, A receives a private signal $s^r \in \mathcal{S}^r$ and sends $m^r \in \mathcal{M}^r$.[9]

6. All the messages and signal of the relevant mechanism are made public and the final transfers are executed accordingly.

A principal's (pure) strategy in the subgame $G_\Gamma(\gamma)$ is a renegotiation offer $\gamma^r \in \Gamma \cup \{\emptyset\}$. An agent's (behavioral) strategy in $G_\Gamma(\gamma)$, which we denote $\lambda(\gamma)$, associates to any $\gamma^r \in \Gamma \cup \{\emptyset\}$ and $s \in \mathcal{S}$ a probability distribution over E. Further, for any history (e, s, γ^r) with $\gamma^r \neq \{\emptyset\}$, λ specifies a probability distribution over the participation choices in γ, and, for any $(e, s, \{\emptyset\})$ or (e, s, γ^r, n), a probability distribution over the messages $m \in \mathcal{M}$. Finally, $\lambda(\gamma)$ specifies a probability distribution over the messages in γ^r, at any continuation history of (e, s, γ^r, y) in which, given the extensive form of γ^r, it is required.

A mediated mechanism $\gamma \in \Gamma_0$ can indeed be thought as randomly drawing and secretly offering menus of contracts to the agent before her effort decision. In fact, each $s \in \mathcal{S}$ can be interpreted as an index to a menu of lotteries whose items are $\{\tau(s, m)\}_{m \in \mathcal{M}}$. After accepting γ, the agent learns which s has been offered to her, and by sending $m \in \mathcal{M}$ she selects an item $\tau(s, m)$. The principal does not know at the renegotiation stage which menu $s \in \mathcal{S}$ has been offered to the agent.

The extensive form of $G_\Gamma(\gamma)$ is instead as follows if $\gamma \in \Gamma_1$:

1. A chooses $e = H$ with probability $x \in [0, 1]$ and $e = L$ with probability $1 - x$.

2. Without observing e, P makes a renegotiation offer $\gamma^r = \{\mathcal{M}^r, \mathcal{S}^r, \sigma^r, \tau^r\} \in \Gamma \cup \{\emptyset\}$.

3. A send a message $m \in \mathcal{M}$ in the initial mechanism γ.

4. The mechanism γ extracts a signal $s \in \mathcal{S}$ according to the distribution $\sigma(m) \in \Delta(\mathcal{S})$.

5. After privately observing s, A accepts or rejects γ^r by declaring $\rho \in \{y, n\}$.

6. If $\rho = y$, A sends $m^r \in \mathcal{M}^r$ and receives a private signal $s^r \in \mathcal{S}^r$.[10]

7. All the communications in the relevant mechanism are made public and the associated transfers are executed.

[9]The order of the actions depends on which mechanism is offered by the renegotiating principal.
[10]As in footnote 9, the order depends on the extensive form induced by γ^r after $\rho = y$.

A principal's (pure) strategy in $G_\Gamma(\gamma)$, where $\gamma \in \Gamma_1$, is a renegotiation offer $\gamma^r \in \Gamma \cup \{\emptyset\}$. An agent's (behavioral) strategy in G_Γ, which we denote $\lambda(\gamma)$, associates to any $\gamma \in \Gamma$ a probability distribution over efforts. Further, for any history (e, γ^r), λ specifies a probability distribution over the messages in γ, and, for any (e, γ^r, m, s) such that $\gamma^r \neq \{\emptyset\}$, a probability distribution over the participation choices in γ^r. Finally, λ specifies a probability distribution over the messages in γ^r, at any continuation history in which, given the extensive form of γ^r, it is required.

Comparing the subgame $G_\Gamma(\gamma)$ with $\gamma \in \Gamma_1$ to the subgame $G^r(\gamma_c)$ in FT reveals that a mediated mechanism transforms the simultaneous decision of A about her message m and acceptance decision ρ into a sequential one, comprising three substages: first, A sends a message m, she then privately observes the random signal s, and finally accepts or rejects the renegotiation offer. The crucial idea is that A receives her (private) signal s *before* deciding about accepting P's renegotiation offer.

As in FT, the game G_Γ is an extensive form game with imperfect information, as P does not observe A's effort choice, when making his renegotiation offer. Consequently, we focus on the perfect Bayesian equilibria of G_Γ.

As the subgame $G_\Gamma(\gamma)$ is also an extensive form game with imperfect information, a perfect Bayesian equilibrium of G_Γ also prescribes a perfect Bayesian equilibrium for the subgame $G_\Gamma(\gamma)$. That is, P chooses an optimal mechanism γ, given that the players continuation strategies constitute a perfect Bayesian equilibrium (henceforth equilibrium) of $G_\Gamma(\gamma)$.

In the next sections, we show that mediated mechanisms allow to fully overcome the threat of renegotiation. That is, we show that *in any* equilibrium of G_Γ, P obtains the second-best payoff V^{SB}, A picks $e = H$ with probability one, and there is no incentive for renegotiation.

More in detail, we first show that, by offering a mediated mechanism $\hat{\gamma} \in \Gamma_0$, the principal can obtain any payoff arbitrarily close to the second-best level V^{SB} at *some* equilibrium of the subgame $G_\Gamma(\hat{\gamma})$. Note that a mechanism in the class Γ only changes the extensive form assumed in FT slightly, that is, by including a private disclosure to the agent that takes place at the ex ante stage. The timing of the agent's participation and reporting decisions is instead unaltered with respect to G^r.[11]

[11]Another difference is the ability of the principal to offer stochastic mechanisms; however, the inclusion of lotteries over transfers is shown in the Discussion section of this paper to be irrelevant in the original FT framework, and thus, it does not constitute *per se* an extension of the FT model, until the private disclosures are added to the picture.

Then, we exhibit a mechanism $\gamma^* \in \Gamma_1$ which implements the second-best allocation exactly, and we also show that this allocation is the only one to be supported in an equilibrium of the overall game G_Γ since, for any payoff V lower but arbitrarily close to V^{SB}, there exists $\gamma' \in \Gamma_1$ that yields V as the only continuation payoff of the principal.

Thus, the introduction of mediation allows to reach different conclusions according to the level of sophistication that is allowed in the interplay between the agent's reporting and participation decisions. In particular, our analysis suggest that an approximate and partial implementation of the second-best can be achieved by adding a minimal piece of endogenous asymmetric information to the standard framework in FT; however, for the exact and unique implementation of the second-best, the principal must be able to design mechanisms requiring the agent to report *after* observing a renegotiated offer, and *before* accepting or rejecting it.

3.2 Virtual Implementation of the Second Best

We prove first two preliminary results, which will be instrumental to characterize the principal's payoff at the renegotiation stage.

Lemma 1 *There exists a $\bar{\pi} \in (0,1)$ such that, for each $\pi \in (0,\bar{\pi})$, one can find a $U^j(\pi) \in (U^0,\infty)$ that satisfies:*

$$(1-\pi)V^{SB} + \pi V^{IC}(U^j(\pi)) = Y^H - \Phi(U^j(\pi)+d). \tag{4}$$

There could be multiple values of $U \in (U^0,\infty)$ that verify (4). We henceforth refer to $U^j(\pi)$ as the value characterized in the proof, which is the smallest $U \in (U^0,\infty)$ yielding the result.[12]

We now establish a second result:

Lemma 2 *For each pair $\left(\pi \in (0,\bar{\pi}), U^j(\pi)\right)$, there is a unique $U^k(\pi) \in (U^j(\pi),\infty)$ such that:*

$$(1-\pi)(Y^H-\Phi(U^0+d))+\pi\left[\frac{1}{2}V^{IC}(U^k(\pi))+\frac{1}{2}V^{IC}(2U^j(\pi)-U^k(\pi))\right] = (1-\pi)V^{SB}+\pi V^{IC}(U^j(\pi)). \tag{5}$$

[12]Consider in fact H, U^* and U^1 as defined in the appendix. For all $\pi \in (0,\bar{\pi})$, $U^j(\pi)$ is shown to be the unique $U \in [U^*,U^1]$ satisfying (16) and thus (9). Since $H(U)<0$ for any $U \in [U^0,U^*)$, and since $U^1 > U^*$, there is no other value $U \in [U^0,U^j(\pi))$ satisfying (16).

We are now ready to show that mediated communication allows to (approximately) implement the second-best allocation yielding V^{SB} to the principal.

Let $\hat{\gamma} = \{\hat{\mathcal{S}}, \hat{\mathcal{M}}, \hat{\sigma}, \hat{\tau}\} \in \Gamma_0$ be such that:

1. $\hat{\mathcal{S}} = \{\alpha, \beta\}$, $\hat{\mathcal{M}} = \{a, b\}$

2. $\hat{\sigma}(\alpha) = 1 - \pi$, $\hat{\sigma}(\beta) = \pi$, with $\pi \in (0, \bar{\pi})$, as in Lemma 1

3. If α realizes, then $\hat{\tau}(\alpha, m) = c^{IC}(U^0) = c^{SB}$ with probability one, for every $m \in \{a, b\}$

4. If β realizes, then the final transfers depend on the agent's message:

   - if $m = a$, then $\hat{\tau}(\beta, a) = c^{IC}(U^j(\pi))$ with probability one,
   - if $m = b$, then

$$\hat{\tau}(\beta, b) = \begin{cases} c^{IC}(U^k(\pi)) & \text{with probability } \frac{1}{2} \\ c^{IC}(U^l(\pi)) & \text{with probability } \frac{1}{2}, \end{cases}$$

with $U^j(\pi)$ and $U^k(\pi)$ defined in Lemma 1 and 2, and $U^l(\pi) = 2U^j(\pi) - U^k(\pi)$.

The mechanism can be interpreted as follows: with probability $1-\pi$, only the second-best contract c^{SB} is offered. With probability π, the agent is offered a menu, which leaves her free to choose between two items. The first one is an incentive-compatible contract yielding the (expected) utility $U^j(\pi) > U^0$ to her, the second item is a randomization over the two incentive-compatible contracts $c^{IC}(U^k(\pi))$ and $c^{IC}(U^l(\pi))$. One should observe that, upon selecting this second item, i.e. sending the message $m = b$, the agent's expected payoff is

$$\frac{1}{2}U^k(\pi) + \frac{1}{2}U^l(\pi) = U^j(\pi), \tag{6}$$

for any $e \in \{L, H\}$. That is, upon receiving the signal β, the agent is indifferent over reporting $m \in \{a, b\}$ in the mechanism, for any effort choice.

The subgame $G_\Gamma(\hat{\gamma})$ unfolds as follows:

1. A observes the signal $s \in \{\alpha, \beta\}$.

2. After seeing s, A takes $e = H$ with probability $x(s) \in [0, 1]$

3. P offers a renegotiation mechanism $\gamma^r \in \Gamma \cup \{\emptyset\}$.

4. A takes a (possibly random) participation decision $\rho \in \{y, n\}$ if $\gamma^r \neq \{\emptyset\}$.

5. If $\rho = y$, A observes s^r and sends a possibly random report $m^r \in \mathcal{M}^r$. If $\rho = n$ or $\gamma^r = \{\emptyset\}$, A sends a possibly random report $m \in \{a, b\}$.

6. All communications are made public and the transfer from the relevant mechanism are realized.

The following result holds:

Proposition 1 *Fix any $\pi \in (0, \bar{\pi})$. The subgame $G_\Gamma(\hat{\gamma})$ has an equilibrium in which:*

i. A plays $e = H$ with probability one

ii. P does not renegotiate with probability one

iii. P gets the payoff V^{SB} with probability $(1 - \pi)$.

We still need to show that the P's equilibrium payoff can be made arbitrarily close to the second-best one. This is implied by the following

Corollary 1 *Let $\pi \in (0, \bar{\pi})$. Then,*

$$\lim_{\pi \to 0} \pi V^{IC}(U^j(\pi)) + (1 - \pi)V^{SB} = V^{SB}. \tag{7}$$

3.3 Exact Implementation of the Second Best

We now show that there exists $\gamma^* \in \Gamma_1$ that allows to obtain the second-best allocation at an equilibrium of $G_\Gamma(\gamma^*)$.

Given the definition of $V_e^{FI} : \mathbb{R} \to \mathbb{R}$ and $V_e^{IC} : \mathbb{R} \to \mathbb{R}$ from Section 2, the following lemma is key for establishing our main result.

Lemma 3 *There exists $U^n \in (U^0, \infty)$ such that for all $U \geq U^n$ and all $e \in E$:*

$$V_e^{IC}(U^0) > \max\left\{V_e^{FI}(U), \, \frac{1}{2}V_e^{FI}(2U^0 - U) + \frac{1}{2}V_e^{IC}(U)\right\}. \tag{8}$$

Lemma 3 implies a threshold U^n such that, for any $e \in E$, P prefers the incentive-compatible contract, $c^{IC}(U^0)$, to any full-insurance contract that leaves $U \geq U^n$ to A. Moreover, for any $U \geq U^n$, P also prefers the contract $c^{IC}(U^0)$ to a 50/50 lottery between the full-insurance contract leaving $2U^0 - U$ to A, and the incentive-compatible one leaving her U.

As we shall discuss, Lemma 8 allows to construct a mediated mechanism γ^* such that P attains the left hand side of (8) when he does not renegotiate, while the right hand side of

(8) corresponds to what P expects from the best possible renegotiation offer that A accepts with a strict positive probability. Hence, the inequality in (8) implies that P is better off by not renegotiating.

Let $\gamma^* = \{\mathcal{M}^*, \mathcal{S}^*, \sigma^*, \tau^*\}$ be such that $\mathcal{M}^* = \{a, b\}$ and $\mathcal{S}^* = \{\alpha, \beta\}$. The signals are extracted according to $\sigma^* : \mathcal{M}^* \to \Delta(\mathcal{S}^*)$ with:

$$\sigma^*(\alpha|m) = \sigma^*(\beta|m) = \frac{1}{2} \qquad \text{for each } m \in \{a, b\}.$$

The decision rule $\tau^* : \mathcal{M}^* \times \mathcal{S}^* \to \mathbb{R}^2$ is such that:

$$\tau^*(a, \alpha) = \tau^*(a, \beta) = c^{SB}; \quad \tau^*(b, \alpha) = c^{IC}\left(2U^0 - U^n\right); \quad \tau^*(b, \beta) = c^{IC}(U^n).$$

Hence, γ^* shares with the mechanisms in FT the restriction to only two messages for A, i.e. $|\mathcal{M}^*| = 2$. By contrast, it selects one of the two signals with equal probability and privately discloses it to A. Although, in general, the distribution of the signal may depend on the message m, the specific mediated mechanism γ^* does not exploit this feature. Yet, γ^* is a mediated mechanism in the sense of Rahman and Obara (2010) because it conditions its final transfers on some information *privately* sent to A.

Proposition 2 *The second-best allocation* (H, c^{SB}) *is supported in an equilibrium of the subgame* $G_\Gamma(\gamma^*)$*.*

Since P cannot obtain more in the game with renegotiation than without renegotiation, Proposition 2 implies that the mediated renegotiation game has *an* equilibrium in which the possibility of renegotiation does not constrain final outcomes. The result stands in stark contrast to that in FT, who do not consider mediated mechanisms.

To establish Proposition 2, observe that, by reporting $m = a$ in γ^*, A gets the second-best contract c^{SB}, which makes $e = H$ an optimal choice. In the absence of renegotiation, this yields U^0 to A and V^{SB} to P. Hence, it suffices to exhibit a profile of continuation strategies that support these behaviors in an equilibrium of $G_\Gamma(\gamma^*)$.

Let m^r_e denote some A's optimal message $m \in \mathcal{M}^r$ when she accepted a renegotiation offer $\gamma^r \neq \{\emptyset\}$ and she chose an effort $e \in E$.[13] In addition, let $\hat{U}^r_e$ denote her corresponding payoff. That is,

$$m^r_e \in \arg\max_{m \in \mathcal{M}^r} \sum_{s \in \mathcal{S}^r} \sigma^r(s|m) U_e(\tau^r(m, s)) \quad \text{and} \quad \hat{U}^r_e = \sum_{s \in \mathcal{S}^r} \sigma^r(s|m^r_e) U_e(\tau^r(m^r_e, s)). \qquad (9)$$

[13] We only consider renegotiated mechanism that let the agent report before receiving a message; the reasoning can be straightforwardly extended to the case in which s is sent before m, since, also in this case, any non-degenerate distribution of payoff-relevant signals violates the *full insurance* requirement.

By construction, sending m_e^r is sequentially rational for A following any history (e, γ^r, m, s, y).

We now describe the strategies $\{\lambda(\gamma^*), \gamma^r(\gamma^*)\}$ supporting (H, c^{SB}) in an equilibrium of $G_\Gamma(\gamma^*)$: P's strategy is not to renegotiate, i.e. $\gamma^r(\gamma^*) = \{\emptyset\}$, while A's strategy $\lambda(\gamma^*)$ is as follows:

1. A chooses $e = H$ with probability one.
2. Her messages in γ^*, and her subsequent participation decisions in γ^r, depend on the history (e, γ^r) as follows:
   - (i) If $\gamma^r = \{\emptyset\}$, and for any γ^r such that $\hat{U}_e^r \leq 2U^0 - U^n$, A sends $m = a$ in γ^*, followed by $\rho = n$.
   - (ii) For any $\gamma^r \neq \{\emptyset\}$ such that $\hat{U}_e^r \in (2U^0 - U^n, U^n]$, A sends $m = b$ in γ^*, followed by $\rho = y$ when $s = \alpha$, and by $\rho = n$ when $s = \beta$.
   - (iii) For any $\gamma^r \neq \{\emptyset\}$ such that $\hat{U}_e^r > U^n$, A sends $m = b$ in γ^*, followed by $\rho = y$ for any received signal.
3. For any history $(e, \gamma^r \neq \{\emptyset\}, m, s, y)$, A sends m_e^r.

We show that the strategy profile $\{\lambda(\gamma^*), \gamma^r(\gamma^*)\}$ together with P's belief that A picked $e = H$ with probability $x = 1$ constitutes an equilibrium of $G_\Gamma(\gamma^*)$.

First note that the only non-trivial information set in $G_\Gamma(\gamma^*)$ is at the renegotiation stage, where P offers γ^r. The only belief that is consistent with the strategy profile $\{\lambda(\gamma^*), \gamma^r(\gamma^*)\}$ is, indeed, $x = 1$, as $\lambda(\gamma^*)$ prescribes A to pick $e = H$. Observe that if the strategies $\{\lambda(\gamma^*), \gamma^r(\gamma^*)\}$ are played, then P obtains V^{SB} and A obtains U^0.

We develop our argument in two lemmas. The first one refers to the A's behavior in $G_\Gamma(\gamma^*)$.

Lemma 4 *The agent's strategy $\lambda(\gamma^*)$ is sequentially rational given any history (e, γ^r).*

Proof. We already noted that sending m_e^r is sequentially rational for any history $(e, \gamma^r \neq \{\emptyset\}, m, s, y)$. Next, consider any history $(e, \gamma^r \neq \{\emptyset\})$. It is optimal for A to send $m = a$ in γ^* whenever

$$\max\{U^0, \hat{U}_e^r\} \geq \frac{1}{2} \max\left\{2U^0 - U^n, \hat{U}_e^r\right\} + \frac{1}{2} \max\{U^n, \hat{U}_e^r\}, \tag{10}$$

where $\hat{U}_e^r$ is defined in (9). The left(right)-hand side of (10) is her continuation payoff after sending $m = a(b)$. The following holds:

(i) If $\hat{U}_e^r \leq 2U^0 - U^n \vee \gamma^r = \{\emptyset\}$, then (10) is satisfied because it reduces to $U^0 \geq U^0$ since $\hat{U}_e^r \leq 2U^0 - U^n < U^0$, where the latter inequality follows from $U^n > U^0$. Sending $m = a$ in γ^*, followed by $\rho = n$, as prescribed by $\lambda(\gamma^*)$, is hence optimal.

(ii) If $\hat{U}_e^r \in (2U^0 - U^n, U^n]$, then, upon sending $m = b$, it is optimal for A to choose $\rho = y$ when $s = \alpha$ (as rejection leads to $2U^0 - U^n < \hat{U}_e^r$), and $\rho = n$ when $s = \beta$ (as rejection leads to $U^n \geq \hat{U}_e^r$). We next argue that sending $m = b$ in γ^*, as prescribed by $\lambda(\gamma^*)$, is optimal. That is, the sign of the inequality in (10) is reversed, where we note that, due to $\hat{U}_e^r \in (2U^0 - U^n, U^n]$, its RHS reduces to $\hat{U}_e^r/2 + U^n/2$. Hence, we only need to show that

$$\max\{U^0, \hat{U}_e^r\} \leq \frac{1}{2}\hat{U}_e^r + \frac{1}{2}U^n. \tag{11}$$

To get the result, it is sufficient to observe that:

(a) If $\hat{U}_e^r < U^0$, then (11) is satisfied since $\hat{U}_e^r > 2U^0 - U^n$.

(b) If $\hat{U}_e^r \geq U^0$, then (11) is satisfied since $\hat{U}_e^r \leq U^n$.

(iii) If $\hat{U}_e^r \in (U^n, \infty)$, then we have $U^0 < U^n < \hat{U}_e^r$, implying that A is indifferent between $m = a$ and $m = b$, followed by $\rho = y$ for any $s \in \{\alpha, \beta\}$. In particular, as prescribed by $\lambda(\gamma^*)$, sending $m = b$ in γ^*, and then accepting to participate in γ^r for any received signal is optimal. ■

The next lemma considers P's behavior in $G_\Gamma(\gamma^*)$. In this subgame, P makes a renegotiation offer γ^r, given his belief that $x = 1$, and anticipating A's continuation strategy derived from $\lambda(\gamma^*)$.

Lemma 5 *The strategy $\gamma^r(\gamma^*) = \{\emptyset\}$ is a principal's best response to his (Bayes-consistent) belief $x = 1$, and to the agent's strategy $\lambda(\gamma^*)$.*

Proof. We first note that P can improve on any renegotiation offer $\gamma^r \neq \{\emptyset\}$ that does not achieve full insurance to A. Hence, any optimal renegotiation offer involves full insurance, i.e., γ^r is such that $\tau^r(m,s) = (U_H^r(m,s), U_H^r(m,s))$ for any $(m,s) \in \mathcal{M}^r \times \mathcal{S}^r$ with the interpretation that it yields the payoff $U_H^r(m,s)$ to A when she picks $e = H$. Since any renegotiation that does condition transfers non-trivially on A's private signal, implies that A is not fully insured, we need to consider only $\gamma^r \in \Gamma$ such that $\mathcal{S}^r = \{s^r\}$ and $\tau^r(m, s^r) = (U_H^r(m, s^r), U_H^r(m, s^r))$ for any $m \in \mathcal{M}^r$. But then there is also no loss in considering offers such that $\mathcal{M}^r$ is a singleton, as P correctly anticipates that for any $|\mathcal{M}^r| > 1$, A sends some $m^r \in \arg\max_{m \in \mathcal{M}^r} U_H^r(m, s^r)$, implying that P can do just as well by letting $\mathcal{M}^r = \{m^r\}$. Thus, there is no loss in assuming $\mathcal{S}^r = \{s^r\}$, $\mathcal{M}^r = \{m^r\}$, and $\tau^r(m_1, s_1) = (U_H^r(m_1, s_1), U_H^r(m_1, s_1))$, which implies that any γ^r can be characterized by the number $\hat{U}^r = U_H^r(m^r, s^r) \in \mathbb{R}$, expressing the payoff that the renegotiation offer leaves to A when she picks $e = H$.

We next verify that, for any $\hat{U}^r \in \mathbb{R}$, P's expected payoff does not exceed $V^{SB} = V_H^{IC}(U^0)$, his utility when not renegotiating. We distinguish three cases:

(i) If $\hat{U}^r \leq 2U^0 - U^n$ then $\lambda(\gamma^*)$ prescribes $(m = a, \rho = n)$ and P gets $V_H^{IC}(U^0)$.

(ii) If $\hat{U}^r \in (2U^0 - U^n, U^n]$ then $\lambda(\gamma^*)$ prescribes ($m = b$, $\rho = y$ when $s = \alpha$, and $\rho = n$ when $s = \beta$), and P gets

$$\frac{1}{2}V_H^{FI}(\hat{U}^r) + \frac{1}{2}V_H^{IC}(U^n) < \frac{1}{2}V_H^{FI}\left(2U^0 - U^n\right) + \frac{1}{2}V_H^{IC}(U^n) < V_H^{IC}(U^0), \tag{12}$$

where the first inequality follows from the fact that V_H^{FB} is decreasing, and the second one from Lemma 3.

(iii) If $\hat{U}^r > U^n$ then $\lambda(\gamma^*)$ prescribes ($m = b$, $\rho = y$) for any received signal, and P gets

$$V_H^{FI}(\hat{U}^r) < V_H^{FI}(U^n) < V_H^{IC}(U^0) \tag{13}$$

where, again, the first inequality follows from the fact that each V_e^{FB} is decreasing, and the second one from Lemma 3.

Thus, P cannot gain by offering any $\gamma^r \neq \{\emptyset\}$. ■

We complete the proof of Proposition 2 by considering A's effort choice. Given P's strategy $\gamma^r(\gamma^*) = \{\emptyset\}$, this is straightforward, as A does not expect γ^* to be renegotiated. In particular, she expects U^0 from either effort level, because $U_L(c^{SB}) = U_H(c^{SB}) = U^0$, so that choosing $e = H$ is indeed optimal.

Thus, as claimed in Proposition 2, the profile $\{\lambda(\gamma^*), \gamma^r(\gamma^*)\}$ together with P's belief $x = 1$ at his non-trivial information set, form a perfect Bayesian equilibrium of $G_\Gamma(\gamma^*)$, in which A chooses $e = H$ with probability one, and P obtains $V^{SB} = V_H^{IC}(U^0)$.

The presented verification of Proposition 2 proves that the mediated mechanism γ^* makes any renegotiation offer unprofitable to P. It is, in particular, useful to describe A's behavior towards any γ^r involving a full-insurance contract leaving her a utility U^0. Recall that it is this offer that upsets the second best for non-mediated contracts, implying that such contracts are unable to yield an equilibrium in which A picks $e = H$ with probability 1. We instead show that γ^* induces A to reject any such contract upon receiving the signal $s = \beta$, which occurs with probability 1/2. The rejection probability of 1/2 makes these offers unprofitable to P, because whenever A rejects the offer, the relevant contracting terms are highly unfavorable to him.

Indeed, for γ^* to implement the second best, the P's anticipation of a probabilistic rejection of his offer is crucial. This also explains why the mediated mechanism must send the signal s privately to A rather than announcing it publicly. Because γ^* guarantees that, at the renegotiation stage, P is unaware of the signal realization $s \in \mathcal{S}$, it is also crucial that P is unable to condition the renegotiation offer on s. But if the mechanism would reveal s publicly, P would then be able to condition the renegotiation offer on the signal, thereby undoing the probabilistic rejection.

3.4 Unique Implementation of the Second Best

Proposition 2 shows that the mediated mechanism γ^* induces a subgame that supports the second-best allocation at equilibrium. Because this outcome yields A the utility U^0, it is also incentive compatible for her to accept γ^* at stage 2, as she cannot strictly gain by rejecting it. Moreover, P cannot attain a payoff greater than V^{SB}, in the game without renegotiation. This then shows that *an* equilibrium exists in the overall game G_Γ that yields the second-best allocation.

From the Myersonian mechanism design perspective that the principal can pick not only the mechanism but also the equilibrium to be played, the presence of *an* equilibrium yielding the second-best allocation provides a satisfactory answer to its implementability. However, taking a stricter implementation perspective, one may worry that G_Γ may also admit *other* equilibrium outcomes.

Indeed, the mechanism γ^* makes A indifferent over her messages as well as over her effort choices. As a consequence, γ^* can be shown to implement a continuum of allocations for the subgame $G_\Gamma(\gamma^*)$. One may, for example, show that any $x \in [0, 1]$ together with P not renegotiating (i.e. $\gamma^r = \{\emptyset\}$) can be supported by an equilibrium of $G_\Gamma(\gamma^*)$, in which A sends $m = a$ on the equilibrium path, and $m = b$ if any renegotiation occurs.

That is, γ^* does not *uniquely* implement the second-best allocation. Yet, in spite of this multiplicity, this allocation is the only one to be supported at equilibrium in the game G_Γ.

Our argument parallels that developed in Remark 1 for the standard contracting game without renegotiation. That is, we exhibit a mediated mechanism $\gamma' \neq \gamma^*$ which *uniquely* implements an allocation yielding P a payoff arbitrarily close to V^{SB}. Specifically, the following holds:

Proposition 3 *The renegotiation game G_Γ has a unique equilibrium allocation, which coincides with the second-best one (H, c^{SB}).*

The proof of Proposition 3 constructs a mechanism γ' by perturbing γ^* in such a way that, in the subgame $G_\Gamma(\gamma')$, for any belief $x \in [0,1]$, choosing not to renegotiate is the unique best response of P to any sequentially rational behavior of A. This, in turn, guarantees that $e = H$ is the unique optimal choice.

Proposition 3 shows that, in terms of implementation, our mediated mechanisms are more powerful than the mediated contracts in Rahman and Obara (2010), because they allow a *full* rather than only a *virtual* implementation of the principal's optimal outcome. Because our full implementation implies the uniqueness of the corresponding equilibrium allocation, it also guarantees that our results do not depend on any specific equilibrium selection criterion, neither at the ex-ante, nor at the interim stage.

4 Discussion

In this section, we put our results in perspective, clarifying the strategic role of the asymmetric information generated by mediated mechanisms.

Random vs. Mediated Mechanisms. Because γ^* conditions the final transfers on the random signal s, it effectively induces a random contract. Given this observation, it is natural to ask whether there is a random but *non-mediated* mechanism, i.e. a map associating any agent's message to a random contract, which allows to implement the second-best allocation. Indeed, throughout their analysis, FT restrict attention to the *deterministic* mechanisms in the class C, which prevents them from showing whether random mechanisms are welfare-enhancing. The following lemma, provides however a negative answer, thereby showing that P cannot gain by committing to lotteries over contracts.

Lemma 6 *Consider the game $\tilde{G}^r$, which coincides with G^r, with the exception that the set of available mechanisms C is now enlarged to $\tilde{C}$, which includes all the stochastic revelation mechanisms $\gamma_{\tilde{c}} : E \to \Delta(\mathbb{R}^2)$. Then, $\tilde{G}^r$ admits only one equilibrium allocation, which coincides with that of G^r.*

To understand the difference between a non-degenerate random mechanism $\gamma_{\tilde{c}} \in \tilde{C}$ and the role of randomness in our mediated mechanism γ^*, note that a mechanism $\gamma_{\tilde{c}} \in \tilde{C}$ is random for both P and A, in the sense that neither party can condition any of their decisions on the realization of the randomness. By contrast, γ^* is random for P, but not for A, because A can condition her choice whether to accept some renegotiation offer on the realization of the random component. Indeed, γ^* crucially exploits this feature.

Mediated Mechanisms and Information Storage. The mechanism γ^* is designed such that any signal s privately sent to A is publicly revealed only if she rejects the renegotiation offer γ^r. Indeed, our construction crucially exploits this feature. If γ^r could be made contingent on the realized s, then the renegotiation can be fine-tuned to match the effective reservation utility of the agent, thereby overcoming the obstacles raised by mediation.[14] In other words, γ^* is designed to store some information, which gets irrevocably lost whenever it is renegotiated upon. This makes our mediated mechanisms close to the "smart contracts" proposed by Georgiadis-Harris et al. (2023), which guarantee the privacy of the buyer's reported willingness to pay. In general terms, the idea that part of the communication taking place in a mechanism cannot be exploited by the subsequent trading proposals has been extensively applied in the limited commitment literature.[15]

Alternative Communication Protocols. In this paper, the communication protocol, i.e. the sequence of the messages to be sent and that of the signals to be received from any mechanism, is taken as given by both players. The specific protocol we adopt, which allows to uniquely implement the second-best allocation, successfully prevents renegotiation. One may ask whether, at the renegotiation stage, the principal may instead benefit from an alternative mode of communication. Specifically, he may be willing to post his offer γ^r *after* the agent communicates in the original mechanism γ^*. This may, in principle, mitigate the punishments he is subject to, at the renegotiation stage, through the agent's behavior in the mediated mechanism. Yet, even in such a scenario, one can construct a richer mediated mechanism, which incorporates a further degree of delegation to the agent. Specifically, designing a mechanism which gives the agent the power to choose *at which stage* to send messages (and subsequently receive signals), would give her an incentive to do that after γ^r has been posted, so to exploit the additional market information. This, we argue, allows to

[14] A formal proof of the profitability of renegotiation offers contingent on the realized signal s in the subgame $G_\Gamma(\gamma^*)$ is available in the Appendix.

[15] In Doval and Skreta (2022), the mechanism can garble the reports made by the buyer in each period. This turns out to reduce the set of available deviations for the next-period seller, which has a welfare-enhancing effect.

extend Proposition 2 to such a richer strategic scenario.[16]

Appendix 1: Proofs of the Main Results

Proof of Lemma 1. Consider first the function $H : (U^0, \infty) \to \mathbb{R}$, such that:

$$H(U) = \frac{V^{SB} - Y^H + \Phi(U+d)}{V^{SB} - V^{IC}(U)}, \tag{14}$$

which denominator is strictly positive since $V^{SB} > V^{IC}(U)$ over its domain. One should observe that:

1. $H(U) < 1$. Indeed, the convexity of Φ guarantees that $V^{IC}(U) < Y^H - \Phi(U+d)$ for any $U \in (U^0, \infty)$.

2. To determine the sign of $H(U)$, we only need to consider its numerator. Recalling that

$$V^{SB} = Y^H - p^H \Phi\left(U^0 + \frac{1-p^L}{\Delta p} d\right) - (1-p^H)\Phi\left(U^0 - \frac{p^L}{\Delta p} d\right)$$

the numerator of (14) can be written as

$$N(U) = \Phi(U+d) - p^H \Phi\left(U^0 + \frac{1-p^L}{\Delta p} d\right) - (1-p^H)\Phi\left(U^0 - \frac{p^L}{\Delta p} d\right), \tag{15}$$

that is continuous, strictly increasing, and strictly convex in U, which guarantees that $\underset{U\to\infty}{lim} N(U) = \infty$. In addition, we have:

$$N(U^0) = \Phi(U^0+d) - p^H \Phi\left(U^0 + \frac{1-p^L}{\Delta p} d\right) - (1-p^H)\Phi\left(U^0 - \frac{p^L}{\Delta p} d\right) < 0,$$

where the inequality follows from the convexity of Φ. This guarantees that there exists a unique finite $U^* \in (U^0, \infty)$ such that $N(U^*) = 0$. That is:

$$\begin{cases} H(U) < 0 & \text{if } U \in (U^0, U^*) \\ H(U) = 0 & \text{if } U = U^* \\ H(U) > 0 & \text{if } U \in (U^*, \infty) \end{cases}$$

3. H is continuously differentiable since, for all $U \in (U^0, \infty)$, the numerator is continuously differentiable and the denominator is continuosly differentiable, different from zero, and has first derivative different from zero.

[16] A detailed analysis of this setting, and of the generalized mediated mechanism one may construct, is available in the Appendix (Remark 10).

Next, observe that:

$$H'(U^*) = \frac{\Phi'(U^*)[V^{SB} - V^{IC}(U^*)] + V^{sb\prime}(U^*)[V^{SB} - Y^H + \Phi(U^*)]}{[V^{SB} - V^{IC}(U^*)]^2} = \frac{\Phi'(U^*)}{V^{SB} - V^{IC}(U^*)} > 0,$$

where the last equality follows from the fact that $V^{SB} - Y^H + \Phi(U^*) = N(U^*) = 0$.

We now argue that there exists a finite value $U^1 \in (U^*, \infty)$ such that H is defined, continuous and strictly increasing on the closed interval $[U^*, U^1]$.

To establish the result, two cases should be distinguished.

a) $H'(U) > 0$ for all $U \in (U^*, \infty)$. In this case, let U^1 be any $U \in (U^*, \infty)$, which implies that $H' > 0$ on (U^*, U^1). Since H is continuous on $[U^*, U^1]$, the Mean Value Theorem guarantees that it is also strictly increasing on this interval.

b) $H'(U)$ is *not* strictly positive for all $U \in (U^*, \infty)$. Then, since $H'(U^*) > 0$ and H' is continuous on (U^*, ∞), H' must have at least a zero on this interval. Take U^1 to be the smallest $U \in (U^*, \infty)$ such that $H'(U) = 0$. Since H' is continuous on $[U^*, U^1]$, with $H'(U^*) > 0$ and $H'(U^1) = 0$, and given the definition of U^1, we have that $H'(U) > 0$ for all $U \in (U^*, U^1)$. Once again, the continuity of H guarantees that this function is also strictly increasing on $[U^*, U^1]$.

Denote $\bar{\pi} = H(U^1)$. We now show that, for any $\pi \in (0, \bar{\pi})$, one can find a $U^j(\pi) \in (U^0, \infty)$ such that

$$\pi = H(U^j(\pi)), \tag{16}$$

which is a reformulation of (9). Indeed, recalling that H is continuous on $[U^*, U^1]$ with $H(U^*) = 0$ and $H(U^1) = \bar{\pi}$, the Intermediate Value Theorem guarantees the existence of some $U^j(\pi) \in [U^*, U^1]$ satisfying (16). In particular, since H is strictly increasing on $[U^*, U^1]$, $U^j(\pi)$ is uniquely defined on that interval. ■

Proof of Lemma 2. Define

$$W^{IC}(U) \doteq Y_H - V^{IC}(U)$$

to be the expected monetary transfer induced by the contract $c^{IC}(U)$. By subtracting Y^H from both sides and changing signs, equation (5) can be rewritten as

$$(1-\pi)\Phi(U^0+d) + \pi \left[\frac{1}{2}W^{IC}(U) + \frac{1}{2}W^{IC}(2U^j(\pi) - U)\right] = (1-\pi)W^{IC}(U^0) + \pi W^{IC}(U^j(\pi)). \tag{17}$$

Consider a pair $\left(\pi \in (0, \bar{\pi}), U^j(\pi)\right)$, and let the function $Z : [U^j(\pi), \infty) \to \mathbb{R}$ be such that[17]

$$Z(U) = (1-\pi)\Phi(U^0 + d) + \pi \left[\frac{1}{2}W^{IC}(U) + \frac{1}{2}W^{IC}(2U^j(\pi) - U)\right].$$

Since the RHS of (17) does not depend on U, to establish the result, we only need to show that there is a unique $U = U^k(\pi) > U^j(\pi)$ such that

$$Z(U^k(\pi)) = (1-\pi)W^{IC}(U^0) + \pi W^{IC}(U^j(\pi)).$$

One should observe that:

a) $Z(U^j(\pi)) = (1-\pi)\Phi(U^0 + d) + \pi W^{IC}(U^j(\pi)) < (1-\pi)W^{IC}(U^0) + \pi W^{IC}(U^j(\pi))$, where the inequality follows from the convexity of Φ.

b) Z is continuously differentiable on $[U^j(\pi), \infty)$, because it is the weighted sum of continuously differentiable functions. Also, it is strictly increasing and convex in $(U^j(\pi), \infty)$, which follows from the first and second derivatives of Z being positive for any $U > U^j(\pi)$. In fact:

$$\frac{\partial Z(U)}{\partial U} = \frac{\pi}{2}\frac{\partial W^{IC}(U)}{\partial U} - \frac{\pi}{2}\frac{\partial W^{SB}(2U^j(\pi) - U)}{\partial U} > 0,$$

which is satisfied since $\frac{\partial W^{IC}(U)}{\partial U}$ is increasing and $U > 2U^j(\pi) - U$ for any $U \in (U^j(\pi), \infty)$.

The second derivative of Z is :

$$\frac{\partial^2 Z(U)}{\partial U^2} = \frac{\pi}{2}\frac{\partial^2 W^{IC}(U)}{\partial U^2} + \frac{\pi}{2}\frac{\partial^2 W^{SB}(2U^j(\pi) - U)}{\partial U^2} > 0,$$

where the inequality follows from the convexity of W^{SB}.

c) $\lim_{U \to \infty} Z(U) = +\infty$, which is implied by Z being continuous, strictly increasing and convex on $(U^j(\pi), \infty)$.

Taken together, $(a) - (c)$ guarantee the existence of a unique $U^k(\pi) \in (U^j(\pi), \infty)$ satisfying (5). ■

[17] Note that $W^{IC}(2U^j(\pi) - U)$ is defined for any $U \in [U^j(\pi), \infty)$ since $W^{IC}(U)$ is defined for any $U \in (-\infty, +\infty)$ (p. 10).

Proof of Proposition 1. The proof is developed in three steps.

Step 1: Equilibrium Strategies. We construct an equilibrium of $G_\Gamma(\hat{\gamma})$ in which the principal does not renegotiate, i.e. $\gamma^r = \{\emptyset\}$.[18] The agent's strategy is such that she chooses $x(\alpha) = x(\beta) = 1$, i.e. she plays H with probability one, for any received signal. Her reporting strategy, instead, depends on whether renegotiation has taken place. Specifically:

1. If renegotiation does not occur, $\lambda(\hat{\gamma})$ prescribes to report $m = a$ with probability one in the original mechanism γ for any $s \in \hat{\mathcal{S}}$ and for any $e \in \{L, H\}$,

2. If the principal renegotiates, i.e. $\gamma^r \neq \{\emptyset\}$, $\lambda(\hat{\gamma})$ prescribes the following:

   - If $s = \alpha$, and for any $e \in \{L, H\}$, the agent selects $\rho = y$ and reports some $m^r \in \underset{m \in \mathcal{M}^r}{argmax}\, U(e, \tau^r(m))$ in γ^r if and only if $U(e, \tau^r(m^r)) > U^0$. Otherwise, she reports $m = b$ in the original mechanism γ.

   - If $s = \beta$, and for any $e \in \{L, H\}$, the agent selects $\rho = y$ and reports any $m^r \in \underset{m \in \mathcal{M}^r}{argmax}\, U(e, \tau^r(m))$ in γ^r if and only if $U(e, \tau^r(m^r)) > U^j(\pi)$. Otherwise, she reports $m = b$ in the original mechanism γ.

One can check that, if players stick to the strategies above, then the conditions $(i) - (iii)$ are satisfied. This in turn implies that the principal's (expected) payoff is

$$\hat{V}(\pi) \doteq \pi V^{IC}(U^j(\pi)) + (1 - \pi) V^{SB}.$$

Observe that the agent's strategy incorporates a potential punishment against the principal's attempts to renegotiate. Indeed, off-the-equilibrium-path, and upon receiving $s = \beta$, the agent may hinder a renegotiating offer $\gamma^r \neq \{\emptyset\}$ by reporting $m = b$ in the original mechanism γ, which leads to the inefficient lottery $\tau(\beta, b)$.

Step 2: The Agent's Sequential Rationality. We now show that the above strategy is sequentially rational for the agent.

Suppose first that the principal does not renegotiate, i.e. $\gamma^r = \{\emptyset\}$. When receiving $s = \alpha$ and upon choosing $e = H$, the agent is indifferent between reporting $m = a$ and $m = b$. Indeed, in any such case, she gets her reservation utility U^0. This shows that reporting $m = a$ is optimal. Likewise, the optimality of reporting $m = a$ in γ when receiving $s = \beta$ and upon choosing $e = H$, follows from (6), which guarantees that, in this case, she is indifferent between reporting $m = a$ and $m = b$. Her corresponding payoff is $U^j(\pi) > U^0$.

[18]To save notation, we denote $\gamma^r = \{\emptyset\}$ the principal's decision to abstain from renegotiation. In this case, the agent of type s may only trade by selecting an item $\tau(s, m) \in \{\tau(s, m)\}_{m \in \mathcal{M}}$ in the original mechanism.

Finally, we argue that taking H with probability one is a best response at stage 2 for any realized $s \in \{\alpha, \beta\}$. Indeed, if $s = \alpha$, the agent reports $m = a$, and c^{SB} is implemented. The contract belongs to the second-best frontier, which guarantees that $e = H$ is an optimal choice for the agent. If instead $s = \beta$, and $m = a$ is reported, $c^{IC}(U^j(\pi))$ is implemented. This contract also belongs to the second-best frontier, which makes $e = H$ an optimal choice.

Suppose next that the principal renegotiates, i.e. $\gamma^r \neq \{\emptyset\}$. It is immediate to see that the above reporting strategy is sequentially rational for the agent when she receives $s = \alpha$ and for any $e \in \{L, H\}$. If, instead, she receives $s = \beta$, chooses $e \in \{L, H\}$, and the principal renegotiates to γ^r such that $U(e, \tau^r(m^r)) \leq U^j(\pi)$, then (6) guarantees that it is optimal to declare $\rho = n$ and report $m = b$ in γ. By construction, declaring $\rho = y$ and reporting any $m^r \in \underset{m \in \mathcal{M}^r}{argmax}\, U(e, \tau^r(m))$ in γ^r is optimal whenever $U(e, \tau^r(m^r)) > U^j(\pi)$.

Step 3: The Principal's Incentives to Renegotiate. We now consider the problem faced by the principal at the renegotiation stage. If the principal's beliefs are consistent, he assigns probability one to the agent choosing H, and probability π $(1 - \pi)$ to the agent receiving the private signal β (α).

Given the agent's reporting strategy, the payoff that the principal obtains by not renegotiating is $\hat{V}(\pi) = (1-\pi)V^{SB} + \pi V^{IC}(U^j(\pi))$. We now show that there is *no* renegotiation γ^r yielding the principal a payoff strictly greater than $\hat{V}(\pi)$.

To characterize an optimal renegotiating offer γ^r, i.e. a P's best-response at stage 3, one should observe that the standard revelation principle applies at the renegotiation stage. That is, there is no loss of generality in restricting to mechanisms such that $|\mathcal{S}^r| = 1$, and $\mathcal{M}^r = \{\alpha, \beta\}$: A receives no private signals, and her reports in γ^r only consists of the private information generated by γ. We denote (U_g^α, U_b^α) the state-contingent payoff that γ^r guarantees to the agent when she reports α, and (U_g^β, U_b^β) that associated to the report β.

To yield a positive profit, γ^r must attract at least one type $s \in \{\alpha, \beta\}$ of A. We shall then consider two situations, depending on whether γ^r attracts both types, or only type α.[19]

To attract both types, γ^r should guarantee to the agent the expected utility $U^j(\pi)$, otherwise, upon receiving $s = \beta$, A would rather report in the original mechanism γ. The P's optimal renegotiation program is therefore:

[19]Since $U^j(\pi) > U^0$ is the agent's utility in the mechanism γ when $s = \beta$, there is no renegotiation γ^r which only attracts type β of the agent.

$$\max_{U_g^\alpha, U_b^\alpha, U_g^\beta, U_b^\beta} \quad Y^H - \pi[p^H \Phi(U_g^\beta) + (1-\pi^H)\Phi(U_b^\beta)] - (1-\pi)[p^H \Phi(U_g^\alpha) + (1-\pi^H)\Phi(U_b^\alpha)]$$

$$\text{s.t.:} \quad p^H U_g^\alpha + (1-p^H)U_b^\alpha \geq U^0 + d$$

$$p^H U_g^\beta + (1-p^H)U_b^\beta \geq U^j(\pi) + d$$

$$p^H U_g^\alpha + (1-p^H)U_b^\alpha \geq p^H U_g^\beta + (1-p^H)U_b^\beta$$

$$p^H U_g^\beta + (1-p^H)U_b^\beta \geq p^H U_g^\alpha + (1-p^H)U_g^\alpha$$

The two incentive-compatibility constraints are simultaneously satisfied if and only if:

$$p^H U_g^\alpha + (1-p^H)U_b^\alpha = p^H U_g^\beta + (1-p^H)U_b^\beta \tag{18}$$

Since there is no sorting condition and thus no incentive-compatibility reason for P to allocate any amount of risk to A, P's optimal choice is to offer full insurance to both types, i.e. $U_g^\alpha = U_b^\alpha = U^\alpha$, $U_g^\beta = U_b^\beta = U^\beta$.[20] This implies that (18) is satisfied letting $U^\beta = U^\alpha$. Since the participation constraint of type β must also bind, we get $U^\beta = U^\alpha = U^j(\pi) + d$.

The optimal γ^r is the full-insurance, pooling contract $c^{FI}(U^j(\pi)) = (U^j(\pi)+d, U^j(\pi)+d)$. The corresponding payoff to P is $Y^H - \Phi(U^j(\pi) + d)$.

By not renegotiating, P achieves the payoff $\pi V^{IC}(U^j(\pi)) + (1-\pi)V^{SB}$. It then follows from Lemma 1 that he is indifferent between renegotiating via the pooling contract $c^{FI}(U^j(\pi))$, and abstaining from renegotiation.

We next consider the case in which the renegotiating offer only attracts the agent of type α. Given the above reporting strategy, type β will report $m = b$ in the original mechanism, γ upon rejecting γ^r, which leads to implement the lottery $\tau(\beta, b)$ with probability π. It is straightforward to check that, in this case, the optimal γ^r is the full-insurance contract $c^{FI}(U^0) = \Phi(U^0 + d, U^0 + d)$. The corresponding payoff to the principal is:

$$(1-\pi)(Y^H - \Phi(U^0 + d)) + \pi \left(\frac{1}{2} V^{IC}(U^k(\pi)) + \frac{1}{2} V^{IC}(U^l(\pi)) \right)$$

Since $U^l(\pi) = 2U^j(\pi) - U^k(\pi)$, Lemma 2 guarantees that the principal is indifferent between renegotiating through the pooling contract $c^{FI}(U^j(\pi))$, and abstaining from any renegotiation. ■

Proof of Corollary 1. Since V^{SB} does not depend on π, the term $(1-\pi)V^{SB}$ converges to V^{SB} as π approaches zero. Consider next the term $\pi V^{IC}(U^j(\pi))$. The proof of Lemma

[20]Offers with random contracts can be excluded from the analysis as noted in Chade and Schlee (2012), for a similar reason: they introduce costly and unnecessary uncertainty as they do not help the principal to sort out types.

1 shows that there is a $U^1 \in (U^*, \infty)$ such that H is strictly increasing on $[U^*, U^1]$. This implies that its inverse $H^{-1} : [0, H(U^1)] \to [U^*, U^1]$ exists, and it is continuous on $[0, H(U^1)]$. In particular, we have $H^{-1}(0) = U^*$, and $\lim_{\pi\to 0} H^{-1}(\pi) = U^*$. In the proof, we let $U^j(\pi) = H^{-1}(\pi)$ for any $\pi \in (0, \bar{\pi})$, which yields $\lim_{\pi\to 0} U^j(\pi) = U^*$. Since $V^{IC}(U)$ is continuous at U^*, one can write ve $\lim_{\pi\to 0} V^{IC}(U^j(\pi)) = V^{IC}(U^*)$, which is a finite term. Thus, $\lim_{\pi\to 0} \pi V^{IC}(U^j(\pi)) = 0$. ■

Proof of Lemma 3. For a given $e \in E$, define the function $\tilde{V}_e : [U^0, \infty) \to \mathbb{R}$ as

$$\tilde{V}_e(U) \doteq \frac{1}{2} V_e^{FI}(2U^0 - U) + \frac{1}{2} V_e^{FI}(U).$$

The function satisfies the following properties:

a) $\tilde{V}_e(U)$ is well-defined, continuous, and twice differentiable for $U \in [U^0, \infty)$, because $\Phi(U)$ and, thus, $V_e^{FI}(U)$, are defined for every $U \in (-\infty, +\infty)$ and, moreover, are continuous, and twice differentiable.

b) $\tilde{V}_e(U)$ is strictly decreasing since

$$\frac{\partial \tilde{V}_e(U)}{\partial U} = \frac{1}{2}\frac{\partial V_e^{FI}(U)}{\partial U} - \frac{1}{2}\frac{\partial V_e^{FI}(2U^0 - U)}{\partial U} < 0,$$

where the inequality obtains since $U > 2U^0 - U$ for any $U \in [U^0, \infty)$, and because $V_e^{FI}(U)$ is concave so that $\partial V_e^{FI}/\partial U$ is decreasing.

c) $\tilde{V}_e(U)$ is strictly concave since

$$\frac{\partial^2 \tilde{V}_e(U)}{\partial U^2} = \frac{1}{2}\frac{\partial^2 V_e^{FI}(U)}{\partial U^2} + \frac{1}{2}\frac{\partial^2 V_e^{FI}(2U^0 - U)}{\partial U^2} < 0,$$

where the inequality follows because $\partial^2 V_e^{FI}(U)/\partial U^2 < 0$.

d) It follows from (b) and (c) that $\lim_{U\to\infty} \tilde{V}_e(U) = -\infty$.

e) For each $e \in E$, there is a $\underline{U}_e \in (U^0, \infty)$ such that

$$V_e^{IC}(U^0) = \tilde{V}_e(\underline{U}_e) \quad \text{and} \quad V_e^{IC}(U^0) > \tilde{V}_e(U) \quad \forall U \in (\underline{U}_e, \infty).$$

This holds since $\tilde{V}_e(U^0) = V_e^{FI}(U^0) > V_e^{IC}(U^0) > \lim_{U\to\infty} \tilde{V}_e(U) = -\infty$, where the first inequality follows from the convexity of Φ. Because $\tilde{V}_e(U)$ is continuous, the intermediate value theorem guarantees that there is a $\underline{U}_e \in (U^0, \infty)$: $\tilde{V}_e(\underline{U}_e) = V_e^{IC}(U^0)$. Because $\tilde{V}_e(U)$ is strictly decreasing, we have $\tilde{V}_e(U) < \tilde{V}_e(\underline{U}_e) = V_e^{IC}(U^0)$ for all $U > \underline{U}_e$.

It follows from (e) that for any $U^n > \max\{\underline{U}_H, \underline{U}_L\}$, we have

$$V_e^{IC}(U^0) > \tilde{V}_e(U^n). \tag{19}$$

Since $U^n > U^0 \Leftrightarrow U^n > 2U^0 - U^n$, and given that $\tilde{V}_e(U)$ is decreasing and concave, we have:

$$\tilde{V}_e(U^n) = \frac{1}{2}V_e^{FI}(2U^0 - U^n) + \frac{1}{2}V_e^{FI}(U^n) > max\{V_e^{FI}(U^n), \frac{1}{2}V_e^{FI}(2U^0 - U^n) + \frac{1}{2}V_e^{IC}(U^n)\}. \tag{20}$$

Taken together, (19) and (20) imply (8). ■

Proof of Proposition 3. The proof is developed in two steps.

1. We construct a mechanism γ', which uniquely implements an allocation such that $e = H$ and P's payoff is arbitrarily close to V^{SB}.

Take any $\varepsilon \in (0, \bar{\varepsilon})$, with $\bar{\varepsilon} > 0$ and denote:

$$c_\varepsilon^{SB} = \left(U^0 + \frac{(1-p_L)d + (1-p_H)\varepsilon}{p_H - p_L},\ U^0 - \frac{p_L d + p_H \varepsilon}{p_H - p_L}\right).$$

One can check that c_ε^{SB} leaves a payoff U^0 to A if she selects $e = H$, and $U^0 - \varepsilon$ if $e = L$. Observe that P's payoff $V_e(c_\varepsilon^{SB})$ is continuous in ε for each $e \in E$.

The mechanism $\gamma' = \{\mathcal{M}' = \mathcal{M}^*, \mathcal{S}' = \mathcal{S}^*, \sigma' = \sigma^*, \tau'\}$, coincides with γ^*, except for the rule τ', which, for a given ε, is such that:

(i) If A sends $m = a$, then $\tau'(a, \alpha) = \tau'(a, \beta) = c_\varepsilon^{SB}$.

(ii) If A sends $m = b$, and receives $s = \beta$, then $\tau'(b, \beta) = c^{IC}(U^n)$.

(iii) If A sends $m = b$, and receives $s = \alpha$, then $\tau'(b, \alpha) = c^{IC}(2U^0 - U^n - 3\varepsilon)$.

We now consider $G_\Gamma(\gamma')$, and show that $\gamma^r = \{\emptyset\}$ is the unique best response of P, for any sequentially rational behavior of A, and for any belief $x \in [0, 1]$.

Denote $\Lambda(\gamma')$ the set of A's sequentially rational strategies in $G_\Gamma(\gamma')$. Let us characterize the sequentially rational participation and reporting behaviors induced by some $\lambda(\gamma') \in \Lambda(\gamma')$ starting from the terminal nodes of $G_\Gamma(\gamma')$.

In any history (e, γ^r, m, s, y) such that $\gamma^r \neq \{\emptyset\}$, A sends any (distribution of) m_e^r as characterized in Proposition 2, and obtains $\hat{U}_e^r$.

Consider now A's participation behavior in any (e, γ^r, m, s) with $\gamma^r \neq \{\emptyset\}$. If $m = b$ and $s = \alpha$, $\rho = y$ is optimal if $\hat{U}_e^r \geq 2U^0 - U^n - 3\varepsilon$, while $\rho = n$ is optimal if the opposite weak inequality holds. Similarly, if $m = b$ and $s = \beta$, $\rho = y$ is optimal if $\hat{U}_e^r \geq U^n$ and $\rho = n$ if

$\hat{U}_e^r \leq U^n$.[21] If $m = a$ and $e = H(L)$, regardless of $s \in \{\alpha, \beta\}$, $\rho = y$ is optimal if $\hat{U}_H^r \geq U^0$ ($\hat{U}_L^r \geq U^0 - \varepsilon$), and $\rho = n$ is optimal if the opposite weak inequality holds.

Consider next A's reports in γ^* after any (e, γ^r). Given the above participation behavior, it is optimal for her to report $m = a$ in γ^*, if $e = H$, whenever

$$\max\{U^0, \hat{U}_H^r\} \geq \frac{1}{2} \max\left\{\hat{U}_H^r, 2U^0 - U^n - 3\varepsilon\right\} + \frac{1}{2} \max\{\hat{U}_H^r, U^n\} \tag{21}$$

while $m = b$ is optimal when the opposite weak inequality holds. If $e = L$, (21) becomes

$$\max\{U^0 - \varepsilon, \hat{U}_L^r\} \geq \frac{1}{2} \max\left\{\hat{U}_L^r, 2U^0 - U^n - 3\varepsilon\right\} + \frac{1}{2} \max\{\hat{U}_L^r, U^n\}. \tag{22}$$

Consider now P's behavior at the renegotiation stage and assume, to start with, that he holds a deterministic belief $x \in \{0, 1\}$ over A's effort. As argued in the proof of Proposition 2 (Lemma 5), for any $e \in E$, it is optimal for him to choose either $\gamma^r = \{\emptyset\}$ or a renegotiated offer $\gamma^r \in \Gamma$ such that $\mathcal{M}^r = \{m_1\}$ and $\mathcal{S}^r = \{s_1\}$ are singletons, and the full insurance transfers $\tau^r(m_1, s_1) = c_e^{FI}(\hat{U}_e^r)$ are implemented. Thus, any optimal renegotiated offer is characterized by the number $\hat{U}_e^r \in (-\infty, +\infty)$, that is, the expected payoff it yields to A when she chooses $e \in E$. It follows that:

1. If $\hat{U}_H^r < 2U^0 - U^n \vee \hat{U}_L^r < 2U^0 - 2\varepsilon - U^n \vee \gamma^r = \{\emptyset\}$, A's sequentially rational behavior is unique, and coincides with ($m = a, \rho = n$ for all $s \in \{\alpha, \beta\}$), which yields $V_e(c_\varepsilon^{SB}(U^0))$ for $e \in \{L, H\}$ to P. Note that $\rho = n$ is optimal after sending $m = a$, since

   $$\hat{U}_H^r < 2U^0 - U^n < U^0 \quad \text{and} \quad \hat{U}_L^r < 2U^0 - 2\varepsilon - U^n < U^0 - \varepsilon$$

   where both inequalities follow from $U^n > U^0$. To see why $m = a$ is preferred to $m = b$, two cases must be considered:

   (a) If $\hat{U}_H^r \leq 2U^0 - U^n - 3\varepsilon$, then (21) becomes $U^0 \geq U^0 - \frac{3}{2}\varepsilon$; likewise, if $\hat{U}_L^r \leq 2U^0 - U^n - 3\varepsilon$, then (22) becomes $U^0 - \varepsilon \geq U^0 - \frac{3}{2}\varepsilon$, and both inequalities are strictly satisfied since $\varepsilon > 0$.

   (b) If $\hat{U}_H^r \in (2U^0 - U^n - 3\varepsilon, 2U^0 - U^n)$, then (22) becomes $\hat{U}_H^r \leq 2U^0 - U^n$; likewise, if $\hat{U}_L^r \in (2U^0 - U^n - 3\varepsilon, 2U^0 - 2\varepsilon - U^n)$, then (22) becomes $\hat{U}_L^r \leq 2U^0 - 2\varepsilon - U^n$, and both inequalities are strictly satisfied by construction.

[21] The participation behavior following $m = b$ is optimal for each $e \in E$, since both $\tau'(b, \alpha)$ and $\tau'(b, \beta)$ implement contracts on the incentive-compatibility frontier.

2. If $\hat{U}_H^r = 2U^0 - U^n \ \vee \ \hat{U}_L^r = 2U^0 - 2\varepsilon - U^n$, A is indifferent between ($m = a, \rho = n$ for all $s \in \{\alpha, \beta\}$), and sending $m = b$, followed by $\rho = y(n)$ when $s = \alpha(\beta)$ since, in this case, (21) and (22) hold as equalities. For any randomization over A's decisions, P's payoff is a convex combination of $V_L(c_\varepsilon^{SB})$ and $\frac{1}{2}V_L^{FI}(2U^0 - 2\varepsilon - U^n) + \frac{1}{2}V_L^{IC}(U^n)$, if $e = L$, or $V_H(c_\varepsilon^{SB})$ and $\frac{1}{2}V_H^{FI}(2U^0 - U^n) + \frac{1}{2}V_H^{IC}(U^n)$, if $e = H$.

3. If $\hat{U}_H^r \in (2U^0 - U^n, U^n) \ \vee \ \hat{U}_L^r \in (2U^0 - 2\varepsilon - U^n, U^n)$, then A's sequentially rational behavior is unique, and coincides with $m = b$, followed by $\rho = y$ when $s = \alpha$, and by $\rho = n$ when $s = \beta$. In this case, both (21) and (22) are violated. The corresponding payoff to P is $\frac{1}{2}V_e^{FI}(\hat{U}_e^r) + \frac{1}{2}V_e^{IC}(U^n)$, for $e \in E$.

4. If $\hat{U}_e^r = U^n$, A is indifferent between sending $m = a$, followed by $\rho = y$ for all $s \in \{\alpha, \beta\}$, and $m = b$, followed by $\rho = y$ when $s = \alpha$ and any (randomization over) $\rho \in \{y, n\}$ when $s = \beta$. For any mixture over her message and participation decisions, P obtains a convex combination between $V_e^{FI}(U^n)$ and $\frac{1}{2}V_e^{FI}(U^n) + \frac{1}{2}V_e^{IC}(U^n)$, for $e \in E$.

5. If $\hat{U}_e^r \in (U^n, \infty)$, A is indifferent between sending $m = a$ and $m = b$, followed by $\rho = y$ for any $s \in \{\alpha, \beta\}$. In any such situation, P obtains $V_e^{FI}(\hat{U}_e^r)$, for $e \in E$.

The above remarks guarantee that, if $e = H$, the following inequalities are sufficient for $\gamma^r = \{\emptyset\}$ to be P's unique best response:

$$V_H(c_\varepsilon^{SB}) - \frac{1}{2}V_H^{FI}\left(2U^0 - U^n\right) - \frac{1}{2}V_H^{IC}(U^n) > 0, \tag{23}$$

and

$$V_H(c_\varepsilon^{SB}) - V_H^{FI}(U^n) > 0. \tag{24}$$

Observe that, if $\varepsilon = 0$, (23) and (24) are satisfied because they coincide with (12) and (13), respectively. Since $V_H(c_\varepsilon^{SB})$ is continuous in ε, there is a $\varepsilon^H > 0$ such that (23) and (24) are satisfied for any $\varepsilon \in (0, \varepsilon^H)$. If, instead, $e = L$, P to strictly prefers not to renegotiate if

$$V_L(c_\varepsilon^{SB}) - \frac{1}{2}V_L^{FI}\left(2U^0 - 2\varepsilon - U^n\right) - \frac{1}{2}V_L^{IC}(U^n) > 0 \tag{25}$$

and

$$V_L(c_\varepsilon^{SB}) - V_L^{FI}(U^n) > 0. \tag{26}$$

Again, since $V_L(c_\varepsilon^{SB})$ is continuous in ε, there is a $\varepsilon^L > 0$ such that (25) and (26) are satisfied for any $\varepsilon \in (0, \varepsilon^L)$. Denoting $\bar{\varepsilon} \doteq min\{\varepsilon^L, \varepsilon^H\}$ allows to conclude that, if P holds a degenerate belief, and for any $\varepsilon \in (0, \bar{\varepsilon})$, he cannot gain by renegotiating.

Consider next the general case $x \in [0,1]$. Denote P's expected payoff by not renegotiating

$$V_x(c_\varepsilon^{SB}) = xV_H(c_\varepsilon^{SB}) + (1-x)V_L(c_\varepsilon^{SB}).$$

By offering $\gamma^r \neq \{\emptyset\}$, he would instead get

$$V_x^*(\gamma^r, \lambda(\gamma')) = xV_H^*(\gamma^r, \lambda(\gamma')) + (1-x)V_L^*(\gamma^r, \lambda(\gamma')),$$

which is P's continuation payoff, incorporating the optimal behavior of A following the history (e, γ^r), according to $\lambda(\gamma')$. Yet, as shown in the first part of the proof, for any $e \in E$, there is no γ^r yielding P a payoff above $V_e(c_\varepsilon^{SB})$. This in turn prevents him from getting an expected payoff greater than $V_x(c_\varepsilon^{SB})$, for every $(x, 1-x)$.

We finally consider A's effort choice. Since A perfectly anticipates that $m = a$ will be sent, leading to either $\tau'(a, \alpha) = \tau'(a, \beta) = c_\varepsilon^{SB}$, she strictly prefers choosing $e = H$ with probability one. The corresponding payoff to P is therefore $V_H(c_\varepsilon^{SB})$, which can be made arbitrarily close to $V^{SB} = V_H^{IC}(U^0)$ by choosing a sufficiently small ε. Specifically, one can check that

$$\lim_{\varepsilon \to 0} V_H(c_\varepsilon^{SB}) = V^{SB}. \tag{27}$$

2. We now prove that G_Γ admits an equilibrium, and that, in any equilibrium, (H, c^{SB}) is implemented.

To show equilibrium existence, it is enough to consider the strategies exhibited in the proof of Proposition 2. If P posts γ^*, and, players stick to the equilibrium strategies $(\lambda(\gamma^*), \gamma^r = \{\emptyset\})$ in the subgame $G_\Gamma(\gamma^*)$, he gets his maximal payoff V^{SB}, and has therefore no incentive to deviate to any $\gamma' \neq \gamma^*$.[22]

To establish the uniqueness result, we proceed by contradiction. Suppose that there is an equilibrium in G_Γ yielding P a payoff $V^* < V^{SB}$. Then, given (27) there is a mechanism γ' and an arbitrarily small ε which yield P a payoff $V_H(c_\varepsilon^{SB}) \in (V^*, V^{SB})$ in the unique equilibrium of $G_\Gamma(\gamma')$. This generates a contradiction. ■

Proof of Lemma 6. For any stochastic mechanism $\gamma_{\tilde{c}} \in \tilde{C}$, define $\gamma_{\tilde{c}}(e) = \tilde{c}_e$ and let

$$\tilde{U}_e \doteq p_e \mathbb{E}[U_g|\tilde{c}_e] + (1-p_e)\mathbb{E}[U_b|\tilde{c}_e]$$

[22]To simplify exposition, we do not provide a full description of the players' continuation strategies. In particular, we do not specify their behaviors in any subgame $G_\Gamma(\gamma')$, with $\gamma' \neq \gamma^*$.

be the A's expected payoff after taking the effort $e \in E$, and truthfully reporting it in $\gamma_{\tilde{c}}$. Consider the subgame $G^r(\gamma_{\tilde{c}})$, and suppose that $e = H$ is chosen with probability $x \in [0,1]$. The revelation principle guarantees that the maximal payoff attainable by P by a renegotiation offer $\gamma^r \in \tilde{C}$ is the value of the program $P(x, \tilde{U}_H, \tilde{U}_L)$:

$$V^*(x, \tilde{U}_H, \tilde{U}_L) = \max_{\gamma^r_{\tilde{c}} \in \tilde{C}} \quad Y(x) - x[p_H \mathbb{E}(\Phi(U_g)|c^r_H) + (1-p_H)\mathbb{E}(\Phi(U_b)|c^r_H)] \\ - (1-x)[p_L \mathbb{E}(\Phi(U_g)|c^r_L) + (1-p_L)\mathbb{E}(\Phi(U_b)|c^r_L)] \tag{28}$$

$$\text{s.t.:} \quad p_H \mathbb{E}(U_g|c^r_H) + (1-p_H)\mathbb{E}(U_b|c^r_H) \geq \tilde{U}_H \tag{IRC_H}$$

$$p_L \mathbb{E}(U_g|c^r_L) + (1-p_L)\mathbb{E}(U_b|c^r_L) \geq \tilde{U}_L \tag{IRC_L}$$

$$p_H \mathbb{E}(U_g|c^r_H) + (1-p_H)\mathbb{E}(U_b|c^r_H) \geq p_H \mathbb{E}(U_g|c^r_L) + (1-p_H)\mathbb{E}(U_b|c^r_L) \tag{ICC_H}$$

$$p_L \mathbb{E}(U_g|c^r_L) + (1-p_L)\mathbb{E}(U_b|c^r_L) \geq p_L \mathbb{E}(U_g|c^r_H) + (1-p_L)\mathbb{E}(U_b|c^r_H) \tag{ICC_L}$$

where $Y(x) = xY_H + (1-x)Y_L$. The following two results hold:

Lemma 7 *$P(x, \tilde{U}_H, \tilde{U}_L)$ admits a unique solution, which is deterministic.*

Proof. See Chade and Schlee (2012, Proposition 2).

Denote $\gamma^r(\gamma_{\tilde{c}}, x)$ the unique solution to $P(x, \tilde{U}_H, \tilde{U}_L)$.

Lemma 8 *For any $\gamma_{\tilde{c}} \in \tilde{C}$ and $x \in [0,1]$ there is a deterministic $\gamma_c \in C$ such that $\gamma^r(\gamma_{\tilde{c}}, x) = \gamma^r(\gamma_c, x)$.*

Proof. Given $\gamma_{\tilde{c}} \in \tilde{C}$, we construct the deterministic mechanism γ_c yielding the transfers $U^e_\omega = \mathbb{E}(U_\omega|\tilde{c}_e)$ for each $(e, \omega) \in E \times \{g, b\}$. Thus, for any $x \in [0,1]$, the optimal renegotiation offer in $G(\gamma_c)$ obtains again from solving $P(x, \tilde{U}_H, \tilde{U}_L)$. ■

Finally, if γ_c is constructed from $\gamma_{\tilde{c}}$ as in the proof of Lemma 8, the following holds:

Lemma 9 *The subgames $G^r(\gamma_c)$ and $\tilde{G}^r(\gamma_{\tilde{c}})$ have the same (perfect-Bayesian) equilibrium allocations.*

Proof. Consider the subgame $\tilde{G}^r(\gamma_{\tilde{c}})$, and let $x \in [0,1]$ be the equilibrium distribution over efforts. Let $\tilde{G}^r(\gamma_c)$ be the subgame induced by the mechanism γ_c, which is obtained from $\gamma_{\tilde{c}}$ as in the proof of Lemma 8. It follows that, in either subgame, P's renegotiation offer is $\gamma^r(\gamma_{\tilde{c}}, x) = \gamma^r(\gamma_c, x)$, which is accepted by A, who truthfully reports her former effort.[23] Furthermore, the transfers corresponding to the unique solution of $P(x, \tilde{U}_H, \tilde{U}_L)$ are

[23] See Fudenberg and Tirole (1990, p. 1295).

implemented. Thus, playing $e = H$ with probability $x \in [0,1]$ is sequentially rational for A in $\tilde{G}^r(\gamma_c)$, which implies that $G^r(\gamma_c)$ and $G^r(\gamma_{\tilde{c}})$ have the same PBE allocations. ■

To conclude the proof, denote x^{FT} the equilibrium probability of $e = H$ characterized by FT, and U^{FT} the equilibrium rent of A. Lemma 9 implies that the upper bound $V^{FT} = V^*(x^{FT}, U^{FT}, U^{FT})$ of P's payoffs characterized by FT for the deterministic game G^r is also an upper bound in $\tilde{G}^r$. In addition, in the game $\tilde{G}^r$, P can achieve V^{FT} as the unique continuation payoff by offering any of the mechanisms characterized in Fudenberg and Tirole (1990, Proposition 3.4). Thus, the unique equilibrium's payoff of P in $\tilde{G}^r$ is V^{FT} as in G^r, and the same distributions over efforts and transfers are implemented. ■

Appendix 2: Extensions

Renegotiation with public signals. Consider the mechanism $\gamma^* = \{\mathcal{M}^*, \mathcal{S}^*, \sigma^*, \tau^*\}$ characterized in Proposition 2, where $\mathcal{M}^* = \{a, b\}$, $\mathcal{S}^* = \{\alpha, \beta\}$, $\sigma^*(\alpha|m) = \sigma^*(\beta|m) = \frac{1}{2}$ for all $m \in \mathcal{M}^*$, and:

$$\tau^*(a, \alpha) = \tau^*(a, \beta) = c^{SB}$$
$$\tau^*(b, \alpha) = c^{IC}(2U^n - U^0), \quad \tau^*(b, \beta) = c^{IC}(U^n)$$

with U^n defined as in Lemma 3. We now assume that the realization $s \in \{\alpha, \beta\}$ is public.

This raises an important question: shall we enlarge the set of available renegotiated offers to allow P to design transfers conditionally on the realized public signals? We show that, if the answer is positive, there is a renegotiated offer γ^r which guarantees him a payoff strictly above V^{SB} regardless of the agent continuation play.

Consider the subgame $G_\Gamma^{Pub}(\gamma^*)$, which starts after $\gamma = \gamma^*$ is offered and accepted:

1. A takes $e \in \{L, H\}$.

2. P offers $\gamma^r \in \Gamma^r = \{\mathcal{M}^r, \mathcal{S}^r, \sigma^r, \tau^r\}$, with $\tau^r : \mathcal{M}^r \times \mathcal{S}^r \times \mathcal{S} \to \Delta C$, allowing to condition on the realization $s \in \mathcal{S}$.

3. A sends $m \in \mathcal{M}^*$.

4. The signal $s \in \mathcal{S}^*$ is realized according to σ^*.

5. If $\rho = y$, A sends $m^r \in \mathcal{M}^r$, the signal $s^r \in \mathcal{S}^r$ is realized according to σ^r, and the transfers $\tau^r(s^r, m^r, s)$ are implemented. If $\rho = n$ or $\gamma^r = \{\emptyset\}$, the transfers $\tau^*(m, s)$ are implemented.

Observe that, if $\gamma^r = \{\emptyset\}$, it is optimal for A to send $m = a$, as already shown in Proposition 2. This leads to the implementation of c^{SB} and yields P the payoff V^{SB}.

We now argue that there is no equilibrium of G_Γ^{Pub} in which $x = 1$ and renegotiation is prevented. In particular, we exhibit a $\hat{\gamma}^r$ which, under the belief $x = 1$, yields P a payoff strictly greater than V^{SB} under any sequentially rational continuation play.

To simplify the exposition, let $\hat{\gamma}^r$ be such that $|\hat{\mathcal{S}}^r| = |\hat{\mathcal{M}}^r| = 1$, and suppress any dependence of $\hat{\tau}^r$ on the unique (m^r, s^r). The decision rule of $\hat{\gamma}^r$ is:

$$\hat{\tau}^r(s) = \begin{cases} c_H^{FI}(U^0 + \eta) & \text{if } s = \alpha; \\ c^{IC}(\underline{U}) & \text{if } s = \beta, \end{cases}$$

where $\underline{U} < 2U^0 - U^n$,[24] and $\eta > 0$ is small enough to guarantee that:

- $U^0 + \eta < U^n$;
- $V_H^{FI}(U^0 + \eta) > V^{SB}$.

We first characterize the optimal participation and reporting behavior of A at any history such that $e = H$. Specifically, we show that, at any $(H, \hat{\gamma}^r, m, s)$, A only accepts $\hat{\gamma}^r$ if $(m = a, s = \alpha)$, and that, at any $(H, \hat{\gamma}^r)$, the unique optimal choice of A is to report $m = a$.

Start from the participation behavior. Suppose first $(m = a, s = \alpha)$: by rejecting, A gets U^0, and by accepting, $\hat{\tau}^r(\alpha) = c_H^{FI}(U^0 + \eta)$, which yields $U^0 + \eta > U^0$. Thus, it is strictly optimal for her to accept the offer. If instead $(m = a, s = \beta)$, A gets U^0 in the original mechanism and $\underline{U} < U^0$ in the renegotiated one, and is thus strictly better off by rejecting. Suppose instead that $m = b$. Then, if $s = \alpha$, A is better off by rejecting since $U_H(c^{IC}(U^n)) > U_H(c_H^{FI}(U^0 + \eta))$ by construction of η. In this case, she obtains U^n. If $s = \beta$, she will also reject $\hat{\gamma}^r$ since $U_H(c^{IC}(2U^0 - U^n)) > U_H(c^{IC}(\underline{U}))$, hence obtaining $2U^0 - U^n$.

To see that it is (strictly) optimal for A to report $m = a$, note that, if $m = a$, she gets $U^0 + \frac{\eta}{2} > U^0$. If, instead, $m = b$, her expected payoff is $\frac{2U^0 - U^n}{2} + \frac{U^n}{2} = U^0$. Thus, given her participation behavior, A prefers the report $m = a$ to $m = b$ at any $(H, \hat{\gamma}^r)$.

Given the A's behavior at any $(H, \hat{\gamma}^r)$, P's payoff from any offer $\hat{\gamma}^r$ under $x = 1$ is

$$\frac{1}{2} V^{FI}(U^0 + \eta) + \frac{1}{2} V^{SB} > V^{SB}$$

where the inequality holds by construction of η. Thus, P strictly prefers offering $\hat{\gamma}^r$ than abstaining from renegotiation, which allows to conclude that there is no equilibrium of $G_\Gamma^{Pub}(\gamma^*)$, in which $x = 1$ and no renegotiation takes place.

[24] It does not really matter that $\tau^r(\alpha)$ is an incentive-compatible contract. What is important is that $U_H(\tau^r(\beta)) < 2U^0 - U^n$.

Endogenous timing of the agent's report. Consider the mechanism $\mu = \{\mathcal{M}^\mu, \mathcal{S}^\mu, \sigma^\mu, \tau^\mu\}$ such that $\mu \notin \Gamma$ and

$$\mathcal{M}^\mu = \{a, b, \emptyset\}, \quad \mathcal{S}^\mu = \{\alpha, \beta\}, \quad \sigma^\mu = \left(\frac{1}{2}, \frac{1}{2}\right).$$

The mechanism μ is a modified version of $\gamma' = \{\mathcal{M}', \mathcal{S}', \sigma', \tau'\}$ from Proposition 3. If μ is offered by P and accepted by A, it induces the following extensive form game G_μ:

1. A selects $e \in E$.
2. A sends $m_1 \in \{a, b, \emptyset\}$.
3. If $m_1 \neq \emptyset$, $s \in \{\alpha, \beta\}$ is extracted from the distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$ and disclosed privately to A.
4. P proposes $\gamma^r \in \Gamma$.
5. If $m_1 = \emptyset$, A sends $m_2 \in \{a, b, \emptyset\}$, and $m_2 = \emptyset$ otherwise.
6. If $(m_1 = \emptyset, m_2 \neq \emptyset)$, $s \in \{\alpha, \beta\}$ is extracted from the distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$ and is disclosed privately to A.
7. A takes the participation decision $\rho \in \{y, n\}$.
8. There are three possible situations:
   - If $\gamma^r \neq \{\emptyset\}$ and $\rho = y$, A sends $m^r \in \mathcal{M}^r$, $s^r \in \mathcal{S}^r$ is extracted from σ^r and $\tau^r(m^r, s^r)$ is implemented.
   - If $\gamma^r = \{\emptyset\}$ or $\rho = n$, and if $m_1 = m_2 = \emptyset$, A sends $m_3 \in \{a, b\}$, then $s \in \{\alpha, \beta\}$ is extracted from the distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$, and the decision rule $\tau^\mu(m_1, m_2, m_3, s)$ is implemented.
   - If $\gamma^r = \{\emptyset\}$ or $\rho = n$, and $m_1 \neq \emptyset$ or $m_2 \neq \emptyset$, A sends $m_3 = \emptyset$ and the decision rule $\tau^\mu(m_1, m_2, m_3, s)$ is implemented.

For a given vector of A's messages (m_1, m_2, m_3) let $m_j \in (m_1, m_2, m_3)$ be the only one different from $\emptyset$, and assume that $\tau^\mu(m_1, m_2, m_3, s) = \tau'(m_j, s)$. Note that μ extends the optimal mechanism γ' characterized in the proof of Proposition 3, by giving A the freedom to decide, at each step $i = 1, 2, 3$ of the *interim* stage, whether to send the message $m_i \in \{a, b\}$ or to stay silent $(m_i = \emptyset)$.[25] A *must* speak, only once, in the mechanism μ, i.e. if

[25] We assume, for simplicity of exposition and without loss of generality, that A cannot send any message at the ex-ante stage, that is, before taking $e \in E$.

$m_1 = m_2 = \emptyset$,A is forced to send a nonempty message $m_3 \in \{a, b\}$ at the last stage of the game, but, if at some point she sends a nonempty message, her future messages must be empty. Note also that A learns the realization $s \in \{\alpha, \beta\}$ of the private signal as soon as she sends a nonempty message $m_i \neq \emptyset$.

A behavioral strategy for P in G_μ is a distribution over the set of the renegotiated offers Γ. In fact, since the message and the signal are exchanged privately, at the renegotiation stage, the principal does not know whether communication has taken place or not in the original mechanism, and thus, he cannot condition his renegotiation on such information. This crucially implies that the renegotiating principal cannot prevent the agent from *waiting* until a renegotiation offer before making his report in the original mechanism.

A behavioral strategy $\lambda(\mu)$ of A specifies an effort probability $x \in [0, 1]$ at the initial history (μ); a distribution over $m_1 \in \mathcal{M}^\mu$ for each (μ, e), a distribution over the participation decisions $\rho \in \{y, n\}$ at any $(\mu, e, m_1, s, \gamma^r)$ with $m_1 \neq \emptyset$ and $\gamma^r \neq \{\emptyset\}$, followed by a distribution over $m^r \in \mathcal{M}^r$ at any history such that $\rho = y$. It also involves a distribution over $m_2 \in \mathcal{M}^\mu$ at any $(\mu, e, \emptyset, \gamma^r)$, followed by a distribution over ρ at any $(\mu, e, \emptyset, \gamma^r, m_2, s)$ and a distribution over $m^r \in \mathcal{M}^r$ at any continuation $(\mu, e, \emptyset, \gamma^r, m_2, s, y)$ with $\gamma^r \neq \{\emptyset\}$. Finally, it involves a distribution over $m_2 \in \mathcal{M}^\mu$ at any $(\mu, e, \emptyset, \gamma^r)$ with $\gamma^r = \{\emptyset\}$, a distribution over ρ at any $(\mu, e, \emptyset, \gamma^r, \emptyset)$ with $\gamma^r \neq \{\emptyset\}$, a distribution over $m_3 \in \mathcal{M}^\mu$ at the continuation histories such that $\rho = n$, and a distribution over $m^r \in \mathcal{M}^r$ at the continuation histories such that $\rho = y$.

One can then show the following:

Lemma 10 *G_μ admits an equilibrium which supports the same equilibrium allocation as that of $G_\Gamma(\gamma')$.*

Proof. We show that there exists an equilibrium in which P abstains from renegotiating, A chooses $e = H$ selecting $(m_1 = \emptyset, m_2 = a)$ with probability one on the equilibrium path; also, A sends $m_1 = \emptyset$ off-the-equilibrium path when $e = L$, and, following any history such that $\gamma^r = \{\emptyset\}$ and $m_1 = \emptyset$, she sends $m_2 = a$. Also, at any history (μ, e, m_1, γ^r) such that $m_1 = \emptyset$ and $\gamma^r \neq \{\emptyset\}$, A send $m_2 \in \mathcal{M}^\mu$ according to the following rule:

(i) If $\gamma^r = \{\emptyset\}$, and for any γ^r such that $\hat{U}_H^r \leq 2U^0 - U^n \vee \hat{U}_L^r \leq 2U^0 - 2\varepsilon - U^n$, A sends $m = a$ in μ.

(ii) For any $\gamma^r \neq \{\emptyset\}$ such that $\hat{U}_H^r \in (2U^0 - U^n, U^n] \vee (2U^0 - U^n - 2\varepsilon, U^n]$, A sends $m = b$ in μ.

(iii) For any $\gamma^r \neq \{\emptyset\}$ such that $\hat{U}^r_e > U^n$, A sends $m = b$ in μ.

Moreover, at any history such that $\gamma^r \neq \emptyset$ and $m_j \in \{m_1, m_2\}$, she selects a participation decision for each $s \in \{\alpha, \beta\}$ according to the following rule:

(i) If $m_j = a$ and for each $s \in \{\alpha, \beta\}$, $\rho = y$ if $\hat{U}^r_H \geq U^0$ ($\hat{U}^r_L \geq U^0 - \varepsilon$) and $\rho = n$ otherwise.

(ii) If $m_j = b$ and $s = \alpha$, $\rho = y$ if $\hat{U}^r_e \geq 2U^0 - U^n - 3\varepsilon$ and $\rho = n$ otherwise.

(iii) If $s = \beta$, $\rho = y$ if $\hat{U}^r_e \geq U^n$ and $\rho = n$ otherwise.

At any history such that $m_1 = m_2 = \emptyset$, A participates in the renegotiated mechanism if and only if $\hat{U}^r_H \geq U^0$ ($\hat{U}^r_L \geq U^0 - \varepsilon$).[26] Finally, at any terminal history such that $m_1 = m_2 = \emptyset$ in which she is asked to report some $m_3 \in \mathcal{M}^\mu$ into μ, she sends $m_3 = a$, while, at any terminal history in which she participates in a renegotiated mechanism, she optimally sends some $m^r_e \in \mathcal{M}^r$ as defined in the proof of Proposition 2.

Note first that, given the A's behavior, the sequential rationality of P's behavior has already been shown in Proposition 3. The same argument applies to the A's effort and participation behaviors at any history such that $m_j \in \{m_1, m_2\}$.

We check the sequential rationality of the other A's decisions, starting from the terminal nodes. It can be seen that, at any terminal history in which A sends $m_3 \neq \emptyset$ (or $m_2 \neq \emptyset$ with $\gamma^r = \{\emptyset\}$), her only sequentially rational decision is to send $m_3 = a$. Since any such history is such that either $\rho = n$ or $\gamma^r = \{\emptyset\}$, and thus A has to stick to the original mechanism, by sending $m_j = a$ she avoids the penalty $-\frac{3}{2}\varepsilon$ associated to the report $m_j = b$ in μ.

Consider now A's participation behavior when $m_2 = \emptyset$. Since $s \in \{\alpha, \beta\}$ is payoff-irrelevant when $m_j = a$, the fact that at any history $(\mu, e, \emptyset, \gamma^r, m_2, s)$ it is sequentially rational to select $\rho \in \{y, n\}$, implies that the same participation decision $\rho \in \{y, n\}$ is optimal at any $(\mu, e, \emptyset, \gamma^r, \emptyset)$ for the same renegotiated offer γ^r. But then, the participation behavior associated to $m_2 = \emptyset$ is sequentially rational, since it is equivalent to the one constructed when $m_2 = a$, which has been shown to be sequentially rational in Proposition 3.

Let us now turn to A's choice of $m_2 \in \mathcal{M}^\mu$ at any $(\mu, e, \emptyset, \gamma^r)$. Since A is indifferent between sending $m_2 = \emptyset$ and $m_2 = a$ for any sequentially rational continuation play, she has no incentive to deviate from $m_2 = a$ to $m_2 = \emptyset$ when prescribed by the behavioral strategy we construct. Furthermore, deviating from $m_2 = b$ to $m_2 = \emptyset$ yields her no strictly profitable deviation since $m_2 = \emptyset$ is equivalent to $m_2 = a$, and thus, the existence of a profitable

[26] This is equivalent to her participation behavior when $j \in \{1, 2\}$ and $m_j = a$ for each $s \in \{\alpha, \beta\}$.

deviation to $m_2 = \emptyset$ would imply the existence of a deviation from $m = b$ to $m = a$ in the original game studied in Proposition 3. Thus, the A's message behavior corresponding to $m_2 \in \mathcal{M}^\mu$ is sequentially rational. Finally, consider her behavior concerning m_1. Since $\gamma^r = \emptyset$ at equilibrium, it is payoff-irrelevant for A to send the optimal message $m_j = a$ at the stages $i \in \{1, 2\}$, and thus, there is no profitable deviation from the behavior that we assume, in which $m_1 = \emptyset$ and $m_2 = a$. ■

It is also noteworthy that no equilibrium in pure strategies exists such that $m_1 \neq \emptyset$, at least for values of U^n that are large enough in the mechanism μ.[27] In fact, if the equilibrium strategy of the A's endogenous type $e \in E$ is $m_1 = a$, P's optimal response is to give full insurance to this type, leading A to pick $m_j = b$ as shown in Proposition 3. Also, if A picks $m_1 = b$, for large enough values of U^n, P optimally proposes a full insurance contract targeted only to the type (e, α), which means that by sending $m = b$, A expects the same payoff as in the absence of renegotiation. But then, as as shown in Proposition 3, $m_j = a$ is the A's unique optimal report. Thus, γ' provides an incentive to coordinate on equilibria in which m_j is sent after γ^r is posted and before ρ is taken.

References

Asseyer, Andreas, "Collusion and delegation under information control," *Theoretical Economics*, 2020, *15* (4), 1547–1586.

Attar, Andrea, Eloisa Campioni, Thomas Mariotti, and Alessandro Pavan, "Keeping Agents in the Dark: Private Disclosures in Competing Mechanisms," *Mimeo, Toulouse School of Economics*, May 2023.

Bester, Helmut and Roland Strausz, "Contracting with imperfect commitment and noisy communication," *Journal of Economic Theory*, 2007, *136* (1), 236–259.

Chade, Hector and Edward Schlee, "Optimal insurance with adverse selection," *Theoretical Economics*, 2012, *7* (3), 571–607.

Doval, Laura and Vasiliki Skreta, "Mechanism design with limited commitment," *Econometrica*, 2022, *90* (4), 1463–1500.

Fudenberg, Drew and Jean Tirole, "Moral hazard and renegotiation in agency contracts," *Econometrica*, 1990, *58* (6), 1279–1319.

[27]Indeed, our conjecture is that, if U^n is such that $V_H^{FI}(U^n) < \frac{1}{2}V_H^{FI}(2U^0 - U^n) + \frac{1}{2}V_H^{IC}(U^n)$ and $V_L^{FI}(U^n) < \frac{1}{2}V_L^{FI}(2U^0 - 2\varepsilon - U^n) + \frac{1}{2}V_L^{IC}(U^n)$, the principal obtains V^{SB} at any equilibrium of G_μ.

Georgiadis-Harris, Alkis, Thomas Brzustowski, and Balasz Szentes, "Smart Contracts and the Coase Conjecture," *American Economic Review*, 2023, *113(5)*, 1334–1359.

Holmström, Bengt, "Moral hazard and observability," *The Bell journal of economics*, 1979, pp. 74–91.

Lomys, Niccolo and Takuro Yamashita, "A Mediator Approach to Mechanism Design with Limited Commitment," Technical Report, Toulouse School of Economics 2022.

Myerson, Roger B, "Optimal coordination mechanisms in generalized principal-agent problems," *Journal of mathematical economics*, 1982, *10* (1), 67–81.

Ortner, Juan and Sylvain Chassang, "Making Corruption Harder: Asymmetric Information, Collusion, and Crime," *Journal of Political Economy*, 2018, *126* (5), 2108–2133.

Rahman, David and Ichiro Obara, "Mediated partnerships," *Econometrica*, 2010, *78* (1), 285–308.

Shavell, Steven, "Risk sharing and incentives in the principal and agent relationship," *The Bell Journal of Economics*, 1979, pp. 55–73.

Strausz, Roland, "Mediated contracts and mechanism design," *Journal of Economic Theory*, 2012, *147* (3), 1280–1290.

von Negenborn, Colin and Martin Pollrich, "Sweet lemons: Mitigating collusion in organizations," *Journal of Economic Theory*, 2020, *189* (C), 1280–1290.