WORKING PAPERS

N° 1522

February 2025

"Mediated Renegotiation"

Andrea Attar, Lorenzo Bozzoli, Roland Strausz

Toulouse School of Economics

Mediated Renegotiation

Andrea Attar, Lorenzo Bozzoli, Roland Strausz*

February 12, 2025

Abstract

We propose a novel approach to contract renegotiation with asymmetric information, introducing mediated mechanisms that generate additional private information to deter renegotiation. These mechanisms prevent any renegotiation, upholding second-best optimality as the unique equilibrium outcome. Thus, the inefficiencies typically associated with the threat of renegotiation are completely offset by the design of mediated mechanisms. We formally illustrate this result in the canonical framework of Fudenberg and Tirole (1990). We explicitly show that these mediated mechanisms can be decentralized by smart contracts, running on a public blockchain, which guarantees that our results do not require any trustworthy third party. (*JEL* D43, D82, D86)

*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: strauszr@hu-berlin.de). We thank Eloisa Campioni, Dino Gerardi, Johannes Horner, Fahad Khalil, Daniel Krähmer, Elliot Lipnowslki, Alessandro Pavan, Soenje Reiche, Francois Salanié, Steve Tadelis, and Takuro Yamashita for very thoughtful comments. We also thank seminar audiences at Berkeley University, Bonn University, Northwestern University, Università degli Studi di Roma "Tor Vergata", Toulouse School of Economics, Washington University, Yale University, as well as conference participants at the 2024 Conference on Mechanism and Institution Design (Budapest), at the 2024 Conference in honor of Françoise Forges, and at the 2024 Game Theory and Information Economics Conference (Rio de Janeiro) for many useful discussions. Andrea Attar acknowledges financial support from the Agence Nationale de la Recherche (ANR) (project ANR-23-CE26-0006), and from Ministero dell' Università e della Ricerca, (project PRIN-2022-PXE3B7). Roland Strausz acknowledges financial support from the European Union through the ERC-grant PRIVDIMA (project number 101096682) and the Deutsche Forschungsgemeinschaft through CRC-TRR 190 (project number 280092119).

1 Introduction

The threat of renegotiation is ubiquitous in contracting, embodying the problem of collective opportunism that inherently emerges when dealing with incentive problems. As first pointed out by Dewatripont (1989), this opportunism arises because contracts that optimally resolve incentive problems typically do so by implementing allocations that prove inefficient ex-post. Consequently, when contracting parties are unable to credibly commit to refraining from renegotiating away ex-post inefficiencies, they find themselves at a disadvantage from an ex-ante perspective.

In this paper, we present the novel insight that the threat of renegotiation can be fully mitigated by mediated mechanisms that carefully design the timing of communication. Specifically, mediated mechanisms uniquely implement the optimal second-best allocation that is implemented in the absence of any renegotiation.

The result indicates that the threat of renegotiation is a by-product of implicit restrictions on the set of feasible mechanisms, rather than of legal restrictions on the enforceability of contractual clauses to prevent contract modifications. Overcoming this threat does not require to introduce sophisticated mechanisms, nor to modify the standard extensive form of renegotiation games—it merely requires the mechanism to communicate the outcome of a random coin toss after it receives a private binary message about the occurrence of a renegotiation offer.

Our optimal mechanism eliminates renegotiation by making it prohibitively costly. It incentivizes a contracting party to privately report receiving a renegotiation offer before deciding whether to accept it. Upon receiving such a report, the mechanism randomly improves or worsens the prevailing offer and privately communicates this outcome. This binary lottery satisfies three properties: i) it induces truthful reporting; ii) it ensures renegotiation offers are accepted only if the lottery worsens the original proposal to the reporting party; iii) it makes the subsequent acceptance behavior prohibitively expensive to the proposing party in expected terms. Crucially, private communication prevents the renegotiating party from conditioning their offer on whether it was reported.

To derive our results formally, we take as a reference Fudenberg and Tirole's (1990) classical analysis of moral hazard with renegotiation: a principal incentivizes a risk-averse agent to provide effort. Because any ex-ante efficient contractual arrangement imposes risk on the agent, the problem of collective opportunism arises once effort is sunk. Hence, contracting parties are susceptible to renegotiation after effort has been chosen.

The impossibility to prevent renegotiation is often seen as reflecting a conflict between

economic efficiency and legal doctrine. Courts generally refuse to enforce no-renegotiation clauses, viewing them as violations of the freedom of contract principle.[1] Our result reconciles this conflict: while freedom of contract precludes direct prevention of renegotiation, the legal doctrine of duress—which protects parties' ability to seek advice before signing modified contracts—provides the tools needed for our mechanism to work.

A critique leveled against mediated mechanisms is their dependence on a trustworthy third party, as they generally require the exchange of private messages back and forth between the parties and the mechanism. We negate this critique by implementing the mechanism indirectly via smart contracts using off-the-shelf cryptographic techniques. We show explicitly how these allow fully transparent, verifiable implementation that is robust against manipulation while preserving the required private communication.

Related Literature. Our work contributes to the literature on contract renegotiation, which, starting with Dewatripont (1989), focuses on optimal, renegotiation-proof contracts. While Fudenberg and Tirole (1990) study renegotiation under moral hazard, the threat of renegotiation has been also analyzed under incomplete information (e.g., Hart and Tirole, 1988; Laffont and Tirole, 1990).

Bolton (1990) points out that, regardless of the specific informational assumptions, optimally preventing renegotiation requires to introduce (or to maintain) some degree of private information at the renegotiation stage. When mechanisms are not mediated, this private information can only be generated by the agent, requiring her to mix between her reports and/or efforts at equilibrium. Such a random behavior implies allocative costs, which makes it impossible to attain the second-best outcome that obtains in the absence of renegotiation. We construct, instead, a mediated mechanism which generates private information for the agent, without imposing any random behavior. This allows to successfully prevent any renegotiation, and to uniquely implement the second-best allocation. As we discuss in detail in Section 5, these insights are general and extend to a wide range of renegotiation settings, beyond the moral hazard one in Fudenberg and Tirole (1990).

In traditional approaches to mechanism design (Myerson, 1982, 1986; Forges, 1986), mediated mechanisms play a welfare-enhancing role, allowing to correlate agents' independent actions. Rahman and Obara (2010) provide an instance of this effect in a partnership

[1]For instance, the US Code on contract law under Title 42,§1981 declares the right of all persons to "the making, performance, modification, and termination of contracts". Jolls (1997) and Davis (2006) cite multiple applications of this code voiding contractual clauses limiting collective renegotiation. A notable example is Beatty v. Guggenheim Exploration Co. 225 N.Y. 380, 1919, where in his judgement Justice Cardozo voided an explicit contractual clause forbidding future modification stating that "Those who make a contract, may unmake it. The clause which forbids a change, may be changed like any other."

framework, and show how mediated mechanisms make possible to reconcile individual incentives with budget balance by *virtually* implementing an efficient allocation. We emphasize a different role of mediated mechanisms: by endogenously generating some private information, they allow to delegate to the agent the punishment against any subsequent renegotiation. This *off-equilibrium* effect guarantees a full rather than virtual implementation of the second-best allocation.

The idea that a principal designs a mediated mechanism with the aim of reducing the deviations available to his future self has been recently considered in the limited-commitment literature. Bester and Strausz (2007), Doval and Skreta (2022), and Lomys and Yamashita (2022) exploit this channel of communication to formulate a revelation principle. Rather than fully characterizing the set of available communication mechanisms, our objective is to construct a specific mechanism that fully mitigates the threat of renegotiation.

Brzustowski et al. (2023) and Doval and Skreta (2024) focus on a specific equilibrium allocation and analyze its implications for the Coase conjecture. Yet, these papers obtain mediated mechanisms that do not achieve second-best efficiency, while no uniqueness result is established. We consider instead a renegotiation framework, in which, until both parties agree on a new offer, the agent retains access to the options incorporated in the original mechanism, since, unlike under limited-commitment, the initial contract can only be voided under the mutual consent of its participants. In this context, we are the first to show the power of mechanisms which send private signals to the agent. We show that exploiting this channel of communication allows to fully mitigate the threat of renegotiation regardless of any equilibrium selection criterion that goes beyond the standard concept of Perfect Bayesian equilibrium.

Because one can also frame the renegotiation problem as one of a designer competing with its myopic future self to incentivize a common agent, our results are also connected to those of the common agency literature. Indeed, the random offer that, in our construction, successfully prevents renegotiation plays a role similar to that of *latent* contracts under common agency. These contracts, which are offered but not traded in equilibrium, are key to deter deviations and support additional equilibrium allocations both under moral hazard (Bisin and Guaitoli, 2004; Attar et al., 2019) and under adverse selection (Attar et al., 2011, 2022).

Finally, our paper contributes to a nascent literature that studies the concrete use of smart contract and blockchain technologies as a practical device to implement specific allocation

mechanisms.[2] It sees the main advantage of such implementation for situations in which there are incentives to manipulate the mechanism, as, for instance, in Akbarpour and Li (2020). Concrete implementations via smart contracts have been mainly developed in the context of auction mechanisms (see Omar et al., 2021; Roughgarden, 2021, among others). More recently, Brzustowski et al. (2023) point out the ability of smart contracts to indirectly implement mechanisms that receive private messages, but that do not send any signal to the contracting parties. We add to this the insight that smart contracts can also indirectly implement mechanisms which allow for private communication in the (more involved) reverse direction. Thus, we provide the second step for showing how, based on current technologies, mediated mechanisms can be fully implemented by smart contracts, which dispense with the need of a mediator or third party, and thus eliminate any risk of manipulation.

The remaining of the paper is organized as follows. In Section 2, we illustrate the Fudenberg and Tirole (1990) benchmark model. In Section 3, we introduce mediated mechanisms, and analyze the corresponding renegotiation game. Section 4 presents the implementation via smart contracts. In Section 5, we discuss how our insights apply to alternative renegotiation settings. Most proofs are in the Appendix.

2 The Benchmark

We consider the canonical framework of Fudenberg and Tirole (1990) (FT, henceforth), in which a risk-neutral principal (he) incentivizes a risk-averse agent (she), who takes an unobservable effort. There are two outputs (states), a good one g and a bad one b, where $g > b > 0$. 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 expected output $Y_e \doteq p_e g + (1 - p_e)b$.

Payoffs and Allocations. The agent's utility is additively separable in income $w \in \mathbb{R}$ and effort $e \in E$, so that we 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 is unbounded over its domain, i.e., $\lim_{w \to -\infty} U(w) = -\infty$ and $\lim_{w \to \infty} U(w) = \infty$. Consequently, the inverse Φ of u is well–defined for any $u \in \mathbb{R}$ with $\Phi'(u) > 0$ and $\Phi''(u) > 0$. The low effort cost is normalized to $D(e = L) = 0$ and $D(e = H) = d > 0$.

For any $e \in E$, final payoffs are determined by the state-contingent transfers that the principal makes to the agent. A contract is a pair $(w_g, w_b) \in \mathbb{R}^2$ of such transfers. Because it

[2] See Chapter 6 in Townsend (2020) for a recent overview.

is often more convenient to represent a contract in terms of the induced utilities it provides to the agent, 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.[3]

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 the agent is risk-averse, while the principal is risk-neutral, any Pareto-efficient allocation exhibits full insurance. For any $e \in E$, let $c_e^{FI}(U) \doteq (U + D(e), U + D(e))$ denote the full-insurance contract that yields the agent 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

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

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

Because the principal has all bargaining power, the optimal contract with observable effort implements the efficient allocation that yields the principal his maximal payoff while still guaranteeing the agent her outside option U^0. We refer to this outcome as the first-best. Thus, the first-best contract is $c^{FB} \doteq c_H^{FI}(U^0)$, yielding $V^{FB} \doteq V_H^{FI}(U^0)$ to the principal, and $U = U^0$ to the agent.[4]

If, instead, effort is unobservable, any feasible allocation must be incentive-compatible. Then, the optimal contract for the principal induces $e = H$ and gives at least U^0 to the agent.

We refer to this contract as the second-best; it solves:

$$\begin{aligned} \arg\max_{c \in \mathbb{R}^2} \quad & V_H(c) = p_H(g - \Phi(U_g)) + (1 - p_H)(b - \Phi(U_b)) && (1) \\ \text{s.t.} \quad & p_H U_g + (1 - p_H)U_b - d \geq p_L U_g + (1 - p_L)U_b. && (2) \\ & p_H U_g + (1 - p_H)U_b - d \geq U^0. && (3) \end{aligned}$$

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

[4] 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.

At a solution, incentive constraint (2) binds. Accordingly, let $c^{IC}(U) \doteq (U_g^{IC}(U), U_b^{IC}(U))$ denote the contract on the incentive-compatibility frontier leaving expected payoff U to A:

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

It is convenient to define, for each $e \in E$, the function $V_e^{IC} : \mathbb{R} \to \mathbb{R}$, which denotes the principal's payoff when the agent takes $e \in E$ and $c^{IC}(U)$ is implemented:

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

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

The Renegotiation Game. FT point out that the second-best allocation (H, c^{SB}) is *interim* inefficient, i.e., after effort is chosen but before output is realized. This leads FT to analyze an extension G^r of the game G, in which the principal can renegotiate away any such inefficiency by offering a new contract at the *interim* stage. Because, at this stage, the agent is privately informed about her effort choice, FT let the principal design a revelation mechanism, which specifies a contract for each effort announced by the agent. Specifically, a (deterministic) revelation mechanism is a mapping $\gamma_c : E \to \mathbb{R}^2$, and C denotes the set of all such mechanisms. The game G^r unfolds as follows:

1. The principal offers a mechanism $\gamma_c \in C$.
2. The agent accepts or rejects γ_c. If the agent rejects, the game ends and outside options accrue. If the agent accepts, the game continues as follows:
3. The agent chooses $e = H$ with probability $x \in [0,1]$ and $e = L$ with probability $1 - x$.
4. Without observing e, the principal makes a renegotiation offer $\gamma_c^r \in C \cup \{\emptyset\}$, where $\emptyset$ represents the principal's decision not to renegotiate.
5. The agent 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 follow from $\gamma_c^r(m)$.

Any mechanism γ_c that the agent accepts yields a subgame $G^r(\gamma_c)$ as of stage 3. In any such subgame, choosing $x = 1$ is not part of an equilibrium. To see this, suppose

the agent takes $e = H$ with probability one. Then, the principal's best reply is to offer the full-insurance contract $c_H^{FI}(U^0)$ in stage 4 that is accepted by the agent. But against this renegotiation offer, the agent 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, the agent takes $e = H$ with probability $x^{FT} < 1$.

In their analysis, FT restrict attention to revelation mechanisms, which, by construction, do not incorporate any *private* communication with the agent.[5] We next argue that this restriction is critical. Specifically, we show that if the principal can use mediated mechanisms that allow private communication with the agent, then the *unique* equilibrium allocation remains (H, c^{SB}).

3 Mediated Mechanisms

We frame contract renegotiation in the dynamic mechanism design tradition as developed by Forges (1986) and Myerson (1986). Thus, to control the moral hazard problem and the threat of renegotiation, the principal has, in addition to setting the conditional transfers, the power to select a *communication protocol*. The communication protocol does not only prescribe the available set of messages, but also the specific timing of communication between parties.

To demonstrate that the threat of renegotiation can be fully mitigated by design, it suffices to exhibit a communication protocol, which the principal can select to achieve the second-best allocation as the unique equilibrium one.[6] Specifically, the protocol we consider requires the agent to report in a mechanism only once, right after the renegotiation offer has been made, straight after which, the mechanism sends back a signal to the agent. Only after this communication phase between the agent and the mechanism, the agent makes the decision to accept or reject the principal's renegotiation offer.[7]

We let Γ represent the class of feasible mechanisms, which is such that the timing of communication is fixed according to the communication protocol described in the previous

[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] This exercise differs from that of characterizing a canonical structure of communication, to establish a version of the revelation principle, which is at the centre-stage of the Forges (1986) and Myerson (1986) perspective.

[7] As discussed in the introduction, the legal doctrine of duress, protecting a party's ability to seek advice before signing a new or modified contract, effectively grants the principal de-jure authority to mandate this communication phase.

paragraph, but all other aspects of the mechanism can be determined at the principal's discretion. A *mediated* mechanism $\gamma = \{\mathcal{M}, \mathcal{S}, \sigma, \tau\}$ in the set Γ consists of a (finite) set of messages $\mathcal{M}$ sent privately from the agent, a finite set of signals $\mathcal{S}$ privately received by the agent 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 utility pair (U_H, U_L) to any combination of messages and signals. It is *mediated* in the sense that the transfers are made contingent on the private communications exchanged between the agent and the mechanism itself, thereby leaving the principal in the dark at the renegotiation stage.

3.1 The Mediated Renegotiation Game

The principal's fixed timing allows us to formulate the *mediated renegotiation game* G_Γ. In this game, the principal can offer any mechanism in Γ, but is susceptible to the threat of renegotiation considered by FT; at the renegotiation stage, he may offer any other mechanism to the agent as an alternative. The mediated renegotiation game G_Γ itself unfolds as follows:

1. The principal offers a mechanism $\gamma \in \Gamma$.
2. The agent accepts or rejects γ. If the agent rejects, the game ends and outside options accrue. If the agent accepts, the game continues as follows:
3. The agent chooses $e = H$ with probability $x \in [0, 1]$ and $e = L$ with probability $1 - x$.
4. Without observing e, the principal makes a renegotiation offer $\gamma^r = \{\mathcal{M}^r, \mathcal{S}^r, \sigma^r, \tau^r\} \in \Gamma \cup \{\emptyset\}$, where $\emptyset$ represents the principal's decision not to renegotiate.
5. The agent sends a private message $m \in \mathcal{M}$ in the initial mechanism γ.
6. The mechanism γ extracts private signal $s \in \mathcal{S}$ according to distribution $\sigma(m) \in \Delta(\mathcal{S})$.
7. After privately observing s, the agent accepts or rejects γ^r by declaring $\rho \in \{y, n\}$.
8. If $\rho = n$, transfers occur according to $\gamma(m, s)$, after message m and realized signal s are publicly revealed. If $\rho = y$, the agent sends $m^r \in \mathcal{M}^r$, receives a private signal $s^r \in \mathcal{S}^r$, and transfers occur according to $\gamma^r(m^r, s^r)$, after message m^r and realized signal s^r are publicly revealed.

A (pure) strategy for the principal in G_Γ is a mechanism γ, followed by a renegotiation offer γ^r. The agent's (behavioral) strategy in G_Γ, which we denote λ, involves three parts. First, it associates to any $\gamma \in \Gamma$ a probability distribution over efforts. Further, for

any history (e, γ^r), λ specifies a probability distribution over 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 messages in γ^r, at any history (e, γ^r, m, s, y) such that $\gamma^r \neq \{\emptyset\}$.

In line with FT, we take perfect Bayesian equilibrium as our solution concept. We denote $G_\Gamma(\gamma)$ the subgame induced by $\gamma \in \Gamma$ as of stage 3. We let $\lambda(\gamma)$ represent the agent's (continuation) strategy in $G_\Gamma(\gamma)$, while the principal's strategy is a renegotiated mechanism $\gamma^r \in \Gamma$. 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, the principal chooses an optimal mechanism γ, given that the players continuation strategies constitute a perfect Bayesian equilibrium (henceforth equilibrium) of $G_\Gamma(\gamma)$.

Remark 1 *The sequence of physical decisions, including the participation ones, in G_Γ coincides with that considered in FT. Indeed, for $|\mathcal{M}| = 2$ and $|\mathcal{S}| = 1$, mechanisms in Γ reduce to mechanisms in C as defined in FT. Moreover, comparing the subgame $G_\Gamma(\gamma)$ to the subgame $G^r(\gamma_c)$ in FT reveals that a mediated mechanism transforms the simultaneous decision of the agent about her message m and acceptance decision ρ into a sequential one, comprising three substages: first, the agent sends message m, she then privately observes random signal s, and finally accepts or rejects the renegotiation offer. Crucially, the agent receives her (private) signal s before deciding about accepting the renegotiation offer.*

Remark 2 *The game G_Γ also preserves the commitment assumptions outlined in FT. In particular, both the communication and the decision rule of the original mechanism γ are halted mid-execution if and only if the parties mutually agree on a renegotiated mechanism γ^r. This in turn implies that γ^r may affect the relevant communication protocol only after the agent has accepted to renegotiate. So even though we allow the renegotiation offer γ^r to be also mediated, the communication protocol of any optimal renegotiation offer will be trivial and an optimal γ^r is, just as in FT, a menu of insurance contracts.*

3.2 Implementing the Second Best

We next identify a mediated mechanism $\gamma^* \in \Gamma$ that fully mitigates the threat of renegotiation. That is, we show that *in any* equilibrium of G_Γ, the principal obtains the second-best payoff V^{SB}, the agent picks $e = H$ with probability one, and there is no incentive for renegotiation. Specifically, we first exhibit a mechanism $\gamma^* \in \Gamma$ which implements the

second-best allocation, and then show that this allocation is the unique equilibrium outcome of the overall game G_Γ.

To fully mitigate the threat of renegotiation, the mechanism is to impose a costly punishment in expected terms on the principal when he makes a "relevant" renegotiation offer to the agent. The agent is to trigger this punishment by truthfully reporting to the mechanism such a renegotiation attempt. Upon receiving this report, the mechanism is then to improve or worsen contracting terms randomly so that the agent rejects the renegotiation offer when, from the principal's perspective, it worsens his contracting terms. Hence, the punishment itself must be carefully calibrated so that, on the one hand, it is incentive compatible for the agent to report truthfully, while, on the other hand, it ensures that the punishment on the principal deters him from making such an offer in the first place.

While there are many ways to implement this idea, the concrete way we do so is by setting up the original contract so that if the agent reports about the renegotiation attempt, the mechanism simply randomizes between demanding the principal to provide an extra utility of ΔU to the agent or by demanding that the agent is compensated by a utility of ΔU less. Randomizing between these two options with equal probability, ensures that, for any ΔU, it is incentive compatible for the agent not to trigger the random change if the principal does not make a renegotiation offer.

On the other hand, when ΔU is large enough, the agent will indeed reject the renegotiation offer when the outcome of the lottery demands the principal to provide her with ΔU utility more. As intended, this rejection acts as a punishment to the principal, which is triggered with probability 1/2. Consequently, the principal is better off refraining from making a renegotiation offer when ΔU is large enough.

The next lemma confirms this formally and is hence key for establishing our main result.

Lemma 1 *There exists $\Delta U \in (0, \infty)$ such that for all $e \in E$:*

$$V_e^{IC}(U^0) > \max\left\{V_e^{FI}(U^0 + \Delta U),\ \frac{1}{2}V_e^{FI}(U^0 - \Delta U) + \frac{1}{2}V_e^{IC}(U^0 + \Delta U)\right\}. \tag{4}$$

The lemma states that, for any $e \in E$, the principal prefers the second-best contract, $c^{SB} = c^{IC}(U^0)$, to a full-insurance contract that leaves an extra utility of ΔU to the agent. Additionally, the principal prefers c^{SB} to a 50-50 lottery between the full-insurance contract leaving ΔU less to the agent, and the incentive-compatible one leaving the agent an extra utility ΔU. This in turn allows to construct a mediated mechanism γ^* such that the principal attains the lefthand side of (4) when he does not renegotiate, while the righthand side of (4)

corresponds to what the principal expects from the best possible renegotiation offer that the agent accepts with a strict positive probability.

Let $\gamma^* = \{\mathcal{M}^*, \mathcal{S}^*, \sigma^*, \tau^*\}$ be such that $\mathcal{M}^* = \{N, R\}$ and $\mathcal{S}^* = \{h, t\}$. The signals are extracted according to $\sigma^* : \mathcal{M}^* \to \Delta(\mathcal{S}^*)$ with:

$$\sigma^*(h|m) = \sigma^*(t|m) = \frac{1}{2} \qquad \text{for each } m \in \{N, R\}.$$

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

$$\tau^*(N,h) = \tau^*(N,t) = c^{SB}; \quad \tau^*(R,h) = c^{IC}\left(U^0 - \Delta U\right); \quad \tau^*(R,t) = c^{IC}(U^0 + \Delta U).$$

Intuitively, the mediated mechanism sets the second best contract c^{SB} as the "default" but allows the agent to trigger a random "counter-offer" by sending $m = R$, indicating that the principal made a renegotiation offer. The counter-offer gives the agent either an extra utility of ΔU or a utility of ΔU less. On average, it yields the agent the same utility as the default c^{SB}, but is more expensive in terms of wages. To the principal, the counter-offer is random, whereas to the agent, the realized counter-offer is revealed after sending $m = R$ but before she has to decide about the principal's renegotiation offer.

Hence, γ^* shares with mechanisms in FT the restriction to only two messages for the agent. By contrast, it extends on FT by selecting one of two signals with equal probability and privately disclosing it to the agent. Although, in general, the distribution of the signal may depend on the message m, the specific mediated mechanism γ^* does not exploit this feature. Effectively, the signal represents a 50-50 coin toss about whose realization—head or tail—A is informed privately.

Proposition 1 *The second-best allocation* (H, c^{SB}) *is supported in an equilibrium of the subgame* $G_\Gamma(\gamma^*)$.

Since the principal cannot obtain more in a game with renegotiation than without any renegotiation, Proposition 1 implies that the game G_Γ 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 1, observe that, by reporting $m = N$ in γ^*, the agent gets the second-best contract c^{SB}, which makes $e = H$ an optimal choice. In the absence of renegotiation, this yields U^0 to the agent and V^{SB} to the principal. Hence, it suffices to exhibit a profile of continuation strategies that support these behaviors in an equilibrium of $G_\Gamma(\gamma^*)$.

Let $m_e^r \in \mathcal{M}^r$ denote an agent's optimal message when she accepts renegotiation offer γ^r having chosen an effort $e \in E$. In addition, let $\hat{U}_e^r$ denote her corresponding payoff. That is:

$$m_e^r \in \arg\max_{m \in \mathcal{M}^r} \sum_{s \in \mathcal{S}^r} \sigma^r(s|m) U_e(\tau^r(m,s)) \quad \text{and} \quad \hat{U}_e^r = \sum_{s \in \mathcal{S}^r} \sigma^r(s|m_e^r) U_e(\tau^r(m_e^r,s)). \tag{5}$$

By construction, sending m_e^r is sequentially rational for the agent 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 $G_\Gamma(\gamma^*)$. The principal's strategy is not to renegotiate, i.e. $\gamma^r(\gamma^*) = \{\emptyset\}$, while the agent's strategy $\lambda(\gamma^*)$ is as follows:

1. The agent chooses $e = H$ with probability one.

2. Her messages in γ^* and subsequent participation decisions in γ^r, depend on the history (e, γ^r) as follows:

   (i) For any $e \in E$, if $\gamma^r = \{\emptyset\}$, and for any γ^r such that $\hat{U}_e^r \leq U^0 - \Delta U$, the agent sends $m = N$ in γ^*, followed by $\rho = n$.

   (ii) For any $e \in E$ and any $\gamma^r \neq \{\emptyset\}$ such that $\hat{U}_e^r \in (U^0 - \Delta U, U^0 + \Delta U]$, the agent sends $m = R$ in γ^*, followed by $\rho = y$ when $s = h$, and by $\rho = n$ when $s = t$.

   (iii) For any $e \in E$ and any $\gamma^r \neq \{\emptyset\}$ such that $U_e^r > U^0 + \Delta U$, the agent sends $m = R$ in γ^*, followed by $\rho = y$ for any $s \in \{h, t\}$.

3. For any history $(e, \gamma^r \neq \{\emptyset\}, m, s, y)$, the agent sends m_e^r.

We show that the strategies $\{\lambda(\gamma^*), \gamma^r(\gamma^*)\}$, together with the principal's belief that the agent picked $e = H$ with probability $x = 1$, constitute an equilibrium of $G_\Gamma(\gamma^*)$.

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

We develop our argument in two lemmas. First, we characterize the agent's equilibrium behavior in $G_\Gamma(\gamma^*)$.

Lemma 2 *The agent's strategy $\lambda(\gamma^*)$ is sequentially rational.*

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

$$\max\{U^0, \hat{U}_e^r\} \geq \frac{1}{2} \max\left\{U^0 - \Delta U, \hat{U}_e^r\right\} + \frac{1}{2} \max\{U^0 + \Delta U, \hat{U}_e^r\}, \tag{6}$$

where $\hat{U}_e^r$ is defined in (5). The left(right)-hand side of (6) is her continuation payoff after sending $m = N(R)$. The following holds:

(i) If $\hat{U}_e^r \leq U^0 - \Delta U \vee \gamma^r = \{\emptyset\}$, then (6) is satisfied because it reduces to $U^0 \geq U^0$ since $\hat{U}_e^r \leq U^0 - \Delta U < U^0$. Sending $m = N$ in γ^*, followed by $\rho = n$, as prescribed by $\lambda(\gamma^*)$, is hence optimal.

(ii) If $\hat{U}_e^r \in (U^0 - \Delta U, U^0 + \Delta U]$, then, upon sending $m = R$, it is optimal for the agent to choose $\rho = y$ when $s = h$ (as rejection leads to $U^0 - \Delta U < \hat{U}_e^r$), and $\rho = n$ when $s = t$ (as rejection leads to $U^0 + \Delta U \geq \hat{U}_e^r$). We next argue that sending $m = R$ in γ^*, as prescribed by $\lambda(\gamma^*)$, is optimal. That is, the sign of the inequality in (6) is reversed, where we note that, due to $\hat{U}_e^r \in (U^0 - \Delta U, U^0 + \Delta U]$, its RHS reduces to $\hat{U}_e^r/2 + (U^0 + \Delta U)/2$. Hence, we only need to show that

$$\max\{U^0, \hat{U}_e^r\} \leq \frac{1}{2}\hat{U}_e^r + \frac{1}{2}(U^0 + \Delta U). \tag{7}$$

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

(a) If $\hat{U}_e^r < U^0$, then (7) rewrites as $U^0 - \Delta U \leq \hat{U}_e^r$, which is satisfied by assumption.

(b) If $\hat{U}_e^r \geq U^0$, then (7) rewrites as $\hat{U}_e^r \leq U^0 + \Delta U$, which is satisfied by assumption.

(iii) If $\hat{U}_e^r \in (U^0 + \Delta U, \infty)$, then we have $U^0 < U^0 + \Delta U < \hat{U}_e^r$, implying that the agent is indifferent between $m = N$ and $m = R$, followed by $\rho = y$ for any $s \in \{h, t\}$. In particular, as prescribed by $\lambda(\gamma^*)$, sending $m = R$ in γ^*, and then accepting to participate in γ^r for any received signal is optimal. ■

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

Lemma 3 *P's strategy $\gamma^r(\gamma^*) = \{\emptyset\}$ is a best response to his (Bayes-consistent) belief $x = 1$, and to the agent's strategy $\lambda(\gamma^*)$.*

Proof. We first argue that the principal can improve on any renegotiation that does not achieve full insurance to the agent. Indeed, since the agent is risk-averse and the principal holds a degenerate belief $x = 1$, any optimal renegotiation offer γ^r involves full insurance, i.e., for any $(m, s) \in \mathcal{M}^r \times \mathcal{S}^r$, there exists $U_H^r(m, s)$ such that $\tau^r(m, s) = (U_H^r(m, s), U_H^r(m, s))$, with the interpretation that $\tau^r(m, s)$ yields the payoff $U_H^r(m, s)$ to the agent when she picks $e = H$, regardless of the output level.

Since any renegotiation that does condition transfers non-trivially on the agent's private signal implies that the agent is not fully insured, we only need to consider offers γ^r such that $\mathcal{S}^r = \{s^r\}$. But then there is also no loss in considering offers such that $\mathcal{M}^r$ is a singleton, as the principal correctly anticipates that for any $|\mathcal{M}^r| > 1$, the agent randomizes over messages $m^r \in \arg\max_{m \in \mathcal{M}^r} U_H^r(m, s^r)$, implying that the principal can do just as well by letting $\mathcal{M}^r = \{m^r\}$ and implementing his preferred full-insurance contract. Thus, there is no loss in assuming $\mathcal{S}^r = \{s^r\}$, $\mathcal{M}^r = \{m^r\}$, and $\tau^r(m^r, s^r) = (U_H^r(m^r, s^r), U_H^r(m^r, s^r))$, which implies that any γ^r can be characterized by the payoff $\hat{U}^r = U_H^r(m^r, s^r) \in \mathbb{R}$ that the renegotiation offer leaves to the agent when she picks $e = H$. We next verify that, for any $\hat{U}^r \in \mathbb{R}$, the principal's expected payoff does not exceed $V^{SB} = V_H^{IC}(U^0)$, his utility when not renegotiating. We distinguish three cases:

(i) If $\hat{U}^r \leq U^0 - \Delta U$ then $\lambda(\gamma^*)$ prescribes $(m = N, \rho = n)$ and the principal gets V^{SB}.

(ii) If $\hat{U}^r \in (U^0 - \Delta U, U^0 + \Delta U]$ then $\lambda(\gamma^*)$ prescribes ($m = R$, $\rho = y$ when $s = h$, and $\rho = n$ when $s = t$), and the principal gets

$$\frac{1}{2}V_H^{FI}(\hat{U}^r) + \frac{1}{2}V_H^{IC}(U^0 + \Delta U) < \frac{1}{2}V_H^{FI}\left(U^0 - \Delta U\right) + \frac{1}{2}V_H^{IC}(U^0 + \Delta U) < V^{SB}, \quad (8)$$

where the first inequality follows from V_H^{FI} decreasing, and the second from Lemma 1.

(iii) If $\hat{U}^r > U^0 + \Delta U$ then $\lambda(\gamma^*)$ prescribes ($m = R$, $\rho = y$) for any received signal, and the principal gets

$$V_H^{FI}(\hat{U}^r) < V_H^{FI}(U^0 + \Delta U) < V^{SB} \quad (9)$$

where the first inequality follows from V_H^{FI} decreasing, and the second from Lemma 1.

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

We complete the proof of Proposition 1 by considering the agent's effort choice. Given the principal's strategy $\gamma^r(\gamma^*) = \{\emptyset\}$, this is straightforward, as the agent 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 1, strategies $\{\lambda(\gamma^*), \gamma^r(\gamma^*)\}$ together with the principal's belief $x = 1$ at his non-trivial information set, form a perfect Bayesian equilibrium of $G_\Gamma(\gamma^*)$, in which the agent chooses $e = H$ with probability one, and the principal obtains V^{SB}.

The presented verification of Proposition 1 proves that the mediated mechanism γ^* makes *any* renegotiation unprofitable to the principal. It is, in particular, instructive to consider the agent's behavior towards the renegotiation offer $\gamma^{FI}(U^0)$ that in FT undermines any

equilibrium in which the agent picks $e = H$ with probability 1. Indeed, the mediated mechanism γ^* induces the agent to reject $\gamma^{FI}(U^0)$ after receiving $s = t$, which occurs with probability 1/2. The principal's anticipation of this random rejection makes these offers unprofitable to him, since, in case of rejection, he faces highly unfavorable contracting terms.

Random vs. Mediated Mechanisms. Because γ^* conditions the final transfers on the random signal s, it effectively induces a random contract. It is therefore 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.

We claim that the answer is negative. To see this intuitively, note that in line with Chade and Schlee (2012), the optimal renegotiation offer against any probability distribution over the agent's efforts is *deterministic*. Then, simple backward-induction reasoning guarantees that the principal cannot gain by committing ex-ante to a random mechanism.[8]

At this point, it is important to clarify that a non-degenerate random mechanism is random for both the principal and the agent, in the sense that neither party can condition any of their decisions on the realization of the randomness. By contrast, our mediated mechanism γ^* is random for the principal, but not for the agent, because the agent can condition her choice whether to accept some renegotiation offer on the realization of the random component. Indeed, γ^* crucially exploits this feature to generate the agent's private information.

An Alternative Communication Protocol. We derived our result by focusing on the principal picking a specific communication protocol, requiring the agent to send a message and receive a signal after obtaining a renegotiation offer but before making her acceptance decision. In this protocol, the agent's communication in the original mechanism cannot be prevented by the principal when renegotiating. We see the protocol as reflecting the legal doctrine of duress, protecting parties' ability to seek advice before signing modified contracts.

Yet, we point out that this specific protocol is not strictly necessary to our implementation. Indeed, an alternative protocol that also implements the second best is one that allows the agent to actually choose *at which stage* to communicate in the original mechanism. Such a protocol leads to a larger game, but at any equilibrium of this enlarged game, the agent chooses to send messages, and, subsequently, to receive signals *only* after a renegotiation offer γ^r is made, so to make an efficient use of her information. Thus, although the principal may want to offer γ^r *after* the agent communicates in the original mechanism γ^*, the agent has instead an incentive to postpone the timing of her communication, so to activate her

[8]The result is formally established in Appendix B.

preferred option in γ^*. As a consequence, we are able to extend Proposition 1 to such a richer strategic scenario.[9] Indeed, any restriction on the agent's freedom to communicate with third parties before agreeing to a renegotiation offer could be seen as a form of contractual duress, which violates contract law.

3.3 Unique Implementation of the Second Best

Proposition 1 shows that the mediated mechanism γ^* induces a subgame supporting the second-best allocation at equilibrium. Because this outcome yields to the agent the payoff U^0, it is also incentive-compatible for her to accept γ^* at stage 2, as she cannot strictly gain by rejecting it. Moreover, the principal 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 (continuation) 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 the agent 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 $G_\Gamma(\gamma^*)$. In particular, any $x \in [0,1]$ together with the principal not renegotiating (i.e. $\gamma^r = \{\emptyset\}$) can be supported by an equilibrium of $G_\Gamma(\gamma^*)$. Consequently, γ^* does not *uniquely* implement the second-best allocation. Yet, although the subgame $G_\Gamma(\gamma^*)$ admits multiple equilibrium allocations, the next proposition shows that only the second-best one is supported at equilibrium in the overall game G_Γ.

Proposition 2 *The renegotiation game G_Γ has a unique equilibrium allocation, which coincides with the second-best one (H, c^{SB}).*

The proof of Proposition 2 constructs a mechanism γ_ε by perturbing γ^* in such a way that, in the subgame $G_\Gamma(\gamma_\varepsilon)$, for any belief $x \in [0,1]$, choosing not to renegotiate is the unique best response of the principal to any sequentially rational behavior of the agent. This, in turn, guarantees that $e = H$ is the unique optimal choice.

[9] The result is formally established in Appendix B.

4 A Decentralized Smart Contract Implementation

Our analysis provides the insight that the inefficiencies typically associated with the threat of renegotiation result from restrictions made on the set of feasible mechanisms. Such restrictions may however be reasonable if the mechanisms that fully mitigate the threat of renegotiation require an excessive degree of complexity. Indeed, a common critique leveled against mediated mechanisms is that, in contrast to non-mediated ones, they often require a trustworthy third party for their implementation. We negate this critique by explicitly demonstrating how existing blockchain technologies allow to implement our efficient outcome via a decentralized smart contract, de-facto dispensing with the need for any third party.[10] In so doing, we show that the approach suggested by Brzustowski et al. (2023) extends to a large class of mediated mechanisms.

To develop our analysis, we start with acknowledging an important limitation of current technologies: smart contracts are currently unable to send private signals to players.[11] This prevents a direct implementation of our mechanism γ^*: if signal s were public, then, for instance, the *signal-conditional* renegotiation offer $\gamma^r = (\gamma^r_t, \gamma^r_h) = (c^{FI}(U^0 + \varepsilon), c^{FI}(-\infty))$, with $\varepsilon > 0$ small, undermines γ^* to implement the second-best allocation. This is so, because this conditional offer triggers the agent's message $m = N$, and her acceptance of γ^r in case $s = t$, yielding the principal a myopic gain and thus invalidating the second-best outcome.

Preventing Renegotiation with Public Signals. We next show that the privacy of the signal s can be dispensed with by appropriately adapting the mechanism γ^*. Consider, in particular, the modified mechanism γ^{**} with *three* private messages, i.e. $\mathcal{M}^{**} = \{N, R_1, R_2\}$, and two *public* signals, i.e. $\mathcal{S}^{**} = \{h, t\}$, extracted with probability $1/2$ each. Let the decision

[10] While smart contracts are technically immutable at the code level, this property alone does not make them renegotiation-proof from a contract theoretical perspective. A crucial vulnerability arises when smart contracts rely on external accounts for executing conditional transfers, as these accounts must maintain sufficient funds for the contract's execution. This makes such contract directly susceptible to renegotiation: contracting parties can effectively trigger a renegotiation by deliberately depleting the external accounts that the smart contract depends on for execution. Although the technical immutability of smart contracts could theoretically prevent this through complete internalization of all funds, such a solution would require each party to lock up, at the contract's initiation, sufficient capital to cover any possible conditional transfer, both on- and off-path, i.e., including any funds needed for indirect enforcement. The capital costs of such full internalization would typically be prohibitive, as these funds would need to remain locked for the contract's entire duration, creating substantial opportunity costs.

[11] Brzustowski et al. (2023) can sidestep this issue, because they consider mechanisms that do not require sending private signals. Indeed, the ability for smart contracts to send signals privately, is currently under active development. See for instance, work on fully homomorphic encryption such as Zama-AI on Github.

rule τ^{**} be

$$\begin{aligned}
&\tau^{**}(N,h) = \tau^{**}(N,t) = c^{IC}(U^0) = c^{SB};\\
&\tau^{**}(R_1,t) = c^{IC}\left(U^0 + \Delta U\right); \quad \tau^{**}(R_1,h) = c^{IC}(U^0 - \Delta U);\\
&\tau^{**}(R_2,t) = c^{IC}\left(U^0 - \Delta U\right); \quad \tau^{**}(R_2,h) = c^{IC}(U^0 + \Delta U).
\end{aligned}$$

Effectively, γ^{**} allows the agent the option between two random counter-offers, which only differ by the face of the coin flip that leads to the better or worse contract.[12]

Analogously to the proof of Proposition 1, one can show that γ^{**} implements the second-best allocation.[13] In particular, γ^{**} induces the agent to respond to a relevant renegotiation offer by picking the counter-offer that with probability 1/2, grants her the payoff $U^0 + \Delta U$ and leads her to reject this offer. Just as under γ^*, this punishes the principal so that he is better off not renegotiating.

Observe, in addition, that the modified mechanism γ^{**} still requires that the agent's message m remains private if she accepts the principal's renegotiation offer. To see this, note that if m were public, then the *message-conditional* renegotiation offer $\gamma^r = (\gamma^r_N, \gamma^r_{R_1}, \gamma^r_{R_2}) = (c^{FI}(U^0 + \varepsilon), c^{FI}(-\infty), c^{FI}(-\infty))$ undermines γ^{**} to implement the second-best allocation.

We now argue that, despite its dependence on the messages' privacy, γ^{**} is implementable via a decentralized smart contract on a transparent blockchain.

Preventing Renegotiation with Decentralized Smart Contracts. In general terms, smart contracts are self-executing programs that run on a decentralized network, automatically executing the terms of an agreement between parties. While these contracts can receive and send messages, it is crucial to note that all these messages are publicly recorded on the blockchain.[14]

Because the privacy of the agent's message plays a crucial role in the mediated mechanism γ^{**}, the public nature of smart contract interactions presents a challenge for indirectly implementing γ^{**}. However, by combining smart contracts with the off-the-shelf cryptographic "commit-and-reveal" technique, we are able to overcome the challenge. Doing so allows us to

[12]Hence, implementation with an observable signal s comes at the complexity cost of an extra message m.

[13]This is shown formally in Appendix B.

[14]For an extensive definition of a smart contract see Szabo (1996) and Catalini and Gans (2020) for a discussion of potential economic applications for smart contracts. We here emphasize however that, in general, an enforcement of smart contracts depends on the shadow of the law. To see this in our specific context of γ^{**}, note that because its transfers condition on the realized output value $Y \in \{g, b\}$, the realized output value must somehow be reported to the smart contract. This can be done by, for instance, the principal, but only the verifiability by a court ensures that the principal will do so truthfully, anticipating its prohibitively large punishment when misreporting.

emulate the storage of messages on the blockchain that are initially secret and only publicly revealed later on.

The commit-and-reveal technique is a cryptographic protocol that allows a party to commit to a value without disclosing it immediately.[15] The technique involves two distinct phases: the initial "commit phase" and the subsequent "reveal phase". In the commit phase, it uses a cryptographic hash function to create a hashed output (the "commitment") of the original message together with a random seed. This hash function has the property of being one-way and collision-resistant, meaning it's computationally infeasible to derive the original message from the hash without knowing the random seed or to find two random seeds so that two different messages generate the same hash. The party thus submits this commitment to the smart contract, effectively "locking in" her choice without revealing it. In the reveal phase, the party can publicly reveal the choice by disclosing the random seed publicly. This is so because the disclosure of the random seed allows a verification of the chosen message, as only this random seed and the chosen message generate the hash value, and collision-resistance implies that the party cannot find a seed that together with another message generates the publicly recorded hashed value.

By using this commit-and-reveal technique, we can emulate γ^{**}'s property of recording a secret message that is only revealed at a later time, despite the public nature of the blockchain. During the commit phase, the message remains hidden, known only to the agent, while the commitment is publicly recorded on the blockchain. Later, during the reveal phase, the agent can choose to disclose the original message. The smart contract can then verify that this revealed message indeed corresponds to the earlier commitment by applying the same hash function and comparing the result to the stored commitment. This process ensures that the message was not altered since the commitment while maintaining its secrecy until the designated reveal time.

An Explicit Example of a Smart Contract. We close this section with an example of a smart contract for a fully parameterized version of our framework using the commit-and-reveal technique. In particular, let the CRRA utility function $u(w) = \sqrt{w}$ describe the agent's preferences over transfers, implying that the monetary equivalent is $\Phi(u) = u^2$. Let $U^0 = 10$ be the agent's reservation utility. The cost of high effort is $d = 2$ with success probability $p_H = 3/4$, while for low effort the probability is $p_L = 1/4$, i.e., $\Delta p = 1/2$. The output is $g = 1300$ in the good state, and $b = 100$ in the bad one. Hence, $y_H = 1000$ and $y_L = 400$.

[15] See Narayanan et al. (2016, Chapter 1) for a more in-depth introduction to cryptographic hash functions and the reveal-and-commit technique.

```
pragma solidity ^0.8.0;
contract CommitRevealTransfer {
  address constant AddressP = 0x362CbcC7a9955332e61d47c107543398C3D25261;
  address constant AddressA = 0x818CbcC8de183AED16f850B17c300DB40a4544Eb;
  uint256 constant TG = 169;  uint256 constant TGH = 225;  uint256 constant TGT = 121;
  uint256 constant TB = 81;  uint256 constant TBH = 121;  uint256 constant TBT = 49;
  bytes32 public HASHCOMMIT;  string public S;  string public Y;
  bool public isCommitted;  bool public isRevealed;  bool public isYSent;
  constructor() {
    require(msg.sender == AddressP, "Only AddressP can deploy");
  }
  function commit(bytes32 _hashCommit) external {
    require(msg.sender == AddressA, "Only AddressA can commit");
    require(!isCommitted, "Already committed");
    HASHCOMMIT = _hashCommit;
    isCommitted = true;
  }
  function generateS() internal {
    require(isCommitted, "Waiting for commit");
    S = block.timestamp % 2 == 0 ? "Head" : "Tail";
  }
  function sendY(string calldata _Y) external {
    require(msg.sender == AddressP, "Only AddressP can send Y");
    require(isCommitted, "Waiting for commit");
      require(keccak256(abi.encodePacked(_Y)) == keccak256(abi.encodePacked("G")) ||
          keccak256(abi.encodePacked(_Y)) == keccak256(abi.encodePacked("B")), "Y must be G
          or B");
    Y = _Y;
    isYSent = true;
    generateS();
  }
  function reveal(string calldata _message, string calldata _salt) external {
    require(msg.sender == AddressA, "Only AddressA can reveal");
    require(isYSent, "Waiting for Y");
    require(!isRevealed, "Already revealed");
    require(keccak256(abi.encodePacked(_message, _salt)) == HASHCOMMIT, "Invalid reveal");
      require(keccak256(abi.encodePacked(_message)) == keccak256(abi.encodePacked("N")) ||
          keccak256(abi.encodePacked(_message)) == keccak256(abi.encodePacked("R1")) ||
          keccak256(abi.encodePacked(_message)) == keccak256(abi.encodePacked("R2")), "
          Invalid message");
    isRevealed = true;
    uint256 transferAmount = determineTransferAmount(_message);
    payable(AddressA).transfer(transferAmount);
  }
  function determineTransferAmount(string memory _message) internal view returns (uint256) {
    if (keccak256(abi.encodePacked(_message)) == keccak256(abi.encodePacked("N"))) {
      return keccak256(abi.encodePacked(Y)) == keccak256(abi.encodePacked("G")) ? TG : TB;
    } else if (keccak256(abi.encodePacked(_message)) == keccak256(abi.encodePacked("R1"))) {
      if (keccak256(abi.encodePacked(Y)) == keccak256(abi.encodePacked("G"))) {
        return keccak256(abi.encodePacked(S)) == keccak256(abi.encodePacked("Head")) ? TGH :
            TGT;
      } else {
        return keccak256(abi.encodePacked(S)) == keccak256(abi.encodePacked("Head")) ? TBH :
            TBT;
      }
    } else {
      if (keccak256(abi.encodePacked(Y)) == keccak256(abi.encodePacked("G"))) {
        return keccak256(abi.encodePacked(S)) == keccak256(abi.encodePacked("Head")) ? TGT :
            TGH;
      } else {
        return keccak256(abi.encodePacked(S)) == keccak256(abi.encodePacked("Head")) ? TBT :
            TBH;
      }
    }
  }
  receive() external payable {
    require(msg.sender == AddressP, "Only AddressP can send Ether");
  }
}
```

Figure 1: The smart contract implementing the mediated mechanism γ^{**} with a reveal-and-commit technique based on the keccak-256 hash function in Solidity.

It is easy to check that $\Delta U = 2$ together with the parameterized example satisfies (4) and yields the mediated contract γ^{**} with the following transfers

$$\begin{array}{ll} \tau^{**}(N,h) = (169,81); & \tau^{**}(N,t) = (169,81) \\ \tau^{**}(R_1,h) = (225,121); & \tau^{*}(R_1,t) = (121,49) \\ \tau^{**}(R_2,h) = (121,49); & \tau^{*}(R_2,t) = (225,121). \end{array}$$

Figure 1 presents an explicit smart contract that implements the mediated mechanism γ^{**} over the Ethereum blockchain using the commit-and-reveal technique for our numerical example. The smart contract is written in Solidity, which is currently the most common language for Ethereum smart contracts. We use its current version 0.8.0.

To allow the agent to send a secret (hashed) message $m \in \{N, R_1, R_2\}$ with a random seed σ, the smart contract implements the commit-and-real technique as previously discussed, based on the public keccak-256 hash function. After sending the hashed message the smart contract responds with sending either the signal $s \in \{h, t\}$ in a (quasi)-random fashion by recording the realized signal publicly on the blockchain. After the signal s is sent, the principal reports the realized output level $Y \in \{g, b\}$. Finally, the agent is to report the seed σ to the smart contract by which the smart contract can recover the original message m so that it can make the transfers according to τ^{**}.

We set up the contract such that if the agent does not reveal the seed σ honestly, this is interpreted as tearing up the original contract and accepting a renegotiated one, $(\rho = y)$, so that the smart contract stops in that no transfers flow and message m stays hidden.

These steps are detailed enough to correctly generate the corresponding smart contract in Solidity with an AI-Assistant such as Claude.AI. This contract is presented in Figure 1.

5 Mitigating Alternative Forms of Renegotiation

As thoroughly summarized by Bolton (1990), standard approaches to renegotiation share the idea that any optimal, renegotiation-proof mechanism should generate private information at the renegotiation stage, thereby leaving the renegotiating principal in the dark when he tries to renegotiate upon the the original decision rule. In all these approaches, such private information is generated by inducing an agent to randomize at equilibrium. This in turn requires to make her indifferent over several alternatives, which, for incentive-compatibility reasons, imposes an allocative cost.

For the specific context of Fudenberg and Tirole (1990), we have formally shown that mediated mechanisms allow to spare this incentive-compatibility cost. They generate private information costlessly, without any need for the agent's random behaviors. In addition, a

mediated mechanism needs to generate the relevant uncertainty only off-equilibrium. Hence, in equilibrium, the contracting parties do not need to incur any indirect costs associated with private information, such as an increased randomness on a risk-averse agent.

We next discuss the extension of this result to other settings of contract renegotiation. In Fudenberg and Tirole (1990), a monopolist provides incentives to an agent who takes a non-observable effort. Since the mediated mechanism effectively operates *after* effort is chosen, our results do not depend on the assumption of a binary effort, and easily extend to the continuum of effort case.[16] Moreover, our solution does not depend on the principal's objective of maximizing revenue. Indeed, a mediated mechanism similar to that we construct in Section 3 can be also exploited by a utilitarian planner to implement a second-best insurance policy under the threat of renegotiation thereby addressing the government failure emphasized by Netzer and Scheuer (2010).

More generally, mediated mechanisms allow to restore the full-commitment equilibrium allocations in contexts of incomplete information rather than moral hazard. To make this concrete, consider a version of the procurement model in Laffont and Tirole (1990). A buyer (principal) contracts over 2 periods with a privately informed seller (agent), who produces a single good through a convex-cost technology. The seller's marginal cost θ_i, with $i = 1, 2$, is her private information, and $\theta_2 > \theta_1$. The second-best involves rationing: type θ_2 ends up selling less than her first-best quantity. This allocation is however fragile to renegotiation: in the last period, the buyer can make a new offer to exploit all gains from trades on θ_2. Thus, renegotiation-proofness requires that the agent's type is not perfectly revealed after her first purchase: renegotiation slows down the speed of information revelation, upholding private information vis-a-vis the principal. Following the logic developed in Section 3, we can construct a mediated mechanism that successfully prevents renegotiation. In particular, the mechanism makes available a new set of contracts, which induce a random participation of θ_2 off-equilibrium. Reflecting the arguments above, the random participation effectively generates private information off-equilibrium, attaining renegotiation-proofness costslessly.[17] Importantly, the agent's risk-neutrality does not prevent the effectiveness of a mediated mechanism. This paves the way to extend our approach to the renegotiation setting of Hart and Tirole (1988), who analyze the Coase-conjecture under the assumption of linear preferences for both trading parties.

[16] This is the extension that Fudenberg and Tirole (1990) consider in their Section 5.A.

[17] Compared with our mediated mechanism γ^*, the construction also guarantees that it is not incentive compatible for type θ_1 to activate the implied punishments. A specific description of such a mechanism is available from the authors.

Appendix A

This appendix collects the proofs of Lemma 1 and Proposition 2.

Proof of Lemma 1. 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

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

for any $U \in (U^0, \infty)$, where the inequality obtains since $U > 2U^0 - U$, 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

$$V_e^{IC}(U^0) > \tilde{V}_e(U^n). \tag{10}$$

Since $U^n > U^0 \Leftrightarrow U^n > 2U^0 - U^n$, it follows from $V_e^{FI}(U)$ decreasing and Φ convex:

$$\tilde{V}_e(U^n) = \frac{1}{2}V_e^{FI}(2U^0 - U^n) + \frac{1}{2}V_e^{FI}(U^n) > max\{V_e^{FI}(U^n), \frac{1}{2}V_e^{FI}(2U^0 - U^n) + \frac{1}{2}V_e^{IC}(U^n)\}. \tag{11}$$

Taking $\Delta U = U^n - U^0 > 0$ together with both (10) and (11) imply (4). ■

Proof of Proposition 2. We construct a mechanism γ_ε that uniquely implements $e = H$ and yields a principal's payoff arbitrarily close to V^{SB}.

Define for any $\varepsilon \in (0, \bar{\varepsilon})$ with $\bar{\varepsilon} > 0$, the contract

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

Note that c_ε^{SB} yields the agent the payoff U^0 if she selects $e = H$, and $U^0 - \varepsilon$ if $e = L$.

Mechanism $\gamma_\varepsilon = \{\mathcal{M}^*, \mathcal{S}^*, \sigma^*, \tau_\varepsilon\}$ coincides with γ^*, except for τ_ε:

$$\tau_\varepsilon(N,h) = \tau_\varepsilon(N,t) = c_\varepsilon^{SB};\ \tau_\varepsilon(R,t) = c^{IC}(U^0 + \Delta U);\ \tau_\varepsilon(R,h) = c^{IC}(U^0 - \Delta U - 3\varepsilon).$$

We consider the subgame $G_\Gamma(\gamma_\varepsilon)$, and construct $\bar{\varepsilon} > 0$ so that, for any belief $x \in [0,1]$ and any $\varepsilon \in (0, \bar{\varepsilon})$, the principal is strictly worse off from any renegotiation offer that the agent accepts with a strictly positive probability.

Given γ_ε, let $\Lambda(\gamma_\varepsilon)$ denote the set of the agent's sequentially rational strategies in $G_\Gamma(\gamma_\varepsilon)$. Starting from the terminal nodes of $G_\Gamma(\gamma_\varepsilon)$, we characterize $\Lambda(\gamma_\varepsilon)$.

Recalling (5), note that in any history (e, γ^r, m, s, y) with $\gamma^r \neq \emptyset$, the agent sends any (distribution of) m_e^r that satisfies the lefthand side of (5), expecting to obtain $\hat{U}_e^r$ from accepting γ^r as expressed in the righthand side of (5).

Given $\hat{U}_e^r$ and rule τ_ε, we derive the agent's optimal acceptance behavior $(\rho(h), \rho(t))$:

(a) For $(e,m) = (H,R)$ and $(e,m) = (L,R)$, we have

$$\rho(h) \in \begin{cases} \{y\} & \text{if } \hat{U}_e^r > U^0 - \Delta U - 3\varepsilon; \\ \{n\} & \text{if } \hat{U}_e^r < U^0 - \Delta U - 3\varepsilon; \\ \{n,y\} & \text{if } \hat{U}_e^r = U^0 - \Delta U - 3\varepsilon; \end{cases} \text{ and } \rho(t) \in \begin{cases} \{y\} & \text{if } \hat{U}_e^r > U^0 + \Delta U; \\ \{n\} & \text{if } \hat{U}_e^r < U^0 + \Delta U; \\ \{n,y\} & \text{if } \hat{U}_e^r = U^0 + \Delta U. \end{cases}$$

(b) For $(e,m) = (H,N)$, we have

$$\rho(h) \in \begin{cases} \{y\} & \text{if } \hat{U}_H^r > U^0; \\ \{n\} & \text{if } \hat{U}_H^r < U^0; \\ \{n,y\} & \text{if } \hat{U}_H^r = U^0; \end{cases} \text{ and } \rho(t) \in \begin{cases} \{y\} & \text{if } \hat{U}_H^r > U^0; \\ \{n\} & \text{if } \hat{U}_H^r < U^0; \\ \{n,y\} & \text{if } \hat{U}_H^r = U^0. \end{cases}$$

(c) For $(e,m) = (L,N)$, we have

$$\rho(h) \in \begin{cases} \{y\} & \text{if } \hat{U}_L^r > U^0 - \varepsilon; \\ \{n\} & \text{if } \hat{U}_L^r < U^0 - \varepsilon; \\ \{n,y\} & \text{if } \hat{U}_L^r = U^0 - \varepsilon; \end{cases} \text{ and } \rho(t) \in \begin{cases} \{y\} & \text{if } \hat{U}_L^r > U^0 - \varepsilon; \\ \{n\} & \text{if } \hat{U}_L^r < U^0 - \varepsilon; \\ \{n,y\} & \text{if } \hat{U}_L^r = U^0 - \varepsilon. \end{cases}$$

Given $(\rho(h), \rho(t))$, we derive her optimal messaging behavior for $e \in E$ and an offer $\gamma^r \neq \emptyset$, yielding $\hat{U}_e^r$ if accepted. For $e = H$, $m = N$ is optimal if

$$\max\{U^0, \hat{U}_H^r\} \geq \frac{1}{2}\max\left\{\hat{U}_H^r, U^0 - \Delta U - 3\varepsilon\right\} + \frac{1}{2}\max\{\hat{U}_H^r, U^0 + \Delta U\}, \tag{12}$$

while $m = R$ is optimal if the opposite weak inequality holds. For $e = L$, $m = N$ is optimal if

$$\max\{U^0 - \varepsilon, \hat{U}_L^r\} \geq \frac{1}{2}\max\left\{\hat{U}_L^r, U^0 - \Delta U - 3\varepsilon\right\} + \frac{1}{2}\max\{\hat{U}_L^r, U^0 + \Delta U\}, \tag{13}$$

while $m = R$ is optimal if the opposite weak inequality holds.

Having characterized $\Lambda(\gamma_\varepsilon)$, we consider the principal's behavior at the renegotiation stage.

First, suppose the principal holds a deterministic belief $x \in \{0, 1\}$ over the agent's effort. In this case, the first argument in the proof of Lemma 3 implies that we can characterize any renegotiated offer that the principal considers optimal by some $\hat{U}_x^r \in (-\infty, +\infty)$, representing the agent's utility that the principal expects given his belief $x \in \{0, 1\}$. Using the agent's sequentially rational behavior as derived above by substituting $\hat{U}_H^r = \hat{U}_1^r$ and $\hat{U}_L^r = \hat{U}_0^r$, we derive the payoff that the principal himself expects from $\hat{U}_x^r$:

1. For $\hat{U}_1^r < U^0 - \Delta U$, the principal expects payoff $V_H(c_\varepsilon^{SB}(U^0))$ and for $\hat{U}_0^r < U^0 - \Delta U - 2\varepsilon$, the principal expects payoff $V_L(c_\varepsilon^{SB}(U^0))$. This follows because the principal expects the agent to consider her strategy $(m, \rho(t), \rho(h)) = (N, n, n)$ uniquely optimal. To see this, note that conditional on sending $m = N$, $\rho(t) = \rho(h) = n$ is strictly optimal, because

$$\hat{U}_1^r < U^0 - \Delta U < U^0 \quad \text{and} \quad \hat{U}_0^r < U^0 - \Delta U - 2\varepsilon < U^0 - \varepsilon.$$

   To see why the principal expects the agent to strictly prefer $m = N$ over $m = R$, consider the two subcases:

   (a) If $\hat{U}_1^r \leq U^0 - \Delta U - 3\varepsilon$, then (12) with $\hat{U}_H^r = \hat{U}_1^r$ becomes $U^0 \geq U^0 - \frac{3}{2}\varepsilon$; likewise, if $\hat{U}_0^r \leq U^0 - \Delta U - 3\varepsilon$, then (13) with $\hat{U}_L^r = \hat{U}_0^r$ becomes $U^0 - \varepsilon \geq U^0 - \frac{3}{2}\varepsilon$. Both inequalities are strictly satisfied since $\varepsilon > 0$.

   (b) If $\hat{U}_1^r \in (U^0 - \Delta U - 3\varepsilon, U^0 - \Delta U)$, then (12) with $\hat{U}_H^r = \hat{U}_1^r$ becomes $\hat{U}_H^r \leq U^0 - \Delta U$; likewise, if $\hat{U}_0^r \in (U^0 - \Delta U - 3\varepsilon, U^0 - \Delta U - 2\varepsilon)$, then (13) with $\hat{U}_L^r = \hat{U}_0^r$ becomes $\hat{U}_0^r \leq U^0 - \Delta U - 2\varepsilon$. Both inequalities are strictly satisfied in case (b) by assumption.

2. For $\hat{U}_1^r = U^0 - \Delta U$ or $\hat{U}_0^r = U^0 - \Delta U - 2\varepsilon$, the principal expects the agent to consider only her strategies $(m, \rho(h), \rho(t)) = (N, n, n)$ and $(m, \rho(h), \rho(t)) = (R, y, n)$ optimal, as, in this case, (12) and (13) both hold with equality. For any randomization over the agent's decisions, the principal expects a payoff that is a convex combination of $V_L(c_\varepsilon^{SB})$ and $\frac{1}{2}V_L^{FI}(U^0 - \Delta U - 2\varepsilon) + \frac{1}{2}V_L^{IC}(U^0 + \Delta U)$ for $x = 0$, and of $V_H(c_\varepsilon^{SB})$ and $\frac{1}{2}V_H^{FI}(U^0 - \Delta U) + \frac{1}{2}V_H^{IC}(U^0 + \Delta U)$ for $x = 1$.

3. For $\hat{U}_1^r \in (U^0 - \Delta U, U^0 + \Delta U)$ or $\hat{U}_0^r \in (U^0 - \Delta U - 2\varepsilon, U^0 + \Delta U)$, the principal expects the agent to consider only $(m, \rho(h), \rho(t)) = (R, y, n)$ optimal because with $(\hat{U}_H^r, \hat{U}_L^r) = (\hat{U}_1^r, \hat{U}_0^r)$ both (12) and (13) are violated. Hence, the principal expects the payoff $\frac{1}{2}V_H^{FI}(\hat{U}_1^r) + \frac{1}{2}V_H^{IC}(U^0 + \Delta U)$ for $x = 1$, and payoff $\frac{1}{2}V_L^{FI}(\hat{U}_0^r) + \frac{1}{2}V_L^{IC}(U^0 + \Delta U)$ for $x = 0$.

4. For $\hat{U}_1^r = U^0 + \Delta U$ or $\hat{U}_0^r = U^0 + \Delta U$, the principal expects the agent to consider exactly the 3 strategies $(m, \rho(h), \rho(t)) = (N, y, y)$, $(m, \rho(h), \rho(t)) = (R, y, y)$, and $(m, \rho(h), \rho(t)) = (R, y, n)$ optimal. For any mixture over these strategies, the principal obtains a convex combination between $V_H^{FI}(U^0 + \Delta U)$ and $\frac{1}{2}V_H^{FI}(U^0 + \Delta U) + \frac{1}{2}V_H^{IC}(U^0 + \Delta U)$ for $x = 1$; and between $V_L^{FI}(U^0 + \Delta U)$ and $\frac{1}{2}V_L^{FI}(U^0 + \Delta U) + \frac{1}{2}V_L^{IC}(U^0 + \Delta U)$ for $x = 0$.

5. For $\hat{U}_e^r \in (U^0 + \Delta U, \infty)$, the principal expects the agent to consider exactly strategies $(m, \rho(h), \rho(t)) = (N, y, y)$ and $(m, \rho(h), \rho(t)) = (R, y, y)$ optimal. For any mixture over these strategies, the principal obtains $V_H^{FI}(\hat{U}_1^r)$ for $x = 1$ and $V_L^{FI}(\hat{U}_0^r)$ for $x = 0$.

From the above it follows that with belief $x = 1$, the following inequalities guarantee that the principal believes to be strictly worse off from a renegotiation offer that the agent accepts with a strictly positive probability:

$$V_H(c_\varepsilon^{SB}) - \frac{1}{2}V_H^{FI}\left(U^0 - \Delta U\right) - \frac{1}{2}V_H^{IC}(U^0 + \Delta U) > 0, \tag{14}$$

and

$$V_H(c_\varepsilon^{SB}) - V_H^{FI}(U^0 + \Delta U) > 0. \tag{15}$$

Observe that, if $\varepsilon = 0$, (14) and (15) are satisfied because they coincide with (8) and (9), respectively. Since $V_H(c_\varepsilon^{SB})$ is continuous in ε, there is a $\varepsilon^H > 0$ such that (14) and (15) are satisfied for any $\varepsilon \in (0, \varepsilon^H)$. If, instead, $x = 0$, the principal believes to be strictly worse off from the agent accepting a renegotiation offer with a strict positive probability when

$$V_L(c_\varepsilon^{SB}) - \frac{1}{2}V_L^{FI}\left(U^0 - \Delta U - 2\varepsilon\right) - \frac{1}{2}V_L^{IC}(U^0 + \Delta U) > 0 \tag{16}$$

and

$$V_L(c_\varepsilon^{SB}) - V_L^{FI}(U^0 + \Delta U) > 0. \tag{17}$$

Again, since $V_L(c_\varepsilon^{SB})$ is continuous in ε, there is a $\varepsilon^L > 0$ such that (16) and (17) are satisfied for any $\varepsilon \in (0, \varepsilon^L)$. Defining $\bar{\varepsilon} \doteq min\{\varepsilon^L, \varepsilon^H\}$ implies that if the principal holds a degenerate belief, then, for any $\varepsilon \in (0, \bar{\varepsilon})$, he believes that he is strictly worse off from a renegotiation offer that the agent accepts with a strictly positive probability.

We next argue that the polar cases $x \in \{0, 1\}$ as studied above, imply that, also for an intermediate belief $x \in (0, 1)$, the principal expects to be strictly worse off from the agent accepting a renegotiation offer with strictly positive probability. To see this, note that the principal's expected payoff by not renegotiating is linear in x:

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

Moreover, note that by offering $\gamma^r \neq \emptyset$ and some sequentially rational behavior $\lambda \in \Lambda(\gamma_\varepsilon)$ by the agent, he would instead get

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

As the agent's behavior is independent of the principal's belief x, this is also linear in x. We however already obtained that if λ is such that it implies a strict positive probability of accepting the renegotiation offer then

$$V_H(c_\varepsilon^{SB}) > V_H^*(\gamma^r, \lambda) \text{ and } V_L(c_\varepsilon^{SB}) > V_L^*(\gamma^r, \lambda).$$

Thus, the suboptimality of renegotiation extends to intermediate beliefs $x \in (0, 1)$.

Finally, we consider the agent's effort choice. Since the agent anticipates that the principal does not make a renegotiation offer that makes her accept it with positive probability, $(e, m) = (H, N)$ is strictly optimal among all $(e, m) \in \{H, L\} \times \{R, N\}$. Hence, the principal's payoff in $G_\Gamma(\gamma_\varepsilon)$ is $V_H(c_\varepsilon^{SB})$. Consequently, in any equilibrium of G_Γ, the principal must obtain at least the payoff V^{SB} as

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

Given that the principal cannot obtain more than what he can get without any possibility of renegotiation, the principal obtains exactly the payoff V^{SB} in any equilibrium. The analysis in Section 1 guarantees that only c^{SB} yields V^{SB} to the principal without violating constraints (2) and (3). ■

Appendix B

This appendix develops three relevant extensions, and collects the proofs of additional results.

Irrelevance of Random Mechanisms in Fudenberg and Tirole (1990)

We here formalize the claim that random mechanism play no role in the FT construction.

Lemma 4 *Let $\tilde{G}^r$ be a game which coincides with G^r, but enlarges the set of available mechanisms C to $\tilde{C}$, which also includes all stochastic mechanisms $\gamma_{\tilde{c}} : E \to \Delta(\mathbb{R}^2)$. Then, $\tilde{G}^r$ has only one equilibrium allocation, which coincides with that in G^r.*

Proof of Lemma 4. For any stochastic mechanism $\gamma_{\tilde{c}} \in \tilde{C}$, define $\gamma_{\tilde{c}}(e) = \tilde{c}_e$ and let

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

be the the agent'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 the principal by a renegotiation offer $\gamma^r \in \tilde{C}$ is the value of the program $P(x, \tilde{U}_H, \tilde{U}_L)$:

$$\begin{aligned} V^*(x, \tilde{U}_H, \tilde{U}_L) = \max_{\gamma^r_{\tilde{c}} \in \tilde{C}} \quad & Y(x) - x[p_H \mathbb{E}(\Phi(U_g)|c^r_H) + (1-p_H)\mathbb{E}(\Phi(U_b)|c^r_H)] \\ & - (1-x)[p_L \mathbb{E}(\Phi(U_g)|c^r_L) + (1-p_L)\mathbb{E}(\Phi(U_b)|c^r_L)] \end{aligned} \tag{18}$$

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

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

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

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

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

Claim 1 *$P(x, \tilde{U}_H, \tilde{U}_L)$ admits a unique solution, which is deterministic.*

Proof. See Chade and Schlee (2012, Proposition 1).

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

Claim 2 *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 Claim 2, the following holds:

Claim 3 *The subgames $G^r(\gamma_c)$ and $\tilde{G}^r(\gamma_{\tilde{c}})$ have the same 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 Claim 2. It follows that, in either subgame, the principal's renegotiation offer is $\gamma^r(\gamma_{\tilde{c}}, x) = \gamma^r(\gamma_C, x)$, which is accepted by the agent, who truthfully reports her former effort.[18] Furthermore, the transfers corresponding to the unique solution of $P(x, \tilde{U}_H, \tilde{U}_L)$ are implemented. Thus, playing $e = H$ with probability $x \in [0,1]$ is sequentially rational for the agent 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 the agent. Claim 3 implies that the upper bound $V^{FT} = V^*(x^{FT}, U^{FT}, U^{FT})$ of the principal'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 $\tilde{G}^r$, the principal 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 the principal in $\tilde{G}^r$ is V^{FT} as in G^r, and the same distributions over efforts and transfers are implemented. ■

Endogenous Timing of the Agent's Report

We here show that Proposition 1 extends to scenarios, in which the agent can send her report before or after the principal makes a renegotiation offer.

Consider the mechanism $\mu = \{\mathcal{M}^\mu, \mathcal{S}^\mu, \sigma^\mu, \tau^\mu\}$ such that $\mu \notin \Gamma$ and

$$\mathcal{M}^\mu = \{N, R, \emptyset\}, \quad \mathcal{S}^\mu = \{h, t\}, \quad \sigma^\mu = \left(\frac{1}{2}, \frac{1}{2}\right).$$

The mechanism μ is a modified version of $\gamma^* = \{\mathcal{M}^*, \mathcal{S}^*, \sigma^*, \tau^*\}$ from Proposition 1. If μ is offered by the principal and accepted by the agent, it induces the following extensive form game G_μ:

1. The agent selects $e \in E$.

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

2. The agent sends $m_1 \in \{N, R, \emptyset\}$.

3. If $m_1 \neq \emptyset$, $s \in \{h, t\}$ is extracted from the distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$ and disclosed privately to the agent.

4. The principal proposes $\gamma^r \in \Gamma$.

5. If $m_1 = \emptyset$, the agent sends $m_2 \in \{N, R, \emptyset\}$, and $m_2 = \emptyset$ otherwise.

6. If $(m_1 = \emptyset, m_2 \neq \emptyset)$, $s \in \{h, t\}$ is extracted from the distribution $\left(\frac{1}{2}, \frac{1}{2}\right)$ and is disclosed privately to the agent.

7. The agent takes the participation decision $\rho \in \{y, n\}$.

8. There are three possible situations:

   - If $\gamma^r \neq \{\emptyset\}$ and $\rho = y$, the agent 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 $m_1 = m_2 = \emptyset$, the agent sends $m_3 \in \{N, R\}$, then $s \in \{h, t\}$ 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 $m_2 \neq \emptyset$, the agent sends $m_3 = \emptyset$ and the decision rule $\tau^\mu(m_1, m_2, m_3, s)$ is implemented.

For a given vector of the agent'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 2, by giving the agent the freedom to decide, at each step $i = 1, 2, 3$ of the *interim* stage, whether to send the message $m_i \in \{N, R\}$ or to stay silent ($m_i = \emptyset$).[19] Furthermore, the agent *must* speak, only once, in the mechanism μ, i.e. if $m_1 = m_2 = \emptyset$, the agent is forced to send a nonempty message $m_3 \in \{N, R\}$ 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 the agent learns the realization $s \in \{h, t\}$ of the private signal as soon as she sends a nonempty message $m_i \neq \emptyset$.

A behavioral strategy for the principal 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

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

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 is made, before reporting m in the original mechanism.

A behavioral strategy $\lambda(\mu)$ of the agent 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 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 5 *The second-best allocation (H, c^{SB}) is supported in an equilibrium of game G_μ.*

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

(i) If $\gamma^r = \{\emptyset\}$, and for any γ^r such that $\hat{U}_e^r \leq U^0 - \Delta U$, the agent sends $m = N$ in μ.

(ii) For any $\gamma^r \neq \{\emptyset\}$ such that $\hat{U}_e^r \in (U^0 - \Delta U, U^0 + \Delta U]$, the agent sends $m = R$ in μ.

(iii) For any $\gamma^r \neq \{\emptyset\}$ such that $\hat{U}_e^r > U^0 + \Delta U$, the agent sends $m = R$ 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 $s \in \{h, t\}$ according to the following rule:

(i) If $m_j = N$ and for each $s \in \{h, t\}$, $\rho = y$ if $\hat{U}_e^r \geq U^0$ and $\rho = n$ otherwise.

(ii) If $m_j = R$ and $s = h$, $\rho = y$ if $\hat{U}_e^r \geq U^0 - \Delta U$ and $\rho = n$ otherwise.

(iii) If $m_j = R$ and $s = t$, $\rho = y$ if $\hat{U}_e^r \geq U^0 + \Delta U$ and $\rho = n$ otherwise.

At any history such that $m_1 = m_2 = \emptyset$, the agent participates in the renegotiated mechanism if and only if $\hat{U}_e^r \geq U^0$.[20] 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 $m_3 = N$, while, at any terminal history in which she participates in a renegotiated mechanism, she optimally sends some $m_e^r \in \mathcal{M}^r$ as defined in the proof of Proposition 1.

Note first that, given the the agent's behavior, the sequential rationality of the principal's behavior has already been shown in Proposition 1. The same argument applies to the the agent'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 the agent's decisions, starting from the terminal nodes. At any history in which the agent sends $m_3 \neq \emptyset$ (as well as $m_2 \neq \emptyset$ with $\gamma^r = \{\emptyset\}$), the agent is indifferent between $m_j \in \{N, R\}$ as she obtains the expected payoff U^0 from μ regardless of the report she selects. Thus, it is sequentially rational for her to send $m_j = N$ as we construct.

Consider now the agent's participation behavior when $m_2 = \emptyset$. Since in this case the agent's strategy prescribes $m_3 = N$, and since $s \in \{h, t\}$ is payoff-irrelevant when $m_j = N$, the fact that at any history $(\mu, e, \emptyset, \gamma^r, N, s)$ 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 $m_2 = N$, which has been shown to be sequentially rational in Proposition 1.

Let us now turn to the agent's choice of $m_2 \in \mathcal{M}^\mu$ at any $(\mu, e, \emptyset, \gamma^r)$. Since the agent is indifferent between sending $m_2 = \emptyset$ and $m_2 = N$ as just argued, she has no incentive to deviate from $m_2 = N$ to $m_2 = \emptyset$ when prescribed by the behavioral strategy we construct. Furthermore, deviating from $m_2 = R$ to $m_2 = \emptyset$ yields her no strictly profitable deviation since $m_2 = \emptyset$ is equivalent to $m_2 = N$, and thus, the existence of a profitable deviation to $m_2 = \emptyset$ would imply the existence of a deviation from $m = R$ to $m = N$ in the original game studied in Proposition 1. Thus, the the agent'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 the agent to send the optimal message $m_j = N$ 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 $m_2 = N$. ■

It is also noteworthy that no equilibrium in pure strategies exists such that $m_1 \neq \emptyset$, at least for values of ΔU that are large enough in the mechanism μ. In fact, if the equilibrium

[20] This is equivalent to her participation behavior when $j \in \{1, 2\}$ and $m_j = N$, for each $s \in \{h, t\}$.

strategy of the the agent's endogenous type $e \in E$ is $m_1 = N$, the principal's optimal response is to give full insurance to this type, leading the agent to pick $m_j = R$ as shown in Proposition 1. Also, if the agent picks $m_1 = R$, for large enough values of $U^0 + \Delta U$, the principal optimally proposes a full-insurance contract targeted only to the type (e, h), which means that by sending $m = R$, the agent expects the same payoff as in the absence of renegotiation. But then, as as shown in Proposition 1, $m_j = N$ is the the agent's unique optimal report. Thus, the mechanism γ^* from Proposition 1 provides an incentive to coordinate on equilibria in which m_j is sent after γ^r is posted and before ρ is taken.

Renegotiation with Public Signals

We here show that privacy of the signals is not needed to achieve our efficiency result. Specifically, we show that the mechanism γ^{**} as defined in Section 4 supports the second-best allocation (H, c^{SB}) at equilibrium. To argue this, first consider the subgame $G_{\Gamma}^{Pub}(\gamma^{**})$, which starts after γ^{**} is offered and accepted:

1. The agent takes $e \in \{L, H\}$.
2. The principal offers $\gamma^r = \{\mathcal{M}^r, \mathcal{S}^r, \sigma^r, \tau^r\}$, 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. The agent privately sends $m \in \mathcal{M}^{**} = \{N, R_1, R_2\}$.
4. The signal $s \in \mathcal{S}^{**} = \{h, t\}$ realizes publicly according to σ^{**}.
5. If $\rho = y$, the agent sends $m^r \in \mathcal{M}^r$, the signal $s^r \in \mathcal{S}^r$ is realized according to σ^r, and the transfers $\tau^r(m^r, s^r, s)$ are implemented. If $\rho = n$ or $\gamma^r = \{\emptyset\}$, the transfers $\tau^{**}(m, s)$ are implemented.

Observe that the game $G^{Pub}(\gamma^{**})$ involves a larger strategy space for the renegotiating principal, who can now make his offer γ^r contingent on the realized signal $s \in S$. Although this may in principle create new incentives to renegotiate, the mechanism γ^{**} allows to fully mitigate any renegotiation threat. Indeed, the following holds:

Lemma 6 *The allocation (H, c^{SB}) is supported in an equilibrium of $G^{Pub}(\gamma^{**})$.*

Proof. Observe first that a strategy for the principal in $G^{Pub}(\gamma^{**})$ is a signal-contingent renegotiated offer γ^r. An agent's behavioral strategy λ consists of a randomization $(1-x, x)$ over $e \in E$ at the initial history of $G^{Pub}(\gamma^{**})$, a randomization over messages in $\mathcal{M}^{**}$ at

each history (e, γ^r), a randomization over participation decisions $\rho \in \{y, n\}$ at each history (e, γ^r, m, s) where $\gamma^r \neq \{\emptyset\}$ and a randomization over messages in $\mathcal{M}^r$ at each history $(e, \gamma^r, m, s, \rho)$ such that $(\gamma^r \neq \emptyset, \rho = y)$.

For any signal $s \in \mathcal{S}^{**}$ extracted in γ^{**}, let $m_e^r(s) \in \mathcal{M}^r$ be an optimal message when the agent accepts γ^r having chosen the effort $e \in E$. Following (5), we denote $\hat{U}_e^r(s)$ the corresponding payoff. That is:

$$\hat{U}_e^r(s) = \sum_{s^r \in \mathcal{S}^r} \sigma^r(s^r | m_e^r(s), s) U_e(\tau^r(m_e^r(s), s^r, s)) \quad \forall s \in \mathcal{S}^{**}. \tag{19}$$

We now construct a PBE of $G^{Pub}(\gamma^{**})$ which implements the allocation (H, c^{SB}).

The principal's equilibrium strategy is not to renegotiate, i.e. $\gamma^r(\gamma^{**}) = \{\emptyset\}$, while the agent's strategy $\lambda(\gamma^{**})$ is as follows:

1. The agent 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) For any $e \in E$, for $\gamma^r = \{\emptyset\}$, and for any γ^r such that

   $$\begin{aligned} &\frac{1}{2}\max\{U^0, \hat{U}_e^r(h)\} + \frac{1}{2}\max\{U^0, \hat{U}_e^r(t)\} \geq \\ &\quad \max\Big\{\frac{1}{2}\max\{U^0 - \Delta U, \hat{U}_e^r(h)\} + \frac{1}{2}\max\{U^0 + \Delta U, \hat{U}_e^r(t)\}, \\ &\qquad \frac{1}{2}\max\{U^0 + \Delta U, \hat{U}_e^r(h)\} + \frac{1}{2}\max\{U^0 - \Delta U, \hat{U}_e^r(t)\}\Big\} \end{aligned} \tag{20}$$

   the agent sends $m = N$ in γ^{**}, followed by $\rho = n$. Observe that the lhs of (20) corresponds to the agent's expected payoff of reporting $m = N$ in γ^{**}, followed by her (optimal) signal-contingent participation decisions. The rhs of (20) characterizes the payoff corresponding to the best alternative report.

   (ii) For any $e \in E$, and for any $\gamma^r \neq \{\emptyset\}$ such that (20) is *not* satisfied, the agent sends $m = R_1$ in γ^{**} whenever

   $$\begin{aligned} &\frac{1}{2}\max\{U^0 - \Delta U, \hat{U}_e^r(h)\} + \frac{1}{2}\max\{U^0 + \Delta U, \hat{U}_e^r(t)\} \geq \\ &\quad \frac{1}{2}\max\{U^0 + \Delta U, \hat{U}_e^r(h)\} + \frac{1}{2}\max\{U^0 - \Delta U, \hat{U}_e^r(t)\}, \end{aligned} \tag{21}$$

   followed by the (optimal) signal-contingent participation $\rho = y$ whenever $\hat{U}_e^r(s) \geq U_e(\tau^{**}(m, s))$, and $\rho = n$ otherwise, for any $s \in \mathcal{S}^{**}$.

(iii) For any $e \in E$, and for any $\gamma^r \neq \{\emptyset\}$ such that (20) and (21) are *not* satisfied, the agent sends $m = R_2$ in γ^{**} followed by the (optimal) signal-contingent participation described in (ii).

3. For any history $(e, \gamma^r \neq \{\emptyset\}, m, s, y)$, the agent sends $m_e^r(s)$.

It is immediate to check that the agent's strategy $\lambda(\gamma^{**})$ is sequentially rational. In particular, it is optimal for her to choose $e = H$ with probability one since, on the equilibrium path, the incentive-compatible transfers $c^{SB} = c^{IC}(U^0)$ are executed.

To conclude the proof, it remains to check that, given $\lambda(\gamma^{**})$, there is no renegotiated offer $\gamma^r \neq \{\emptyset\}$ yielding the principal a strictly higher payoff than V^{SB}. We partition the set of available renegotiated offers according to the reports that $\lambda(\gamma^{**})$ induce in the mechanism γ^{**}.

Observe first that, for any γ^r such that $\lambda(\gamma^{**})$ prescribes to report $m = R_1$ in γ^{**}, the principal's payoff cannot exceed

$$V^R = \frac{1}{2}V_H^{FI}(U^0 - \Delta U) + \frac{1}{2}V_H^{FI}(U^0 + \Delta U),$$

that is, the payoff providing full insurance to the agent conditional on each realized signal. In this case, Lemma 1 guarantees that $V^{SB} > V^R$. Thus, the principal prefers not to renegotiate than renegotiating an offer which induces the report $m = R_1$. A symmetric argument applies to any γ^r such that $\lambda(\gamma^{**})$ prescribes to report $m = R_2$ in γ^{**}. In any such case, one can also check that the principal cannot achieve a payoff greater than V^R.

Thus, for any profitable renegotiation γ^r, the agent's equilibrium strategy $\lambda(\gamma^{**})$ must prescribe to report $m = N$ in γ^{**}. That is, given (20), and since $e = H$, one should have:

$$\begin{aligned}\frac{1}{2}\max\{U^0, \hat{U}_H^r(h)\} + \frac{1}{2}\max\{U^0, \hat{U}_H^r(t)\} \geq \\ \max\Big\{\frac{1}{2}\max\{U^0 - \Delta U, \hat{U}_H^r(h)\} + \frac{1}{2}\max\{U^0 + \Delta U, \hat{U}_H^r(t)\}, \\ \frac{1}{2}\max\{U^0 + \Delta U, \hat{U}_H^r(h)\} + \frac{1}{2}\max\{U^0 - \Delta U, \hat{U}_H^r(t)\}\Big\}.\end{aligned} \tag{22}$$

We now argue that (22) is satisfied only if one of the following two conditions is met:

$$\hat{U}_H^r(s) < U^0 \ \forall s \in \mathcal{S}^{**} \wedge \hat{U}_H^r(s) \geq U^0 + \Delta U \ \forall s \in \mathcal{S}^{**}. \tag{23}$$

To see this, suppose that (23) does not hold, which leads to consider three cases.

- If $\hat{U}_H^r(t) < U^0$ and $\hat{U}_H^r(h) \geq U^0$, then the lhs of (22) is $\frac{1}{2}\hat{U}_H^r(h) + \frac{1}{2}U^0$ and its rhs is at least $\frac{1}{2}\hat{U}_H^r(h) + \frac{1}{2}(U^0 + \Delta U)$, which obtains for $m = R_1$. The latter is strictly greater than the former, which violates (22).

- If $U^0 \leq \hat{U}_H^r(t) < U^0 + \Delta U$, then the lhs of (22) is $\frac{1}{2}\max\{U^0, \hat{U}_H^r(h)\} + \frac{1}{2}\hat{U}_H^r(t)$. Suppose now that $\hat{U}_H^r(h) < U^0 + \Delta U$: the value of the rhs is at least $\frac{1}{2}(U^0 + \Delta U) + \frac{1}{2}\hat{U}_H^r(t)$, which obtains for $m = R_2$. The latter is strictly greater than the former, which violates (22). In the mutually exclusive case $\hat{U}_H^r(h) \geq U^0 + \Delta U$, the value of the rhs is at least $\frac{1}{2}\hat{U}_H^r(h) + \frac{1}{2}(U^0 + \Delta U)$, which obtains for $m = R_1$, which leads to violate (22) again.

- If $\hat{U}_H^r(t) \geq U^0 + \Delta U$, and $\hat{U}_H^r(h) < U^0 + \Delta U$, the lhs of (22) is $\frac{1}{2}\max\{U^0, \hat{U}_H^r(h)\} + \frac{1}{2}\hat{U}_H^r(t)$, and the rhs is at least $\frac{1}{2}(U^0 + \Delta U) + \frac{1}{2}\hat{U}_H^r(t)$, which obtains for $m = R_2$. The latter is strictly greater than the former, which violates (22)

Thus, following a renegotiation γ^r, $\lambda(\gamma^{**})$ prescribes $m = N$ and only if (23) holds. Two cases must then be considered:

(i) If $\hat{U}_H^r(s) < U^0 \ \forall s \in \mathcal{S}^{**}$, then (22) rewrites $U^0 \geq U^0$, and is thus satisifed with equality. Thus, $\lambda(\gamma^{**})$ prescribes to report $m = N$ in γ^{**} and to choose $\rho = n$, which yields the principal the same profit V^{SB} obtained without renegotiation.

(ii) If $\hat{U}_H^r(s) \geq U^0 + \Delta U \ \forall s \in \mathcal{S}^{**}$, then (22) rewrites $U^0 + \Delta U \geq U^0 + \Delta U$, and is thus satisfied with equality. Thus, $\lambda(\gamma^{**})$ prescribes to report $m = N$ in γ^{**}. In addition, for any such γ^r, the agent is guaranteed the payoff $U^0 + \Delta U$ in the continuation play, which implies that the principal's payoff cannot exceed $V_H^{FI}(U^0 + \Delta U)$, which is not greater than V^{SB}, as shown in Lemma 1.

Thus, the principal's strategy $\gamma^r(\gamma^{**}) = \{\emptyset\}$ is sequentially rational. ■

References

Akbarpour, Mohammad and Shengwu Li, "Credible Auctions: A Trilemma," *Econometrica*, March 2020, *88* (2), 425–467.

Attar, Andrea, Catherine Casamatta, Arnold Chassagnon, and Jean-Paul Decamps, "Multiple Lenders, Strategic Default, and Covenants," *American Economic Journal: Microeconomics*, May 2019, *11* (2), 98–130.

_ , Thomas Mariotti, and François Salanié, "Nonexclusive Competition in the Market for Lemons," *Econometrica*, 2011, *79(6)*, 1869–1918.

_ , _ , and _ , "Regulating Insurance Markets: Multiple Contracting And Adverse Selection," *International Economic Review*, August 2022, *63* (3), 981–1020.

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

Bisin, Alberto and Danilo Guaitoli, "Moral Hazard and Nonexclusive Contracts," *RAND Journal of Economics*, 2004, *35(2)*, 306–328.

Bolton, Patrick, "Renegotiation and the dynamics of contract design," *European Economic Review*, May 1990, *34* (2-3), 303–310.

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

Catalini, Christian and Joshua S. Gans, "Some simple economics of the blockchain," *Communications of the ACM*, 2020, *63* (7), 80–90.

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

Davis, Kevin E., "The Demand For Immutable Contracts: Another Look At The Law And Economics Of Contract Modifications," *New York University Law Review*, May 2006, *81*, 487–549.

Dewatripont, Mathias, "Renegotiation and Information Revelation Over Time: The Case of Optimal Labor Contracts," *The Quarterly Journal of Economics*, 1989, *104* (3), 589–619.

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

_ **and** _ , "Optimal mechanism for the sale of a durable good," *Theoretical Economics*, 2024, *19* (2).

Forges, Francoise, "An approach to communication equilibria," *Econometrica: Journal of the Econometric Society*, 1986, pp. 1375–1385.

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

Hart, Oliver D. and Jean Tirole, "Contract Renegotiation and Coasian Dynamics," *The Review of Economic Studies*, 1988, *55* (4), 509–540.

Jolls, Christine, "Contracts as Bilateral Commitments: A new Perspective on Contract Modification," *Journal of Legal Studies*, 1997, *26*, 203–237.

Laffont, Jean-Jacques and Jean Tirole, "Adverse Selection and Renegotiation in Procurement," *The Review of Economic Studies*, 1990, *57* (4), 597–625.

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.

_ , "Multistage Games with Communication," *Econometrica*, March 1986, *54* (2), 323–358.

Narayanan, Arvind, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder, *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*, Princeton University Press, 2016.

Netzer, Nick and Florian Scheuer, "Competitive markets without commitment," *Journal of political economy*, 2010, *118* (6), 1079–1109.

Omar, Ilhaam A., Haya R. Hasan, Raja Jayaraman, Khaled Salah, and Mohammed Omar, "Implementing decentralized auctions using blockchain smart contracts," *Technological Forecasting and Social Change*, 2021, *168* (C).

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

Roughgarden, Tim, "Transaction Fee Mechanism Design," Papers, arXiv.org June 2021.

Szabo, Nick, "Smart Contracts: Building Blocks for Digital Markets," 1996. Accessed on October 3, 2024.

Townsend, Robert M., *Distributed Ledgers: Design and Regulation of Financial Infrastructure and Payment Systems*, MIT Press, 2020.