Contract Negotiation and the Coase Conjecture: A Strategic Foundation for Renegotiation-Proof Contracts *

Bruno Strulovici

Northwestern University

December 8, 2016

Abstract

What does contract negotiation look like when some parties hold private information and negotiation frictions are negligible? This paper analyzes this question and provides a foundation for renegotiation-proof contracts in this environment. The model extends the framework of the Coase conjecture to situations in which the quantity or quality of the good is endogenously determined and to more general environments in which preferences are nonseparable in the traded goods. As frictions become negligible, all equilibria converge to a unique outcome which is separating, efficient, and straightforward to characterize.

*I am grateful for comments from Mehmet Ekmekci, Drew Fudenberg, Lucas Maestri, Jim Malcomson, Paul Milgrom, Alessandro Pavan, Marcin Peski, Ariel Rubinstein, Larry Samuelson, and Asher Wolinsky, discussions by Takuo Sugaya and Alex Wolitzky, as well as the editor and four anonymous referees, Mikhail Safronov's excellent research assistance, and reactions from seminar participants at Princeton University, Caltech, UCSD, Oxford University, the École Polytechnique Fédérale de Lausanne, the Séminaire Roy in Paris, the University of Illinois at Urbana-Champaign, the London School of Economics, the University of Essex, the University of Pennsylvania, Duke University, the 2013 Cowles Conference in Economic Theory, Kyoto University (KIER), the Stanford Institute for Theoretical Economics (SITE 2013), the Asian Meeting of the Econometric Society (2013), Columbia University, Western University, the University of Toronto, the 2014 Winter Meetings of the Econometric Society, the Toulouse School of Economics, and NYU Stern. This research is supported by an NSF Grant (No.1151410) and a fellowship from the Alfred P. Sloan Foundation.

1 Introduction

Real negotiations contain a puzzle: On the one hand, parties often try to withhold private information until a deal is reached. On the other hand, the very fact of agreeing to a deal, and the timing of this agreement, reveals to some extent the parties' real stakes in the negotiations. How is this new information incorporated in the final agreement, when players can freely renegotiate their contract?

This puzzle does not arise in the best-known models of negotiation, because of their specific structure. In Coase's model of a durable-good monopolist (Coase (1972)), for instance, buyers need only one unit of the good. Any sale is therefore efficient, and further negotiations are pointless regardless of the information revealed about the buyer by the timing of his purchase. Similarly, when players with privately known patience bargain over splitting a pie (Rubinstein (1985)), any split of the pie is ex-post efficient and further negotiations are pointless regardless of what players learn about each other's patience.

In richer contractual environments, however, an initial agreement may be inefficient. If, for example, a buyer may value multiple units or various qualities of a good, his acceptance to buy the good early in the negotiation process may reveal a preference for more units or higher qualities of the good, spurring further negotiations before the final agreement. Moreover, this perspective affects both parties' incentives from the outset of negotiations.

This paper's objective is to characterize the possible outcomes of negotiation in these richer contractual environments, when the ability to refine contracts is (almost) unrestricted. It provides a dynamic resolution of the above puzzle by describing the gradual succession of agreements leading to the final outcome, and answers the following questions: How gradual are the agreements? How fast do they incorporate private information? How much efficiency is lost in the process? How does the outcome depend on the buyer's type and on the seller's belief about it? And on the initial relationship between these parties?

The analysis is based on an explicit protocol of negotiation in which one party—the agent—possesses private information while the other party—the principal—makes the offers. The framework is broader than the buyer-seller model with quasilinear preferences and linear cost. For example, the goods traded may be complements.

The main result is that all equilibria of this negotiation game converge to a unique outcome, which is fully separating and efficient, as negotiation frictions become negligible. The outcome is renegotiation-proof in the sense that no surplus may be gained from renegotiating it further. Seen in this light, the paper provides a strategic foundation for renegotiation-proof contracts, as well as a dynamic implementation, without commitment, of efficient allocations.

In the negotiation protocol considered here, the principal can propose new contracts or changes to a previously accepted contract at any time, including after the agent has accepted or rejected a previous proposal. The principal is thus unable to commit *not* to renegotiate the contract. Time is divided into negotiation rounds: In each round, the principal can propose various changes to the current agreement. The agent then accepts a proposal or rejects all changes. This exchange captures in a stylized fashion the idea of a gradual agreement formation in real negotiations; here, any agreement is interpreted as a (possibly oral) bilateral contract, binding unless both parties agree to replace it by another contract.[1]

For the final outcome to be well defined, negotiations must end somehow, and the protocol relies on a particular concept of negotiation friction: at the end of each round, negotiations are exogenously interrupted with a fixed probability, η, in which case the current agreement is implemented. This interruption may be interpreted in various ways. For example, suppose that parties are negotiating a risk-sharing contract, each dimension of which concerns a state of the world. Then, the realization of the state of the world (or its public announcement) makes any further negotiation moot. In a sales contract application, interruption may be coming from a third party, supplier or customer, demanding a commitment or service which requires the immediate implementation of the contract. The interruption probability captures the negotiation friction. When η is equal to 1, the protocol reduces to full commitment since the first proposal is also the last one. In this case, it is well-known that the principal typically distorts the allocation of some types of the agent, causing ex post inefficiency. This paper's interest lies in the opposite case, in which negotiation frictions are negligible (η goes to 0). This case should be interpreted as parties having arbitrarily frequent opportunities to negotiate with each other: the time interval between consecutive rounds is so small that parties become unlikely to be exogenously interrupted in any such interval. Under this inter-

[1] In practice, reneging on an agreement is costly even if it was made orally or informally—indeed, many jurisdictions recognize oral contracts as legally binding. Reneging on an informal agreement bears other costs, as it damages the reputation of the reneging party and puts a strain on further negotiations.

pretation, the time interval between rounds is of order η. Therefore, even though it takes an increasing number rounds for negotiations to stop as η goes to zero, the expected stopping time of negotiation is independent of η and thus uniformly bounded.[2]

While the ability to freely modify past agreements seems necessary to guarantee an ex post efficient outcome, establishing that it is sufficient involves complex issues. To appreciate the difficulty, consider again the standard durable-good monopolist. Efficiency in that context means that the good is sold without any delay, and was established by Gul, Sonnenschein, and Wilson (1986) as the discount rate between consecutive periods goes to zero.[3] The proof is sophisticated even in this considerably simpler contractual environment, where each contract amounts to a single posted price. The key question is to determine whether the seller can benefit from distorting the allocation of the low-valuation buyer by inefficiently delaying the sale, in order to extract some rent from the high-valuation buyer. In richer environments, the question is more complex because i) any initial agreement may be followed by further negotiations (e.g., contractual covenants, increases in quantities or qualities), ii) the principal may benefit from proposing multiple new contracts at each round instead of a single one,[4] iii) the agent can randomize over all such contracts resulting in complex belief dynamics, and iv) utility may be nonlinear and non-separable in the contract components.

Due to the complexity of the analysis, the model focuses on a binary information struc-

[2]Precisely, suppose that the time interval between two rounds is equal to $a \times \eta$ for some $a > 0$. Then the interruption time has mean a regardless of η and becomes approximately exponentially distributed with parameter $1/a$ as η goes to zero.

[3]The result is shown for the "gap" case and the "no gap" case under some Lipschitz condition on the distribution of types, for weak Markov equilibria (see also Sobel and Takahashi (1983) and Fudenberg, Levine, and Tirole (1985)). Ausubel and Deneckere (1989) show that the conjecture can fail when more general equilibria are allowed. The analysis of the Coase conjecture has been extended to various environments: interdependent values (Deneckere and Liang (2006)), an incoming flow of new buyers (Fuchs and Skrzypacz (2010)), and outside options for the buyer (Board and Pycia (2013)). In Board and Pycia, the seller can extract all the surplus owing to positive selection, unlike the present model, which is closer to negative selection. Skreta (2006) takes a mechanism design approach and shows the optimality of price posting. All these papers assume that the buyer can buy only one unit of the good, a single quality of the good is available, and utility functions are quasilinear.

[4]For example, the principal may propose one contract for each type of the agent, or propose multiple almost identical contracts as a communication device to emulate cheap talk.

ture: the agent can be of two types, and the corresponding utility functions satisfy a standard single-crossing condition. As long as the types of the agent have not been fully separated, there are strictly positive gains from renegotiation, and the paper's main result shows that all these gains are realized as frictions become negligible: *all* Perfect Bayesian Equilibria (PBEs) converge to a unique outcome, which is separating and efficient.

The type-specific contracts to which all PBE outcomes converge are straightforward to characterize and determine graphically. Unlike the full-commitment case (but like the Coase conjecture, which it generalizes in this respect), these contracts are independent of the initial (non degenerate) belief that the principal holds about the type of the agent. They do depend on the initial contract—or absence thereof, which is formally equivalent to a 'null' contract—which may lie in three possible regions of the contract space. In the "No-Rent" region, the principal extracts all surplus of renegotiation regardless of the agent's type. In the other two regions, there is one region-specific type ("L", say) who gains nothing from negotiation while the other type ("H") gets a positive rent.

The fact that almost-efficient contracts are proposed immediately implies that renegotiation plays a relatively minor role *in equilibrium*, even though the *possibility* of renegotiation has a major impact on the outcome. This suggests that, empirically, one should not infer that renegotiation is impossible or difficult in practice just because the observed renegotiation activity seems negligible. Instead, negotiation may be feasible and cheap but largely internalized in the very first contracts that are proposed.

Section 2 presents the setting and main results. Section 3 compares the results with the Coase conjecture. The main arguments are given in Sections 4 and 5. Section 6 discusses the related literature and extensions of the framework.

2 Setting and Overview of the Results

There are two players, a principal (P) and an agent (A), who negotiate a contract lying in some compact and convex subset $\mathcal{C}$ of $\mathbb{R}^2$ whose components are denoted x_1 and x_2.

A has a utility function $u_\theta : \mathcal{C} \to \mathbb{R}$, where $\theta \in \{L, H\}$ denotes his type, and P has a cost function $Q : \mathcal{C} \to \mathbb{R}$. The functions u_L, u_H, and Q are twice continuously differentiable and have strictly positive derivatives with respect to x_1 and x_2; u_L and u_H are concave and Q is

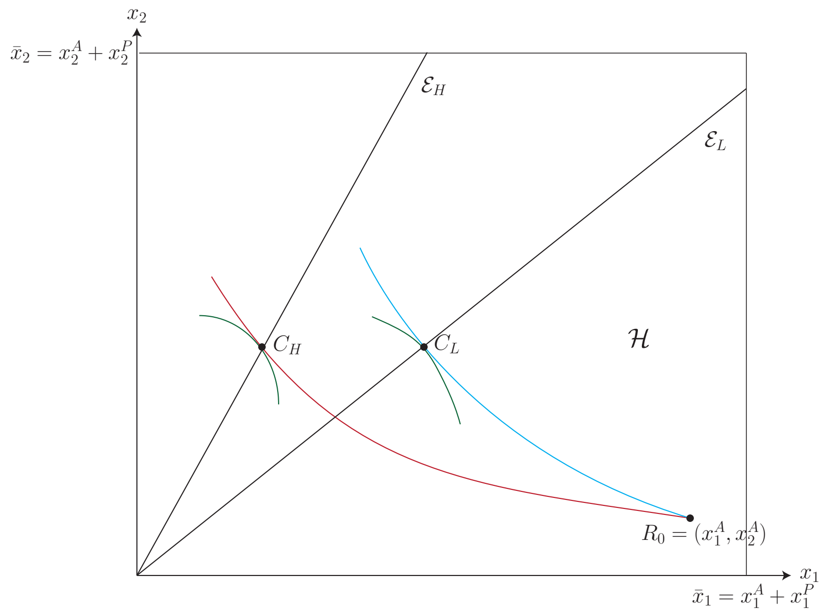

Figure 1: Setting (trade interpretation)

convex. Although it is convenient to think of the agent's type as being "high" (H) or "low" (L) to connect the results to the existing literature, the role played by each type is in fact determined by the initial contract, as explained below.

A contract $C = (x_1, x_2) \in \mathcal{C}$ is θ-*efficient* if it minimizes the principal's cost among all contracts in $\mathcal{C}$ providing θ with some given utility level. For each θ, let $\mathcal{E}_\theta$ denote the set of θ-efficient contracts in the interior of $\mathcal{C}$; θ's iso-utility curve and P's iso-cost curve are tangent at any such contract. It is assumed that the efficiency curves $\mathcal{E}_L, \mathcal{E}_H$ are smooth and upward sloping.[5] The setting is represented on Figure 1 in the context of a trade application (other applications are described later in this section). It is also assumed that, given any contract C on $\mathcal{E}_\theta$, P's and θ's indifference curves going through C do not both have a zero curvature at C. The assumption is satisfied, for instance, when the agent has a quasi-linear utility function and his valuation for the good is strictly concave.[6]

[5]These assumptions hold, e.g., if u_H, u_L, and $-Q$ are weakly supermodular in (x_1, x_2) and either u_θ's or $-Q$ is strictly concave in x_2, as explained in Appendix J. Substantively, these assumptions holds as long as the goods are not too strongly substitutes of each other.

[6]Alternatively, the agent's utility could be linear in the good if P's cost function is strictly convex. The assumption does rule out settings in which both A's and P's indifference curves are linear, but even this case can be approximated by an arbitrarily small curvature (cf. Section 3). The assumption guarantees that

The functions u_L and u_H are required to satisfy a standard single-crossing condition: iso-utility curves of L are steeper than those of H at their intersection point. This implies that the efficiency curve $\mathcal{E}_L$ lies to the lower right of $\mathcal{E}_H$. $\mathcal{C}$ can therefore be partitioned into three regions separated by $\mathcal{E}_L$ and $\mathcal{E}_H$. Contracts in the inner region are said to be in the 'No Rent' configuration, while contracts strictly below $\mathcal{E}_L$ (above $\mathcal{E}_H$) are in the 'H-Rent' ('L-Rent') configuration. The set of contracts in the H-Rent configuration is denoted by $\mathcal{H}$. In the trade application, $\mathcal{C}$ represents an Edgeworth box, delimited by the sum of endowments of the agent and the principal. A contract C specifies the final allocation of the agent, the efficiency curve $\mathcal{E}_\theta$ is the 'contract curve' corresponding to type θ, and the status quo R_0 represents the endowment of the agent before any trade.

The Negotiation Game

The game unfolds as follows. First, the agent privately observes his type θ; P has a prior characterized by the probability $\beta_0 = Pr(\theta = H)$. The game starts with a reference "contract" $R_0 \in \mathcal{C}$, which represents the current engagement between the principal and the agent. In many settings this initial contract would simply represent the absence of any past engagement, as in most contracting models. In the durable-good monopolist application, the initial contract is the "no sale" outcome; in the trade application, the initial contract corresponds to the initial endowment of the agent before any trade. Specifying R_0 explicitly is useful for two reasons: First, it will allow us to treat the initial round like any later round in which an agreement has already been made, which simplifies the exposition. Second, there are environments, such as the trade application, in which the initial contract (initial endowment of the agent) plays an important role on the outcome of negotiations.

There are countably many potential rounds, indexed by $n \in \mathbb{N}$. At each round n, P can propose a finite menu M_n of contracts in $\mathcal{C}$. In terms of interpretation, proposing contracts or *changes* to the current contract is formally equivalent; the former formulation is used here.[7]

the distance between the two curves increases quadratically as one moves away from C, a property used to compute a lower bound for the inefficiency of some contracts (see Lemma 15).

[7]One could restrict the principal to propose at most two contracts in each round. Such a restriction is not desirable for several reasons. First, there is no guarantee that proposing only two contracts at each round is without loss of generality. As Bester and Strausz (2001) have shown, the set of implementable outcomes

The agent chooses a contract in M_n or holds on to the last accepted contract, R_n. Any mixed strategy over the choice set $M_n \cup \{R_n\}$ is allowed. The selected contract, R_{n+1}, becomes the new reference. At the end of each round, negotiations are frozen with probability $\eta \in (0,1]$, in which case the last accepted contract, R_{n+1}, is implemented. Otherwise, negotiations move on to the next round. The event of a negotiation freeze will be hereafter referred to as a "breakdown." This term means that *future* negotiations are terminated. However, the last agreement formed before the breakdown (or the initial contract, if no agreement was formed) is still valid. As explained in the Introduction, one may view η as the time interval between two rounds, in which case the expected stopping time of negotiation is independent of η and thus bounded as η goes to zero.

Letting $\{R_n\}$ denote the stochastic process of contracts entering each round n, the agent's expected utility is equal to

$$\mathcal{V}_\theta = E\left[\sum_{n\geq 0}(1-\eta)^n \eta u_\theta(R_{n+1})\right]$$

while P's expected cost is

$$\mathcal{Q} = E\left[\sum_{n\geq 0}(1-\eta)^n \eta Q(R_{n+1})\right].$$

The parameter η represents the *negotiation friction* of the game.[8] The objective of this paper is to characterize the PBEs of the game as η goes to zero. The existence of a PBE is guaranteed by Theorem 1, whose proof is in the Online Appendix (Appendix D).

Theorem 1 *For each $\eta \in (0,1]$, there exists a PBE of the negotiation game.*

To prove this theorem, backward induction techniques cannot be used because there is no non-degenerate belief, no matter how extreme, for which negotiations end in finite time, as

generally requires strictly more "messages" (or contracts) than the number of types of the agent, even in a two-period setting. Here, one must consider all possible continuation equilibria, including incentive *in*efficient ones. Indeed, an inefficient continuation equilibrium may provide incentives at earlier stages of the game and it is precisely the purpose of the analysis to show that the negotiation outcome is efficient, rather than assume this. Second, such a restriction would not necessarily simplify the analysis as the agent would still choose between three contracts (the two contracts offered and the current one), which potentially results in all the issues, such as belief non-monotonicity, which arise when more contracts are allowed. Finally, allowing more contracts paves the way for an extension to three or more agent types, as explained in Appendix L.

[8]There is another interpretation of the setting where η is the discount rate and the parties receive payoffs at each period of the on going relationship. This interpretation is discussed in Section 6.

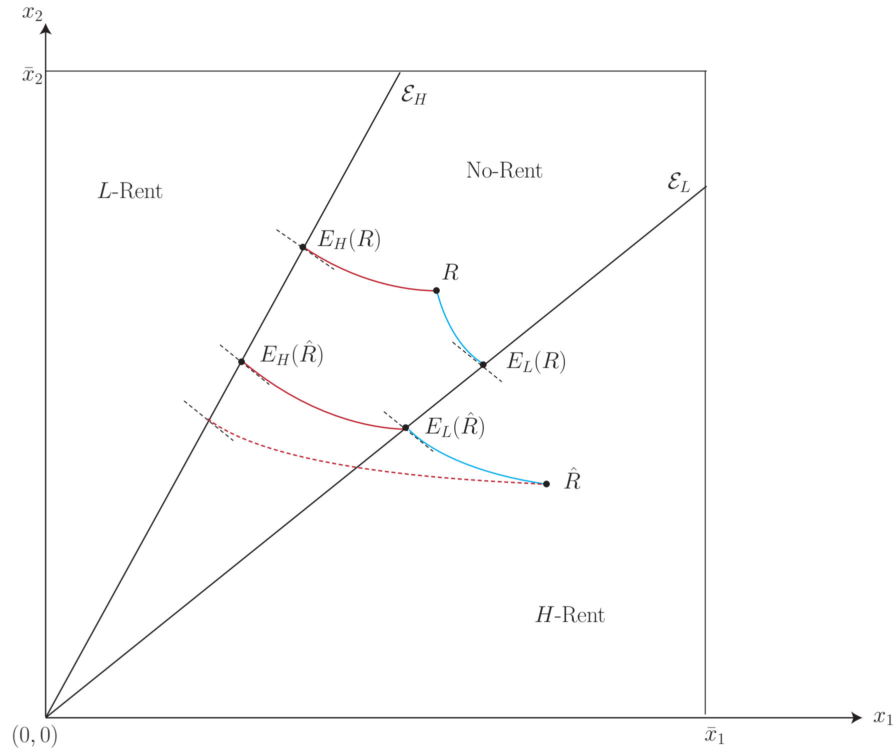

Figure 2: Renegotiation outcomes

will be explained in the analysis below and proved in Appendix K. Instead, the proof builds on a theorem by Harris (1985) for games of perfect information to show the existence of an equilibrium in an auxiliary game between the principal and the high type of the agent, which treats the low type "mechanically." This equilibrium is then used to construct an equilibrium of the negotiation game with private information.

For any contract $R \in \mathcal{C}$, let $E_H(R)$ and $E_L(R)$ denote the least costly pair of H- and L-efficient contracts such that each type $\theta \neq \theta'$ weakly prefers $E_\theta(R)$ to $E_{\theta'}(R)$ and to R. This pair is well defined for each possible configuration of R.[9] Figure 2 represents these definitions for the case of CRRA utility functions and a linear cost function.

Theorem 2, below, requires that no contract arising in equilibrium be jointly efficient

[9]If R is in the No-Rent configuration, $E_\theta(R)$ is simply the θ-efficient contract that gives θ the same utility as R. If R is in the H-Rent configuration, then $E_L(R)$ is similarly defined, while $E_H(R)$ is the H-efficient contract that gives H the same utility as $E_L(R)$. Because that contract gives a strictly higher utility to H than the initial contract R, H must be getting a positive rent in any equilibrium, hence the name of that configuration. A symmetric construction obtains if R is instead in the L-Rent configuration.

for both agent types. The single-crossing property already rules out such contracts in the interior of $\mathcal{C}$. To deal with $\mathcal{C}$'s boundary, say that a contract R_0 is ***regular*** if it is in the No-Rent configuration or if it satisfies the following condition, stated for $R_0 \in \mathcal{H}$ (an analogous condition is required for the L-Rent configuration): for any $R' \in \mathcal{H}$,

$$u_H(E_L(R')) \geq u_H(R_0) \Rightarrow E_L(R') \neq E_H(R') \tag{1}$$

Given any contract R_0, regularity—whose role is explained below—can always be achieved by arbitrarily small perturbations of the utility or the cost functions, as illustrated by Section 3.[10] In Figure 2, all contracts are regular except for the origin.

THEOREM 2 *Consider any regular contract R_0, belief $\beta_0 \in (0,1)$, and $\varepsilon > 0$. There exists $\bar{\eta} > 0$ such that the following statements hold for any $\eta \leq \bar{\eta}$ and corresponding PBE:*
A: The expected utility of each type θ is bounded below by $u_\theta(E_\theta(R_0)) - \varepsilon$.
B: The probability that each type θ gets a contract within a distance[11] ε of $E_\theta(R_0)$ when renegotiation breaks down is greater than $1 - \varepsilon$.

Theorem 2 shows that all PBEs reduce to an essentially unique one, which is efficient, as η becomes arbitrarily small. The proof identifies a feasible strategy for P, which is to "give up" on screening H, and proceeds to show that there is essentially nothing more that P can do. Methodologically, it is interesting to contrast this result with Ausubel and Deneckere (1989) who show that a large range of utilities and profits can be sustained in a Coase conjecture environment with zero gap by punishing any seller deviation with a zero-screening equilibrium. What serves in their setting as a punishment threat for the seller becomes here a tempting deviation for the principal, which prevents the existence of (essentially) any other equilibrium. This model is thus closer to the positive-gap case of the Coase conjecture but the gap here is endogenous and vanishes asymptotically, creating several major challenges for the analysis explained in Section 5.

Theorem 2 implies that P always extracts some surplus from negotiation: When R_0 is in the No-Rent configuration, P extracts, in fact, all the surplus regardless of the agent's type.

[10] These arbitrarily small perturbations are chosen so as to slightly push efficient contracts on $\mathcal{C}$'s boundary into the interior of $\mathcal{C}$, by an arbitrarily small amount.

[11] The statement holds for any norm on $\mathbb{R}^2$.

When R_0 is in the H-Rent configuration, P extracts all the surplus from negotiating with L, and extracts some additional surplus if he faces H, corresponding to a move from $E_L(R_0)$ to $E_H(R_0)$.

Regularity is used as follows: When $R_0 \in \mathcal{H}$, one may show that any contract $R' \in \mathcal{H}$ arising in equilibrium satisfies $u_H(E_L(R')) \geq u_H(R_0)$ (Proposition 1, Part iv) and, hence, that the premise of (1) is always satisfied on path. If R_0 is regular, this implies that any breakdown with a non-degenerate belief yields some inefficiency.

Applications

1. **Durable-Good Monopolist.** A is a buyer with quasi-linear utility $u_\theta(C) = \theta\bar{u}(x_2) + x_1$, where x_2 is the quantity of the good sold by P, x_1 is A's wealth, and $\bar{u}$ is a strictly concave function.[12] The initial contract, R_0, is equal to $(\bar{x}_1, 0)$ where $\bar{x}_1$ is A's initial wealth. P's cost is $Q(x_1, x_2) = cx_2 + x_1$, where $c > 0$ is the marginal cost for producing the good and x_1 captures how much wealth P "leaves" to A.[13]

2. **Labor Contract.** P is a potential employer and A is a worker. $-x_2$ represents A's effort and x_1, his wage. A gets a utility $u_\theta(C) = \theta\psi(-x_2) + x_1$ from contract C, where ψ is a factor entering A's cost of effort, increasing in its argument, and θ is a worker-specific skill entering his cost. The status quo $R_0 = (0, 0)$ represents unemployment. P's profit is $\Pi(x_1, x_2) = -Q(x_1, x_2) = -x_2p - x_1$, where $p > 0$ is the unit price of the good.

3. **Consumption Smoothing and Insurance.** There are two periods and a single good. The dimensions of $\mathcal{C}$ represent A's consumption in each period. P is a social planner or a bank who can help the agent smooth his consumption. The agent's type corresponds to a privately known patience/discount factor, or a distribution parameter describing how likely the agent is to value the good in the second period. For example, $u(x_1, x_2)$ may be equal to $v(x_1) + \theta v(x_2)$ or to $v(x_1) + E[w(x_2, \tilde{\rho})|\theta]$ where $\tilde{\rho}$ is a taste shock whose distribution is increasing in θ in the sense of first-order stochastic dominance and where w is supermodular, so that $E[\partial w/\partial x(x_2, \tilde{\rho})|\theta]$ is increasing in θ.[14] R_0 is A's autarkic income stream. $Q(x_1, x_2) =$

[12] The iso-level curves of u_θ have a positive curvature as long as the second derivative of $\bar{u}$ is strictly negative, as is easily checked. A similar observation applies to ψ in the labor contract application.

[13] P's profit is $\Pi(t, x_2) = t - cx_2$, where t is how much the agent pays P. Letting $t = \bar{x}_1 - x_1$, yields the formulation in terms of the cost function Q.

[14] This application is explored in detailed by Strulovici (2013).

$p_1x_1 + p_2x_2$, where p_t is the market price for the good in period t.

4. **Risk Sharing** Each dimension corresponds to a state of the world. The quantity x_i specifies the transfer of a good from P to A if state i is realized. L values the good more in state 1 than state 2, relative to H. Alternatively, the types have the same preferences but have different subjective beliefs, with L assigning a higher probability to the first state of the world than H does.

5. **Trade.** More generally, the model describes a trade environment in which the dimensions of $\mathcal{C}$ represent distinct goods, with x_i denoting the quantity of good i consumed by A. Type L cares more about the first good than the second, relative to H. P (like A) has convex preferences and Q is the negative of a utility function representing P's preferences. R_0 denotes A's initial holdings of the goods.

3 Relation to the Coase conjecture

In the standard Coase conjecture, all buyer types have a single-unit valuation. The set of efficient contracts is the same for all types: it consists of all contracts for which the buyer gets the good, regardless of the sale price.[15] However, when goods are divisible or available in multiple qualities, it becomes rather restrictive to assume that the same contracts which are efficient for one type of the agent are also efficient for the other types. In fact, when efficient contracts lie in the interior of the contract space, the strict single-crossing property implies that whatever contract is efficient for one agent type is inefficient for the other.

This distinction explains why, in this paper, the principal can always extract some surplus from the "high" type, in contrast to the Coase conjecture. It also explains why, when the initial contact lies between the efficiency curves of both types—an impossibility for the standard Coase conjecture, where these curves coincide—the principal can extract *all* the surplus from negotiation. In this sense, the Coase conjecture appears to be non-generic as it relies on the assumption that all agent types share the same efficient outcomes.

However, because the model imposes no lower bound on how "far" the efficiency curves of both types have to be, it is easy to recover the Coase conjecture as a limit of Theorem 2.

[15] This statement concerns the "positive gap" case (lowest buyer valuation exceeds seller cost), which is the relevant comparison here.

To see this, suppose that the first contractual dimension represents the agent's wealth and the second dimension represents the quantity, between 0 and 1, of a divisible good sold to the agent. The initial contract is $(W, 0)$, where W is the agent's initial wealth. An agent of type θ has utility $u(x_1, x_2) = v_\theta x_2 + x_1$ with $v_H > v_L > 0$. The principal incurs a marginal cost $\frac{\partial Q}{\partial x_2}(x_1, x_2) = c(1 - x_2)^{\delta-1}$ for producing a quantity x_2 of the good, where $c < v_L$ and $\delta \in [0, 1]$. When $\delta = 1$, the marginal cost is constant equal to c. The parameter δ is chosen arbitrarily close to 1, so that Q is slightly convex with a strictly positive curvature.

The efficiency curve $\mathcal{E}_H$ is horizontal and characterized by the quantity $x_2^H(\delta)$ solving $v_H = c(1 - x_2^H(\delta))^{\delta-1}$. Likewise, $\mathcal{E}_L$ is characterized by the quantity $x_2^L(\delta)$ such that $v_L = c(1 - x_2^L(\delta))^{\delta-1}$. We have

$$x_2^L(\delta) = 1 - \left(\frac{c}{v_L}\right)^{\frac{1}{1-\delta}} \quad \text{and} \quad x_2^H(\delta) = 1 - \left(\frac{c}{v_H}\right)^{\frac{1}{1-\delta}},$$

so $x_2(\delta, L) < x_2(\delta, H)$ and both converge to 1 as δ goes to 1. The efficiency curves of both types are thus distinct but converge to the same contract curve, characterized by a single unit of good sold to the agent, as δ goes to 1. The setting is represented on Figure 3. Red (blue) curves represent the iso-utility curves of the high (low) type. The boundary of regular contracts is shown on the left of the figure. All contracts of $\mathcal{H}$ lying to the right of this boundary, including R_0, are regular and if necessary one could expand the contract space to the left (or translate the agent's wealth) to make any given contract $R \in \mathcal{H}$ regular.

The Coase conjecture is recovered as follows: if P were sure to face H, he would move to the contract C_H on Figure 3. With uncertainty about the buyer's type, however, Theorem 2 implies that the outcome is given by the contracts $E_H(R_0), E_L(R_0)$, which converge to the same contract as δ goes to 1. Both types of the buyer obtain essentially the same outcome, which is the (almost) sure sale of the good at the same price. The high type gets a rent corresponding to the distance between $E_H(R_0)$ and C_H, while L gets no rent.

Theorem 2 shows which part of the Coase conjecture is robust to a richer contractual environment: the efficiency part continues to hold, but the seller's inability to extract any surplus disappears. The richer contract space also gets rid of the stark discontinuity arising in the Coase conjecture between the gap and no gap cases: With two types, H's rent increases as L's valuation v_L (and hence the equilibrium price) becomes lower. However, when L's valuation reaches P's marginal cost c, turning into the "no gap" case, H's rent suddenly

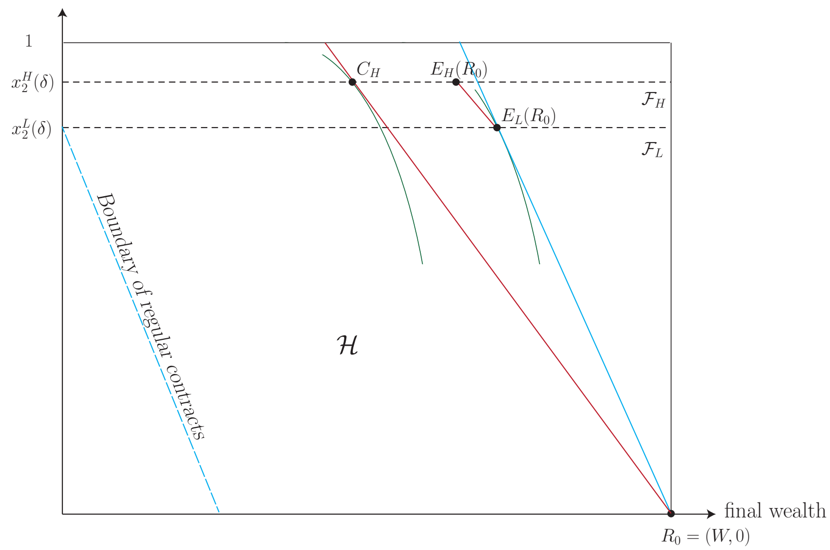

Figure 3: Recovering the standard Coase conjecture

drops to zero and P's profit leaps up from zero to $\beta_0(v_H - v_L)$. Consider a similar exercise in the setting of Theorem 2, where L's efficiency curve is lowered until it goes through R_0. The surplus that P extracts from H then varies continuously, until $\mathcal{E}_L$ goes exactly through R_0 and P extracts all surplus from H.

4 Results holding for all friction levels

This section presents results holding for all values of η which will be used to prove Theorem 2. When P assigns probability 1 to either type of the agent, or when the contract is in the No-Rent configuration, there is a unique continuation PBE: P immediately extracts all the rent from negotiation and efficiency obtains exactly.[16] In other cases, one may compute an upper bound on the rent which the mimicking type (i.e., H, if $R_0 \in \mathcal{H}$) can extract. Intuitively, the rent cannot exceed what this type would get if P "gave up" on screening H by immediately giving the other type (L) his efficient contract. These results are collected in Proposition 1.

[16]Showing that this outcome constitutes an equilibrium is straightforward, but establishing uniqueness is more involved. The proof uses a cost undercutting argument similar to the one applied to utility levels in Rubinstein's (1982) bargaining model with complete information.

Unless indicated otherwise, all results of this section are proved in Appendix B.

PROPOSITION 1 *The following holds for any η and corresponding PBE:*

i) If the prior β puts probability 1 on some type θ, P immediately proposes the θ-efficient contract that leaves θ's utility unchanged, and θ accepts it.

ii) If R_0 is θ-efficient, P immediately proposes $E_{\theta'}(R_0)$ ($\theta' \neq \theta$), and θ' accepts it.

iii) If R_0 is in the No-Rent configuration, P immediately proposes $E_L(R_0)$ and $E_H(R_0)$, and each type θ accepts $E_\theta(R_0)$.

iv) If R_0 is in the H-Rent (L-Rent) configuration, H's (L's) expected utility is bounded above by $u_H(E_H(R_0))$ ($u_L(E_L(R_0))$).

From now on, the analysis focuses on R_0 in the H-Rent configuration, the L-rent configuration case being symmetrical. The next result plays a crucial role for the analysis: At any round n, P can propose the contracts $E_H(R_n)$ and $E_L(R_n)$ and have them accepted by types H and L, respectively. This deviation puts an upper bound on P's continuation cost as a function of the current contract and belief, and will often be referred to as "jumping" or "giving up" (on screening H). Let β_n denote the probability, at the beginning of round n, that P assigns to type H.

LEMMA 1 (JUMP) *If R_n is in the H-Rent configuration and P proposes the contracts $E_H(R_n)$ and $E_L(R_n)$, with $E_H(R_n)$ augmented by an arbitrarily small amount $\varepsilon > 0$, then H accepts $E_H(R_n)$ with probability 1 and L accepts $E_L(R_n)$ with probability 1. Therefore, P's continuation cost is bounded above by $\beta_n Q(E_H(R_n)) + (1-\beta_n)Q(E_L(R_n))$.*

The next result shows that as long as H hasn't revealed himself, on-path contracts R_n all lie in $\mathcal{H}$.

LEMMA 2 *For any $R_0 \in \mathcal{H}$ and PBE, L accepts only contracts in $\mathcal{H}$ throughout negotiations.*

Thus, any accepted contract $C_n \notin \mathcal{H}$ reveals H and can be replaced by the H-efficient contract $\tilde{C}_n$ that gives H the same utility: this reduces P's cost without affecting anyone's incentive. The remainder of the analysis focuses without loss on PBEs in which P only proposes, in each round, contracts in $\mathcal{H}$ and/or the H-efficient contract giving H his continuation utility. Given any PBE, any contract sequence $\{R_n\}$ that is accepted by L with positive probability will be called a **choice sequence**. In equilibrium, the agent follows a choice sequence until, possibly, accepting an H-efficient contract, revealing by that choice that he is of type H. Choice sequences have several important properties, described next.

Let $u_H(n)$ denotes H's continuation utility at the beginning of round n and $w_n = u_H(E_H(R_n)) - u_H(n)$. w_n is the equilibrium rent that P takes away from H compared to immediately giving up on screening him: it is a *rent reduction index*. But, as we shall see, w_n is also closely related to the L-inefficiency of the current contract: the more L-efficient R_n, the smaller w_n has to be and $w_n = 0$ if R_n lies exactly on $\mathcal{E}_L$.

PROPOSITION 2 *Along any choice sequence $\{R_n\}$, i) β_n converges to zero, ii) R_n converges to an L-efficient contract, and iii) w_n converges to zero, as n goes to infinity.*

Proposition 2 shows that full efficiency and screening obtain asymptotically. However, because negotiations break down exogenously in finite time, the key is to determine the speed at which negotiated contracts converge to efficiency relative to the speed at which the exogenous breakdown occurs. In the standard Coase conjecture with two types, one may compute a uniform upper bound on the time at which the sale takes place, which implies equilibrium efficiency as the discount rate goes to zero. However, efficiency cannot be reached in finite time here, as the next result shows (the proof is in Appendix K).

PROPOSITION 3 *For any $R_0 \in \mathcal{H}$, $\eta \in (0, 1]$, $n \in \mathbb{N}$, and PBE starting with belief $\beta > 0$, R_n is not L-efficient. Equivalently, $w_n > 0$ for all n.*

Intuitively, this result comes from the richness of the contract space, and more particularly the fact that negotiated contracts can become arbitrarily close to being L-efficient. In the Coase conjecture with two types, there is a belief threshold $\hat{\beta}$ below which the benefit of screening H becomes negligible relative to the cost of delaying a sale to L, prompting the seller to immediately set the price at L's valuation. Here, instead, P can always propose a contract at a small distance x from the L-efficiency curve, which entails an L-inefficiency of order x^2 but provides a screening benefit of order x for H. For x small enough, this departure from efficiency is always beneficial.

If type H is very unlikely, however, P's expected cost cannot be too different from what he would get if he just ignored H. As a result, P leaves most of the rent to H, which puts the following bound on w_n.

Lemma 3 *There exists $K_w > 0$ such that $w_n \leq K_w \beta_n^{1/3}$ for all n, η, and PBE.*

This bound can be used to prove Theorem 2 for some parameters of the model, but it is too coarse for a general proof. Intuitively, Lemma 3 exploits only the inefficiency loss that P incurs conditional on facing L while trying to screen H—a loss which is of second-order if R is almost L-efficient. The general proof also exploits the loss incurred conditional on facing H, which is of first order, and shows how these two kinds of losses interact and cause P to give up on screening H.

5 Proof of Theorem 2

This section proves that the expected rent, w_0, which P takes away from H relative to H's utility upper bound $u_H(E_H(R_0))$, converges to zero as η goes to zero. The other claims of Theorem 2 are relatively simple corollaries of this result and are proved in Appendix I.

Part I: Block Construction

The proof starts by constructing blocks of rounds such that i) within each block, H is screened with significant probability and ii) w_n shrinks geometrically across blocks. The construction draws from the dynamic screening literature[17] but presents specific challenges.

The first challenge is that the inefficiency caused by a breakdown is endogenous (unlike, e.g., delaying a sale in the Coase conjecture when the gap is positive). Moreover, contracts become asymptotically L-efficient conditional on staying in $\mathcal{H}$, which makes the loss conditional on facing L arbitrarily small and thus hard to exploit. These contracts do entail some non-negligible H-inefficiency, since they are bounded away from $\mathcal{E}_H$, but a contract R_n's H-inefficiency only translates into a loss for P, relative to jumping to $E_H(R_n)$, if the utility that R_n gives H is sufficiently close to what H gets from $E_H(R_n)$. To ensure this, the proof uses the following observation: as a contract's L-inefficiency becomes arbitrarily small, so does the maximum gain from screening H (both quantities are closely related to w_n). Therefore, H-inefficiency must dominate the gain from screening provided that w_n is small

[17] In addition to Gul et al. (1986), see, e.g., Myerson (1991), Abreu and Gul (2000), Abreu and Pearce (2007), and Atakan and Ekmekci (2012, 2014).

enough. The next result captures this intuition. Results of Part I are proved in Appendix F.

Lemma 4 *There exist ε and D positive such that the following holds: if $w_m \le \varepsilon$, then $Q(R_n) \ge Q(E_H(R_m)) + D$ for $n \ge m$.*

Thus, even if R_m gives H a lower utility than the efficient contracts $E_H(R_n)$ arising from round m onwards, it is uniformly more costly to implement R_m than these contracts. Of course, the lemma applies to round 0 only if $w_0 \le \varepsilon$. However, if Theorem 2's conclusion holds for this case, the case $w_0 > \varepsilon$ can be easily ruled out by a discontinuity argument.[18]

Another challenge is that the inefficiency loss identified above is only incurred when facing H, but the probability of facing H becomes increasing smaller over time, conditional on staying in $\mathcal{H}$. This implies, as explained below, that a standard block construction with a fixed number of rounds in each block will not guarantee a geometric decrease in posteriors across blocks. To induce this geometric decrease, we need to offset the rather ineffective loss identified above by a sharper upper bound on P's gain from screening H. This is achieved by controlling H's continuation utility at the beginning of each block: the higher H's continuation utility at the next block, the lower P's expected gains from screening H *from that block onwards.*

Thus suppose that $w_0 \le \varepsilon$, and start Block 1 at round $n_0 = 0$. The size of Block 1 is determined endogenously: Let $\hat{u}_0 = u_H(n_0)$, $\hat{e}_0 = u_H(E_H(R_{n_0}))$, and $\hat{\beta}_0 = \beta_{n_0}$. Also define $\hat{u}_1$ by $\hat{e}_0 - \hat{u}_0 = t(\hat{u}_1 - \hat{u}_0)$, for a parameter $t > 1$ to be set shortly, and let $n_1 = \inf\{n : u_H(n) \ge \hat{u}_1\}$ denote the first round[19] at which H's continuation utility exceeds the threshold $\hat{u}_1$. The first round of Block 2—and, hence, the size of Block 1—is set to n_1.

The number of rounds in Block 1 must sufficiently large to guarantee that the probability of a breakdown and, hence, P's expected loss during the block, is significant. This number can be bounded below using H's continuation utility, which increases by steps of order η between consecutive rounds. Letting $u_H(n)$ denote H's continuation utility at the beginning

[18]Intuitively, w_n's decrements are of order η. If one shows (for any m) that $w_m \le \varepsilon \Rightarrow w_m = O(\eta)$, it follows that w_0 cannot exceed ε because the sequence w_n would otherwise have to drop by an impossibly large amount when it crosses ε. See Lemma 20 for the proof.

[19]Lemma 18 guarantees that H's continuation utility always reaches $\hat{u}_1$ in finite time.

of round n and $\Delta_H = \max_{C,C' \in \mathcal{C}} u_H(C) - u_H(C')$, H's Bellman equation implies the following result.

Lemma 5 *$u_H(n)$ is nondecreasing in n and satisfies $u_H(n+1) - u_H(n) \leq \eta \Delta_H$.*

Therefore, reaching $\hat{u}_1$ requires at least $\underline{n}(1) = \lfloor (\hat{u}_1 - \hat{u}_0)/\eta\Delta_H) \rfloor$ rounds.

To guarantee that significant screening takes place during any given block and, hence, that P's posterior goes down by a suitable factor, another challenge is to bound P's gain conditional on successfully screening the agent. This gain is closely related to the rent w_n that P takes away from H: Lemma 13 shows the existence of a Lipschitz constant a bounding this gain by aw_n.[20]

These observations deliver a bound on the probability $\hat{\mu}_0$ that H rejects all H-efficient contracts until round n_1. Indeed, at round n_0, P can always implement the contracts $(E_H(R_{n_0}), E_L(R_{n_0}))$, by Lemma 1. For this to be suboptimal, the gain from reducing H's rent through screening must outweigh the inefficiency loss from a negotiation breakdown, which yields the following condition.

Lemma 6 *$\hat{\mu}_0$ satisfies*

$$\hat{\beta}_0(1-\hat{\mu}_0)a(\hat{e}_0 - \hat{u}_0) + \hat{\beta}_0\hat{\mu}_0 a(\hat{e}_0 - \hat{u}_1) \geq \hat{\beta}_0\hat{\mu}_0 D\frac{\hat{u}_1 - \hat{u}_0}{\Delta_H}. \tag{2}$$

The last term on the left is of particular importance: it is an upper bound on P's expected gain, relative to jumping at round n_0, from reducing H's rent *after round* n_1. $\hat{\beta}_0\hat{\mu}_0$ is the probability of facing H *and* H staying in $\mathcal{H}$ until round n_1, and $\hat{u}_1$ is the smallest utility that P must provide to H at any round following n_1. This term is where the particular construction pays off: it captures the fact that P's maximal gain from screening H decreases across blocks.[21]

[20] Lipschitz constants are used in the bargaining and reputation literature (see, e.g., Cripps, Dekel, and Pesendorfer (2005)) where they arise naturally from the polyhedral structure of the utility sets. By contrast, one challenge of the present analysis is to deal with the nonlinear structure of model.

[21] Without this decrease, the left-hand side of (2) would reduce to the coarser bound $\hat{\beta}_0 a(\hat{e}_0 - \hat{u}_0)$, which may fail to provide a useful control on $\hat{\mu}_0$. This coarser bound guarantees that $\hat{\mu}_0 \leq \frac{a(\hat{e}_0 - \hat{u}_0)\Delta_H}{D(\hat{u}_1 - \hat{u}_0)}$, but the right-hand side exceeds $a\Delta_H/D$ which may be greater than 1.

Defining t by $t^2 = 1 + \frac{D}{a\Delta_H} > 1$ and rearranging (2) yields

$$\hat{\mu}_0 \leq \frac{1}{1 + D/a\Delta_H} \frac{\hat{e}_0 - \hat{u}_0}{\hat{u}_1 - \hat{u}_0} = \frac{1}{t}.$$

$\hat{\mu}_0$ is only an average rejection probability over all choice sequences which may occur during Block 1, but it implies (Lemma 19) the existence of a "pushdown" choice sequence for which the posterior $\hat{\beta}_1$ at round n_1 satisfies

$$\hat{\beta}_1 \leq \frac{\hat{\mu}_0 \hat{\beta}_0}{\hat{\mu}_0 \hat{\beta}_0 + (1 - \hat{\beta}_0)} \leq \hat{\beta}_0 \frac{t^{-1}}{\beta_0 t^{-1} + (1 - \beta_0)} = g\hat{\beta}_0.$$

where $g = \frac{1}{\beta_0 + (1-\beta_0)t} < 1$, since $t > 1$. β_n thus drops by a fixed factor along the pushdown sequence. We will construct pushdown sequences, one for each block, which succeed one another. The choice sequence obtained from stringing the pushdown sequences together, is such that the probability of facing H decreases geometrically across blocks.

To initiate the second block, define $\hat{e}_1 = u_H(E_H(R_{n_1}))$ and $\hat{u}_2 \in (\hat{u}_1, \hat{e}_1)$ by $\hat{e}_1 - \hat{u}_1 = t(\hat{u}_2 - \hat{u}_1)$, and let $\hat{\mu}_1$ denote the probability, seen from round n_1, that H accepts only contracts in $\mathcal{H}$ until $\hat{u}_2$ is reached. Let n_2 denote the first round at which $\hat{u}_2$ is exceeded. Repeating the earlier analysis, there exists a pushdown choice sequence for Block 2 such that the posterior $\hat{\beta}_2$ after observing this sequence until n_2 satisfies

$$\hat{\beta}_2 \leq \frac{\hat{\mu}_1 \hat{\beta}_1}{\hat{\mu}_1 \hat{\beta}_1 + (1 - \hat{\beta}_1)} \leq \hat{\beta}_1 \frac{t^{-1}}{\hat{\beta}_1 t^{-1} + (1 - \hat{\beta}_1)} \leq g^2 \hat{\beta}_0.$$

By induction, this defines a sequence of blocks k, rounds n_k, values $\hat{u}_k, \hat{e}_k$, and a pushdown sequence running through these blocks such that, letting $\hat{\beta}_k = \beta_{n_k}$,[22]

$$\hat{\beta}_k \leq g^k \hat{\beta}_0. \tag{3}$$

As noted earlier, negotiations do not end endogenously in finite time. Instead, the backward induction argument must be constructed from a round at which the agent has not yet been perfectly screened. This round is obtained by stopping the construction at the end of the first block, K, for which $\hat{w}_K = \hat{e}_K - \hat{u}_K < \bar{W}\eta$, where $\bar{W} = \max\{\frac{t(1+\hat{b}\beta_0/(1-\beta_0))}{t-1}(1 +$

[22]The actual value of $u_H(n_k)$ may be slightly above $\hat{u}_k$, but by no more than $\Delta_H \eta$, by Lemma 5. This maximal overshoot is negligible when computing a lower bound on the number of rounds in each block, because we stop the block construction when $\hat{u}_{k+1} - \hat{u}_k$ is still large relative to $\Delta_H \eta$, as explained below.

$\Delta_H), \frac{\hat{W}+\Delta_H}{t\Delta_H}\}$. The constants $\hat{b}$ and $\hat{W}$ are chosen to give $w_{n_K} = \hat{e}_K - u_H(n_K)$ and $\hat{\beta}_K$ appropriate lower bounds—a key ingredient for the backward induction explained below.[23]

The utility thresholds $\{\hat{u}_k\}_{k\leq K}$ guarantee a geometric decrease of w_n across blocks, which will be used to bound w_0 by backward induction on k.

LEMMA 7 *There exists $c_w > 0$ such that*

$$\begin{aligned} i)\quad & w_0 = \hat{e}_0 - \hat{u}_0 \leq c_w \left(\tfrac{t}{t-1}\right)^K \eta \qquad (4) \\ ii)\quad & w_{n_K} \geq \eta. \end{aligned}$$

To show that w_0 converges to zero with η, one must bound the factor $(t/(t-1))^K$ appearing in (4) and, hence, find an upper bound on K. This is achieved by computing a lower bound $\hat{\beta}$ for $\hat{\beta}_K$: since $\hat{\beta}_k$ is knocked down by a fixed factor g in each block, any lower bound $\hat{\beta}$ yields a bound on the number K of blocks needed to reach $\hat{\beta}$.

One bound for $\hat{\beta}_K$ comes from Lemma 3, which guarantees that $\beta_n \geq \underline{\beta} w_n^3$ for all n, combined with Lemma 7, which implies that $w_{n_K} \geq \eta$. This approach, described in Appendix C, can be used to prove that w_0 converges to 0 whenever the parameter t is high enough.

PROPOSITION 4 *Let $\hat{t}$ denote the unique solution of $3\ln(t-1) = 2\ln(t)$. If $t > \hat{t}$, then for all $\varepsilon > 0$ there exists $\bar{\eta}(\varepsilon)$ such that $w_0 \leq \varepsilon$ for all $\eta \leq \bar{\eta}(\varepsilon)$ and corresponding PBE.*

Since $t^2 = 1 + \frac{D}{a\Delta_H}$, Proposition 4 applies if a is small enough and/or D is large enough. These are the "easy" cases: Intuitively, D bounds below the loss incurred by P when he tries to screen H. The higher this loss, the lower P's incentive to screen H. Conversely, a bounds the maximal gain that P can achieve from reducing H's informational rent. The lower a, the weaker P's incentive to screen H.

This approach has limited power, however, because it builds on Lemma 3, which only exploits the inefficiency loss on type L to bound $\hat{\beta}_K$. Its success hinges on the relative speeds at which β_k and w_k decrease across consecutive blocks, which depends on parameter values.

[23]The block K must exist because $\hat{w}_k$ converges to zero as a result of the contracts' asymptotic efficiency (Proposition 2). The first term defining $\bar{W}$ guarantees that $w_{n_K} \geq \eta$, as shown by Lemma 7. The second one guarantees that the number of rounds in each block $k \leq K$ is bounded below by $\frac{\hat{u}_k - \hat{u}_{k-1} - \Delta_H \eta}{\Delta_H \eta} \geq \frac{1}{t\Delta_H \eta}(\hat{e}_{k-1} - \hat{u}_{k-1} - \Delta_H \eta) \geq \frac{\bar{W}\eta - \Delta_H \eta}{t\eta\Delta_H} > \hat{W}$, which can be made arbitrarily large by choosing $\hat{W}$ appropriately.

Moreover, this limitation cannot be overcome by changing the players' utility representation, improving the 1/3 exponent of Lemma 3, or changing the threshold t used in the construction (cf. Appendix C).

Establishing Theorem 2 for all primitives of the model requires a more sophisticated analysis, which exploits and combines inefficiency losses on both types to derive a sharper bound on $\hat{\beta}_K$. The key is to show the following proposition.

Proposition 5 *There exist positive constants η^* and β^* such that for any PBE associated with $\eta < \eta^*$ and any round N, $\beta_N \leq \beta^*$ implies that $w_N < \eta$.*

Since $w_{n_K} \geq \eta$, Proposition 5 implies that $\beta^* \leq \hat{\beta}_K \leq g^K \beta_0$. This bounds K independently of η by $\frac{\ln(\beta_0/\beta^*)}{\ln(-g)}$. Lemma 7 then implies that $w_0 = O(\eta)$, concluding the proof.

Part II: Proving Proposition 5

Proposition 5 provides a much sharper bound on w_N than the one guaranteed by Lemma 3. Its proof is based on a self-reinforcing phenomenon: roughly put, the smaller β_n gets, and the more *persistent* the inefficiency of the contract R_n becomes—whatever its current level was—which causes β_n to drop faster as more screening is needed to offset the inefficiency, and makes the inefficiency even more persistent, etc. Since all contract sequences are asymptotically efficient, this deceleration is compatible with equilibrium only if the contract R_N at the beginning of the deceleration is sufficiently efficiency and, hence, w_N is sufficiently small.

The link between β_n and inefficiency persistence at the heart of this spiral will stem from two analytical results. First, the L-inefficiency of R_n is well captured by the index w_n. Second, while w_n must converge to zero, its decrements between consecutive rounds must be of order at most $\sqrt{\beta_n}$, owing to the geometry of the problem.[24]

To formalize this phenomenon, the techniques developed in this section require the use of an "orderly" choice sequence, i.e., one along which β_n is decreasing and w_n remains sufficiently small across all rounds. Proving that such a sequence exists is challenging. In this model, contracts and beliefs are not a priori well behaved: some contract choice may

[24]The key step is Lemma 15, which uses analytical geometry near the efficiency curve $\mathcal{E}_L$ and exploits convexity of the payoff functions to derive a quadratic lower bound on the cost function as one moves away from $\mathcal{E}_L$ along L's isoutility curves.

create a spike up in β_n or dramatically increase, at least temporarily, the inefficiency of current contracts. Indeed, until one proves that the equilibrium starting from any contract and belief is essentially unique (which is what Theorem 2 achieves), there may a priori be a whole range of continuation utilities for all players and types, used to sustain very complex menu proposals and contract choices.

The formal analysis starts from the round N appearing in Proposition 5. For any round $n \geq N$ and $R_{n+1} \in M_n \cup \{R_n\}$, let $\mu_n^\theta(R_{n+1})$ denote the probability that θ accepts R_{n+1}. Because P can immediately jump to the efficient contracts, his IC constraint implies that

$$\beta_n a w_n \geq \sum_{R_{n+1} \in (M_n \cup \{R_n\}) \cap \mathcal{H}} \beta_n \mu_n^H(R_{n+1}) \eta D + (1-\beta_n) \mu_n^L(R_{n+1}) \eta (Q(R_{n+1}) - Q(E_L(R_n))) \quad (5)$$

$$= \sum_{R_{n+1} \in (M_n \cup \{R_n\}) \cap \mathcal{H}} \mu_n^L(R_{n+1}) \left[\beta_n \mu_n(R_{n+1}) \eta D + (1-\beta_n) \eta (Q(R_{n+1}) - Q(E_L(R_n)))\right] \quad (6)$$

where $\mu_n(R_{n+1}) = \mu_n^H(R_{n+1})/\mu_n^L(R_{n+1})$.[25] The left-hand side of (5) is an upper bound on P's gain, relative to the immediate jump, from reducing H's rent by w_n. This gain is bounded above by $a(u_H(E_H(R_n)) - u_H(n)) = a w_n$.[26] The first term of the right-hand side captures the inefficiency loss incurred if the agent is of type H, rejects the H-efficient contract in round n, and the breakdown occurs in round n. This loss is bounded below by D as long as $w_n \leq \varepsilon$, which will be true along the choice sequence considered (cf. Lemma 4). The last term is the net loss if the agent is of type L and the breakdown occurs in round n.

The right-hand side of (6) is a convex combination of terms indexed by R_{n+1}, so there must exist some $R_{n+1} \in (M_n \cup \{R_n\}) \cap \mathcal{H}$ such that

$$w_n a \beta_n \geq \beta_n \mu_n(R_{n+1}) \eta D + (1-\beta_n) \eta (Q(R_{n+1}) - Q(E_L(R_n))). \quad (7)$$

[25] We can assume without loss that $\mu_n^L(R_{n+1})$ is strictly positive for all $R_{n+1} \in (M_n \cup \{R_n\}) \cap \mathcal{H}$: First, if some contracts in this set are not chosen with positive probability, a payoff equivalent equilibrium may be constructed without these contracts. Second, if any contract R'_{n+1} is chosen only by H, Proposition 1 implies that H gets the H-efficient contract C that gives him the same utility as R'_{n+1} in the next round. The equilibrium can therefore be modified by having P immediately propose C instead of R'_{n+1}, a change which reduces P's cost and does not affect incentives.

[26] See Lemma 13. The bound is computed using a best-case scenario for P, in which H accepts with probability 1 the H-efficient contract C_n providing utility $u_H(n)$. It is an upper bound on P's gain because C_n is the least costly way of providing H with his continuation utility.

Therefore, there exists a choice sequence that satisfies (7) for all $n \geq N$. The analysis focuses on this sequence, which will be called an **orderly** sequence.

Orderly sequences are useful for the following reason. Let $\mu_n = \mu_n(R_{n+1})$ denote the likelihood ratio associated with observing R_{n+1}. Provided that w_n is small enough, (7) implies that $\mu_n \leq 1$ and thus $\beta_{n+1} \leq \beta_n$. Therefore, β_n decreases along any orderly sequence *as long as w_n remains small enough.* Even if w_N is small, however, there is no a priori guarantee that w_n—and, hence β_n—remain small in subsequent periods. These problems are intertwined: from inequality (18) in Appendix A,

$$w_{n+1} \leq w_n \frac{1 + \alpha\beta_n}{1 - \beta_n}$$

for some constant α. Therefore, if β_n decreases sufficiently fast, w_n cannot increase too much, which guarantees that β_n continues to drop, and so on. This is formalized by the following result. (All results of Part II are proved in Appendix G.)

LEMMA 8 *There exist $\hat{\eta}$, $\hat{\beta}$, and $\hat{w}$ positive such that for any $\eta < \hat{\eta}$ and associated PBE, and any round N such that $\beta_N \leq \hat{\beta}$ and $w_N \leq \frac{D\eta}{2a}$, the following holds for all $n \geq N$: i) β_n is decreasing in n, ii) $\mu_n \leq 3/4$, and iii) $w_n \leq \hat{w}\eta$. Moreover, we have*

$$\mu_n \leq \frac{w_n a}{\eta D}, \tag{8}$$

$$\beta_n w_n a \geq (1 - \beta_n)\eta(Q(R_{n+1}) - Q(E_L(R_n))). \tag{9}$$

Equation (8) shows that if P's potential gain from screening H (proportional to w_n) is small enough (of order η), then H must reveal himself *in each round* with significant probability for the immediate jump not to be a profitable deviation. This happens when P's maximum screening gain *over the entire continuation of the game*, aw_n, is of the same order as the expected loss from a breakdown in a *single* period. The virtue of orderly sequences is to convert this ex ante condition into an ex post drop in P's posterior.

Equation (9) is where the loss on L is exploited: if β_n is small, it imposes an upper bound on the L-inefficiency $Q(R_{n+1}) - Q(E_L(R_n))$ of any contract R_{n+1} in the orderly sequence. This implies that R_{n+1} is very close to $E_L(R_n)$, and it follows that H's continuation utility increases extremely slowly, because his "flow utility" (i.e., the utility he gets from R_{n+1} if a breakdown occurs in the current round) is almost indistinguishable from his continuation utility. Formally, H's Bellman equation implies that

$$u_H(n+1) - u_H(n) = \eta(u_H(n+1) - u_H(R_{n+1})), \tag{10}$$

and almost-L-efficiency of R_{n+1} implies that the difference on the right is arbitrarily small. Since $w_n = u_H(E_H(R_n)) - u_H(n)$, the difference $w_{n+1} - w_n$ is closely related to the left-hand side of (10), which suggests that w_n must decrease very slowly along the orderly sequence. This is formalized by the next result.

LEMMA 9 *There exists a positive constant A_w such that*

$$w_{n+1} \geq w_n - \eta A_w \sqrt{\beta_{n+1}}. \tag{11}$$

Long-run negotiations are thus dominated by two forces: First, the posterior belief β_n must decrease sufficiently fast for the screening activity to be sufficiently profitable (H-loss force). Second, w_n must decrease very slowly—by an order of $\sqrt{\beta_{n+1}}$—because the contract must be almost L-efficient (L-loss force). This leads to the crux of the proof: as β_n decreases through screening, the decrements of w_n become smaller in absolute value, which causes L-inefficiency of the current contracts to persist, and forces to β_n decrease faster for screening to offset the H-inefficiency. This accelerates β_n's drop to zero and decelerates w_n's own drop to 0, which may cause w_n to stall altogether. Since w_n must converge to zero along all sequences, this scenario is only possible if w_N was small enough to begin with.

The first part of this intuition is captured by the following lemma, which says that if w_N is too high relative to $\eta\beta_N$, w_n's *decreasing* factor goes to 1 provided that β_N lies below some threshold. Let $c = \frac{a}{D}A_w$ and $\hat{c} = \frac{4Dc^2}{a}$.

LEMMA 10 (DECELERATION) *If $w_N \geq \hat{c}\eta\beta_N$ and $\beta_N \leq c^{-2}/16$, then* $\liminf_{n\to+\infty} \frac{w_{n+1}}{w_n} \geq 1$.

The second part of the intuition is captured by a mathematical lemma which, applied to the ratio $q_n = \frac{aw_n}{\eta D}$, implies that w_n cannot converge to 0 if its decreasing factor goes to 1.

LEMMA 11 (STALLING) *Consider a positive sequence $\{q_n\}$ and constants $c' > 0$ and $N \in \mathbb{N}$ such that i) $q_n - q_{n+1} \leq c'\sqrt{\Pi_N^n q_k}$ for $n \geq N$, and ii)* $\liminf_n \frac{q_{n+1}}{q_n} \geq 1$. *Then, $\{q_n\}$ does not converge to zero.*

Equation (11) implies that $q_n = aw_n/\eta D$ satisfies premise i). A simple contradiction argument based on the previous lemmas then implies the following proposition. Let $\tilde{\beta} = \min\{\hat{\beta}, c^{-2}/16\}$.

Proposition 6 *There exists $\tilde{\eta} > 0$ such that for any $\eta \leq \tilde{\eta}$,*

$$\beta_N \leq \tilde{\beta} \text{ and } \quad w_N \leq \frac{\eta D}{2a} \qquad \Rightarrow \qquad w_N \leq \hat{c}\eta\beta_N.$$

If we could ignore the premise $w_N \leq \frac{\eta D}{2a}$, this result would prove Proposition 5 with $\beta^* = \min\{\hat{c}^{-1}, \tilde{\beta}\}$. Part III, below, constructs a round that satisfies this premise.

Part III: Bridging Argument

Propositions 5 and 6 deliver a lower bound for β_N at the end of block K when $w_N \leq \frac{D\eta}{2a}$. Intuitively, a similar bound should apply if w_N exceeds $\frac{\eta D}{2a}$: if P is taking more rent away from H, thus incurring losses on L, it better be the case that the probability of facing H is non-negligible. To formalize this intuition, starting from $w_N \geq \frac{\eta D}{2a}$, a natural strategy is to follow a choice sequence along which w_n and β_n are both decreasing. When w_n crosses $\frac{\eta D}{2a}$, Proposition 6 implies that β_n is greater than $\frac{\hat{c}w_n}{\eta}$, which yields a lower bound for β_n and, since the sequence $\{\beta_m\}$ was decreasing, for β_N. The implementation of this idea raises two challenges.

First, when w_n drops below $\frac{\eta D}{2a}$, it may jump to an arbitrarily low level, such as η^2, rendering the lower bound $\frac{\hat{c}w_n}{\eta}$ on β_n useless. To avoid this, one must select a choice sequence along which w_n's decrements are sufficiently small. This will be achieved by exploiting the loss on L: as noted earlier and developed below, $w_{n+1} - w_n$ is closely related to the L-inefficiency of the contract R_{n+1}.

Second, there need not exist a sequence along which β_n and w_n are both decreasing. Instead, we will construct a sequence along which these variables cannot increase "too much," and break up the sequence into blocks of constant size—unlike those of Part I—across which β_n drops by a constant factor.

A similar point arose in Part II, in which we had to focus on orderly sequences, but the approach here is different. Methodologically, it is common in the screening literature to bound a type's average probability of rejection over all possible choice sequence realizations in a block, and then extract from it one sequence which satisfies the appropriate drop in belief. The challenge here and in Part II is that more properties are required of the sequence extracted. In Part II, we imposed a low value on w_N and used it to directly construct an orderly sequence. Here, instead, we must first work globally, by showing that the set of sequences with the desired properties has a high enough probability in each block, and deduce from this that one may extract a sequence *within that set* that also satisfies the standard drop in belief.

To analyze this problem, it suffices to consider the case $w_N \leq \bar{W}\eta$, which was used to determine the last block of Part I.[27] To address the first challenge, we introduce the variable $y_n = u_H(E_H(R_n)) - u_H(R_{n+1})$, represented on Figure 4. y_n provides the control required on the decrements of w_n: subtracting $u_H(E_H(R_n))$ from H's Bellman equation (equation (10) in Part II) and rearranging it yields

$$w_{n+1} = w_n - \eta y_n + \eta w_{n+1} + (1-\eta)(u_H(E_H(R_{n+1})) - u_H(E_H(R_n))). \tag{12}$$

The control of w_n's decrements through y_n stems from the following lemma (all proofs for this part are in Appendix F). Fix a positive integer $\bar{N}$ and a small positive constant $\bar{\varepsilon}$.

Lemma 12 *There exist positive constants k_y, k_w, and k_ε with the following properties. Consider any round $\bar{n}$ such that $w_{\bar{n}} \leq \bar{W}\eta$ and $\beta_{\bar{n}} \leq \bar{\varepsilon}^{\bar{N}}$, and let $\mathcal{S}$ denote the event that the agent chooses in all rounds $n \in \{\bar{n}, \ldots, \bar{n} + \bar{N} - 1\}$ contracts such that $y_n \leq k_y \bar{\varepsilon}^{1/4}$, $\beta_n \leq \bar{\varepsilon}$, and $w_n \leq k_w \eta$. For η small enough, the probability of $\mathcal{S}$ is greater than $(1 - k_\varepsilon \sqrt{\bar{\varepsilon}})^{2\bar{N}}$.*

Lemma 12 implies, for $\bar{\varepsilon}$ and $\beta_{\bar{n}}$ small enough, that choice sequences along which y_n is less than any fixed amount for at least $\bar{N}$ rounds have probability almost 1. We now construct blocks of fixed size $\bar{N}$, to be set shortly. The first block starts at round N, the second one at $N + \bar{N}$, etc. In the first round, $\bar{n}$, of any given block, P's IC constraint implies that, letting $e_{\bar{n}} = u_H(E_H(R_{\bar{n}}))$,

$$\beta_{\bar{n}} a \left\{(1-\mu_{\bar{n}})(e_{\bar{n}} - u_H(\bar{n})) + \mu_{\bar{n}}(e_{\bar{n}} - E[u_H(\bar{n} + \bar{N})])\right\} \geq \beta_{\bar{n}} \mu_{\bar{n}} D \eta \bar{N}$$

[27] If w_N lies above this value, the block construction of Part I can be used to decrease β_n block by block until reaching $\bar{W}\eta$. Any lower bound on the posterior $\hat{\beta}_K$ once $\bar{W}\eta$ is reached also applies to β_N at the beginning of the block construction, since β decreases across blocks.

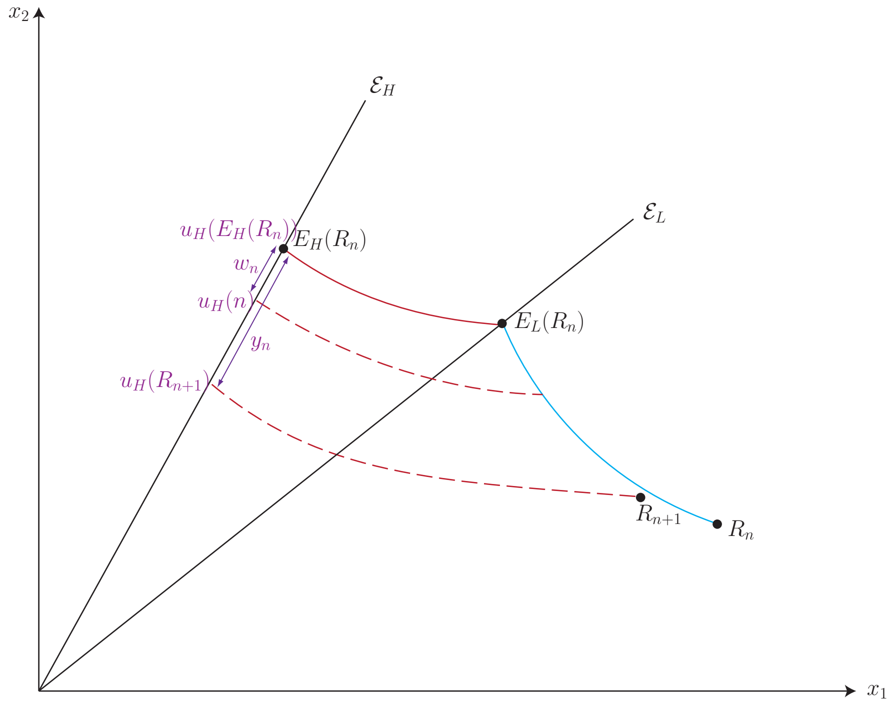

Figure 4: Representation of y_n and w_n.

where $\mu_{\bar{n}}$ is the probability, seen from round $\bar{n}$, that H rejects all H-efficient contracts between rounds $\bar{n}$ and $\bar{n}+\bar{N}$. The argument behind this inequality is similar to the one used in Part I, the main difference being the expectation of $u_H(\bar{n}+\bar{N})$ at the end of the block, which is no longer controlled. The loss bound D is valid by virtue of Lemma 4 and the fact that $w_{\bar{n}}$ is of order η and hence less than ε, for η small enough, and the probability of a breakdown during the block is $1-(1-\eta)^{\bar{N}} \sim \eta\bar{N}$. This implies that

$$\mu_{\bar{n}} \leq \frac{a(e_{\bar{n}} - u_H(\bar{n}))}{a(E[u_H(\bar{n}+\bar{N})] - u_H(\bar{n})) + D\eta\bar{N}} \leq \frac{a\bar{W}}{D\bar{N}}, \tag{13}$$

where the second inequality stems from the inequality $Eu_H(\bar{n}+\bar{N}) - u_H(\bar{n}) \geq 0$, which holds by monotonicity of $u_H(n)$, and the fact that $w_{\bar{n}} = e_{\bar{n}} - u_H(\bar{n}) \leq \bar{W}\eta$.[28]

These blocks are simpler than those of Part I because their size $\bar{N}$ is fixed and H's utility at the end of each block is "free." Two reasons motivate this difference: First, H's utility increments are now so small that they no longer provide a useful lower bound on the number

[28] This inequality holds without loss of generality as explained in Remark 3 of the Online Appendix.

of rounds in each block. Second, P's potential gain from screening H is now so small that it is no longer useful to distinguish how much rent can taken from H across different blocks.[29]

$\bar{N}$ is set so that $a\bar{W}/D\bar{N}$ and, hence, $\mu_{\bar{n}}$, be less than $\frac{1}{4}$. As in Part I, $\mu_{\bar{n}}$ is an average probability over possible choice sequences. And as in Part I, we wish to select a pushdown sequence along which the posterior at the end of the block is pushed down by this probability. However, here we also wish to select the sequence so that y_n remains small. Fortunately, Lemma 12 guarantees that these sequences have probability almost 1 provided that $\bar{\varepsilon}$ is small enough, and hence guarantees the dual goal of having small decrements for w_n and reducing the posterior β_n across blocks. Formally, starting from N, let M denote the first round at which w_n drops below $\eta D/2a$. The next result guarantees that w_M is bounded below and that β_M is bounded above in fixed proportion to β_N.

Proposition 7 (Bridge) *There exist k_β and $\check{\beta}$ such that if $w_N \in (\frac{\eta D}{2a}, \bar{W}\eta)$ and $\beta_N \leq \check{\beta}$, then there is a choice sequence such that 1. $w_M \geq \frac{\eta D}{4a}$ and 2. $\beta_M \leq k_\beta \beta_N$, where $M \geq N$ is the first round such that $w_M \leq \frac{\eta D}{2a}$.*

Proposition 5 follows easily: provided that β_N lies below some appropriate threshold related to k_β and $\check{\beta}$, β_M will be small enough to apply Proposition 6 of Part II to round M. Since, also, w_M is bounded below by Part 1 of Proposition 7, this provides an η-independent lower bound for β_M and, hence, for β_N. The proof is in Appendix H.

6 Discussion

6.1 Related Literature

Contract renegotiation with private information has traditionally been studied from two perspectives. The first one is axiomatic and focuses on "renegotiation-proof" contracts.[30] It essentially *assumes* that renegotiation leads to an efficient contract, ignoring the difficulties

[29]Also note that the bound computed in (13) uses the fact that $w_{\bar{n}} \leq \bar{W}\eta$, but if $w_{\bar{n}}$ only satisfied an η-independent bound, as in Part I, the right-hand side of (13) would be of order $1/\eta$ and thus useless.

[30]See Dewatripont (1989), Maskin and Tirole (1992), Battaglini (2007), Maestri (2015), and Strulovici (2011, 2013). A similar approach has been used to study renegotiation in repeated games with complete information. See, e.g., Bernheim and Ray (1989) and Farrell and Maskin (1989).

which may arise when some party holds private information. The second one focuses on simple renegotiation protocols, in which the principal makes a single take-it-or-leave-it offer. This perspective typically results in *inefficient* contracts.[31] The second perspective seems incomplete: what, in reality, prevents the principal from proposing a new contract after learning the inefficiency of the current contract? Such a restriction amounts to a strong form of commitment for the principal and can even result in full-commitment outcomes.[32] This paper shows that, by dropping the restriction on the number of negotiation rounds, one can reconcile these two perspectives.[33]

While the model's main interpretation concerns the implementation of a single contract arising from a sequence of temporary agreements, it has an alternative interpretation: In each round, the current contract is implemented and parties receive the corresponding payoff, and the interruption probability is reinterpreted as the discount rate between periods.[34] In each period, the principal has an opportunity to renegotiate the contract for future periods. The model is formally equivalent to an infinite-horizon version of Hart and Tirole (1988), with divisible goods as in Laffont and Tirole (1990), arbitrary utility and cost functions, and in which the contract is constant until renegotiated. Thus interpreted, this paper shows that *all* equilibria become efficient as the discount rate goes to zero.

Maestri (2015) analyzes renegotiation-proof equilibria in a divisible-good, binary-type version of the Hart-Tirole framework and proposes a renegotiation-proofness refinement of PBE, formulated recursively: In each period, the contracts offered by the principal maximize his revenue among all renegotiation-proof continuations, given the current contract and

[31]See Hart and Tirole (1988) and Fudenberg and Tirole (1990).

[32]For example, imposing any *finite* number k of negotiation opportunities results in the full-commitment outcome: the principal simply passes the first $k-1$ opportunities to negotiate the contract, and then proposes the full commitment allocation in the last round. Wang (1998) considers a more flexible protocol, in which the principal proposes contracts until an agreement is reached. Such protocol leaves a high commitment power to the principal, since he cannot renegotiate any agreement. Indeed, Wang shows that the principal achieves the full commitment allocation, which is also ex post inefficient. See also Beaudry and Poitevin (1993) and Matthews (1995).

[33]Brennan and Watson (2013) study another friction in the form of explicit renegotiation costs. Another extension of the model would consider general negotiation protocols, as in Rubinstein and Wolinsky (1992).

[34]The formal equivalence appears clearly in the payoff formulas of page 8: $1-\eta$ plays the role of the discount factor, and η is the weight put on the current period.

beliefs. The principal can propose at most two contracts in each period and the probability of facing the high type decreases monotonically as long as this type is not fully revealed. As the discount rate goes to zero, all renegotiation-proof equilibria become efficient. By contrast, this paper implies that all PBEs become efficient, allowing for arbitrarily large menus of contracts, general utility and cost functions, and an infinite horizon, and focusing on contracts which are constant until renegotiated.[35]

Finally, the paper is related to the literature on bargaining and reputation, in which some players are trying to determine whether other players have a "commitment" type.[36] This paper presents specific challenges described in Section 5 and differs from reputation models in other ways: i) the "actions" of the players are endogenous, because the principal sets the contract menu in each round, ii) the state space includes the current contract, in addition to the principal's belief, and iii) all types are strategic.[37]

6.2 Extensions

The analysis has focused on two types. With more types, it is natural to conjecture that ex post efficiency still obtains. Appendix L explains how the techniques developed here can be

[35] Conceptually, the papers' objectives are complementary: while Maestri studies renegotiation-proof equilibria when trade happens over time, the present paper shows—under its main interpretation—that if parties have frequent opportunities to bargain before a one-time trade decision, all equilibria will be approximately efficient. It thus provides a non-cooperative foundation for renegotiation-proof contracts. The technical contributions are also different. For example, an important part of Maestri's analysis lies in its conceptual work to define renegotiation-proof contracts. By contrast, this paper analyzes all equilibria, which creates important challenges described in earlier sections.

[36] Fudenberg and Levine (1989), Schmidt (1993), Abreu and Gul (2000), Cripps et al. (2005), and Atakan and Ekmekci (2012, 2014). Abreu and Pearce (2007) consider a repeated game setting in which players bargain over commitments for the continuation of the game.

[37] The richer state space and the nonlinear geometry of the problem (the parties' utility functions are nonlinear) make the problem particularly difficult to analyze. Some differences can formally be incorporated into the standard reputation framework. For example, the action space of the agent may be assumed to be fixed by setting a large negative default value if the agent chooses anything outside of the principal's proposed set. This formal equivalence does not resolve the substantive differences between the settings.

used to make progress in this direction.[38]

The model featured, under its main interpretation, a single "delivery" time at which the contract is implemented. To provide a foundation for renegotiation-proof contracts with multiple deliveries, one should consider a more general model with multiple "physical" events (payments or efforts are made, exogenous information arrives, etc) and renegotiation protocols like the one studied here inserted between consecutive physical events. Each renegotiation protocol would then pertain to continuation contracts over the remaining horizon, and any event arrival would trigger the end of current negotiations.

Appendices

A Inequalities

This appendix presents key inequalities used in other appendices and derived in Appendix E.

LEMMA 13 (REGULARITY BOUNDS) *There exist positive constants $\underline{a}, a, \underline{b}, b$ such that for any $C, \hat{C} \in \mathcal{E}_H$ such that $u_H(C) < u_H(\hat{C})$,*

$$\underline{a}(u_H(\hat{C}) - u_H(C)) \leq Q(\hat{C}) - Q(C) \leq a(u_H(\hat{C}) - u_H(C)) \tag{14}$$

$$\underline{b}(Q(\hat{E}) - Q(E)) \leq Q(\hat{C}) - Q(C) \leq b(Q(\hat{E}) - Q(E)), \tag{15}$$

where E (resp. $\hat{E}$) is the L-efficient contract that gives H the same utility as C (resp. $\hat{C}$).

Let Q_L denote P's expected cost, view from round n, conditional on facing L.

LEMMA 14 (INCENTIVE BOUNDS) *Given any PBE and choice sequence $\{R_n\}$, there exist positive constants α, γ, b, and $\hat{b}$ such that, for any n,*

$$Q_L \leq Q(E_L(R_n)) + \frac{\beta_n}{(1-\beta_n)} a w_n, \tag{16}$$

$$u_H(E_H(R_{n+1})) - u_H(E_H(R_n)) \leq \frac{\alpha\beta_n}{1-\beta_n} w_n, \tag{17}$$

$$w_{n+1} \leq w_n \left(1 + \frac{\alpha\beta_n}{1-\beta_n}\right), \tag{18}$$

$$u_L(R_n) - u_L(R_{n+1}) \leq \gamma\beta_{n+1} w_{n+1}, \tag{19}$$

$$w_{n+1}(1 + b\beta_{n+1}) \geq w_n - \eta y_n, \tag{20}$$

[38]With three types H, M, and L, for instance, the contracts $\{E_\theta(R_0)\}_{\theta \in \{H,M,L\}}$ would be the least costly θ-efficient contracts which are incentive compatible and individual rational, starting at R_0.

and, for any $n < n'$,

$$u_H(E_H(R_{n'})) - u_H(E_H(R_n)) \geq -\frac{\hat{b}\beta_{n'}}{1-\beta_{n'}} w_{n'}. \quad (21)$$

LEMMA 15 (GEOMETRIC BOUND) *There exists $\underline{q} > 0$ such that for any $C \in \mathcal{E}_L$ and $R \in \mathcal{H}$ such that $u_L(R) = u_L(C)$,*

$$Q(R) - Q(C) \geq \underline{q}(u_H(C) - u_H(R))^2.$$

LEMMA 16 *There exist positive constants k_2 and k_3 such that*

$$y_n^2 \leq k_2[Q(R_{n+1}) - Q(E_L(R_n))] + k_3(\max\{(\beta_n w_n/(1-\beta_n))^2, (\beta_{n+1} w_{n+1})^2\} + \beta_{n+1} w_{n+1}) \quad (22)$$

B Proofs of Section 4

PROOF OF PROPOSITION 1

Part i) Let $\bar{u}$ denote the agent's supremum over his expected utility, given his type θ, over all possible continuation PBEs starting from R_0 at which P puts probability 1 on type θ, and let $u = u_\theta(R_0)$. Suppose by contradiction that $\bar{u} > u$. By time homogeneity, $\bar{u}$ will be the same in the next round if the agent rejects new offers from P in round 0 and P continues to assign probability 1 on facing type θ. In such case, the agent's continuation payoff is bounded above by $\tilde{u} = \eta u + (1-\eta)\bar{u} < \bar{u}$. Consider any PBE that gives θ an expected utility $u_0 \in (\tilde{u}, \bar{u})$ (such PBE must exist, by definition of $\bar{u}$). Suppose that the principal deviates by proposing the θ-efficient contract C that give θ a utility level u' in $(\tilde{u}, u_0)$. By definition of a PBE,[39] P continues to assign probability 1 to type θ after his *own* deviation. If the agent accepts C with probability 1, the deviation is strictly profitable to P since C is the least costly way of providing utility $u' < u_0$ to the agent. If the agent rejects the offer with positive probability, then by Bayes rule, P must continue to assign probability 1 to type θ, which implies that his continuation utility is bounded above by $\bar{u}$. Therefore, the agent's rejection is strictly suboptimal, implying that the agent must accept C with probability 1 and the deviation is profitable.[40] Let $\underline{Q}$ denote the cost of the θ-efficient contract, $\underline{C}$, that

[39] See Fudenberg and Tirole (1991), part iii) of the definition.

[40] The continuation play after P's deviation must be a PBE of the corresponding continuation game. Therefore, if θ's continuation strategy, after P's deviation, is to reject the proposed deviation with positive probability, Bayes rule applies. I am grateful to Marcin Peski for proposing the current version of this argument.

provides utility $\underline{u}$ to θ. Clearly, any PBE must cost exactly $\underline{Q}$, otherwise P has a profitable deviation which is to propose the θ-efficient contract that gives θ slightly more than $\underline{u}$ and costs less than following the PBE. Moreover, the only way of achieving $\underline{Q}$ is to propose $\underline{C}$ in the first round and have it accepted with probability one.

Part ii) Suppose without loss that $\theta = L$ (the opposite case is treated identically). Let $u_L = u_L(R_0)$ and $u_H = u_H(R_0)$. Also let $\bar{u}_H(\beta)$ denote the supremum utility that H can achieve over any continuation PBE starting from R_0 when P assigns probability β to H, and let $\bar{u}_H = \sup_{\beta\in[0,1]} \bar{u}_H(\beta)$. Suppose by contradiction that $\bar{u}_H > u_H$. Then, for any small $\varepsilon > 0$, there exists $\bar{\beta}$ and an associated PBE for which H's continuation utility is above $\bar{u}_H - \varepsilon > u_H$. For that PBE, because L gets at least u_L and R_0 is L-efficient, we have $\bar{Q}_L \geq Q$ where $Q = Q(R_0)$ and $\bar{Q}_L$ is P's expected cost in that PBE conditional on facing θ_L. Since not proposing any new contract is always feasible for P, and costs Q, the continuation cost $\bar{Q}_H$ conditional on facing H must satisfy $\bar{Q}_H \leq Q$ to offset the weakly higher cost conditional on facing L. Suppose that P deviates from that PBE by proposing the H-efficient contract that gives θ_H utility $\bar{u}_H - \varepsilon - \epsilon$, for arbitrarily small ϵ. Because, for small enough ε and ϵ, $\bar{u}_H - \varepsilon - \epsilon > \eta u_H + (1-\eta)\bar{u}_H$, H accepts this proposal with probability 1. For any strategy that θ_L chooses and continuation equilibrium, this proposal strictly reduces P's expected cost (since $\bar{Q}_H \leq Q$), yielding a contradiction. This shows that $\bar{u}_H(\beta) = u_H$ for all β.[41] To conclude, suppose that P proposes the H-efficient contract that gives H utility $u_H + \tilde{\epsilon}$, for $\tilde{\epsilon}$ arbitrarily small. From the previous observation, H must accept that contract regardless of L's strategy. This shows that P can and, hence, does achieve the full-commitment optimal cost under any PBE. This proves Part ii).

Part iii) Suppose that $Q_L \geq Q_H$ where $Q_\theta = Q(E_\theta(R_0))$ (the opposite case is proved symmetrically) and let $\bar{Q}$ denote the maximal expected cost incurred by P over all PBEs and beliefs $\beta \in [0,1]$, starting from R_0.

We start by showing that $\bar{Q} \leq Q_L$. Suppose by contradiction that $\bar{Q} > Q_L$ and consider any PBE that achieves $\bar{Q}$.[42] Now suppose that P deviates by proposing the pair $\tilde{C}_L, \tilde{C}_H$ of

[41] If $\beta = 0$, P does not propose anything new, from i) and L-efficiency of R_0, and the result trivially holds in that case too.

[42] If the supremum $\bar{Q}$ is not achieved, the argument below can easily be adapted by considering a PBE whose expected cost is arbitrarily close to $\bar{Q}$.

contracts such that $\tilde{C}_\theta$ is efficient for θ and costs $\bar{Q}-\varepsilon$ for some ε arbitrarily small compared to η. Those contracts maximize each type's utility subject to costing P at most $\bar{Q}-\varepsilon$. Because these contracts are efficient and incentive compatible, Part ii) guarantees that no type ever chooses the contract meant for the other type. Moreover, no matter what belief and continuation PBE follows rejection of these contracts, P's continuation cost must be less than $\bar{Q}$, by definition of $\bar{Q}$. But this latter bound implies that there must be at least one type θ of the agent who is getting a lower payoff if he rejects $\tilde{C}_\theta$ than if he accepts it: conditional on rejection P has to be spending weakly less on at least one type of the agent than under $\tilde{C}_\theta$ (up to ε, which is negligible compared to η). Moreover, the contract $\tilde{C}_\theta$ maximizes this type's utility subject to P spending less than $\tilde{C}_\theta$. Since rejection leads to a renegotiation breakdown with probability η, giving this type a strictly lower utility than $\tilde{C}_\theta$, it is strictly better for this type to accept $\tilde{C}_\theta$ with probability 1. As a result, a rejection fully reveals that the agent is of the other type. From Part i), that type agent gets $u_\theta(R_0)$ after the rejection, which is strictly less than the utility he gets from $\tilde{C}_\theta$—since this contract maximizes the agent's utility subject to a higher cost than what P incurs with R_0. Therefore, both types accept their contract, and this reduces the cost of the principal strictly below $\bar{Q}$, showing that this is a profitable deviation. Thus, necessarily, $\bar{Q} \leq Q_L$.

Since L gets utility at least $u_L(R_0)$ in any PBE, and Q_L is the least cost for providing this utility, in all PBEs starting with $\beta \in (0,1)$, P must spend weakly less than $\bar{Q}$ on H to guarantee that $\bar{Q} \leq Q_L$. Let $\bar{u}_H$ denote the supremum expected utility that H gets over all PBEs and beliefs $\beta > 0$. Since P spends less than Q_L on H, $\bar{u}_H$ is bounded by the utility $\hat{u}_H$ obtained from the H-efficient contract $\hat{C}_H$ that costs Q_L. We will show that $\bar{u}_H = u_H(E_H(R_0))$. Suppose by contradiction that $\bar{u}_H > u_H(E_H(R_0))$, and consider a PBE that achieves $\bar{u}_H$.[43] The expected cost Q from that PBE must be above $\beta Q(\bar{C}_H) + (1-\beta)Q_L$ where $\bar{C}_H$ is the H-efficient contract that gives utility $\bar{u}_H$ to H. Suppose that P deviates by proposing the contracts $\tilde{C}_L, \tilde{C}_H$ such that $\tilde{C}_L$ is L-efficient and gives utility $u_L(C) + \varepsilon^2$ to L and $\tilde{C}_H$ is H-efficient and gives utility $\bar{u}_H - \varepsilon$ to H, for ε small compared to η. H accepts $\tilde{C}_H$, since rejection leads to a continuation utility bounded above by $\bar{u}_H$ and to a strictly lower payoff in case of a breakdown. L then also accepts the contract since rejecting it would reveal

[43] Again, the proof is easily adapted if the supremum is not achieved, by considering a PBE that gets very close to providing $\bar{u}_H$.

his type and, by Part i), yield a utility of $u_L(E_L(R_0))$. The cost reduction facing H is of order ε relative to $Q(\bar{C}_H)$, while the cost increase facing L is of order ε^2 relative to Q_L. This deviation is thus strictly profitable for ε small enough, which shows that $\bar{u}_H = u_H(E_H(R_0))$. Proceeding as in the end of the proof of Part i), this shows that L's maximal utility across all PBEs for $\beta \in (0, 1)$ is $u_L(E_L(R_0))$.

Part iv) The proof is similar to that of Part iii). Let $\bar{Q}$ denote P's maximal expected cost over all PBEs and beliefs, starting from R_0. We will start by showing that $\bar{Q} \leq Q(E_L)$, where $E_L = E_L(R_0)$. Suppose by contradiction that $\bar{Q}$ is strictly greater than $Q(E_L)$ and achieved for some PBE and belief,[44] and consider the following deviation: P proposes the contracts $\tilde{C}_\theta$ that are efficient for each type and cost $\bar{Q} - \varepsilon$ for ε arbitrarily small. It is easily shown that these contracts are IC, and by a similar argument to the one used in Part iii), rejecting these contracts is a strictly dominated strategy for one of the two types, and hence for both types. This is a strictly profitable deviation for P, yielding a contradiction. Hence, $\bar{Q} \leq Q(E_L)$. Since L's expected utility is at least $u_L(R_0)$ in all PBEs, and providing this utility costs at least $Q_L = Q(E_L)$ to P, this means that P spends at most Q_L on H, in all PBEs, and for all initial beliefs $\beta > 0$. This implies that H's expected utility is bounded above by the utility that he achieves with the H-efficient contract that costs Q_L. We now show that H's expected utility is bounded above by $u_H(E_L)$. Suppose not and consider a PBE that gives H his highest utility across all PBEs and beliefs, which we denote $\bar{u}_H > u_H(E_L)$. The expected cost Q from this PBE must exceed $\beta Q(\bar{C}_H) + (1 - \beta)Q_L$, where $\bar{C}_H$ is the H-efficient contract that gives utility $\bar{u}_H$ to H. Suppose that P deviates by proposing the contracts $\tilde{C}_L, \tilde{C}_H$ such that $\tilde{C}_L$ is L-efficient and gives utility $u_L(R_0) + \varepsilon^2$ to L, and $\tilde{C}_H$ is H-efficient and gives utility $\bar{u}_H - \varepsilon$ to H, for ε arbitrarily small. Because $\tilde{C}_H$ gives strictly more to H than $\bar{u}_H$, H accepts $\tilde{C}_H$ and, hence, L accepts $\tilde{C}_L$. As in Part iii), this deviation is strictly profitable and yields the desired contradiction. Also note that for ε small enough, $\bar{u}_H > u_H(E_L)$, so mimicking L is suboptimal for H.[45]

Proof of Lemma 1

The result follows from Part iv) of Proposition 1: $E_H(R_n)$ plus any small amount gives

[44] As before, one can use a PBE that yields a cost arbitrarily close to $\bar{Q}$, in case it is not exactly achieved.

[45] It is straightforward to show that L does not want to mimic H, since P spends less on H than on L, and L is already getting his maximal utility given the cost that P incurs conditional on facing L.

a strictly higher utility to H than his maximal continuation utility and strictly more utility than $E_L(R_n)$. Therefore, H accepts the contract with probability 1. Since L strictly prefers $E_L(R_n)$ to $E_H(R_n)$ (Part ii) of Proposition 1) and his type is revealed unless he takes the strictly suboptimal contract $E_H(R_n)$, L accepts $E_L(R_n)$. ■

Proof of Lemma 2

Consider any PBE starting with $R_0 \in \mathcal{H}$ and, by contradiction, the first round n such that i) R_n is in $\mathcal{H}$ and ii) L accepts with positive probability a contract R_{n+1} that is in a different configuration. Suppose first that R_{n+1} is in the No-Rent configuration. Then $u_L(n) = u_L(R_{n+1})$, by Part iii) of Proposition 1, which implies that $u_L(R_n) \leq u_L(R_{n+1})$: R_{n+1} is on a weakly higher isoutility curve of u_L than R_n. Moreover, because H can always accept R_{n+1}, $u_H(n) \geq u_H(R_{n+1}) > u_H(E_H(R_n))$, where the strict inequality comes from the fact that u_H is increasing along the isoutility curve of u_L in the direction of $\mathcal{E}_H$.[46] This implies that the continuation cost for P is strictly above $\beta_n Q(E_H(R_n)) + (1-\beta_n)Q(E_L(R_n))$, which contradicts Lemma 1. Now suppose that R_{n+1} is in the L-Rent configuration. Part iv) of Proposition 1 applied to the L-Rent configuration implies that, by choosing R_{n+1}, L gets a continuation utility of at most $u_L(\tilde{E}_L(R_{n+1}))$ where $\tilde{E}_L(\tilde{R})$ is defined—when $\tilde{R}$ is in the L-Rent configuration—similarly to $E_H(R)$ when R is in the H-Rent configuration. Therefore, $u_L(\tilde{E}_L(R_{n+1}))$ must be weakly greater than $u_L(R_n)$. However, notice that when $\tilde{E}_L$ is constructed, we use L's isoutility curve between the efficiency curves $\mathcal{E}_H$ and $\mathcal{E}_L$, which is steeper than H's isoutility curve at $\tilde{E}_L(R_{n+1})$, from the single-crossing property. As can be easily checked graphically, this implies that $u_H(R_{n+1})$ must have been *strictly* greater than $u_H(E_H(R_n))$, contradicting Part iv) of Proposition 1 applied to H. ■

Proof of Proposition 2

i) Observe, first, that negotiation cannot end endogenously at a finite round N with $\beta_n = \beta_N > 0$ and $R_n = R_N \in \mathcal{H}$ for all $n \geq N$. If this were the case, P could strictly reduce his cost at round N by proposing the H-efficient contract $E_H(R_N)$ and have it accepted by H with probability 1, by Part iv) of Proposition 1. Suppose instead that P keeps proposing new contracts until renegotiation is exogenously interrupted, and suppose by contradiction that there is a choice sequence with an associated belief subsequence $\{\beta_{n(k)}\}_{k\in\mathbb{N}}$ that converges

[46] More explicitly, we have $u_H(R_{n+1}) > u_H(E_L(R_{n+1})) \geq u_H(E_L(R_n)) = u_H(E_H(R_n))$.

to $\beta^* > 0$ (so both types accept each contract in this subsequence with strictly positive probability). Let $u_H^* = \sup\{u_H(R_n)\}$ where the supremum is taken over all contracts in the choice sequence. For H to accept R_n with positive probability infinitely often, $u_H(R_n)$ must converge to u_H^* for any subsequence, including along the subsequence $\{n(k)\}$.[47] However, this implies that proposing the H-efficient contract C_H that gives u_H^* to H is a strictly profitable deviation as $\beta_{n(k)}$ gets arbitrarily close to β^*: it does not change P's cost conditional on facing L but it strictly reduces P's expected cost by an amount arbitrarily close to $\beta^*[Q(C_L) - Q(C_H)]$, where C_θ is the θ-efficient contract that provides H with utility u_H^*.[48]

ii) Suppose that there exists $\varepsilon > 0$ and a subsequence of rounds, indexed by m, for which $Q(R_m) - Q(E_L(R_m)) \geq \varepsilon$. For m large enough, β_m converges to zero, from part i), and is thus bounded above by $\frac{\eta\varepsilon}{2\Delta_Q}$, where $\Delta_Q = \max_{C\in\mathcal{C}} Q(C) - \min_{C\in\mathcal{C}} Q(C)$. Therefore, P can deviate by proposing $E_L(R_m), E_H(R_m)$, which are respectively accepted by L and H. This deviation yields an immediate gain of $\eta\varepsilon$ on L and a loss of at most $\frac{\eta\varepsilon}{2}$ on H, given the upper bound on β_m, and is thus strictly profitable. This shows that the limit points of $\{R_n\}$ are all L-efficient. Let $u_L^* = \sup\{u_L(R_n)\}$. There is a subsequence indexed by $\tilde{m}$ for which $u_L(R_{\tilde{m}})$ converges to u_L^*. Moreover, since L can always hold on to any contract R_n along the choice sequence, and thus in particular to the contracts occurring along the subsequence $\{R_{\tilde{m}}\}$, $u_L(R_n)$ must converge to u_L^* for all subsequences. Combining these observations, $\{R_n\}$ must converge to the L-efficient contract $\bar{C}_L$ such that $u_L(\bar{C}_L) = u_L^*$. ∎

iii) Parts i) and ii) have shown that β_n converges to zero and R_n converges to an L-efficient contract as n goes to infinity. This implies that $E_H(R_n)$ gives asymptotically the same utility to H as R_n does and, hence, that w_n converges to zero.[49]

Proof of Lemma 3

[47] Otherwise, there must exist a subsequence of rounds for which $u_H(R_{m+1})$ is bounded above away from u_H^* by some constant $\delta > 0$. However, H's continuation utility, $u_H(m)$, is nondecreasing and becomes arbitrarily close to u_H^*. (Monotonicity comes Lemma 5.) When H's continuation gets within $\varepsilon\eta$ of u_H^* for some ε arbitrarily small, this implies that accepting R_{m+1} causes a loss of order $\eta\delta$, due to the probability of an immediate breakdown, and contradicts the fact that $u_H(m)$ is within $\varepsilon\eta$ of u_H^*.

[48] Since $u_H(R_n)$ gets arbitrarily close to u_H^* and R_n lies in $\mathcal{H}$, $Q(R_n)$ becomes arbitrarily close to (or above) $Q(C_L)$ as n gets large.

[49] Put differently, the contract $\bar{C}_H$ defined in Part iii) of the proof of Proposition 1 satisfies $u_H(\bar{C}_H) = u_H(\bar{C}_L)$.

Without loss, set $n = 0$, $\beta = \beta_0$, $R = R_0$, $\bar{u}_L = u_L(R_0)$, $\bar{u}_H = u_H(R_0)$, and $\bar{Q} = Q(E_L(R_0))$. It suffices to prove the claim for $\beta \leq \acute{\beta}$ for some small threshold $\acute{\beta} \in (0,1)$.[50] Focusing on this case, (16) implies that P's expected cost Q_L conditional on facing L satisfies $Q_L \leq \bar{Q} + \acute{q}\beta$ where $\acute{q} = \frac{a\Delta_H}{1-\acute{\beta}}$. For any small $\epsilon > 0$, to be chosen shortly, let $\mathcal{T}^\epsilon = \{R \in \mathcal{H} : u_H(E_L(R)) - u_H(R) \leq \epsilon\}$ and $\mathcal{D}^\epsilon = \mathcal{H} \setminus \mathcal{T}^\epsilon$. Graphically, $\mathcal{T}^\epsilon$ represents a tube-like region of $\mathcal{H}$ bordering $\mathcal{E}_L$, whose (varying) width is of order ϵ.

Let τ denote the index of the round immediately following the exogenous breakdown—the contract implemented is thus R_τ, recalling that the agent chooses contract R_{n+1} in period n. Let p_L denote the probability that $R_\tau \in \mathcal{D}^\epsilon$, conditional on facing type L, and $u_L^{\mathcal{D}}$ and $u_L^{\mathcal{T}}$ denote L's expected utilities conditional on $R_\tau \in \mathcal{D}^\epsilon$ and $R_\tau \in \mathcal{T}^\epsilon$, respectively. Similarly, let $Q_L^{\mathcal{D}}$ and $Q_L^{\mathcal{T}}$ denote P's expected cost conditional on facing L when $R_\tau \in \mathcal{D}^\epsilon$ and $R_\tau \in \mathcal{T}^\epsilon$, respectively. We have $(1-p_L)u_L^{\mathcal{T}} + p_L u_L^{\mathcal{D}} = u_L(0) \geq \bar{u}_L$ and $(1-p_L)Q_L^{\mathcal{T}} + p_L Q_L^{\mathcal{D}} = Q_L \leq \bar{Q} + \acute{q}\beta$. By definition of $\mathcal{D}^\epsilon$, Lemma 15 implies,[51] whenever $R_\tau \in \mathcal{D}^\epsilon$, that

$$Q(R_\tau) \geq Q(E_L(R_\tau)) + \underline{q}\epsilon^2. \tag{23}$$

Taking expectations in (23) conditional on $R_\tau \in \mathcal{D}^\epsilon$ yields

$$Q_L^{\mathcal{D}} \geq E[Q(E_L(R_\tau))|R_\tau \in \mathcal{D}^\epsilon] + \underline{q}\epsilon^2.$$

Let $E_L(u)$ denote the L-efficient contract that provides L with utility u. By convexity of $Q(\cdot)$ and concavity of $u_L(\cdot)$,

$$E[Q(E_L(R_\tau))|R_\tau \in \mathcal{D}^\epsilon] \geq Q(E_L(u_L^{\mathcal{D}}))$$

since each $E_L(R_\tau)$ gives L the same utility as R_τ and, conditional on lying in $\mathcal{D}^\epsilon$, the contracts R_τ's give L a utility of $u_L^{\mathcal{D}}$ in expectation. Again by convexity of $Q(\cdot)$ and concavity of $u_L(\cdot)$, we have $\bar{Q} \leq p_L Q(E_L(u_L^{\mathcal{D}})) + (1-p_L)Q(E_L(u_L^{\mathcal{T}}))$ since $\bar{Q}$ is the smallest cost for providing

[50] w_n is uniformly bounded as a difference of bounded utilities and we can always increase K_w so that $K_w\acute{\beta}^{1/3}$ exceeds w_n's upper bound.

[51] Here, as well as later in the proof, we are applying an inequality to round τ. The inequality must hold because the variable R_τ is determined before the breakdown, by the agent's contract choice at round $\tau - 1$ before knowing whether the breakdown will occur at the end of that round. More generally, any inequality satisfied by variables determined before the breakdown must hold regardless of whether the breakdown occurs immediately after this determination or later.

L with utility $\bar{u}_L$, while the right-hand side is the cost associated with a particular way of providing L with utility $u_L(0) = p_L u_L^{\mathcal{D}} + (1 - p_L)u_L^{\mathcal{T}} \geq \bar{u}_L$. Combining these inequalities,

$$\bar{Q} + \acute{q}\beta \geq Q_L \geq p_L(Q(E_L(u_L^{\mathcal{D}})) + \underline{q}\epsilon^2) + (1 - p_L)Q(E_L(u_L^{\mathcal{T}})) \geq \bar{Q} + p_L\underline{q}\epsilon^2, \tag{24}$$

which implies that $p_L\underline{q}\epsilon^2 \leq \acute{q}\beta$. Choosing $\epsilon = \beta^{1/3}$, we get $p_L \leq k_p\beta^{1/3}$ where $k_p = \acute{q}/\underline{q} > 0$. The reason for choosing this value of ϵ comes from (27) below: it optimizes the trade-off between a higher probability of R_τ being in $\mathcal{T}^\epsilon$ and a tighter bound on w over $\mathcal{T}^\epsilon$.

Set $\mathcal{T} = \mathcal{T}^{\beta^{1/3}}$ and $\mathcal{D} = \mathcal{D}^{\beta^{1/3}}$ and let p denote the unconditional probability that $R_\tau \notin \mathcal{T}$. We have $p = \beta p_H + (1 - \beta)p_L$, where p_H is the probability that $R_\tau \notin \mathcal{T}$ conditional on the agent being of type H,[52] and thus

$$p \leq \beta + p_L \leq k_p\beta^{1/3} \tag{25}$$

where the last inequality is obtained by increasing the constant k_p derived earlier, whose precise value is unimportant for the proof.

The breakdown time τ has finite expectation: $E[\tau] = 1/\eta$. By the optional sampling theorem, this implies that $E[\beta_\tau] = \beta$. Moreover, by definition of p we have $E[\beta_\tau] = pE[\beta_\tau|R_\tau \notin \mathcal{T}] + (1 - p)E[\beta_\tau|R_\tau \in \mathcal{T}]$. Since $p \leq k_p\beta^{1/3}$, this implies that

$$E[\beta_\tau|R_\tau \in \mathcal{T}] \leq \beta + O(\beta^{4/3}). \tag{26}$$

Suppose $\beta_\tau \leq \bar{\beta}$ for some $\bar{\beta} \in (0, 1)$. Applying (21) to rounds $n = 0$ and $n' = \tau$,

$$u_H(E_H(R_\tau)) - \bar{u}_H \geq -\frac{\hat{b}\bar{\beta}}{1 - \bar{\beta}}w_\tau.$$

When $R_\tau \in \mathcal{T}$, we have

$$w_\tau = u_H(E_H(R_\tau)) - u_H(\tau) \leq u_H(E_L(R_\tau)) - u_H(R_\tau) \leq \epsilon = \beta^{1/3}, \tag{27}$$

where the first inequality comes from $u_H(E_L(R_\tau)) = u_H(E_H(R_\tau))$ and $u_H(\tau) \geq u_H(R_\tau)$.[53] and the second one follows from the definition of $\mathcal{T}$.

[52]If the agent is of type H, R_τ may also be an H-efficient contract, hence the conditioning $R_\tau \notin \mathcal{T}$ rather than $R_\tau \in \mathcal{D}$, as these events are only equivalent for type L.

[53]Recall that $u_H(R_n) \leq u_H(n)$ for all n, since holding to R_n forever after round n is always a feasible strategy for H.

Moreover, for $R_\tau \in \mathcal{T}$ we have $u_H(R_\tau) - u_H(E_H(R_\tau)) \geq -\epsilon = -\beta^{1/3}$. Letting $\bar{k} = 1 + \hat{b}\frac{\bar{\beta}}{1-\bar{\beta}} > 0$, these observations imply that for $\beta_\tau \leq \bar{\beta}$,

$$u_H(R_\tau) \geq \bar{u}_H - \bar{k}\beta^{1/3}. \tag{28}$$

Let $\bar{p}$ denote the probability, conditional on $R_\tau \in \mathcal{T}$, that $\beta_\tau \geq \bar{\beta}$. We have

$$\bar{p}\bar{\beta} \leq \bar{p}E[\beta_\tau|\beta_\tau \geq \bar{\beta}, R_\tau \in \mathcal{T}] + (1-\bar{p})E[\beta_\tau|\beta_\tau < \bar{\beta}, R_\tau \in \mathcal{T}] = E[\beta_\tau|R_\tau \in \mathcal{T}] \leq \beta + O(\beta^{4/3}).$$

Therefore, $\bar{p} \leq \bar{k}\beta$ for some constant $\bar{k} > 0$. Combining this with (25), the unconditional probability that R_τ ends up in $\mathcal{T}$ *and* that $\beta_\tau \leq \bar{\beta}$ is bounded below by

$$(1 - k_p\beta^{1/3})(1 - \bar{k}\beta) = 1 - k_p\beta^{1/3} + O(\beta). \tag{29}$$

Thus, either H or L has probability at least $1 - k_p\beta^{1/3}$ of ending up with a contract in $\mathcal{T}$ and a breakdown belief $\beta_\tau \leq \bar{\beta}$. If it is H, (28) implies that H has a probability at least $1 - k_p\beta^{1/3}$ of getting a utility $u_H(R_\tau) \geq \bar{u}_H - \bar{k}\beta^{1/3}$. If it is L, H can mimic L for the entire game and guarantee himself the same utility with probability at least $1 - k_p\beta^{1/3}$. Either way, H's expected utility at the beginning of the game satisfies $u_H(0) = E[u_H(R_\tau)] \geq p_m(\bar{u}_H - \bar{k}\beta^{1/3}) + (1-p_m)\min_{C\in\mathcal{C}} u_H(C)$ where $p_m \geq 1 - k_p\beta^{1/3}$. The right-hand side is bounded below by $\bar{u}_H - K_w\beta^{1/3}$ for some $K_w > 0$. Since $w_0 = \bar{u}_H - u_H(0)$, the lemma follows.

C Proof of Proposition 4

Lemmas 3 and Lemma 7 imply that $\hat{\beta}_K \geq \underline{\beta}\eta^3$ for some $\underline{\beta} > 0$. The condition $\hat{\beta}_K \leq g^K\beta_0$ then implies that $g^K \geq \underline{\beta}\eta^3/\beta_0$. Defining ρ by $g^{-\rho} = \frac{t}{t-1}$, Lemma 7 implies that

$$w_0 \leq c_w g^{-\rho K} \leq c'_w\eta^{1-3\rho}.$$

This shows that w_0 converges to 0 with η if $\rho < 1/3$. By definition of g and ρ, we have $\rho = \ln(t/(t-1))/\ln(\beta_0 + t(1-\beta_0))$. For β_0 small, ρ is thus close to

$$\rho_0 = 1 - \frac{\ln(t-1)}{\ln t}$$

The RHS is decreases in $t \in (1, +\infty)$ from $+\infty$ to 0. Therefore, there is a threshold $\hat{t}$ above which the condition $\rho_0 < 1/3$ is satisfied, provided that β_0 is less than some level $\beta_0(t) > 0$.

Using the block structure of Part I, starting from any $\beta_0 < 1$, we reach $\beta_0(t)$ in at most K_0 blocks, where K_0 is defined by $g_0^{K_0} = \beta_0/\beta_0(t)$ and $g_0 = 1/(\beta_0 + t(1-\beta_0))$. w_0 is thus bounded above by $\frac{t}{t-1}^{K_0}\eta^{1-3\rho}$, which converges to 0 as η goes to zero, proving the result.

The 1/3 exponent coming from Lemma 3 was chosen optimally, as explained in the lemma's proof. Moreover, any bound of the form $w_n \leq \beta_n^\alpha$ for some $\alpha > 0$ would merely change ρ's threshold; it would not cover all primitives of the model.

It may be possible to improve the value of t chosen in Part I, so as to obtain a lower value of ρ_0. However, the minimal value of ρ_0 over all possible values of t diverges to $+\infty$ as $1+\frac{D}{a\Delta_H}$ goes to 1, as shown in Appendix F.1. Low values of $1+\frac{D}{a\Delta_H}$ create two problems: the factor g controlling the belief decrease goes to 1, becoming useless, and the factor $t/(t-1)$ used to bound w_0 above in Lemma 7 becomes arbitrarily large. Using the approach Proposition 4 seems hopeless when screening incentives are high for all values of t.

Finally, players are expected-utility maximizers whose utility functions may, in some applications, be defined up to an affine transformation. Could one get more from Proposition 4 by changing players' utility representation? The answer is negative: scaling or translating the utility and cost functions has no impact on the key parameter t.[54]

References

Abreu, D., Gul, F. "Bargaining and Reputation," *Econometrica*, Vol. 68, pp. 85–117.

Abreu, D., Pearce, D. "Bargaining, Reputation, and Equilibrium Selection in Repeated Games wiht Contracts" *Econometrica*, Vol. 75, pp. 653–710.

Atakan, A., Ekmekci, M. (2012) "Reputation in Long-Run Relationships," *Review of Economic Studies*, Vol. 79, pp. 451–480.

[54] a is a Lipschitz constant whose unit is "cost-per-util;" D is a cost difference; and Δ_H is a utils difference. Therefore, multiplying Q by a scalar has the effect of multiplying a and D by the same amount, which doesn't affect t. Likewise, multiplying u_H by a scalar has the effect of dividing a and multiplying Δ_H by this scalar, leaving t unchanged. Finally, a, D, and Δ_H are unaffected by translations of the utility and cost functions.

Atakan, A., Ekmekci, M. (2014) “Bargaining and Reputation in Search Markets,” *Review of Economic Studies*, Vol. 81, pp. 1–29.

Ausubel, L., Deneckere, R (1989) “Reputation in bargaining and durable goods monopoly,” *Econometrica*, Vol. 57, pp. 511–531.

Battaglini, M. (2007) “Optimality and Renegotiation in Dymamic Contracting,” *Games and Economic Behavior*, Vol. 60, pp. 213–246.

Beaudry, P., Poitevin, M. (1993) “Signalling and Renegotiation in Contractual Relationships,” *Econometrica*, Vol. 61, pp. 745–782.

Bernheim, B.D., Ray, D. (1989) “Collective Dynamic Consistency in Repeated Games,” *Games and Economic Behavior*, Vol. 1, pp. 295–326.

Bester, L., Strausz, R. (2001) “Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case,” *Econometrica*, Vol. 69, pp. 1077–1098.

Board, S., Pycia, M. (2013) “Outside Options and the Failure of the Coase Conjecture,” *Forthcoming, American Economic Review.*

Brennan, J., Watson, J. (2013) “The Renegotiation-Proofness Principle and Costly Renegotiation,” Working Paper, UCSD.

Coase, R. H. (1972) “Durability and monopoly,” *Journal of Law and Economics*, Vol. 15, pp. 143–149.

Cripps, M., Dekel, E., and W. Pesendorfer (2005) “Reputation with Equal Discounting in Repeated Games with Strictly Conflicting Interests,” *Journal of Economic Theory*, Vol. 121, pp. 259–272.

Deneckere, R., Liang, M-Y (2006) “Bargaining with Interdependent Values,” *Econometrica*, Vol. 74, pp. 1309–1364.

Dewatripont, M. (1989) “Renegotiation and Information Revelation over Time: The Case of Optimal Labor Contracts,” *Quarterly Journal of Economics*, Vol. 104, pp. 589–619.

Farrell, J., Maskin, E. (1989) "Renegotiation in Repeated Games," *Games and Economic Behavior*, Vol. 1, pp. 327–360.

Fuchs, W., Skrzypacz, A. (2010) "Bargaining with Arrival of New Traders," *American Economic Review*, Vol. 100, pp. 802-836.

Fudenberg, D., Levine, D., and J. Tirole (1985) "Infinite Horizon Models of Bargaining with One-Sided Incomplete Information," in *Game Theoretic Models of Bargaining,* ed. by A. Roth. Cambridge University Press, Cambridge, UK.

Fudenberg, D., Levine, D. (1989) "Reputation and Equilibrium Selection in Games with a Patient Player," *Econometrica*, Vol. 57, pp. 759–778.

Fudenberg, D., Tirole, J. (1990) "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica*, Vol. 58, No. 6, pp. 1279–1319.

Fudenberg, D., Tirole, J. (1991) *Game Theory*, MIT Press.

Gul, F., Sonnenschein, L., and R. Wilson (1986), "Foundations of Dynamic Monopoly and the Coase Conjecture," *Journal of Economic Theory*, Vol. 39, pp. 155–190.

Harris, C. (1985) "Existence and Characterization of Perfect Equilibrium in Games of Perfect Information," *Econometrica*, Vol. 53, No. 3, pp. 613–628.

Hart, O., Tirole, J. (1988) "Contract Renegotiation and Coasian Dynamics," *Review of Economic Studies*, Vol. 55, No. 4, pp. 509–540.

Laffont, J.-J., Tirole, J. (1990) "Adverse Selection and Renegotiation in Procurement," *Review of Economic Studies*, Vol. 57, No. 4, pp. 597–625.

Maestri, L. (2015) "Dynamic Contracting under Adverse Selection and Renegotiation," Working Paper, FGV/EPGE.

Maskin, E., Tirole, J. (1992) "The Principal-Agent Relationship with an Informed Principal, II: Common Values," *Econometrica*, Vol. 60, pp. 1–42.

Matthews, S. (1995) "Renegotiation of Sales Contracts," *Econometrica*, Vol. 63, pp. 567–590.

Rubinstein, A. (1982) "Perfect Equilibrium in a Bargaining Model," *Econometrica*, Vol. 50, pp. 97–110.

Rubinstein, A. (1985) "A Bargaining Model with Incomplete Information about Time Preferences," *Econometrica*, Vol. 53, pp. 1151–1172.

Rubinstein, A., Wolinsky, A. (1992) "Renegotiation-Proof Implementation and Time Preferences," *American Economic Review*, Vol. 82, pp. 600-614.

Schmidt, K. (1993) "Reputation and Equilibrium Characterization in Repeated Games with Conflicting Interests," *Econometrica*, Vol. 61, pp. 325–351.

Skreta, V. (2006) "Sequential Optimal Mechanisms," *Review of Economic Studies*, Vol. 73, pp. 1085–1111.

Sobel, J., Takahashi, I. (1983) "A Multistage Model of Bargaining," *Review of Economic Studies*, Vol. 50, pp. 411–426.

Strulovici, B. (2011) "Renegotatiation-Proof Contracts with Moral Hazard and Persistent Private Information," Working Paper, Northwestern University.

Strulovici, B. (2013) "Renegotiation and Persistent Information," Working Paper, Northwestern University.

Wang, G.H. (1998) "Bargaining over a Menu of Wage Contracts," *Review of Economic Studies*, Vol. 65, pp. 295–305.