

STABLE CONTRACTS UNDER RENEGOTIATION

VITALI GRETSCHKO AND ACHIM WAMBACH

ABSTRACT. We study a principal contracting with a privately informed agent, where any contract is subject to renegotiation. Rather than specifying a renegotiation game, we propose conditions that the set of terminal outcomes of renegotiation must satisfy. Our central concept is a stable set of contract-belief pairs defined by external stability (outcomes not in the stable set can be improved to outcomes in the stable set) and internal stability (no strict mutual improvements exist within the stable set). Applied to contracting without commitment, our framework bounds the principal's loss from lacking commitment by characterizing the largest payoff she can obtain when the terminal outcomes must be in a stable set. In a nonlinear contracting environment, this payoff is attained by pooling contracts that satisfy a no-distortion-at-the-bottom property.

JEL classification: C71, D82, D86

Keywords: Mechanism design, contract theory, renegotiation, stability, Coase conjecture

1. INTRODUCTION

Consider the problem of a principal (she) who wishes to contract with a privately informed agent (he). In general, any agreement with a privately informed agent reveals some of his private information. Consequently, the agreement terms, along with the principal's posterior, may allow for mutually beneficial improvements, potentially initiating renegotiation. This

(A1) UNIVERSITY OF MÜNSTER; VITALI.GRETSCHKO@WIWI.UNI-MUENSTER.DE

(A2) ZEW MANNHEIM AND UNIVERSITY OF MANNHEIM; WAMBACH@ZEW.DE

Date: October 18, 2026.

We would like to thank the co-editor Gabriel Carroll, four anonymous referees, Geir Asheim, Atoshi Chowdhury, Jeffrey Ely, Roman Inderst, Daniil Larionov, Jörg Lingers, Georg Nöldeke, Alexei Parakhonyak, Ray Rees, Klaus Schmidt, Nicolas Schutz, Ina Taneva, Thomas Tröger, and seminar participants at the University of Toronto (2023), University of Mannheim (2023), London Business School (2022), University of Bonn (2018), Cologne University (2017), the EEA conference (Lisbon 2017), the EARIE conference (Maastricht 2017), Edinburgh University (2016), University of Mannheim (2016), the IIOC conference (Boston 2017), the SED conference (York 2017), and the VfS conference (Vienna 2017) for their helpful comments and suggestions.

presents an analytical challenge: the renegotiation process can take many forms, and specifying a particular protocol imposes restrictive assumptions that may not generalize across environments.

Rather than specifying particular renegotiation processes, we introduce a framework to identify terminal outcomes, that is, those not subject to change. Specifically, we propose conditions that terminal outcomes must satisfy. Our core contribution is the concept of a stable set: a collection of such terminal outcomes that is jointly comprehensive and mutually consistent. With private information, an outcome must encode not only the contractual terms but also the informational state. Thus, we define outcomes as pairs comprising a contract and the principal’s belief about the agent’s type. By characterizing stable sets, we clarify which agreements resist renegotiation without relying on specific procedural details. Our approach builds on the intuition of set stability, introduced by [von Neumann and Morgenstern \(1944\)](#) for situations without private information. Their notion of stability focuses on identifying sets of outcomes that are jointly resilient to deviations.

To decide whether an outcome is terminal, it is not enough to ask whether some other outcome would make both parties better off. Such an alternative is not a credible endpoint of renegotiation if it would itself be renegotiated. Terminality must therefore be determined jointly: a stable set is a candidate collection of terminal outcomes, and potential renegotiation is evaluated only toward outcomes in that set.

To define stable sets, we first establish what constitutes an improvement. When the principal and the agent consider renegotiating a contract, they do not typically consider a single new contract, but a set of new contracts. Which of these contracts will be realized depends on the agent’s private information. Thus, a potential improvement typically involves

moving to a set of new contract-belief pairs. In such a set, each contract is paired with a new belief about the agent’s type, reflecting the information conveyed by assigning types to contracts. We define such a move as a (weak) improvement if each type of agent receives its optimal contract from this set and is (weakly) better off, while the principal is (weakly) better off in expectation. Building on the definition of an improvement, we define a stable set as one that satisfies two key properties reflecting its self-sustainability against renegotiation.

First, the stable set must be comprehensive: if an outcome is not part of the set, an improvement must exist that leads to outcomes within the set and makes both the principal and the agent weakly better off. In other words, every outcome outside the set must admit an improvement to outcomes inside it. We call this condition external stability. Without external stability, the candidate set would classify some outcomes as unstable without offering a path to stable outcomes.

Second, the stable set must be consistent: no outcome within the set can admit an improvement to other outcomes within the set that strictly benefits both the principal and the agent. We call this condition internal stability. Internal stability ensures that a stable set as a whole constitutes a collection of mutually consistent terminal agreements. Without it, a terminal outcome could be profitably replaced by other terminal outcomes, leaving no justification for treating the former outcome as terminal. Section 2 provides an intuitive exploration of our stability concept and its immediate implications using a canonical principal-agent model.

Our stability concept addresses a central issue in models of contract renegotiation under private information: when multiple new contracts are available, which one will an agent, potentially anticipating renegotiation, choose? A key feature of our framework is that both

internal and external stability are evaluated against improvements toward the candidate stable set, whose outcomes are terminal by definition. That is, no renegotiation takes place after an outcome from a stable set has been reached. This implies that assigning a suboptimal contract to the agent is unjustified, and the principal's beliefs for an improvement must reflect the assumption that the agent receives an optimal contract given his type. We thus require improvements into the stable set to satisfy incentive compatibility and consistent belief updating.

We apply our stability concept to contracting without commitment from the perspective of an outside analyst. The analyst does not specify a mechanism or a renegotiation protocol. Instead, the analyst asks how large the principal's expected payoff can be if the interaction must end at outcomes in a stable set. To answer this question, we work directly with assignments of agent types to contracts and with the principal's corresponding beliefs. We require these assignments to satisfy incentive compatibility and individual rationality and to generate outcomes in a stable set. Such assignments can be interpreted as the outcomes of mechanisms or of renegotiation processes.

We apply our stability concept to a setting akin to that in [Mussa and Rosen \(1978\)](#), characterized by a continuous type space, private values, and nonlinear contracts. This environment applies to many situations where the principal naturally cannot rule out renegotiation, such as procurement, labor contracts, or selling when price and quality matter. As a leading example, we consider a seller offering a good of varying quality to a privately informed buyer. The contracts consist of two dimensions: price and quality.

We show that the largest payoff is attained by an assignment with the following properties: (1) it involves a *countably* infinite number of contracts; (2) each contract is held by a *pool*

of agent types with positive measure; (3) the lowest type in each pool receives an efficient contract, while this contract is inefficient for all higher types in the pool, a property that we call “no distortion at the bottom.” Section 2 provides intuition for this result.

As a second application, we analyze the Coase conjecture (Coase, 1972), which addresses the problem of a seller offering an indivisible good to a buyer who has private information about his valuation.¹ For this setting, we rederive the “gap–no-gap” result (Ausubel and Deneckere, 1989) using our stability concept. We show that if there is a gap between the seller’s cost and the buyer’s lowest valuation—the gap case—the seller can charge no more than the buyer’s lowest valuation. In contrast, if the seller’s cost is at least as high as the buyer’s lowest valuation—the no-gap case—stability does not rule out the monopoly price. This rederivation illustrates that our concept of stability captures essential features of the environment.

We briefly outline the related literature. A detailed literature review can be found in Section 8. This work relates to the literature on the Incomplete Information Core (Wilson, 1978) and its refinements (Forges et al., 2002; Forges and Serrano, 2013; Vohra, 1999; Dutta and Vohra, 2005). The Incomplete Information Core evaluates the stability of an ex ante allocation rule, that is, prior to the specific realization of outcomes. Our approach, in contrast, assesses the stability of specific contract-belief pairs once an outcome has been realized and the principal’s beliefs have been updated based on information revealed by the outcome.²

Conceptually, our stability notion builds on the concept of the von Neumann–Morgenstern stable set (von Neumann and Morgenstern, 1944), which the recent literature has analyzed

¹See also Bulow (1982), Doval and Skreta (2024), Gul et al. (1986), or von der Fehr and Kühn (1995).

²An illustrative example can be found in Section 8.

for farsighted, rational-expectations, and history-dependent variants (Ray and Vohra, 2015; Dutta and Vohra, 2017; Dutta and Vartiainen, 2020). Other cooperative notions of stability in markets with private information such as belief-free and belief-based stable matchings (Liu et al., 2014; Liu, 2020), as well as standards of behavior in monopolistic insurance (Asheim and Nilssen, 1997), motivate our focus on the interaction of stability and beliefs.

A distinct strand uses cooperative concepts to study limited commitment in mechanism design. Vartiainen (2013) shows that, absent commitment, only an English auction without reserve is implementable. This result is corroborated in a noncooperative framework by Liu et al. (2019). Vartiainen (2023) embeds stability in a recursive, history-dependent framework and describes it as a history-dependent extension of our framework.

Our paper contributes to the growing literature on mechanism design without commitment (Bester and Strausz, 2001; Evans and Reiche, 2015; Neeman and Pavlov, 2013; Skreta, 2006, 2015; Doval and Skreta, 2022; Pęski, 2022). In general, those papers use sequentially posted mechanism games, whereas our approach characterizes renegotiation outcomes without relying on a specific game form.

Finally, our paper relates to dynamic bargaining analyses of Mussa and Rosen (1978)-style models. In the limit of vanishing frictions, efficient full separation emerges in the work of Strulovici (2017). Gradual information revelation (Chen, 2020) and relational contracting (Malcomson, 2016, 2025) rationalize pooling through payoffs accumulating over time. Our results provide a static, stability-based explanation for pooling outcomes.

2. INTUITIVE EXPLANATION OF STABILITY

Consider a principal and an agent contracting over a quality $q \in \mathbb{R}_+$ and a price $p \in \mathbb{R}$. The principal's profit from implementing contract (p, q) is $p - c(q)$, and the agent's utility is

$u(q, \theta) - p$, where $\theta \in \Theta$ is the agent's private type. The principal's prior on Θ is μ_0 . This setup is a standard nonlinear contracting problem. Let $q^*(\theta)$ denote the unique efficient quality for type θ , defined as the quality that maximizes $u(q, \theta) - c(q)$. An outcome, the primary object of our analysis, is a pair comprising a contract and the principal's belief about the agent's type. Outcomes combine the terms of trade and the informational environment into a single object.

The question is which outcomes can consistently be treated as terminal, that is, not subject to renegotiation. Our stable-set concept describes a candidate collection of such outcomes through two conditions. External stability requires that every outcome outside the set admit a weak improvement to outcomes within it: excluded outcomes must have a weakly beneficial alternative within the proposed collection of terminal outcomes. Internal stability requires that no outcome inside the set admit a strict improvement to other outcomes within it: included outcomes must be mutually consistent as terminal agreements.

Consider a contract (p, q) that is efficient for type θ , meaning $q = q^*(\theta)$. If such an efficient contract is on the table and the principal believes with certainty that the agent is of type θ , this constitutes an efficient and separating outcome. "Efficient" refers to the efficient quality, and "separating" indicates that there is no uncertainty about the agent's type. Once reached, such an outcome admits no mutually beneficial change, as no additional surplus can be generated. In our framework: every efficient and separating outcome must be in every stable set. Suppose such an outcome did not belong to a stable set. External stability would then require a weak improvement to outcomes within the set. However, since the efficient quality is unique, any change in quality reduces joint surplus, while changing

the price alone makes one party worse off. The only weak improvement therefore leaves the outcome unchanged and cannot lead into a set that excludes it.

Now, consider a separating but inefficient outcome with contract (p, q) . That is, the principal is certain of the agent's type θ , but $q \neq q^*(\theta)$. Both parties can strictly increase their payoffs by setting $q = q^*(\theta)$ and adjusting the price accordingly. Thus, they can improve to an efficient and separating outcome. As argued above, such outcomes are necessarily terminal, and thus such an improvement is credible. Consequently, inefficient and separating outcomes cannot be terminal outcomes of renegotiation. In our framework: suppose a separating and inefficient outcome belongs to a stable set. Internal stability would then preclude an improvement to other outcomes in the stable set that strictly benefits both parties. Yet, such a strict improvement exists.

Now consider a pooling outcome described by a contract (p, q) and a belief whose support is an interval. Suppose first that q is not efficient for any type in the pool. An assignment that gives every type in the pool his efficient quality can be priced so that it is incentive compatible and makes every type and the principal strictly better off. Again, the resulting efficient and separating outcomes are necessarily terminal, and thus such an improvement is credible. The original pooling outcome therefore cannot be terminal. In our framework: the existence of a strict improvement to efficient and separating outcomes violates internal stability.

Now suppose that q is efficient for one of the types in the pool. Here, two observations are crucial. First, consider assigning every type his efficient quality as in the previous paragraph. One type already receives his efficient quality at the pooling outcome, so the principal cannot earn more from this type while leaving him weakly better off. On an interval, incentive

compatibility requires her to earn the same profit across all types in an efficient separating schedule: the price increase between nearby types equals the cost increase along their efficient qualities. Thus, the restriction imposed by this one type applies to the entire pool. The principal cannot strictly increase her expected profit by separating the types efficiently while leaving every type weakly better off.

Second, one might ask why the parties do not instead consider two contracts, one with high quality for types with high willingness to pay and one with low quality for types with low willingness to pay. Such a change may strictly benefit both parties if the new contracts are treated as terminal. But their terminality cannot simply be assumed. If the resulting outcomes are themselves subject to renegotiation, the agent's choice need not depend only on the stated contract terms. Our stable-set concept addresses this issue without specifying a renegotiation process. We identify candidate terminal outcomes jointly and evaluate improvements only toward outcomes in that set. Because these outcomes are treated as terminal, the agent's incentives can be assessed directly from the contract terms.

The argument above does not yet establish that the pooling outcome is terminal: other improvements may be possible, and whether they lead to terminal outcomes remains to be determined. We show below that a suitable collection of pooling outcomes, together with all efficient and separating outcomes, forms a stable set. The pooling outcomes are chosen so that any strict improvement from one of these pools into the set would have to separate its types efficiently. The first observation rules this out. Every outcome outside the set, meanwhile, admits a weak improvement to efficient and separating outcomes. Both internal and external stability are therefore satisfied, showing that these pooling outcomes can consistently be treated as terminal.

3. MODEL

A principal (she) and an agent (he) intend to implement a contract from a set of contracts $\mathcal{W}$. If a contract $w \in \mathcal{W}$ is implemented, the principal's utility is $V(w)$, where $V : \mathcal{W} \rightarrow \mathbb{R}$ is a continuous von Neumann–Morgenstern utility function. To allow for probability measures on $\mathcal{W}$, we assume $\mathcal{W}$ is a Polish space, that is, a completely metrizable and separable topological space. We endow $\mathcal{W}$ with the Borel σ -algebra. The agent's utility is given by $U(w, \theta)$, where $U : \mathcal{W} \times \Theta \rightarrow \mathbb{R}$ is a von Neumann–Morgenstern utility function and depends on the agent's type θ . We assume throughout that $U(w, \theta)$ is continuous in (w, θ) . The agent's type is private information to the agent and is an element of $\Theta \subset \mathbb{R}$. We assume that Θ is a closed set and therefore a Polish space. We endow Θ with the Borel σ -algebra and denote by $\Delta(\Theta)$ the set of all probability measures.

For $\mu \in \Delta(\Theta)$, we denote by $\text{supp}(\mu)$ the support of μ , defined as the closure of the set $\{\theta \in \Theta : \mu(\theta_\varepsilon) > 0 \text{ for all } \varepsilon > 0\}$, where θ_ε denotes the ε -neighborhood of θ , that is, the set of all $\theta' \in \Theta$ such that $|\theta - \theta'| < \varepsilon$.

An outcome, the primary object of our analysis, is a pair $o = (w, \mu)$, with $w \in \mathcal{W}$ denoting the contract and $\mu \in \Delta(\Theta)$ representing the principal's belief about the agent's type. Outcomes integrate the terms of trade and the informational environment into a single formal object.

4. STABLE SETS

We study the properties of terminal outcomes, that is, outcomes that, if reached, will not be renegotiated away. The main challenge is that the agent and the principal may benefit from changing any outcome $o = (w, \mu)$ if they anticipate mutual improvement of w given belief μ .

To define an improvement, we first describe how an outcome o can be replaced consistently with a set of new contract-belief pairs. Each agent type is matched with a contract from this set. Simultaneously, the principal updates her belief about the agent's type to the conditional probability distribution given the assigned contract. The object that records this is an assignment measure: a joint distribution over types and contracts that assigns the types present at o to contracts and, through conditioning, fixes the principal's belief at each contract.

Definition 1 (Assignment measures). *Consider an outcome $o = (w, \mu)$. An assignment measure from o is a probability measure γ_o on $\Theta \times \mathcal{W}$ such that $\gamma_o(\cdot, \mathcal{W}) = \mu(\cdot)$. In other words, the marginal distribution of γ_o over types Θ equals μ .*

Let $W = \text{supp}(\gamma_o(\Theta, \cdot))$ be the support of the marginal distribution of γ_o on $\mathcal{W}$. We fix a version $\mu(\cdot \mid \hat{w})$ of the regular conditional probability derived from γ_o and regard it as part of the assignment.³ The set of resulting outcomes is defined as

$$\mathcal{O}(\gamma_o) = \{ (\hat{w}, \mu(\cdot \mid \hat{w})) : \hat{w} \in W \}.$$

An assignment measure specifies a collection of contract-belief pairs, each associating a possible contract with a belief about the agent's type. It assigns types to contracts and adjusts the principal's beliefs accordingly. The marginal constraint on Θ keeps the induced distribution of types consistent with μ , so the belief at each contract is the conditional

³The existence of the regular conditional probability $\mu(\cdot \mid \hat{w})$ follows from the fact that Θ and $\mathcal{W}$ are Polish spaces. We suppress the dependence on the choice of the version of the conditional probability from the notation. Any two versions agree $\gamma_o(\Theta, \cdot)$ -almost everywhere, in particular at every contract that carries positive probability, so on W they can differ only at contracts of probability zero. Assignment measures with the same γ_o but different versions are distinct objects, and all subsequent quantifiers range over assignment measures in this sense. Whenever we construct an assignment explicitly, the accompanying version of the conditional distribution is understood to be the one specified by that construction.

distribution of the types that receive it. An assignment measure imposes only this consistency of beliefs. We next add incentive considerations and their implications for belief updating.

Stable sets. The core of our characterization is a candidate set of terminal outcomes, which we call a *stable set*. Its members are outcomes that, if reached, will not be renegotiated away. As outlined previously, such a set satisfies two conditions. First, external stability: any outcome outside a given stable set must allow for a (weakly) beneficial improvement into this stable set. Second, internal stability: no outcome within a given stable set may allow for a strictly beneficial improvement to other outcomes within this stable set.

To apply these stability conditions, we first formally define when an assignment measure γ_o from an outcome o constitutes an improvement. A key element of our framework is that we evaluate both external and internal stability against improvements leading into the candidate stable set. These outcomes are treated as terminal: once the parties reach one, there is no renegotiation. This has a crucial implication: nothing justifies assigning a suboptimal contract to an agent type, so each agent type receives its utility-maximizing contract from those in the assignment. This is a standard incentive-compatibility requirement on the assignment of types to contracts, and the principal's beliefs are the conditional type distributions induced by that assignment. We define an improvement as follows:

Definition 2 (Improvement). *Let $o = (w, \mu) \in \mathcal{W} \times \Delta(\Theta)$ be an outcome and γ_o an assignment measure with marginal μ on Θ . We say that γ_o improves $o = (w, \mu)$ if it satisfies the following three conditions:*

- (i) *The agent is better off. That is, for all $\theta \in \text{supp}(\mu)$ there exists a $\hat{w} \in \mathcal{W}$ with $(\theta, \hat{w}) \in \text{supp}(\gamma_o)$ such that*

$$(1) \quad U(\hat{w}, \theta) \geq U(w, \theta).$$

- (ii) *The principal is better off. That is,*

$$(2) \quad \int_{\hat{w} \in \mathcal{W}} V(\hat{w}) d\gamma_o(\Theta, \hat{w}) \geq V(w).$$

- (iii) *The agent does not receive a suboptimal contract. That is, for all $\theta \in \text{supp}(\mu)$ and all $w', \hat{w} \in \text{supp}(\gamma_o(\Theta, \cdot))$, if $U(w', \theta) > U(\hat{w}, \theta)$, it follows that $\theta \notin \text{supp}(\mu(\cdot \mid \hat{w}))$.*

We say γ_o strictly improves o if the inequalities in conditions (i) and (ii) are strict.

Throughout, improvements are evaluated pointwise on $\text{supp}(\mu)$, rather than merely μ -almost surely, because the relevant comparisons are made after the agent has learned his type. Conditional on being type θ , whether an assigned contract is optimal for the agent, and whether it makes him better off than the status quo, does not depend on the ex ante probability of θ . Since $\text{supp}(\mu)$ is closed, boundary support points are included even when they have zero probability under μ . Moreover, we do not assume that V is bounded. The integral in requirement (ii) may be infinite or fail to exist. In either case requirement (ii) fails by convention, so such an assignment measure does not improve o .

In Definition 2, requirements (i) and (ii) reflect that the improvement benefits both the principal and the agent. In environments with transferable utility, any improvement that strictly raises the principal's payoff while leaving the agent weakly better off can be made strictly improving for both via a small transfer. Requirement (iii) accounts for the fact that

the outcomes resulting from γ_o are treated as terminal when they lie in the candidate stable set. In particular, after observing that a contract $\hat{w}$ has been chosen, the principal should not assign positive posterior support to a type θ if another contract w' in the assignment would make that type strictly better off. Thus, for any assignment into a stable set, each type must receive an optimal contract among the contracts in the assignment. We can interpret requirement (iii) as a form of incentive compatibility for the agent: if the assigned contracts are implemented without further change, the agent receives his optimal contract among them. Although the concept of improvement is defined generally, this optimal-choice requirement is justified only when the assignment leads into the candidate stable set. Only in this case are its outcomes terminal.

We now precisely define the conditions that a set of outcomes Ω must satisfy to qualify as a *stable set*.

Definition 3 (Stable set). *A set $\Omega \subset \mathcal{W} \times \Delta(\Theta)$ of outcomes is a stable set if it satisfies the following conditions for every outcome $o \in \mathcal{W} \times \Delta(\Theta)$:*

- (i) External stability: *If o is not in Ω , there exists an assignment measure γ_o with $\mathcal{O}(\gamma_o) \subset \Omega$ that improves o .*
- (ii) Internal stability: *If o is in Ω , there exists no assignment measure γ_o with $\mathcal{O}(\gamma_o) \subset \Omega$ that strictly improves o .*

The definition of a stable set is motivated by the idea that if the agent and the principal agree on an outcome in Ω , they do not expect it to be changed. An improvement into a stable set can thus be read as a record of where renegotiation leaves the parties: a contract for each type and, for the principal, the belief consistent with that assignment.

The distinction between strict and weak improvements reflects the role of external and internal stability as necessary conditions. External stability requires *weak* improvements for outcomes outside the set, while internal stability forbids *strict* improvements within the stable set. When an outcome in the stable set admits a strict improvement into the set, it cannot be terminal: there are outcomes, terminal by hypothesis, that both parties strictly prefer to it. Allowing such an outcome to remain in the stable set would undermine the stable-set interpretation. An improvement that is not strict carries no such force: it leaves either the principal or at least one type in the support of her belief without a strict gain.⁴ Its existence alone therefore does not provide a strict reason to abandon the outcome.

Existence of stable sets. We next show that stable sets exist in a standard quasi-linear environment. Let $\mathcal{W} = \mathbb{R} \times Y$, where a contract $w = (p, y)$ consists of a transfer $p \in \mathbb{R}$ from the agent to the principal and an allocation $y \in Y$, and where Y is a compact Polish space. Utilities are $U((p, y), \theta) = u(y, \theta) - p$ and $V(p, y) = v(y) + p$. For this existence argument, assume that Θ is compact. For every type θ , assume that an efficient allocation exists, and assume that the efficient allocations admit a measurable selection $y^*(\theta) \in \arg \max_{y \in Y} \{u(y, \theta) + v(y)\}$.

Consider the set $\Omega^E = \{((p, y), \mu) : y \in \arg \max_{y' \in Y} \{u(y', \theta) + v(y')\} \text{ for every } \theta \in \text{supp}(\mu)\}$.

Thus, an outcome $((p, y), \mu)$ belongs to Ω^E if and only if its allocation y is efficient for every type in the support of μ . We show that Ω^E is a stable set.

Internal stability. Take an outcome $o = ((p, y), \mu) \in \Omega^E$. For every $\theta \in \text{supp}(\mu)$, the allocation y maximizes total surplus. A strict improvement from an efficient allocation is

⁴In the continuous applications studied below, this convention does not give a boundary type of probability zero an independent role. The relevant indirect utility functions are continuous in type, so a strict loss at a boundary support point implies a strict loss for a neighborhood of nearby types, which has positive probability. If a boundary type is exactly indifferent, the improvement is weak and not strict. This case does not affect our results, because at every outcome of the stable sets we construct, the principal cannot gain strictly from any assignment into the set that leaves every type weakly better off.

impossible because, type by type, the total surplus available after any feasible assignment is at most as large as at y .

External stability. Take any outcome $o = ((p, y), \mu)$ and let $\bar{V} := V(p, y) = v(y) + p$ be the principal's payoff given o . For each type $\theta \in \text{supp}(\mu)$, take the efficient allocation $y^*(\theta)$ and choose the transfer $p^*(\theta) := \bar{V} - v(y^*(\theta))$. Let γ_o be the assignment measure induced by the map $\theta \mapsto (p^*(\theta), y^*(\theta))$ and the type distribution μ .

This assignment has three properties. First, every type is weakly better off. Indeed,

$$u(y^*(\theta), \theta) - p^*(\theta) = u(y^*(\theta), \theta) + v(y^*(\theta)) - \bar{V} \geq u(y, \theta) + v(y) - \bar{V} = u(y, \theta) - p,$$

where the inequality follows from the efficiency of $y^*(\theta)$. Second, the principal is exactly indifferent for every new outcome: $v(y^*(\theta)) + p^*(\theta) = \bar{V}$. Third, as the principal's payoff is constant and the new allocations are efficient for the types that receive them, each type receives a utility-maximizing contract in the assignment.

Finally, every outcome induced by γ_o lies in Ω^E . To see this, observe that types who induce the same contract are assigned the same allocation y^* , and that allocation is efficient for each of those types. The induced joint support is compact, since the type space and the closure of the selected allocation range are compact and $p^*(\theta) = \bar{V} - v(y^*(\theta))$. Its projection on the type space is therefore $\text{supp}(\mu)$. Moreover, as u and v are continuous, efficiency and the equality $p + v(y) = \bar{V}$ are closed conditions, so they hold at every point of the joint support. Thus every support type has a weakly improving support contract. We choose a version of the conditional beliefs whose support, at every supported contract, contains only types paired with that contract in the joint support. Hence every resulting outcome belongs

to Ω^E . We have therefore constructed, from every outcome o , an assignment measure with $\mathcal{O}(\gamma_o) \subset \Omega^E$ that weakly improves o .

Some useful results. The preceding existence argument is closely related to the observation that an outcome admitting no improvement to different outcomes must belong to every stable set. We now state this implication explicitly and pair it with a second result that will be useful in the applications below.

Lemma 1. *If, for some outcome o and for every γ_o that improves o , it holds that $\mathcal{O}(\gamma_o) = \{o\}$, then o is in every stable set.*

If o is the only element of $\mathcal{O}(\gamma_o)$ for every γ_o that improves it and o is not in a stable set, external stability would be violated as there would not exist an assignment measure to outcomes in the stable set that improves o .

Lemma 1 demonstrates that it follows from the properties of stable sets, and does not need to be separately assumed, that all unimprovable outcomes are in every stable set. However, as we show in our applications below, stable sets are not restricted to unimprovable outcomes: they may include ex post inefficient outcomes.

Lemma 2. *Let Ω be a stable set. For any outcome $o = (w, \mu)$, if there exists an outcome $o' = (w', \mu)$ such that $V(w') > V(w)$ and $U(w', \theta) > U(w, \theta)$ for all θ in the support of μ , then o is not in Ω .*

If a potential outcome exists where a single contract strictly improves the payoffs of both the principal and every agent type in the support of the belief, then the initial outcome cannot belong to a stable set. If the resulting outcome is itself in Ω , this implies that the original outcome cannot be in Ω . If the new outcome is not in Ω , then by external stability,

an assignment measure with outcomes in Ω must exist that improves it. But this same assignment measure would then strictly improve the original outcome, contradicting internal stability.

5. APPLICATION TO CONTRACTING WITHOUT COMMITMENT

This section applies our stability concept to contracting without commitment. We adopt the perspective of an outside analyst. Rather than modeling how the principal and the agent negotiate, the analyst works directly with assignments of agent types to contracts and with the principal's associated beliefs. Starting from the principal's prior and an initial contract, the analyst asks for the largest expected payoff among assignments that satisfy incentive compatibility and individual rationality and whose induced outcomes lie in a stable set. These assignments are analytical reduced forms. They can be interpreted as arising from mechanisms or from unspecified renegotiation processes. For every process whose terminal assignment satisfies these restrictions, the maximal value is an upper bound on the principal's payoff and therefore yields a lower bound on her loss from lacking commitment.

Terminal assignments. The principal's prior over the agent's type is $\mu_0 \in \Delta(\Theta)$ and is common knowledge between the agent and the principal. The initial contract is $w_0 \in \mathcal{W}$, which the agent can receive if there is no agreement. A *terminal assignment* is a candidate reduced-form description of how the interaction may end and is represented by a triple $(\mathcal{Z}, \varphi, \sigma)$. Here $\mathcal{Z}$ is a Polish space of indices endowed with the Borel σ -algebra, with each index z labeling one possible outcome.⁵ The measurable map $\varphi : \mathcal{Z} \rightarrow \mathcal{W}$ associates a contract $\varphi(z)$ with each index. The joint probability measure σ on $\Theta \times \mathcal{Z}$ records the distribution of the agent's type and the index. Its marginal on Θ equals the prior: $\sigma(A, \mathcal{Z}) =$

⁵In general, different z can label the same final contract but yield different beliefs.

$\mu_0(A)$ for every Borel $A \subset \Theta$. The principal's belief at index z is the regular conditional probability $\mu_0(\cdot \mid z)$ derived from σ . Existence of this regular conditional probability follows from the Polish-space structure of Θ and $\mathcal{Z}$.⁶ We write $\sigma_{\mathcal{Z}}$ for the marginal of σ on $\mathcal{Z}$. The outcome associated with index z is the pair $(\varphi(z), \mu_0(\cdot \mid z))$. The principal's expected payoff from a terminal assignment is

$$\Pi(\mathcal{Z}, \varphi, \sigma) = \int_{\Theta \times \mathcal{Z}} V(\varphi(z)) d\sigma(\theta, z).$$

A terminal assignment records which types are associated with which contract and the principal's belief at each index.

Restrictions on terminal assignments. We call a terminal assignment $(\mathcal{Z}, \varphi, \sigma)$ *admissible* if the following three conditions hold.

(IC) **Incentive Compatibility:** For σ -almost every (θ, z) , $z \in \arg \max_{z' \in \mathcal{Z}} U(\varphi(z'), \theta)$.

(IR) **Individual Rationality:** For σ -almost every (θ, z) , $U(\varphi(z), \theta) \geq U(w_0, \theta)$.

(ST) **Stability:** There exists a stable set Ω such that, for $\sigma_{\mathcal{Z}}$ -almost every z , $(\varphi(z), \mu_0(\cdot \mid z)) \in \Omega$.

Condition (IC) states that a type's assigned contract is best for him among the contracts in the assignment. It is the counterpart, for a terminal assignment, of requirement (iii) in Definition 2: no type is associated with a contract when another contract in the assignment would make him strictly better off. Condition (IR) states that no type is assigned a contract worse than w_0 , which he can secure by not participating; it is the counterpart of requirement (i) in Definition 2, evaluated at the initial outcome (w_0, μ_0) . Condition (ST) is the restriction

⁶See Durrett (2019), Theorem 4.1.17.

due to the lack of commitment: every outcome generated with positive probability must lie in one common stable set.

The maximal payoff. Accordingly, the largest expected payoff consistent with these restrictions is

$$V^* = \sup_{(\mathcal{Z}, \varphi, \sigma)} \Pi(\mathcal{Z}, \varphi, \sigma),$$

where the supremum is taken over all admissible terminal assignments for which Π is defined. Relative to the full-commitment payoff, V^* determines the smallest loss implied by incentive compatibility, individual rationality, and stability: conditions (IC) and (IR) are the constraints of the principal's full-commitment problem, so V^* is at most the full-commitment payoff. We call an admissible terminal assignment that attains V^* *payoff-maximizing*.

Mechanism interpretation. A terminal assignment has a familiar mechanism interpretation. We use the mechanism interpretation for intuition and to describe payoff-maximizing terminal assignments in a familiar way. Interpret $\mathcal{Z}$ as a set of messages, φ as an allocation rule, and σ as the distribution generated by each type's choice of message. Then conditions (IC) and (IR) are the usual incentive-compatibility and participation constraints, and $\mu_0(\cdot \mid z)$ is the principal's posterior after observing message z . A mechanism, together with behavior that satisfies incentive compatibility and individual rationality and induces outcomes in a stable set, thus generates an admissible terminal assignment. Conversely, every admissible terminal assignment can be read as the reduced form of such a mechanism or, more generally, of an unspecified game or renegotiation process.

6. APPLICATION TO NONLINEAR CONTRACTS

This section characterizes the maximal payoff V^* and a payoff-maximizing terminal assignment in a nonlinear-contracting environment. We introduce the setup, state the main result, and prove it through a series of lemmas.

Setup. The principal and the agent want to implement a two-dimensional contract $w = (p, q)$ with price $p \in \mathbb{R}$ and quality $q \in \mathbb{R}_+$. Her profit from a single contract (p, q) is

$$V(w) = p - c(q).$$

Assume $c \in \mathcal{C}^2$ and write c_q and c_{qq} for the first and second derivatives of c . Assume c is increasing and strictly convex: $c_q \geq 0$ and $c_{qq} > 0$.

The agent's type θ is drawn from $\Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$. The agent's utility is

$$U(w, \theta) = u(q, \theta) - p.$$

Assume $u \in \mathcal{C}^2$ and write u_q , u_{qq} , u_θ , and $u_{q\theta}$ for the corresponding partial and cross-partial derivatives of u . Assume $u_q > 0$ and $u_{qq} \leq 0$, so that $u(\cdot, \theta)$ is strictly increasing and concave in q , and that u satisfies single crossing: $u_\theta \geq 0$ and $u_{q\theta} \geq 0$ with strictness if $q > 0$. A higher type therefore receives greater utility and higher marginal utility from a given $q > 0$. We use the normalization $u(0, \theta) = 0$. Let $q^*(\theta)$ denote the efficient quality level for type θ , defined by

$$q^*(\theta) \in \arg \max_{q \geq 0} \{u(q, \theta) - c(q)\}.$$

Assume further that for every $\theta \in \Theta$, $u_q(0, \theta) > c_q(0)$ and $\lim_{q \rightarrow \infty} [c_q(q) - u_q(q, \theta)] > 0$. Since $u_{qq}(q, \theta) \leq 0$ and $c_{qq}(q) > 0$, $u(\cdot, \theta) - c(\cdot)$ is strictly concave. Therefore, our assumptions

imply that $q^*(\theta)$ exists, is unique, and satisfies $q^*(\theta) > 0$. That is, the efficient quality itself is strictly positive. Its slope is also strictly positive, as shown below. The efficient quality is characterized by the first-order condition

$$(3) \quad c_q(q^*(\theta)) = u_q(q^*(\theta), \theta).$$

By the implicit function theorem,

$$q_\theta^*(\theta) = \frac{u_{q\theta}(q^*(\theta), \theta)}{c_{qq}(q^*(\theta)) - u_{qq}(q^*(\theta), \theta)} > 0,$$

where the inequality follows from $u_{q\theta} > 0$ at the positive quality $q^*(\theta)$, $c_{qq} > 0$, and $u_{qq} \leq 0$.

The principal holds a prior belief $\mu_0 \in \Delta(\Theta)$ over the agent's type. The prior μ_0 admits a density f with respect to the Lebesgue measure on $[\underline{\theta}, \bar{\theta}]$, where f is continuous and satisfies $0 < \underline{f} \leq f(\theta) \leq \bar{f} < \infty$ for all $\theta \in [\underline{\theta}, \bar{\theta}]$. In particular, μ_0 has full support on $[\underline{\theta}, \bar{\theta}]$ and is atomless.⁷ For this application, we specialize the initial contract introduced above to $w_0 = (0, 0)$.

While the model accommodates a range of settings, we interpret it throughout as “selling when price and quality matter.”⁸ In this interpretation, the principal is the seller and the agent is the buyer. The contract (p, q) specifies the price and quality of the good. The seller incurs cost $c(q)$ to produce quality q , and a buyer of type θ derives utility $u(q, \theta)$ from consuming it. Higher types enjoy greater total and marginal utility.

The main result. We now characterize a payoff-maximizing admissible terminal assignment. The argument proceeds in two steps: we derive necessary restrictions on the terminal

⁷These assumptions, in particular, rule out discrete distributions.

⁸Other interpretations include procurement from a monopolistic supplier, labor contracts, regulation of a monopolist, and financial contracts; see Section 2.15 of [Laffont and Martimort \(2009\)](#) for an overview.

outcomes that may appear in an admissible assignment and then construct a stable set containing the outcomes of an assignment that attains an upper bound on the principal's payoff. Stating the result requires the following terminology.

Definition 4. *Define μ^θ as the probability measure that puts probability 1 on type θ . That is, for all measurable sets A , $\mu^\theta(A) = 1$ whenever $\theta \in A$ and $\mu^\theta(A) = 0$ otherwise. We use the following terminology:*

- (1) *We call an outcome of the form $((p, q), \mu^\theta)$ a separating outcome. Efficient and separating outcomes are denoted by $((p, q^*(\theta)), \mu^\theta)$.*
- (2) *We call $((p, q), \mu)$ a pooling outcome if it is not a separating outcome.*

We will show that there is a payoff-maximizing terminal assignment that partitions the type space into a countably infinite sequence of intervals $[\theta_n^*, \theta_{n+1}^*)$, $n \in \mathbb{N}_0$, with $\theta_n^* < \theta_{n+1}^*$ and $\theta_n^* \rightarrow \bar{\theta}$. Within each interval, every type receives the same contract (p_n^*, q_n^*) . The quality $q_n^* = q^*(\theta_n^*)$ is efficient for the interval's lowest type. We call this property “no distortion at the bottom.” The price is pinned down by indifference: $p_0^* = u(q^*(\underline{\theta}), \underline{\theta})$, and for $n \geq 1$ the price p_n^* leaves type θ_n^* indifferent between (p_n^*, q_n^*) and (p_{n-1}^*, q_{n-1}^*) . The highest type $\bar{\theta}$ is handled separately, through an efficient separating contract.

This terminal assignment has index set $\mathcal{Z}^* = \mathbb{N}_0 \cup \{z_{\bar{\theta}}\}$ with $\varphi^*(n) = (p_n^*, q_n^*)$ for $n \in \mathbb{N}_0$ and $\varphi^*(z_{\bar{\theta}})$ an efficient separating contract for $\bar{\theta}$. Every type $\theta \in [\theta_n^*, \theta_{n+1}^*)$ is associated with index n , and type $\bar{\theta}$ with index $z_{\bar{\theta}}$. Let σ^* denote the resulting joint distribution of types and indices. The terminal outcomes (the pooling outcomes $o_n^* = ((p_n^*, q_n^*), \mu_0(\cdot | [\theta_n^*, \theta_{n+1}^*)))$) together with all efficient separating outcomes form a stable set Ω^* . We will prove the following proposition, which characterizes the maximal payoff and an admissible terminal assignment that attains it.

Proposition 1. *The supremum of the principal's expected payoff over admissible terminal assignments is attained and equals the value of the cutoff problem*

$$\max_{\{\theta_n\}_{n \geq 0}} \sum_{n=0}^{\infty} \left[u(q^*(\underline{\theta}), \underline{\theta}) + \sum_{m=1}^n [u(q^*(\theta_m), \theta_m) - u(q^*(\theta_{m-1}), \theta_m)] - c(q^*(\theta_n)) \right] \mu_0([\theta_n, \theta_{n+1}))$$

subject to

$$\theta_0 = \underline{\theta}, \quad \theta_n < \theta_{n+1} \text{ for all } n, \quad \lim_{n \rightarrow \infty} \theta_n = \bar{\theta}.$$

For any solution $\{\theta_n^*\}_{n \geq 0}$, define $q_n^* = q^*(\theta_n^*)$, $p_0^* = u(q^*(\underline{\theta}), \underline{\theta})$,

$$p_n^* = u(q^*(\underline{\theta}), \underline{\theta}) + \sum_{m=1}^n [u(q^*(\theta_m^*), \theta_m^*) - u(q^*(\theta_{m-1}^*), \theta_m^*)],$$

and $\mu_n^*(\cdot) = \mu_0(\cdot \mid [\theta_n^*, \theta_{n+1}^*))$. Then the terminal assignment with index set $\mathcal{Z}^* = \mathbb{N}_0 \cup \{z_{\bar{\theta}}\}$, contracts $\varphi^*(n) = (p_n^*, q_n^*)$ for $n \in \mathbb{N}_0$ and $\varphi^*(z_{\bar{\theta}}) = (p_{\bar{\theta}}, q^*(\bar{\theta}))$, $p_{\bar{\theta}} := \lim_{n \rightarrow \infty} p_n^*$, in which every type in $[\theta_n^*, \theta_{n+1}^*)$ is associated with index n and the highest type $\bar{\theta}$ with index $z_{\bar{\theta}}$, is admissible and attains this supremum.

Its positive-mass terminal outcomes are

$$o_n^* = ((p_n^*, q^*(\theta_n^*)), \mu_0(\cdot \mid [\theta_n^*, \theta_{n+1}^*))) , \quad n \in \mathbb{N}_0.$$

There exists a stable set containing these outcomes; for example, the union of all efficient separating outcomes and the pooling outcomes $\{o_n^*\}_{n \geq 0}$ is a stable set.

We call this terminal assignment the *cutoff assignment* of the solution $\{\theta_n^*\}_{n \geq 0}$, and we call the outcomes at its indices of positive probability its positive-mass terminal outcomes. In the mechanism interpretation, this terminal assignment is generated by a menu of countably

many contracts, one for each pooling interval, together with an efficient contract for the highest type: each type takes the contract of its interval.

Figure 1 illustrates the quality received by each type under the terminal assignment of Proposition 1 for $V(p, q) = p - 0.5q^2$, $U((p, q), \theta) = \theta q - p$, $\Theta = [1, 2]$, and μ_0 uniform on $[1, 2]$. The efficient quality is $q^*(\theta) = \theta$, and the second-best quality with commitment is $q^{SB}(\theta) = 2\theta - 2$.

For this linear-quadratic example, rescaling the remaining distance to the highest type 2 after any cutoff leaves the continuation problem unchanged up to a positive factor, when the payoff is measured as the gain over the efficient separation of all types. The solution of the cutoff problem therefore repeats at every cutoff: the distance from the next cutoff to 2 is the same fraction r of the current distance,

$$2 - \theta_{n+1}^* = r(2 - \theta_n^*).$$

The first-order condition for this repeated one-step problem gives $r = \frac{\sqrt{3}-1}{2}$. Since $\theta_0^* = 1$, the cutoffs are $\theta_n^* = 2 - r^n$. Thus the quality under the terminal assignment of Proposition 1 is constant on each interval $[\theta_n^*, \theta_{n+1}^*)$ and equal to $q_n^* = q^*(\theta_n^*) = \theta_n^*$.

The proof of Proposition 1 proceeds by a series of lemmas. We rule out outcomes that cannot belong to any stable set, bound the payoff of every admissible terminal assignment, construct an admissible terminal assignment that attains the bound, and exhibit a stable set that contains its outcomes.

We first provide necessary conditions on outcomes contained in stable sets. The lemma shows that stability already restricts sharply which outcomes can be terminal: a separating outcome can belong to a stable set only if it is efficient, while a pooling outcome can belong

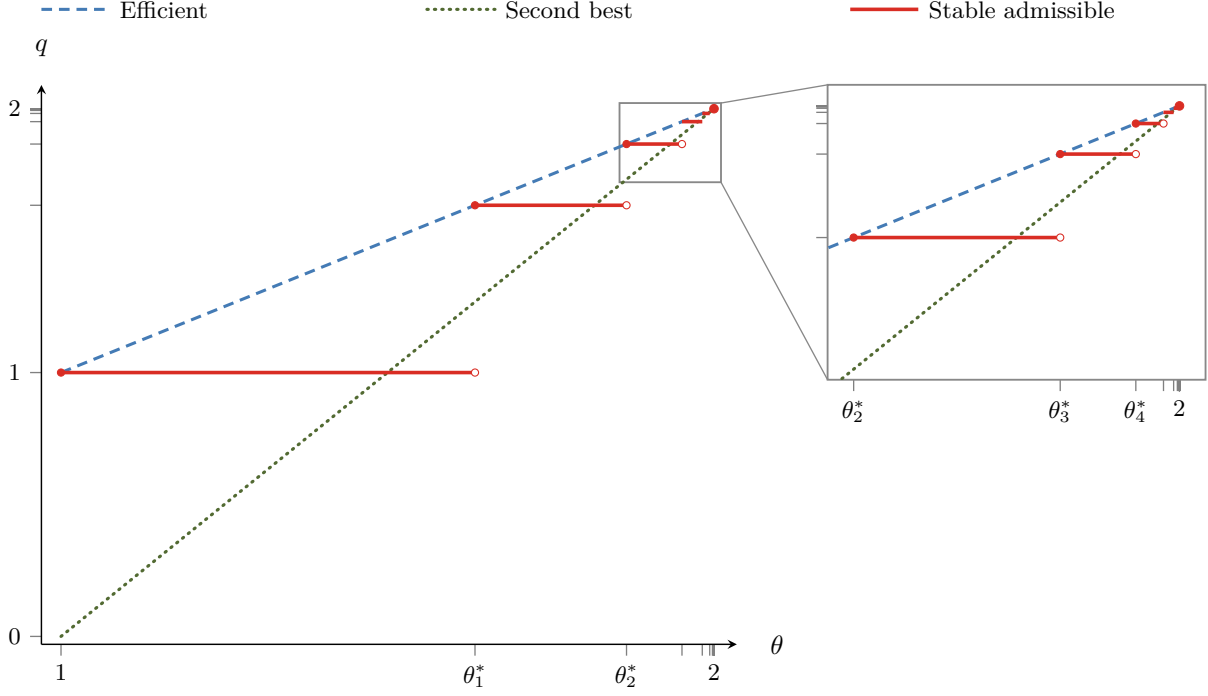


FIGURE 1. Comparison of the efficient quality levels (blue) with the quality levels under the second-best solution with commitment (green) and the quality levels under the terminal assignment of Proposition 1 (red).

to a stable set only if the types it pools form a connected interval and the contract specifies a quality which is efficient for at least one type in the pool.

Lemma 3 (Necessary conditions for outcomes in stable sets). *The following restrictions for outcomes in stable sets hold.*

- (1) *For every stable set Ω , if a separating outcome $((p, q), \mu^\theta)$ belongs to Ω , then it is efficient: $q = q^*(\theta)$.*
- (2) *For every stable set Ω , if a pooling outcome $((p, q), \mu)$ belongs to Ω , then $\text{supp}(\mu)$ is connected and there exists $s \in \text{supp}(\mu)$ such that $q = q^*(s)$.*

With the necessary conditions in hand, we can pin down the shape of any admissible terminal assignment. Although an admissible terminal assignment may in principle involve

an elaborate index set and associate a type with several indices, the next lemma shows that nothing is lost by describing it type by type, assigning each type a single price-quality pair. Incentive compatibility then delivers two familiar properties: higher types receive weakly higher quality, and the set of types receiving any given contract is an interval. Up to a null set, the type space thereby consists of disjoint pooling intervals, ordered from low quality to high quality, and of separated types who receive their efficient quality.

Lemma 4 (Interval representation). *Let $(\mathcal{Z}, \varphi, \sigma)$ be an admissible terminal assignment. Up to μ_0 -null sets, and after merging indices that carry the same contract, the assignment of contracts to types can be described by a deterministic direct allocation, a function $(P, Q) : \Theta \rightarrow \mathbb{R} \times \mathbb{R}_+$. The deterministic direct allocation is incentive compatible in the standard sense (no type prefers another type's contract), and its quality assignment Q is nondecreasing in type. Consequently, for every contract (p, q) held with positive μ_0 -probability, the set*

$$I(p, q) = \{\theta \in \Theta : (P(\theta), Q(\theta)) = (p, q)\}$$

is an interval, up to endpoints and μ_0 -null sets. We call $I(p, q)$ a pooling interval. Moreover, for every positive-mass pooling interval $I(p, q) = [a, b]$, there exists $s \in [a, b]$ such that $q = q^*(s)$. Outside the positive-mass pooling intervals, μ_0 -almost every type receives its efficient quality, $Q(\theta) = q^*(\theta)$.

The next lemma establishes a “no distortion at the bottom” property. Within any pooling interval $[a, b]$ whose types hold the same contract, the principal's payoff rises if the quality is replaced by the one that is efficient for the lowest type a , namely $q^*(a)$. The intuition is direct. Suppose instead that the pool's quality were efficient for some interior type $s > a$.

That quality is then too high for every type in $[a, s)$, in the sense that the marginal cost of producing it exceeds the marginal value to those types. Lowering the quality and reducing the price such that type a is indifferent increases the profit from all types in $[a, b)$, because the cost saving exceeds the price reduction, and also lowers the rent of all types above b .

Lemma 5 (No distortion at the bottom). *Let $(\mathcal{Z}, \varphi, \sigma)$ be an admissible terminal assignment and let (P, Q) be its deterministic direct allocation from Lemma 4. Let $\{[a_i, b_i]\}_{i \in \mathcal{I}}$ be the family of positive-mass pooling intervals. By Lemma 4, for each $i \in \mathcal{I}$ the common quality on $[a_i, b_i)$ can be written as $q^*(s_i)$ for some $s_i \in [a_i, b_i]$.*

Replacing, for all $i \in \mathcal{I}$ with $s_i > a_i$, the pool quality $q^*(s_i)$ by $q^*(a_i)$ and adjusting prices accordingly yields a new direct allocation $(\widehat{P}, \widehat{Q})$ that is incentive compatible, individually rational, satisfies the restrictions of Lemma 3, and gives the principal a weakly higher payoff, which is strictly higher if $s_i > a_i$ for at least one i .

Taken together, the preceding three lemmas reduce the payoff-maximization problem to a choice of cutoff sequence, that is, a sequence that satisfies the constraints of the cutoff problem in Proposition 1. Lemma 3 restricts which outcomes can appear in a stable set. Lemma 4 shows that any admissible terminal assignment can be described by a function in which the types holding each contract form an interval or receive their efficient contract. Lemma 5 shows that the principal's payoff rises if the quality within each interval is the efficient level for its lowest type. Once the interval boundaries fix the quality levels, incentive compatibility determines the prices through the recursive indifference conditions in Proposition 1. The payoff of every admissible terminal assignment is therefore bounded by the value of a relaxed problem, in which the set of interval boundaries may be any closed set of types. The next lemma formalizes this reduction and establishes that it is tight: every solution of

the relaxed problem consists of the highest type and an increasing sequence of cutoffs that converges to it, so the cutoff problem provides an upper bound on the principal's payoff and admits a solution. Every solution of the cutoff problem defines an admissible terminal assignment that attains this bound. In particular, this terminal assignment does not include efficient and separating contracts with positive prior probability.

Lemma 6 (Cutoff reduction, existence, and implementation). *The supremum of the principal's payoff over all admissible terminal assignments is no larger than the value of the cutoff problem displayed in Proposition 1. The cutoff problem admits a solution. For every solution $\{\theta_n^*\}_{n \geq 0}$, the associated cutoff assignment with the recursive prices displayed in Proposition 1 is incentive compatible and individually rational, attains the cutoff value, and has positive-mass terminal outcomes contained in a stable set. Hence the cutoff value equals the supremum of the principal's payoff over admissible terminal assignments.*

The preceding four lemmas, together with the stable set and the contract of the highest type that the proof of Lemma 6 constructs, establish Proposition 1. In the mechanism interpretation, the terminal assignment of Proposition 1 is a cutoff mechanism: it pools types into countably many intervals accumulating at $\bar{\theta}$, gives each interval the quality efficient for its lowest type, and determines prices through the indifference at the boundaries of the intervals.

7. APPLICATION TO THE COASE CONJECTURE

This section considers the Coase conjecture as a special case of our framework. The Coase conjecture posits that if a seller cannot commit not to sell an indivisible durable good in the future, she can charge at most the buyer's lowest valuation. [Ausubel and Deneckere \(1989\)](#)

show that the conjecture holds as long as the seller's cost is strictly below the buyer's lowest valuation (the gap case). However, they also show that when the seller's cost is at least as high as the buyer's lowest valuation (the no-gap case), she can still charge the monopoly price even without commitment. Our approach rederives this gap–no-gap distinction using admissible terminal assignments. In the gap case, stable-set membership rules out no-trade outcomes, so every admissible terminal assignment puts probability one on trade, and the seller's price is bounded by the buyer's lowest valuation. In the no-gap case, we construct a stable set that contains the terminal outcomes generated by the monopoly posted price, so the monopoly payoff is consistent with stability. As in [Ausubel and Deneckere \(1989\)](#), the stability conditions are not strong enough to narrow the set of possible outcomes to a single one.

Setup. Consider a monopolist selling a single indivisible good to a buyer. A contract is a payoff-relevant pair $w = (p, q)$, where $p \in \mathbb{R}$ is the price and $q \in \{0, 1\}$ indicates whether the good is exchanged ($q = 1$) or not ($q = 0$). The seller incurs a cost of c to produce the good. The seller's utility function is thus given by

$$V(w) = p - cq.$$

The buyer's valuation $\theta \in \Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_+$ for consuming the good is private information.

The buyer's utility is thus given by

$$U(w, \theta) = \theta q - p.$$

The seller's prior about the buyer's valuation is given by $\mu_0 \in \Delta(\Theta)$ with full support on Θ .

For this application, the initial contract is $w_0 = (0, 0)$. We study two cases: the gap case

($c < \underline{\theta}$) and the no-gap case ($c \geq \underline{\theta}$). In the gap case, a gap exists between the seller's cost and the buyer's lowest valuation, making trade efficiency common knowledge. In the second case, the buyer may have a valuation for which trade is inefficient.

The gap case. Suppose $c < \underline{\theta}$, so that trade is efficient for every buyer type. The key implication of stability is that no no-trade outcome can belong to a stable set. It follows that every admissible terminal assignment puts probability one on trade, so the seller can charge at most the buyer's lowest valuation.

Proposition 2. *Let $c < \underline{\theta}$. If Ω is a stable set and $((p, q), \mu) \in \Omega$, then $q = 1$. Consequently, every admissible terminal assignment puts probability one on trade, and the maximal seller payoff over admissible terminal assignments is $\underline{\theta} - c$. It is attained by the terminal assignment with $\mathcal{Z} = \{1\}$, $\varphi(1) = (\underline{\theta}, 1)$, and every buyer type associated with index 1.*

In the mechanism interpretation, this terminal assignment is generated by a posted price $\underline{\theta}$ that is accepted by all buyer types. The argument uses only a necessary implication of stable-set membership. Any no-trade outcome can be strictly improved for the seller and for every buyer type by a single trade contract, so Lemma 2 rules it out of every stable set. This delivers the upper bound for all admissible terminal assignments. The terminal assignment generated by the posted price $\underline{\theta}$ attains the bound and is admissible, because the set of all trade outcomes is itself a stable set. The result is consistent with the literature on the Coase conjecture: the monopolist effectively competes against herself, driving the price down to the buyer's lowest valuation.⁹

The no-gap case. Now suppose $c \geq \underline{\theta}$ and $c < \bar{\theta}$. We do not characterize all stable sets. It is enough to construct one stable set that contains the terminal outcomes generated by

⁹See for example [Fudenberg et al. \(1985\)](#).

the monopoly posted price. Since every admissible terminal assignment satisfies conditions (IC) and (IR), it is feasible in the seller's full-commitment problem. The monopoly payoff is therefore an upper bound on the seller's payoff over admissible terminal assignments, and the stable-set construction below shows that this bound is attained.

Proposition 3. *Let $p^M \in \arg \max_p \mu_0([p, \bar{\theta}]) (p - c)$ be a monopoly price,¹⁰ and suppose that $p^M > c$. Let $\mu^M = \mu_0(\cdot \mid [\underline{\theta}, p^M])$ denote the seller's posterior after rejection of the monopoly price; under full support, $\text{supp}(\mu^M) = [\underline{\theta}, p^M]$. Define Ω^M as the set consisting of the following outcomes:*

- (1) *all trade outcomes $((p, 1), \mu)$ with $\min \text{supp}(\mu) \geq c$;*
- (2) *all no-trade outcomes $((p, 0), \mu)$ with $\max \text{supp}(\mu) \leq c$;*
- (3) *the no-trade outcome $((0, 0), \mu^M)$.*

Then Ω^M is a stable set. Consequently, the maximal seller payoff over admissible terminal assignments equals the monopoly payoff $\mu_0([p^M, \bar{\theta}]) (p^M - c)$. It is attained by the terminal assignment with $\mathcal{Z} = \{0, 1\}$, $\varphi(0) = (0, 0)$, $\varphi(1) = (p^M, 1)$, in which types $\theta < p^M$ are associated with index 0 and types $\theta \geq p^M$ with index 1.

The set Ω^M consists of all ex post efficient outcomes together with one additional, inefficient no-trade outcome: the rejection outcome generated by the monopoly posted price. The efficient outcomes are the trade outcomes whose buyer valuations all exceed the cost (item 1) and the no-trade outcomes whose buyer valuations all fall below it (item 2). External stability follows because any outcome can be improved by separating the buyer types at the cost threshold, assigning a trade contract to types above cost and a no-trade contract to

¹⁰The maximum is attained: $p \mapsto \mu_0([p, \bar{\theta}])$ is nonincreasing and left-continuous, so $\mu_0([p, \bar{\theta}]) (p - c)$ is upper semicontinuous on the compact interval $[c, \bar{\theta}]$, on which any maximizer lies.

types at or below cost. Internal stability is immediate for the efficient outcomes. The only additional outcome to check is the rejection outcome $((0, 0), \mu^M)$: any strict improvement into Ω^M would have to serve types arbitrarily close to c from above at a price no greater than c , which leaves the seller no strict gain. In the mechanism interpretation, the associated mechanism is a posted price p^M .

Analogous results, in particular the possibility of inefficient monopoly pricing, hold in a discrete-type model, but only when the seller's cost coincides with one of the buyer's discrete valuations. When it does, the monopoly price can be sustained through a stable set like the one in Proposition 3. When the cost instead falls strictly between two valuations, every outcome in a stable set is efficient, and inefficient monopoly pricing cannot be sustained. This dependence on whether the cost lands on the price grid of buyer valuations is also observed by [von der Fehr and Kühn \(1995\)](#) in a noncooperative model (see their discussion starting on page 794).

8. RELATIONSHIP TO THE LITERATURE

[Wilson \(1978\)](#) introduced the Incomplete Information Core, one of the earliest formal treatments of stability under incomplete information.¹¹ This literature analyzes how different assumptions about information sharing among privately informed agents influence potential collective deviations. The Incomplete Information Core evaluates the stability of an ex ante allocation rule, that is, prior to the specific realization of outcomes. Our approach, in contrast, assesses the stability of specific contract-belief pairs after an outcome realizes and the principal's beliefs update based on information the outcome reveals.

¹¹For an overview of more recent results, see [Forges et al. \(2002\)](#) or [Forges and Serrano \(2013\)](#).

We illustrate the differences between our stability concept and the Incomplete Information Core with a simplified example of the Coase conjecture, adapted from [Liu \(2020\)](#). We employ the Credible Core of [Dutta and Vohra \(2005\)](#), which, like [Vohra \(1999\)](#), incorporates incentive compatibility. Consider a seller with zero cost and a buyer whose valuation is either 1 or $\varepsilon \in (0, 1/2)$, each with prior probability $1/2$. This is the “gap case” of the Coase conjecture.

Under the Credible Core, the monopoly pricing rule, where only the high-type buyer trades at price 1, is stable. To understand why, consider a potential deviation to serve the low-type buyer, requiring a price $p \leq \varepsilon$. Because the high-type buyer would also accept price p , incentive compatibility implies that the seller’s belief does not change when making such an offer. Her expected payoff from this deviation is therefore p . Since $p \leq \varepsilon < 1/2$ and the expected profit from asking the monopoly price is $1/2$, the seller has no incentive to deviate. Thus, the monopoly pricing rule is not blocked.

In contrast, our approach evaluates the stability of outcomes after they realize and beliefs update. We show that, in this “gap case,” any outcome in a stable set in our framework must involve trade with the low-type buyer. A “no-trade” outcome reveals the buyer’s low type, assuming the high type trades at the monopoly price. Trade can strictly improve such an outcome. The key distinction is that our framework allows ex post evaluation of potential improvements, using beliefs updated from observed outcomes.

The stable set of [von Neumann and Morgenstern \(1944\)](#) has received renewed interest in the recent literature. For example, [Ray and Vohra \(2015\)](#) extend the notion of a stable set to a farsighted dominance relation that captures dynamic processes arising from sequential outcome improvements. [Dutta and Vohra \(2017\)](#) refine the concept by requiring players to hold rational expectations about continuation paths of improvements, and [Dutta and](#)

Vartiainen (2020) incorporate history dependence in their concept. The continuing interest in von Neumann–Morgenstern stability underscores the simplicity and power of the underlying idea. Thus, it is natural to build on the idea of set stability in environments with private information, as an abstraction for renegotiation, and to use it to bound the principal’s payoff under limited commitment.

Beyond the von Neumann–Morgenstern notion, other cooperative approaches to stability exist. Liu (2020) analyzes stability in a two-sided market with asymmetric information. Stability is defined with respect to a matching between market participants, accompanied by corresponding beliefs. He constructs beliefs to deter pairwise deviations from the matching, while ensuring individual rationality. This approach complements the belief-free approach to stability of Liu et al. (2014), in which a pair blocks a matching only if the deviation is profitable under all reasonable beliefs. Asheim and Nilssen (1997) consider a monopolistic insurance market with a discrete type space. They employ the concept of a standard of behavior, which resembles the definition of a stable set in this manuscript. As in our case, their framework proves useful in deriving clear results in an otherwise complex environment.

Cooperative solution concepts have also been used to study the implications of limited commitment in mechanism design. Vartiainen (2013) analyzes auctions without commitment, where the principal repeatedly proposes new mechanisms, unbound by previous outcomes. He introduces a cooperative solution concept tailored to this environment and shows that this approach implements only an English auction without a reserve price.¹² More recently, Vartiainen (2023) also employs recursive reasoning to characterize renegotiation-proofness and proposes a history-dependent stability concept. He discusses our stable-set framework in this context, describing his concept as a history-dependent extension of our framework.

¹²This aligns with the noncooperative approach of Liu et al. (2019) to a similar problem.

[Neeman and Pavlov \(2013\)](#), like us, abstract from specific renegotiation games. However, unlike us, they argue that for a mechanism's outcomes to be renegotiation-proof under any procedure, no Pareto improvements should be possible after the mechanism is played. The conceptual difference between our approach and theirs is significant. Their approach permits all Pareto-improving outcomes to be blocking, even if those outcomes are themselves subject to renegotiation. In contrast, we consider outcomes potentially blocking only if they are themselves in a stable set. In particular, in contrast to the approach of [Neeman and Pavlov \(2013\)](#), our approach does not exclude inefficient outcomes.

Our paper also relates to the vast literature on mechanism design without commitment, which models limited commitment using games of sequentially posted mechanisms. For example, [Bester and Strausz \(2001\)](#), [Evans and Reiche \(2015\)](#), [Skreta \(2006\)](#), and [Skreta \(2015\)](#) consider games where a designer posts a new mechanism a finite number of times and solves the problem via backward induction. [Doval and Skreta \(2022\)](#) derive a revelation principle for an infinite sequence of mechanisms. [Pęski \(2022\)](#) adopts an approach intermediate between models with finite and infinite sequences of posted mechanisms. He assumes that the principal and the agent bargain over the mechanism in a random-offer bargaining game. Once they agree on a mechanism, they are committed to its outcome. Regarding the Coase conjecture, he shows that the principal can implement the optimal mechanism. The results of [Gretschko and Wambach \(2017\)](#) complement the ideas presented here. In that paper, we assume a discrete type space and analyze a renegotiation game where the principal may repeatedly propose new mechanisms. Stable sets, we show, arise as outcomes of a perfect Bayesian equilibrium of this game.

[Strulovici \(2017\)](#) shows that in a specific infinite-horizon bargaining protocol over nonlinear contracts, efficiency and full separation arise in any perfect Bayesian equilibrium, given binary types and vanishing friction. In our nonlinear-contracting environment, the same stability argument implies that only efficient and separating outcomes can belong to a stable set with a finite type space. Since this applies to any finite number of types, not only two, it provides a stable-set counterpart to his result within this environment. Furthermore, our approach highlights a conceptual distinction between discrete and continuous type spaces: in the continuous case, pooling and thus inefficient outcomes can belong to a stable set.

[Chen \(2020\)](#) models a dynamic process where initial pooling reflects the expected payoffs from gradual information revelation and surplus extraction over multiple periods. In that setting, contracts are executed in each period. In contrast, our framework focuses on the stability of a single, ultimately implemented contract. This yields different implications for optimal pooling. For example, we obtain a “no-distortion-at-the-bottom” property. This arises because the relevant counterfactual entails a direct shift to efficient outcomes, not a gradual convergence through a sequence of contracts. As [Malcomson \(2016\)](#) and [Malcomson \(2025\)](#) note, in applications, agents are often grouped and treated identically over time, for instance, employees in pay grades or suppliers in categories. [Malcomson \(2016\)](#) offers a dynamic relational-contracting rationale for this practice. Our results provide an alternative explanation grounded in stability.

REFERENCES

ALIPRANTIS, C. D. AND K. C. BORDER (2006): *Infinite dimensional analysis: a hitchhiker’s guide*, Springer.

- ASHEIM, G. B. AND T. NILSSEN (1997): “Insurance monopoly and renegotiation,” *Economic Theory*, 9, 341–354.
- AUSUBEL, L. M. AND R. J. DENECKERE (1989): “Reputation in bargaining and durable goods monopoly,” *Econometrica*, 57, 511–531.
- BESTER, H. AND R. STRAUZ (2001): “Contracting with imperfect commitment and the revelation principle: The single agent case,” *Econometrica*, 69, 1077–1098.
- BULOW, J. I. (1982): “Durable-goods monopolists,” *Journal of Political Economy*, 90, 314–332.
- CHEN, W. (2020): “Essays in mechanism design,” Ph.D. thesis, Stanford University.
- COASE, R. (1972): “Durability and monopoly,” *Journal of Law and Economics*, 15, 143–149.
- DOVAL, L. AND V. SKRETA (2022): “Mechanism design with limited commitment,” *Econometrica*, 90, 1463–1500.
- (2024): “Optimal mechanism for the sale of a durable good,” *Theoretical Economics*, 19, 865–915.
- DURRETT, R. (2019): *Probability: theory and examples*, vol. 49, Cambridge University Press.
- DUTTA, B. AND H. VARTIAINEN (2020): “Coalition formation and history dependence,” *Theoretical Economics*, 15, 159–197.
- DUTTA, B. AND R. VOHRA (2005): “Incomplete information, credibility and the core,” *Mathematical Social Sciences*, 50, 148–165.
- (2017): “Rational expectations and farsighted stability,” *Theoretical Economics*, 12, 1191–1227.
- EVANS, R. AND S. REICHE (2015): “Contract design and non-cooperative renegotiation,” *Journal of Economic Theory*, 157, 1159–1187.

- FORGES, F., E. MINELLI, AND R. VOHRA (2002): “Incentives and the core of an exchange economy: a survey,” *Journal of Mathematical Economics*, 38, 1–41.
- FORGES, F. AND R. SERRANO (2013): “Cooperative games with incomplete information: Some open problems,” *International Game Theory Review*, 15, 1340009.
- FUDENBERG, D., D. LEVINE, 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, 73–98.
- GRETSCHKO, V. AND A. WAMBACH (2017): “Contract (re-) negotiation with private and common values,” *ZEW-Centre for European Economic Research Discussion Paper*.
- GUL, F., H. SONNENSCHN, AND R. WILSON (1986): “Foundations of dynamic monopoly and the Coase conjecture,” *Journal of Economic Theory*, 39, 155–190.
- LAFFONT, J.-J. AND D. MARTIMORT (2009): *The Theory of Incentives*, Princeton University Press.
- LIU, Q. (2020): “Stability and Bayesian consistency in two-sided markets,” *American Economic Review*, 110, 2625–2666.
- LIU, Q., G. J. MAILATH, A. POSTLEWAITE, AND L. SAMUELSON (2014): “Stable matching with incomplete information,” *Econometrica*, 82, 541–587.
- LIU, Q., K. MIERENDORFF, X. SHI, AND W. ZHONG (2019): “Auctions with limited commitment,” *American Economic Review*, 109, 876–910.
- MALCOMSON, J. M. (2016): “Relational incentive contracts with persistent private information,” *Econometrica*, 84, 317–346.
- (2025): “Grouping agents with persistent types,” *Games and Economic Behavior*, 152, 150–174.

- MILGROM, P. AND I. SEGAL (2002): “Envelope theorems for arbitrary choice sets,” *Econometrica*, 70, 583–601.
- MUSSA, M. AND S. ROSEN (1978): “Monopoly and product quality,” *Journal of Economic Theory*, 18, 301–317.
- NEEMAN, Z. AND G. PAVLOV (2013): “Ex-post renegotiation-proof mechanism design,” *Journal of Economic Theory*, 148, 473–501.
- PEŃSKI, M. (2022): “Bargaining with mechanisms,” *American Economic Review*, 112, 2044–2082.
- RAY, D. AND R. VOHRA (2015): “The farsighted stable set,” *Econometrica*, 83, 977–1011.
- SKRETA, V. (2006): “Sequentially optimal mechanisms,” *Review of Economic Studies*, 73, 1085–1111.
- (2015): “Optimal auction design under non-commitment,” *Journal of Economic Theory*, 159, 854–890.
- STRULOVICI, B. (2017): “Contract negotiation and the Coase conjecture: a strategic foundation for renegotiation-proof contracts,” *Econometrica*, 85, 585–616.
- VARTIAINEN, H. (2013): “Auction design without commitment,” *Journal of the European Economic Association*, 11, 316–342.
- (2023): “Recursively renegotiation-proof mechanisms,” *Mimeo*.
- VOHRA, R. (1999): “Incomplete information, incentive compatibility, and the core,” *Journal of Economic Theory*, 86, 123–147.
- VON DER FEHR, N.-H. M. AND K.-U. KÜHN (1995): “Coase versus Pacman: Who eats whom in the durable-goods monopoly?” *Journal of Political Economy*, 103, 785–812.

VON NEUMANN, J. AND O. MORGENSTERN (1944): *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.

WILSON, R. (1978): “Information, efficiency, and the core of an economy,” *Econometrica*, 46, 807–816.

APPENDIX A. OMITTED PROOFS

Proof of Lemma 2. Suppose, to the contrary, that there exist $o \in \Omega$ and $o' = (w', \mu)$ with $V(w') > V(w)$ and $U(w', \theta) > U(w, \theta)$ for all $\theta \in \text{supp}(\mu)$. In this case, internal stability implies that o' is not in Ω . External stability implies that there exists an assignment measure $\gamma_{o'}$ with $\mathcal{O}(\gamma_{o'}) \subset \Omega$ that improves o' . Thus, $\int_{\mathcal{W}} V(\hat{w}) d\gamma_{o'}(\Theta, \hat{w}) \geq V(w')$. As $\gamma_{o'}$ improves o' , it also improves o . Together with $\int_{\mathcal{W}} V(\hat{w}) d\gamma_{o'}(\Theta, \hat{w}) \geq V(w') > V(w)$, this violates internal stability. Thus, it follows that $o \notin \Omega$.

Proof of Lemma 3. We first prove (1). Let $o = ((p, q), \mu^\theta) \in \Omega$ be separating. Suppose $q \neq q^*(\theta)$. Choose $\varepsilon > 0$ such that $\varepsilon < [u(q^*(\theta), \theta) - c(q^*(\theta))] - [u(q, \theta) - c(q)]$. Then the contract $(p + u(q^*(\theta), \theta) - u(q, \theta) - \varepsilon, q^*(\theta))$ strictly benefits both the principal and type θ . By Lemma 2, $o \notin \Omega$, a contradiction. Hence $q = q^*(\theta)$.

Now let $o = ((p, q), \mu) \in \Omega$ be pooling and for notational convenience set $K = \text{supp}(\mu)$. Write $\pi_0 := p - c(q)$ for the principal's payoff from the pooling contract, and define $g(\theta) := [u(q^*(\theta), \theta) - c(q^*(\theta))] - [u(q, \theta) - c(q)]$ as the surplus gain from replacing q by the efficient quality for type θ . Since $q^*(\theta)$ uniquely maximizes $u(\hat{q}, \theta) - c(\hat{q})$, we have $g(\theta) \geq 0$, with equality if and only if $q = q^*(\theta)$. An efficient contract for type θ that gives the principal a fixed payoff x can be written as $(c(q^*(\theta)) + x, q^*(\theta))$, and type θ 's utility gain from this contract relative to the pooling contract is $g(\theta) - (x - \pi_0)$.

We first show that $q = q^*(s)$ for some $s \in K$. Suppose not. Then $g(\theta) > 0$ for all $\theta \in K$. Since K is compact, there is $d := \min_{\theta \in K} g(\theta) > 0$. Assign each $\theta \in K$ the efficient contract $(c(q^*(\theta)) + \pi_0 + d/2, q^*(\theta))$. Each type gains at least $d/2$, and the principal gains $d/2$ in every state. All new contracts give the principal the same payoff. Each type optimally chooses the contract with his efficient quality by the optimality of $q^*(\theta)$. As q^* is strictly increasing, the induced outcomes are efficient and separating. Efficient and separating outcomes are unimprovable and, by Lemma 1, they belong to every stable set. Thus there exists an assignment measure that strictly improves o to outcomes that are in every stable set Ω , contradicting internal stability. Hence there exists $s \in K$ such that $q = q^*(s)$. Since q^* is strictly increasing, this type is unique.

It remains to show that K is connected. Suppose not. Then there is an open gap (a, b) such that $K \cap (a, b) = \emptyset$ and both $K \cap [\underline{\theta}, a]$ and $K \cap [b, \bar{\theta}]$ are nonempty. Let K_1 be the one of these two sets that contains s , and let K_2 be the other. Both have positive μ -measure. Since K_2 is compact and does not contain s , $\eta := \min_{\theta \in K_2} g(\theta) > 0$. Moreover, compactness and strict optimality of $q^*(\theta)$ imply that there is $\rho > 0$ such that, for all $\theta \in K_2$ and $\hat{\theta} \in K_1$, $[u(q^*(\theta), \theta) - c(q^*(\theta))]] - [u(q^*(\hat{\theta}), \theta) - c(q^*(\hat{\theta}))]] \geq \rho$. Choose

$$0 < \varepsilon < \min\{\eta, \rho\} \quad \text{and} \quad 0 < \delta < \min\{\rho - \varepsilon, \varepsilon \frac{\mu(K_2)}{\mu(K_1)}\}.$$

Assign types in K_1 the contracts $(c(q^*(\theta)) + \pi_0 - \delta, q^*(\theta))$, and types in K_2 the contracts $(c(q^*(\theta)) + \pi_0 + \varepsilon, q^*(\theta))$. Types in K_1 gain $g(\theta) + \delta > 0$. Types in K_2 gain $g(\theta) - \varepsilon \geq \eta - \varepsilon > 0$. The principal's expected gain is $-\delta\mu(K_1) + \varepsilon\mu(K_2) > 0$.

It remains only to check incentive compatibility. If two types lie on the same side, their assigned contracts give the principal the same payoff, so incentive compatibility follows from

the optimality of $q^*(\theta)$. If $\theta \in K_1$ and $\hat{\theta} \in K_2$, then type θ 's own contract has both his efficient quality and a lower principal payoff by $\varepsilon + \delta$, so he does not prefer the contract assigned to $\hat{\theta}$. Finally, if $\theta \in K_2$ and $\hat{\theta} \in K_1$, then the gain from choosing his own efficient quality rather than $q^*(\hat{\theta})$ is at least ρ , while the extra principal payoff is only $\varepsilon + \delta < \rho$. Thus type θ also prefers his own contract.

Hence the constructed assignment is incentive compatible, strictly improves o , and leads only to efficient separating outcomes. These outcomes belong to Ω by Lemma 1, contradicting internal stability. Therefore $K = \text{supp}(\mu)$ is connected.

Proof of Lemma 4. Write $\varphi(z) = (p(z), q(z))$. By condition (ST), there is a $\sigma_{\mathcal{Z}}$ -null set $N \subset \mathcal{Z}$ such that, for every index $z \in \mathcal{Z} \setminus N$, the terminal outcome $(\varphi(z), \mu_0(\cdot \mid z))$ lies in a stable set and therefore satisfies the restrictions in Lemma 3, in particular that some type receives his efficient quality. Indices in N carry no probability, so they do not affect the principal's payoff, and dropping them only weakens condition (IC). We therefore ignore them. Hence, for every $z \in \mathcal{Z} \setminus N$, the quality $q(z)$ lies in the compact set $q^*(\Theta) = \{q^*(\theta) : \theta \in \Theta\}$.

Define the indirect utility generated by the terminal contracts in the assignment: $R(\theta) = \sup_{z \in \mathcal{Z} \setminus N} \{u(q(z), \theta) - p(z)\}$. Since $\sigma(\Theta \times N) = 0$, condition (IC) implies that, for σ -almost every (θ, z) , the index z lies in $\mathcal{Z} \setminus N$ and attains this supremum. We call an index $z \in \mathcal{Z} \setminus N$ that attains $R(\theta)$ an optimal index for type θ ; thus, for μ_0 -almost every θ , the indices associated with type θ are optimal.

Because $q^*(\Theta)$ is compact and u_θ is continuous, the functions $\theta \mapsto u(q(z), \theta) - p(z)$, $z \in \mathcal{Z} \setminus N$, are Lipschitz with a common constant. Because the supremum is attained at μ_0 -almost every type, R is finite at some type. By standard envelope arguments (Milgrom and Segal, 2002, Theorems 1 and 2), R is therefore Lipschitz, hence differentiable μ_0 -almost everywhere,

and at every type θ in the interior of Θ at which R is differentiable, $R'(\theta) = u_\theta(q(z), \theta)$ for every optimal index z .

We next show that, at every such type θ , all optimal indices carry the same contract. If z and z' are both optimal indices for type θ , then $u_\theta(q(z), \theta) = R'(\theta) = u_\theta(q(z'), \theta)$. By strict single crossing, $u_{q\theta} > 0$, the map $q \mapsto u_\theta(q, \theta)$ is strictly increasing. Therefore $q(z) = q(z')$. Since z and z' give type θ the same utility and have the same quality, they must also have the same price: $p(z) = p(z')$. Thus, for μ_0 -almost every type, all optimal indices carry the same contract. Therefore, a type can be associated with several indices only if those indices carry the same contract, or if the type belongs to a null set of tie types. Such multiplicity has no effect on payoffs, incentive constraints, or individual rationality.

We may therefore describe the assignment of contracts to types, up to a μ_0 -null set of types, by a function $\theta \mapsto (P(\theta), Q(\theta))$. For almost every θ , $R(\theta) = u(Q(\theta), \theta) - P(\theta)$. The function (P, Q) is incentive compatible because each assigned contract is optimal among the terminal contracts in the assignment. Merging indices that carry the same contract averages the beliefs attached to them, so (P, Q) need not itself describe an admissible terminal assignment; it is used only to bound the principal's payoff.

Standard arguments using incentive compatibility show that Q is nondecreasing almost everywhere. It remains to show that positive-mass pooling cells are intervals. Fix a contract (p, q) that is assigned to a positive-measure set of types, and let

$$I(p, q) = \{\theta \in \Theta : (P(\theta), Q(\theta)) = (p, q)\}.$$

First consider the set of types assigned quality q : $I_q = \{\theta \in \Theta : Q(\theta) = q\}$. If $\theta < \theta' < \theta''$ are outside the discarded null set and $\theta, \theta'' \in I_q$, monotonicity gives $q = Q(\theta) \leq Q(\theta') \leq Q(\theta'') = q$. Hence I_q is an interval up to endpoints and the discarded null set.

Now suppose two types in I_q were assigned the same quality but different prices. Then every type would strictly prefer the cheaper contract, contradicting incentive compatibility. Therefore, on any positive-mass set of types assigned the same quality, the price must be the same, except possibly on a null set.

Consequently, each positive-mass set of types assigned a given contract coincides, up to endpoints and μ_0 -null sets, with a flat interval of the nondecreasing quality function Q . Since μ_0 is atomless, endpoint assignments do not affect expected payoffs or incentive constraints. Thus every positive-mass pooling interval can be written as $[a, b)$ with $a < b$.

Now consider a contract (p, q) that is held by a μ_0 -null set of types. Then $I(p, q)$ contains at most one type outside the discarded null set: if it contained two such types $\theta < \theta'$, monotonicity of Q would give quality q to every type between them that lies outside the discarded null set, the equal-price argument above would give all of these types the price p , and $I(p, q)$ would have positive μ_0 -mass because μ_0 has full support. Since, by construction of (P, Q) , $\varphi(z) = (P(\theta), Q(\theta))$ for σ -almost every (θ, z) , the disintegration of σ implies that, for $\sigma_{\mathcal{Z}}$ -almost every index z whose contract is held by a μ_0 -null set of types, the posterior $\mu_0(\cdot \mid z)$ is concentrated on this single type. The outcome at such an index is therefore separating, and part (1) of Lemma 3 yields $Q(\theta) = q^*(\theta)$ for μ_0 -almost every type outside the positive-mass pooling intervals.

Finally, fix a positive-mass pooling interval $I(p, q) = [a, b)$ and set $Z_{p,q} := \{z \in \mathcal{Z} \setminus N : \varphi(z) = (p, q)\}$. Since $\mu_0(I(p, q)) > 0$, we have $\sigma_{\mathcal{Z}}(Z_{p,q}) > 0$. Since $\varphi(z) = (P(\theta), Q(\theta))$ for

σ -almost every (θ, z) ,

$$0 = \sigma((\Theta \setminus [a, b]) \times Z_{p,q}) = \int_{Z_{p,q}} \mu_0(\Theta \setminus [a, b] \mid z) d\sigma_Z(z).$$

Hence, for σ_Z -almost every $z \in Z_{p,q}$, $\text{supp}(\mu_0(\cdot \mid z)) \subset [a, b]$. For any such z , Lemma 3 yields $s \in \text{supp}(\mu_0(\cdot \mid z))$ such that $q = q^*(s)$. Thus $s \in [a, b]$, proving the final claim.

Proof of Lemma 5. Denote by $S(q, \theta) = u(q, \theta) - c(q)$ the joint surplus and by $F(\theta) = \mu_0([\underline{\theta}, \theta])$ the distribution function. Let $\{[a_i, b_i]\}_{i \in \mathcal{I}}$ be the family of positive-mass pooling intervals. Since these intervals are disjoint and each has positive μ_0 -mass, the set $\mathcal{I}$ is countable.

Define $\widehat{Q}$ by replacing, on every positive-mass pooling interval $[a_i, b_i]$, the original quality $q^*(s_i)$ by $q^*(a_i)$, and by leaving all other qualities unchanged. The rule $\widehat{Q}$ is nondecreasing. Indeed, outside positive-mass pooling intervals, Lemma 4 implies that μ_0 -almost every type receives its efficient quality. On each pooling interval $[a_i, b_i]$, the modified quality is $q^*(a_i)$. Since q^* is strictly increasing and the pooling intervals are ordered by type, replacing the quality on each interval by the quality efficient for its lower endpoint preserves monotonicity.

Let R be the original rent, that is, the Lipschitz indirect utility defined in the proof of Lemma 4. It satisfies $R(\theta) = u(Q(\theta), \theta) - P(\theta)$ for μ_0 -almost every θ and is defined at every type, in particular at $\underline{\theta}$. Define $\widehat{R}(\theta) = R(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u_{\theta}(\widehat{Q}(t), t) dt$. Set $\widehat{P}(\theta) = u(\widehat{Q}(\theta), \theta) - \widehat{R}(\theta)$. By the standard one-dimensional incentive-compatibility characterization, $(\widehat{P}, \widehat{Q})$ is incentive compatible.

Individual rationality is preserved. Individual rationality of the original assignment holds for μ_0 -almost every type and R is continuous, so $R(\underline{\theta}) \geq u(0, \underline{\theta}) = 0$, and since $\widehat{Q}(t) \geq 0$,

single crossing gives $u_\theta(\widehat{Q}(t), t) \geq u_\theta(0, t)$. Therefore

$$\widehat{R}(\theta) - u(0, \theta) = R(\underline{\theta}) - u(0, \underline{\theta}) + \int_{\underline{\theta}}^{\theta} [u_\theta(\widehat{Q}(t), t) - u_\theta(0, t)] dt \geq 0.$$

For any incentive-compatible allocation with absolutely continuous rent, the principal's expected payoff can be written as

$$\Pi(P, Q) = \int_{\underline{\theta}}^{\bar{\theta}} [f(\theta)S(Q(\theta), \theta) - (1 - F(\theta))u_\theta(Q(\theta), \theta)] d\theta - R(\underline{\theta}).$$

The original and modified allocations have the same lowest-type rent $R(\underline{\theta})$. Hence the payoff difference is

$$\int_{\underline{\theta}}^{\bar{\theta}} [f(\theta)(S(\widehat{Q}(\theta), \theta) - S(Q(\theta), \theta)) - (1 - F(\theta))(u_\theta(\widehat{Q}(\theta), \theta) - u_\theta(Q(\theta), \theta))] d\theta.$$

This integrand vanishes outside the positive-mass pooling intervals.

Consider one such interval $[a_i, b_i]$. Its original quality is $q^*(s_i)$, and its modified quality is $q^*(a_i)$. Define $M_i(\theta) = S(q^*(a_i), \theta) - S(q^*(s_i), \theta)$. Then $M'_i(\theta) = u_\theta(q^*(a_i), \theta) - u_\theta(q^*(s_i), \theta)$.

The contribution of interval $[a_i, b_i]$ to the payoff difference is therefore

$$\int_{a_i}^{b_i} [f(\theta)M_i(\theta) - (1 - F(\theta))M'_i(\theta)] d\theta = (1 - F(a_i))M_i(a_i) - (1 - F(b_i))M_i(b_i).$$

If $s_i = a_i$, then $M_i \equiv 0$, so the payoff difference is zero. If $s_i > a_i$, then $M_i(a_i) > 0$, because $q^*(a_i)$ uniquely maximizes $S(\cdot, a_i)$. Moreover, $M_i(b_i) \leq 0$. To see this, note that $q^*(s_i) > q^*(a_i)$, and increasing differences imply that $S(q^*(s_i), \theta) - S(q^*(a_i), \theta)$ is increasing in θ . It is strictly positive at $\theta = s_i$, hence it is nonnegative at $b_i \geq s_i$. Therefore $M_i(b_i) \leq 0$.

Thus the payoff difference is strictly positive whenever $s_i > a_i$.

Because the positive-mass pooling intervals are disjoint and countable, and the payoff integrand is bounded on the compact type space, the total payoff change is the sum of the interval payoff contributions. Each contribution is nonnegative, and at least one is strictly positive if some positive-mass pool is not already bottom-anchored. Hence the modified allocation gives the principal a weakly higher payoff, and a strictly higher payoff whenever at least one positive-mass pool has $s_i > a_i$.

Finally, the modified allocation satisfies the outcome restrictions in Lemma 3: separated types still receive their efficient qualities, and every positive-mass pooling interval $[a_i, b_i)$ now receives quality $q^*(a_i)$, which is efficient for a type in the posterior support of that interval.

Proof of Lemma 6. Denote by $S(q, \theta) = u(q, \theta) - c(q)$ the joint surplus and by $F(\theta) = \mu_0([\underline{\theta}, \theta])$ the distribution function.

Step 1: Upper bound. We first derive an upper bound on the payoff of any admissible terminal assignment. By Lemmas 3 to 5, the payoff of any admissible terminal assignment is bounded above by the payoff of a relaxed allocation, that is, a function (P, Q) with the following properties: it is incentive compatible and individually rational, every positive-mass pooling set is an interval, μ_0 -almost every type outside those intervals receives its efficient quality (Lemma 4), and every positive-mass pooling interval $[a, b)$ receives quality $q^*(a)$.

Fix such a relaxed allocation and write it as (P, Q) . Define the agent's rent by $R(\theta) = u(Q(\theta), \theta) - P(\theta)$. Incentive compatibility yields the envelope formula $R'(\theta) = u_\theta(Q(\theta), \theta)$ for a.e. θ . Since the allocation is individually rational, we have $R(\underline{\theta}) \geq 0$. Hence $R(\theta) \geq$

$\int_{\underline{\theta}}^{\theta} u_{\theta}(Q(t), t) dt$. The principal's expected payoff Π therefore satisfies

$$\begin{aligned} \Pi &= \int_{\underline{\theta}}^{\bar{\theta}} [P(\theta) - c(Q(\theta))] f(\theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} [S(Q(\theta), \theta) - R(\theta)] f(\theta) d\theta \\ &\leq \int_{\underline{\theta}}^{\bar{\theta}} S(Q(\theta), \theta) f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} \left[\int_{\underline{\theta}}^{\theta} u_{\theta}(Q(t), t) dt \right] f(\theta) d\theta. \end{aligned}$$

By Fubini's theorem, $\int_{\underline{\theta}}^{\bar{\theta}} \left[\int_{\underline{\theta}}^{\theta} u_{\theta}(Q(t), t) dt \right] f(\theta) d\theta = \int_{\underline{\theta}}^{\bar{\theta}} (1 - F(t)) u_{\theta}(Q(t), t) dt$. Thus

$$\Pi \leq \int_{\underline{\theta}}^{\bar{\theta}} [f(\theta) S(Q(\theta), \theta) - (1 - F(\theta)) u_{\theta}(Q(\theta), \theta)] d\theta.$$

We now compactify the cutoff problem. Let $\mathcal{A} = \{A \subset [\underline{\theta}, \bar{\theta}] : A \text{ is compact and } \underline{\theta}, \bar{\theta} \in A\}$. So $\mathcal{A}$ is the set of all compact subsets of $[\underline{\theta}, \bar{\theta}]$ that include the endpoints. We will use this compact set to prove that the maximum of our cutoff problem exists.

For $A \in \mathcal{A}$, define the anchor $a_A(\theta) = \max\{a \in A : a \leq \theta\}$. This is a way to encode the cutoffs for our maximization problem. For example, if $A = \{\underline{\theta}, \theta_1, \theta_2, \bar{\theta}\}$, it segments the type space into three intervals, and the anchor of each type is the lowest type of its interval. In particular, if A has a gap (a, b) , all types in this gap are treated as pooled at the anchor a . Define $H(a, \theta) = f(\theta) S(q^*(a), \theta) - (1 - F(\theta)) u_{\theta}(q^*(a), \theta)$, for $a \leq \theta$ and

$$J(A) = \int_{\underline{\theta}}^{\bar{\theta}} H(a_A(\theta), \theta) d\theta.$$

We can use this notation to bound the profit from admissible terminal assignments. Let the positive-mass pooling intervals of the relaxed allocation be $[a_{\alpha}, b_{\alpha})$ with α from some set, and set $A = [\underline{\theta}, \bar{\theta}] \setminus \bigcup_{\alpha} (a_{\alpha}, b_{\alpha})$. Then $A \in \mathcal{A}$. For types inside a pooling interval, $a_A(\theta) = a_{\alpha}$, while separated types satisfy $a_A(\theta) = \theta$, up to null sets. By construction of the relaxed

allocation, $Q(\theta) = q^*(a_A(\theta))$ for μ_0 -a.e. θ . Therefore $\Pi \leq J(A) \leq \sup_{B \in \mathcal{A}} J(B)$. Taking the supremum over all admissible terminal assignments gives $V^* \leq \sup_{A \in \mathcal{A}} J(A)$.

Step 2: Existence of a maximizer. Next we show that the compactified problem has a maximizer. The hyperspace of all compact subsets of $[\underline{\theta}, \bar{\theta}]$ is compact under the Hausdorff metric.¹³ The subspace $\mathcal{A}$ is closed within this hyperspace.¹⁴ As f is continuous and bounded and therefore F continuous, the function H is continuous and bounded on the compact triangle $\{(a, \theta) : \underline{\theta} \leq a \leq \theta \leq \bar{\theta}\}$.

We claim that J is continuous. Let $A_k \rightarrow A$ in the Hausdorff metric. We show that at every continuity point of the monotone function a_A , $a_{A_k}(\theta) \rightarrow a_A(\theta)$. Indeed, compactness implies $\limsup_{k \rightarrow \infty} a_{A_k}(\theta) \leq a_A(\theta)$: every convergent subsequence of $a_{A_k}(\theta)$ has a limit in A that is no larger than θ . For the reverse inequality, first suppose $a_A(\theta) < \theta$. Hausdorff convergence supplies $x_k \in A_k$ with $x_k \rightarrow a_A(\theta)$; eventually $x_k \leq \theta$, and hence $a_{A_k}(\theta) \geq x_k$. If $a_A(\theta) = \theta$ and $\theta > \underline{\theta}$, continuity of a_A at θ implies that, for every $\varepsilon > 0$, there exists $t < \theta$ such that $a_A(t) > \theta - \varepsilon$. Approximating $a_A(t)$ by points of A_k gives $a_{A_k}(\theta) > \theta - 2\varepsilon$ for all sufficiently large k . At $\theta = \underline{\theta}$, convergence is immediate because every A_k contains $\underline{\theta}$. This proves the reverse inequality. Since a_A is monotone, it has at most countably many discontinuities. Hence $a_{A_k}(\theta) \rightarrow a_A(\theta)$ for a.e. θ . Since H is bounded on the relevant domain, dominated convergence then implies $J(A_k) \rightarrow J(A)$. Thus J is continuous on the compact set $\mathcal{A}$, so by the Weierstrass Extreme Value Theorem (Aliprantis and Border, 2006, Corollary 2.35) there exists $A^* \in \mathcal{A}$ such that $J(A^*) = \max_{A \in \mathcal{A}} J(A)$.

Step 3: Every maximizer is a cutoff sequence. We now show that any maximizer A^* is generated by an ordinary cutoff sequence. To do so, we will show that the maximizer has

¹³This is a direct application of Aliprantis and Border (2006), Theorem 3.88.

¹⁴This is a direct application of Aliprantis and Border (2006), Theorem 2.31.

no accumulation points below $\bar{\theta}$. The main idea is that if such an accumulation point exists we can improve the profit of the principal by pooling some of the resulting contracts into one contract. For this, we will need that the virtual surplus function H is decreasing in the anchor a whenever the anchor is sufficiently close to θ . Fix $\eta > 0$. We claim that there exists $\delta_\eta > 0$ such that, whenever $a \leq r \leq \theta \leq a + \delta_\eta$, $a \leq \bar{\theta} - \eta$, we have $H(a, \theta) \geq H(r, \theta)$, with strict inequality if $r > a$.

To prove the claim, differentiate H with respect to its first argument:

$$H_a(x, \theta) = q_\theta^*(x) [f(\theta)S_q(q^*(x), \theta) - (1 - F(\theta))u_{q\theta}(q^*(x), \theta)].$$

At $\theta = x$, the first-order condition for $q^*(x)$ gives $S_q(q^*(x), x) = 0$. Hence $H_a(x, x) = -q_\theta^*(x)(1 - F(x))u_{q\theta}(q^*(x), x) < 0$ for every $x < \bar{\theta}$. Because H_a is continuous and the function $x \mapsto H_a(x, x)$ is strictly negative on the compact interval $[\underline{\theta}, \bar{\theta} - \eta/2]$, its maximum on that interval is strictly negative. Uniform continuity therefore permits us to choose $\delta_\eta \in (0, \eta/2]$ such that $H_a(x, \theta) < 0$ whenever $x \leq \theta$, $\theta - x \leq \delta_\eta$, and $x \leq \bar{\theta} - \eta/2$. If $a \leq r \leq \theta \leq a + \delta_\eta$ and $a \leq \bar{\theta} - \eta$, then every $x \in [a, r]$ satisfies these three conditions. Integrating $H_a(x, \theta)$ over $x \in [a, r]$ gives the claim, with strict inequality whenever $r > a$.

Suppose, toward a contradiction, that A^* has an accumulation point $L < \bar{\theta}$. Choose $\eta > 0$ such that $L < \bar{\theta} - \eta$. Then we can find $a < s < b$ in A^* , all sufficiently close to L , with $b - a < \delta_\eta$. Remove all anchors strictly between a and b : $\tilde{A} = A^* \setminus (a, b)$. Then $\tilde{A} \in \mathcal{A}$. Outside (a, b) , the anchor map is unchanged. On (a, b) , the new anchor is a . Therefore

$$J(\tilde{A}) - J(A^*) = \int_a^b [H(a, \theta) - H(a_{A^*}(\theta), \theta)] d\theta.$$

For every $\theta \in (a, b)$, $a \leq a_{A^*}(\theta) \leq \theta \leq b < a + \delta_\eta$, so the integrand is nonnegative. For every $\theta \in (s, b)$, we have $a_{A^*}(\theta) \geq s > a$, so the integrand is strictly positive. Hence $J(\tilde{A}) > J(A^*)$, contradicting optimality. Thus A^* has no accumulation point below $\bar{\theta}$.

Next we show that the maximizer has an accumulation point at the top. The main idea is to show that if there is a final cutoff, the type space can profitably be split further. Suppose, for a contradiction, that A^* has a last anchor below $\bar{\theta}$. Then there is some $a < \bar{\theta}$ such that $A^* \cap [a, \bar{\theta}] = \{a, \bar{\theta}\}$. Choose $b \in (a, \bar{\theta})$ and set $\tilde{A} = A^* \cup \{b\}$. Only types $\theta \in [b, \bar{\theta}]$ change anchors, so $J(\tilde{A}) - J(A^*) = \int_b^{\bar{\theta}} [H(b, \theta) - H(a, \theta)] d\theta$. Let $M(\theta) = S(q^*(b), \theta) - S(q^*(a), \theta)$. Then $M'(\theta) = u_\theta(q^*(b), \theta) - u_\theta(q^*(a), \theta)$, and therefore

$$H(b, \theta) - H(a, \theta) = f(\theta)M(\theta) - (1 - F(\theta))M'(\theta) = -\frac{d}{d\theta} [(1 - F(\theta))M(\theta)].$$

Since $F(\bar{\theta}) = 1$, integration gives

$$\int_b^{\bar{\theta}} [H(b, \theta) - H(a, \theta)] d\theta = (1 - F(b))M(b) = (1 - F(b)) [S(q^*(b), b) - S(q^*(a), b)].$$

The last expression is strictly positive: $1 - F(b) > 0$ because $b < \bar{\theta}$ and the prior has full support, and the bracket is strictly positive because $q^*(b)$ uniquely maximizes $S(\cdot, b)$. This contradicts optimality of A^* . Hence A^* has no last anchor below $\bar{\theta}$.

The preceding two conclusions imply that $A^* \cap [\underline{\theta}, \bar{\theta})$ is finite on every interval $[\underline{\theta}, \bar{\theta} - \eta]$, is infinite overall, and has $\bar{\theta}$ as its only accumulation point. Therefore it can be enumerated as $A^* \cap [\underline{\theta}, \bar{\theta}) = \{\theta_0^*, \theta_1^*, \theta_2^*, \dots\}$, where $\theta_0^* = \underline{\theta}$, $\theta_n^* < \theta_{n+1}^*$, $\theta_n^* \rightarrow \bar{\theta}$. Thus every maximizer of the compactified problem is generated by an ordinary cutoff sequence.

Step 4: Cutoff allocations and the cutoff value. We now connect the compactified objective to the cutoff problem in Proposition 1. Let $\underline{\theta} = \theta_0 < \theta_1 < \theta_2 < \dots$, $\theta_n \rightarrow \bar{\theta}$ be any

cutoff sequence. Set $q_n = q^*(\theta_n)$, and define prices by $p_0 = u(q^*(\underline{\theta}), \underline{\theta})$, and, for $n \geq 1$, $p_n = p_{n-1} + u(q^*(\theta_n), \theta_n) - u(q^*(\theta_{n-1}), \theta_n)$. Then $u(q_n, \theta_n) - p_n = u(q_{n-1}, \theta_n) - p_{n-1}$. For adjacent contracts define $D_n(\theta) = [u(q_n, \theta) - p_n] - [u(q_{n-1}, \theta) - p_{n-1}]$. We have $D_n(\theta_n) = 0$, and $D'_n(\theta) = u_\theta(q_n, \theta) - u_\theta(q_{n-1}, \theta) > 0$, because $q_n > q_{n-1}$ and $u_{q\theta} > 0$. Hence contract n is preferred to contract $n-1$ exactly above the cutoff θ_n . Iterating this single-crossing argument shows that every type $\theta \in [\theta_n, \theta_{n+1})$ prefers contract n to every other contract.

Define $Q(\theta) = q_n$, $P(\theta) = p_n$ for $\theta \in [\theta_n, \theta_{n+1})$. Let $R(\theta) = u(Q(\theta), \theta) - P(\theta)$. By construction, $R(\underline{\theta}) = 0$, R is continuous at each cutoff, and $R'(\theta) = u_\theta(Q(\theta), \theta)$ for a.e. θ . Since $u_\theta \geq 0$, we have $R(\theta) \geq 0$, so the allocation is individually rational. The limit $R(\bar{\theta}) = \lim_{\theta \uparrow \bar{\theta}} R(\theta)$ exists and is finite. Assign the highest type the efficient separating contract $(p_{\bar{\theta}}, q^*(\bar{\theta}))$, $p_{\bar{\theta}} = u(q^*(\bar{\theta}), \bar{\theta}) - R(\bar{\theta})$. This extends the incentive-compatible allocation to $\bar{\theta}$. Indeed, $p_n \rightarrow p_{\bar{\theta}}$, and the finite-contract incentive inequalities pass to the limit: type $\bar{\theta}$ weakly prefers its contract to every contract n , and every type $\theta \in [\theta_n, \theta_{n+1})$ weakly prefers contract n to every contract $m > n$ and hence, by continuity of u , also to the limit contract $(p_{\bar{\theta}}, q^*(\bar{\theta}))$.

For this cutoff sequence, let $B = \{\theta_n : n \geq 0\} \cup \{\bar{\theta}\}$. Using the envelope formula and $R(\underline{\theta}) = 0$, the principal's payoff from the cutoff allocation is

$$\begin{aligned} & \int_{\underline{\theta}}^{\bar{\theta}} [f(\theta)S(Q(\theta), \theta) - (1 - F(\theta))u_\theta(Q(\theta), \theta)] d\theta \\ &= \int_{\underline{\theta}}^{\bar{\theta}} H(a_B(\theta), \theta) d\theta = J(B). \end{aligned}$$

Equivalently, this is exactly the cutoff objective displayed in Proposition 1, since the recursive prices above are the prices used there.

Since A^* maximizes J over $\mathcal{A}$ and is generated by an ordinary cutoff sequence, the ordinary cutoff problem admits a solution, and its value is $J(A^*)$. Together with the upper bound already proved, this gives $V^* \leq J(A^*) = \text{value of the cutoff problem}$.

Step 5: The cutoff assignment is admissible. It remains to show that every solution of the cutoff problem defines an admissible terminal assignment that attains this value. Let $\{\theta_n^*\}_{n \geq 0}$ be any maximizing cutoff sequence. Define $q_n^* = q^*(\theta_n^*)$, and let p_n^* be the recursive prices from Proposition 1. Consider the terminal assignment with index set $\mathcal{Z}^* = \mathbb{N}_0 \cup \{z_{\bar{\theta}}\}$, contracts $\varphi^*(n) = (p_n^*, q_n^*)$ for $n \in \mathbb{N}_0$ and $\varphi^*(z_{\bar{\theta}}) = (p_{\bar{\theta}}, q^*(\bar{\theta}))$, and σ^* the joint distribution of the type θ and the index $n(\theta)$, where $n(\theta) = n$ for $\theta \in [\theta_n^*, \theta_{n+1}^*)$ and $n(\bar{\theta}) = z_{\bar{\theta}}$. The belief at index n is $\mu_n^* = \mu_0(\cdot \mid [\theta_n^*, \theta_{n+1}^*))$, and the index $z_{\bar{\theta}}$ carries zero probability. The argument above shows that this terminal assignment satisfies conditions (IC) and (IR), and that its payoff is the cutoff value. It remains to verify condition (ST).

For each n , define the pooling outcome $o_n^* = ((p_n^*, q_n^*), \mu_0(\cdot \mid [\theta_n^*, \theta_{n+1}^*)))$. Now define

$$\Omega^* = \{((p, q^*(\theta)), \mu^\theta) : \theta \in \Theta, p \in \mathbb{R}\} \cup \{o_n^* : n \in \mathbb{N}_0\}.$$

We show that Ω^* is a stable set.

External stability. Let $o = ((p, q), \mu)$ be any outcome. We show that there is an improvement to efficient and separating outcomes. For each $\theta \in \text{supp}(\mu)$, define $\hat{w}(\theta) = (p + c(q^*(\theta)) - c(q), q^*(\theta))$. The principal's payoff from $\hat{w}(\theta)$ is $p + c(q^*(\theta)) - c(q) - c(q^*(\theta)) = p - c(q)$, so the principal is indifferent type by type. The agent of type θ obtains $u(q^*(\theta), \theta) - p - c(q^*(\theta)) + c(q)$, which is weakly larger than $u(q, \theta) - p$, because $q^*(\theta)$ maximizes $u(\cdot, \theta) - c(\cdot)$.

Moreover, the schedule is incentive compatible. For every $\theta, \hat{\theta} \in \text{supp}(\mu)$,

$$\begin{aligned} U(\hat{w}(\theta), \theta) &= u(q^*(\theta), \theta) - c(q^*(\theta)) - (p - c(q)) \\ &\geq u(q^*(\hat{\theta}), \theta) - c(q^*(\hat{\theta})) - (p - c(q)) = U(\hat{w}(\hat{\theta}), \theta), \end{aligned}$$

because $q^*(\theta)$ maximizes $u(\cdot, \theta) - c(\cdot)$. The induced assignment therefore weakly improves o , and all induced outcomes are efficient and separating, hence belong to Ω^* . This proves external stability.

Internal stability. Efficient separating outcomes cannot be strictly improved within Ω^* , because the type is known and total surplus is already maximized.

It remains to consider a pooling outcome o_n^* . Suppose, toward a contradiction, that there is an assignment measure from o_n^* into Ω^* that strictly improves o_n^* . Pooling outcomes from other cutoff intervals can appear only on a set of contracts of probability zero, whatever version of the conditional probability is used: the type marginal of the assignment measure is $\mu_0(\cdot \mid [\theta_n^*, \theta_{n+1}^*))$, so the belief at almost every contract is concentrated on $[\theta_n^*, \theta_{n+1}^*)$, whereas the posterior of every other pooling outcome o_m^* , $m \neq n$, assigns this set probability zero. Moreover, o_n^* cannot appear with strictly positive probability. Indeed, strict improvement requires every type in the source support to have some available contract that gives strictly higher utility. Incentive compatibility then prevents any such type from belonging to the posterior support of the inferior original pooling contract. Hence the original pool cannot appear with positive probability.

Hence almost every contract in the assignment induces an efficient separating outcome for a type in $[\theta_n^*, \theta_{n+1}^*]$. Let W be the support of the marginal distribution of the assignment measure on $\mathcal{W}$. Since q^* is strictly increasing, the quality of an efficient separating contract

reveals the type: the belief at such a contract $(\hat{p}, \hat{q})$ is concentrated on the unique type θ with $q^*(\theta) = \hat{q}$. As the belief at a contract is the distribution of the types that hold it, with probability one under the assignment measure a type θ holds a contract $(\hat{p}, q^*(\theta))$ whose belief is concentrated on θ . Define the agent's rent generated by the contracts in W , $R_W(\theta) = \sup_{(\hat{p}, \hat{q}) \in W} \{u(\hat{q}, \theta) - \hat{p}\}$, for $\theta \in [\theta_n^*, \theta_{n+1}^*]$. For almost every pair $(\theta, (\hat{p}, \hat{q}))$, type θ lies in the support of the belief at $(\hat{p}, \hat{q})$, so property (iii) of Definition 2 implies that no contract in W gives type θ a strictly higher utility. The assigned contract therefore attains the supremum: $R_W(\theta) = u(q^*(\theta), \theta) - \hat{p}$.

Every induced outcome lies in Ω^* , so every contract in W , including contracts of probability zero, has a quality in the compact set $q^*(\Theta)$. Since u_θ is continuous, it is bounded on $q^*(\Theta) \times \Theta$. Thus the functions $\theta \mapsto u(\hat{q}, \theta) - \hat{p}$, $(\hat{p}, \hat{q}) \in W$, are Lipschitz with a common constant. Because the supremum is attained, R_W is finite at some type. By standard envelope arguments (Milgrom and Segal, 2002, Theorems 1 and 2), R_W is therefore Lipschitz on $[\theta_n^*, \theta_{n+1}^*]$, hence differentiable almost everywhere, and, at every $\theta \in (\theta_n^*, \theta_{n+1}^*)$ at which R_W is differentiable, $R'_W(\theta) = u_\theta(\hat{q}, \theta)$ for every contract $(\hat{p}, \hat{q}) \in W$ that attains $R_W(\theta)$.

Now take a type $\theta \in (\theta_n^*, \theta_{n+1}^*)$ at which R_W is differentiable and that is assigned a contract $(\hat{p}, q^*(\theta)) \in W$ attaining $R_W(\theta)$. Almost every type has both properties: since f is bounded away from zero, Lebesgue measure and μ_n^* have the same null sets on $[\theta_n^*, \theta_{n+1}^*]$. The envelope formula gives $R'_W(\theta) = u_\theta(q^*(\theta), \theta)$. By (3), the efficient surplus has the same derivative: $\frac{d}{d\theta} S(q^*(\theta), \theta) = (u_q(q^*(\theta), \theta) - c_q(q^*(\theta)))q_\theta^*(\theta) + u_\theta(q^*(\theta), \theta) = u_\theta(q^*(\theta), \theta)$. Hence $\theta \mapsto S(q^*(\theta), \theta) - R_W(\theta)$ is Lipschitz and has derivative zero almost everywhere. A Lipschitz function with derivative zero almost everywhere is constant. So there is a constant π such that $S(q^*(\theta), \theta) - R_W(\theta) = \pi$ for every $\theta \in [\theta_n^*, \theta_{n+1}^*]$.

The principal's profit at a contract $(\hat{p}, \hat{q})$ is $\hat{p} - c(\hat{q})$. For almost every pair $(\theta, (\hat{p}, \hat{q}))$ in the assignment, it equals $\hat{p} - c(q^*(\theta)) = S(q^*(\theta), \theta) - R_W(\theta) = \pi$. The principal's payoff from the assignment measure therefore equals π . Since V is continuous, the set $\{w \in \mathcal{W} : V(w) = \pi\}$ is closed and has probability one, so it contains the support W . Hence $V(\hat{w}) = \pi$ at every contract $\hat{w} \in W$, including contracts of probability zero.

For μ_n^* -almost every θ , the agent's improvement implies $u(q^*(\theta), \theta) - c(q^*(\theta)) - \pi \geq u(q_n^*, \theta) - p_n^*$. Both sides are continuous in θ , and $\text{supp}(\mu_n^*) = [\theta_n^*, \theta_{n+1}^*]$. Hence the inequality also holds at θ_n^* . Since $q_n^* = q^*(\theta_n^*)$, it follows that $\pi \leq p_n^* - c(q_n^*)$, contradicting the principal's strict gain. Therefore no pooling outcome o_n^* admits a strict improvement into Ω^* .

Thus Ω^* is internally and externally stable. Every index carrying positive probability is some $n \in \mathbb{N}_0$, and its terminal outcome o_n^* lies in Ω^* ; hence condition (ST) holds. The cutoff assignment associated with any maximizing cutoff sequence is therefore an admissible terminal assignment and attains the cutoff value. Since we already proved that no admissible terminal assignment can yield more than the cutoff value, the cutoff value is equal to V^* .

Proof of Proposition 2. Let Ω be any stable set, and fix an arbitrary no-trade outcome $o = ((p, 0), \mu)$. Since $c < \underline{\theta}$, choose $\varepsilon \in (0, \underline{\theta} - c)$ and consider the single trade contract $w' = (p + c + \varepsilon, 1)$. For every $\theta \in \text{supp}(\mu)$, $U(w', \theta) - U((p, 0), \theta) = (\theta - (p + c + \varepsilon)) - (-p) = \theta - c - \varepsilon > 0$, and the seller's payoff satisfies $V(w') - V((p, 0)) = ((p + c + \varepsilon) - c) - p = \varepsilon > 0$. Thus w' strictly improves o for the seller and for every buyer type in the support of μ , so $o \notin \Omega$ by Lemma 2. Since o and Ω were arbitrary, every outcome in any stable set has $q = 1$.

It follows that every admissible terminal assignment puts probability one on trade. Let $(\mathcal{Z}, \varphi, \sigma)$ be admissible. By condition (ST), for $\sigma_{\mathcal{Z}}$ -almost every index z the terminal outcome

lies in a stable set, so $\varphi(z) = (p_z, 1)$ for some price p_z . A buyer of type θ obtains utility $\theta - p_z$ from such a contract. By condition (IC), for σ -almost every (θ, z) we have $\theta - p_z \geq \theta - p_{z'}$, that is, $p_z \leq p_{z'}$, for every index z' with $\varphi(z') = (p_{z'}, 1)$. Hence σ -almost every index carries the same price p^* , the lowest trade price in the assignment. Individual rationality, $\theta - p^* = U(\varphi(z), \theta) \geq U(w_0, \theta) = 0$, applied to types arbitrarily close to $\underline{\theta}$, gives $p^* \leq \underline{\theta}$. Hence the seller's payoff is at most $\underline{\theta} - c$.

Let $\Omega^G = \{((p, 1), \mu) : p \in \mathbb{R}, \mu \in \Delta(\Theta)\}$ be the set of all trade outcomes. The set Ω^G is stable. It is externally stable because any outcome outside Ω^G is a no-trade outcome, which the construction above improves to a trade outcome in Ω^G . It is internally stable because every outcome in Ω^G involves trade: a buyer of type θ obtains $\theta - p'$ from a trade contract $(p', 1)$, so if two trade prices had positive posterior support, every type would strictly prefer the lower one, and property (iii) of Definition 2 would empty the higher price's posterior. An improvement into Ω^G therefore uses a single trade price p' , and strictly improving every buyer type from $((p, 1), \mu)$ requires $p' < p$, which lowers the seller's payoff to $p' - c < p - c$; hence no assignment measure strictly improves both sides. Finally, the terminal assignment with $\mathcal{Z} = \{1\}$, $\varphi(1) = (\underline{\theta}, 1)$, and every type associated with index 1 has the single terminal outcome $((\underline{\theta}, 1), \mu_0) \in \Omega^G$; it satisfies conditions (IC) and (IR), hence is admissible, and yields the payoff $\underline{\theta} - c$. In the mechanism interpretation, it is generated by a posted price $\underline{\theta}$.

Proof of Proposition 3. *External stability.* Let $o = ((p, q), \mu)$ be an outcome that is not in Ω^M . Split the support of μ at the cost threshold, $H = \text{supp}(\mu) \cap (c, \bar{\theta}]$ and $L = \text{supp}(\mu) \cap [\underline{\theta}, c]$, omitting either set if it carries no μ -probability, and consider the contracts $w_H = (p + (1 - q)c, 1)$ and $w_L = (p - qc, 0)$. The seller's payoff from each contract equals $p - qc = V((p, q))$, so the seller is indifferent. For $\theta > c$, the buyer's gain from w_H relative

to (p, q) is $(1 - q)(\theta - c) \geq 0$; for $\theta \leq c$, the buyer's gain from w_L is $q(c - \theta) \geq 0$. Moreover $U(w_H, \theta) - U(w_L, \theta) = \theta - c$, so types above c weakly prefer w_H , types below c prefer w_L , and type c is indifferent; assigning w_H to types in H and w_L to types in L therefore satisfies property (iii) of Definition 2. The induced outcomes $((p + (1 - q)c, 1), \mu(\cdot \mid H))$ and $((p - qc, 0), \mu(\cdot \mid L))$ have $\min \text{supp} \geq c$ and $\max \text{supp} \leq c$, respectively, so both belong to Ω^M . Hence there is an assignment measure γ_o with marginal μ and $\mathcal{O}(\gamma_o) \subset \Omega^M$ that weakly improves o , and Ω^M is externally stable.

Internal stability. Consider first an outcome $o = ((p, q), \mu)$ of the form (1) or (2). It is ex post efficient for every type in the support of its posterior and therefore not strictly improvable.

It remains to consider the rejection outcome $o^M = ((0, 0), \mu^M)$, at which every buyer type and the seller receive zero. Suppose, toward a contradiction, that an assignment measure γ_{o^M} with $\mathcal{O}(\gamma_{o^M}) \subset \Omega^M$ strictly improves o^M . Then every type in $\text{supp}(\mu^M)$ receives strictly positive utility and the seller receives strictly positive expected payoff. A no-trade contract that gives a buyer strictly positive utility must have a negative price, hence a negative seller payoff. Types above c cannot be assigned to a no-trade outcome of the form (2), since their valuation would push that posterior's support above c , nor to o^M , which would give them zero; they must therefore trade. Since all trade contracts have $q = 1$, property (iii) of Definition 2 leaves only the lowest trade price with positive posterior support. Because $\text{supp}(\mu^M) = [\underline{\theta}, p^M]$ with $p^M > c$, the support contains types arbitrarily close to c from above; improving them strictly requires a trade price no greater than c , so the seller's payoff from any trade in the assignment measure is at most zero. Together with the negative payoff from any strictly improving no-trade contract, the seller's expected payoff cannot be strictly positive, a contradiction. Hence o^M is internally stable, and Ω^M is a stable set.

Attainment and the bound. Consider the terminal assignment with $\mathcal{Z} = \{0, 1\}$, $\varphi(0) = (0, 0)$, $\varphi(1) = (p^M, 1)$, in which types $\theta < p^M$ are associated with index 0 and types $\theta \geq p^M$ with index 1. Its rejection outcome is $((0, 0), \mu^M) \in \Omega^M$, and its trade outcome is $((p^M, 1), \mu_0(\cdot \mid [p^M, \bar{\theta}]))$, which lies in Ω^M because $p^M > c$; conditions (IC) and (IR) hold by construction, so the terminal assignment is admissible, and its payoff is $\mu_0([p^M, \bar{\theta}]) (p^M - c)$. Finally, every admissible terminal assignment satisfies conditions (IC) and (IR), hence is feasible in the seller's full-commitment problem, whose optimum in this binary environment is the monopoly posted-price payoff $\max_p \mu_0([p, \bar{\theta}]) (p - c)$. The posted price p^M attains this payoff, so the terminal assignment above attains the monopoly payoff and hence the maximal payoff over admissible assignments. In the mechanism interpretation, the terminal assignment above is generated by a posted price p^M .