STABLE CONTRACTS UNDER RENEGOTIATION

VITALI GRETSCHKO AND ACHIM WAMBACH

Abstract. *We consider a principal contracting with a privately informed agent. Any contract is subject to renegotiation. Instead of modeling a specific renegotiation game, we extend the notion of von Neumann–Morgenstern stability to incorporate private information. We identify stable outcomes that will not be renegotiated and allow the principal to optimize among mechanisms that lead to such stable outcomes. The resulting solution concept provides an effective and easy-to-use tool for analyzing contract design with renegotiation. We apply the solution concept to a setting with nonlinear contracts. The principal's optimal stable outcomes are pooling contracts that satisfy a no-distortion-at-the-bottom property.*

JEL classification: C71, D82, D86

Keywords: Mechanism design, renegotiation, stability, Coase conjecture

1. Introduction

Consider the problem of a principal (she), endowed with all the bargaining power, who wishes to contract with a privately informed agent (he). As a consequence of the revelation principle, one can usually dispense with the details of the particular procedure that the principal may use to close the contract, and instead focus on direct revelation mechanisms (Myerson, 1979). That is, it

(A1) ZEW Mannheim and University of Mannheim; gretschko@zew.de
(A2) ZEW Mannheim and University of Mannheim; wambach@zew.de

We would like to thank the co-editor Jeffrey Ely, three anonymous referees, Geir Asheim, Atoshi Chowdhury, Roman Inderst, Georg Nöldeke, Alexei Parakhonyak, Ray Rees, Klaus Schmidt, Ina Taneva, Thomas Tröger, and seminar participants at the University of Bonn (2018), Cologne University (2017), the EEA conference (Lisbon 2017), the EARIE conference (Maastricht 2017), Edinburgh University (2016), the 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.

is without loss to assume that the agent truthfully reports his private information to the mechanism, and the mechanism determines the optimal contract. This approach is valid only if the principal honors the rules of the proposed mechanism and the agent trusts that this is the case. However, the agent reveals information by playing the mechanism, and the contracts resulting from optimal mechanisms are typically inefficient. Thus, after a mechanism is played, the agent and the principal both can potentially benefit from renegotiating the mechanism's outcome. If the agent anticipates such renegotiation, it may not be optimal for him to fully reveal his private information. More generally, the agent's reporting strategy may depend on how the renegotiation proceeds.

Rather than modeling a specific renegotiation procedure, we combine a mechanism design approach with a cooperative concept based on von Neumann–Morgenstern stability (von Neumann and Morgenstern, 1944), which we extend to a setting with private information. We use stability to identify outcomes that will not be renegotiated, that is, stable outcomes. We then allow the principal to optimize among mechanisms that lead to such stable outcomes. As stable outcomes are final, we can characterize the strategy of the agent in such mechanisms: he will choose messages leading to optimal contracts, given his type.

Without private information, the idea of von Neumann–Morgenstern stability is as follows. Outcomes are either stable or unstable. Only stable outcomes can be endpoints of the renegotiation. Unstable outcomes can be improved by stable outcomes (external stability), while stable outcomes cannot be improved by other stable outcomes (internal stability). One of our contributions is to extend this idea to a Bayesian setting with private information.

With private information, the outcome of a mechanism is a tuple consisting of the contract chosen by the agent and the belief formed by the principal after she observes the agent's choice. We capture potential improvements by considering sets of outcomes generated by a joint probability measure over contracts and agent types. We call such measures *outcome measures.* This is essentially a reduced-form approach to any renegotiation game. Suppose that after a mechanism has generated an outcome, renegotiation proceeds through a specific renegotiation game. The chosen strategies of the principal and the agent in the game generate contracts at the terminal nodes of the game which depend on the agent's type through his strategy. The outcomes of the game can be summarized by a joint probability measure over contracts and agent types.

Our generalization of von Neumann–Morgenstern stability then works as follows. Define a certain set of outcomes as *stable*—that is, the principal and the agent expect these stable outcomes not to be overturned. Now consider outcome measures that are consistent with this stable set—that is, outcome measures that yield only stable outcomes. Call such outcome measures *stable outcome measures.* The distribution of types and contracts in a stable outcome measure is derived from the notion that each type of agent will hold a contract that is optimal for his type, among all contracts in the support of the stable outcome measure. This is because the agent would choose not to hold his optimal contract only if there were a further change yielding a better contract. However, by definition, stable outcomes will not be changed; thus, each type of agent must receive an optimal contract.

As with von Neumann–Morgenstern stability, a stable set is defined by two properties. First, stable outcomes cannot be improved by outcomes of any stable outcome measure (internal stability). Second, any unstable outcome can be improved by outcomes of some stable outcome measure (external stability).

This definition of a stable set also allows us to reason about improvements leading to unstable outcomes. In such a potential improvement, by external stability, the agent would expect any unstable outcome to be changed further, leading to a stable outcome. This, in turn, restricts which outcome measures can be considered as potential improvements in the first place. As the final improvement leads to stable outcomes, internal stability ensures that the corresponding outcome measure does not yield an improvement of the initial outcome.

One of the main advantages of our characterization of stability is that it provides an effective and easy-to-use tool for analyzing the principal's mechanism design problem. The principal chooses a stable set and a mechanism to maximize her utility. She is restricted to mechanisms that yield stable outcomes. The agent optimizes his contract within the chosen mechanism.

We demonstrate our approach by considering two instances of the general model. First, we analyze a setting similar to that of Mussa and Rosen (1978), with a continuous type space, private values, and nonlinear contracts. This setting encompasses many applications in which the principal is naturally unable to rule out renegotiation, such as procurement, labor contracts, or selling when price and quality matter. As our leading example of the setting, we use a seller selling a good of varying quality to a privately informed buyer. The contracts consist of two dimensions: price and quality.

With full commitment, it is optimal for the principal to offer a continuum of contracts, and the agent types fully separate, with only the highest type receiving an efficient contract. This is the well-known "no-distortion-at-the-top" property. Clearly this is not sustainable if the principal is not able to commit. If there is full separation of types with inefficient contracts, the principal will know the agent's type and thus can improve the outcome by offering each type his efficient quality. Such efficient outcomes cannot be improved further and must therefore lie in any stable set.

We show that the set of optimal stable outcomes for the principal has the following properties: (1) the principal offers a *countably* infinite number of contracts; (2) each contract is signed by a *pool* of agent types that is of positive measure; (3) the lowest type in each pool receives an efficient contract, while every other type in each pool receives an inefficient contract, a property that we call “no distortion at the bottom”.

Second, we analyze the Coase conjecture (Coase, 1972), which addresses the problem of a seller selling a durable good to a buyer who has private information about his valuation. For this setting, we rederive the “gap–no gap” result in a simple way. 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 at most a price equal to the buyer’s lowest valuation. If, however, the seller’s cost is above the buyer’s lowest valuation—the no-gap case—the seller can charge the monopoly price. It is reassuring that although our approach abstracts away from a specific game structure, it delivers the same results as the approach of Ausubel and Deneckere (1989), which is based on modeling a renegotiation game.

Relationship to the literature. The stable set of von Neumann and Morgenstern (1944) has distinguished standing as a solution concept in cooperative game theory, and it has received renewed interest in the recent literature. For example, Ray and Vohra (2015) extend the notion of a stable set to a far-sighted dominance relation that captures dynamics arising from a sequence of improvements of an outcome. Dutta and Vohra (2017) refine the concept by requiring that players hold rational expectations about continuation paths of improvements, and Dutta and Vartiainen (2020) consider history dependence. The continuing interest in von Neumann–Morgenstern stability shows the simplicity and power of the underlying idea. Thus, it is natural to extend the notion of stability to private information as an abstraction for renegotiation and combine it with a mechanism design approach to study limited commitment.

Beyond the von Neumann–Morgenstern notion, there are other cooperative approaches to stability. Liu (2020) analyzes stability in a two-sided market with asymmetric information. Stability is defined with respect to a matching of market participants with corresponding beliefs. Beliefs are defined so that pairwise deviations from the matching are deterred and the matching is individually rational. The work of Liu (2020) complements the belief-free approach to stability taken by Liu et al. (2014), who require a matching to be stable under any reasonable belief.[1] Asheim and Nilssen (1997) consider a monopolistic insurance market with a discrete type-space. They use the notion of a standard of behavior, which resembles the definition of a stable set in this manuscript. As in our case, this approach proves to be very useful in deriving clear results for an otherwise complex problem.

Cooperative solution concepts are also used to analyze lack of commitment in mechanism design. Vartiainen (2013) analyzes auctions without commitment, in which the principal is able to propose a new mechanism any number of times without honoring previous outcomes. He introduces a cooperative solution concept for such a situation and demonstrates that an English auction without a reserve price is the only mechanism that is implementable with this approach.[2] In contrast to Vartiainen (2013), we analyze renegotiation rather than lack of commitment; in our set-up, the agent and the principal can each choose to retain any signed contract.

Neeman and Pavlov (2013), like us, abstract away from specific renegotiation games. However, unlike us, they argue that for outcomes of a mechanism

[1] Both approaches complement the seminal notions of an incomplete information core (Wilson, 1978) and the credible core (Dutta and Vohra, 2005).

[2] This is in line with the non-cooperative approach of Liu et al. (2019) to a similar problem.

to be renegotiation-proof under any renegotiation procedure, there must be no Pareto improvements possible after the mechanism has been played. The conceptual difference between our approach and theirs is that the latter permits all Pareto-improving outcomes to be blocking, even if those outcomes are themselves subject to renegotiation, whereas we consider outcomes to be stable potentially blocking only if they are themselves stable. In particular, in contrast to the approach of Neeman and Pavlov (2013), our approach does not exclude inefficient stable outcomes.

Our paper is also related to the vast literature on mechanism design without commitment, which models limited commitment as a game of sequentially posted mechanisms. For example, Bester and Strausz (2001), Evans and Reiche (2015), Skreta (2006), and Skreta (2015) consider games in which a designer can post a new mechanism a finite number of times and solve the problem via backward induction.

The results of Gretschko and Wambach (2017) complement the ideas in this manuscript. In that paper we assume a discrete type space and consider a renegotiation game in which the principal can propose a new mechanism any number of times. We show that stable sets arise as a perfect Bayesian equilibrium of this game.

Strulovici (2017) shows that if in a specific infinite-horizon bargaining protocol over nonlinear contracts friction disappears, efficient and fully separating contracts arise in any perfect Bayesian equilibrium if values are private and the type space is binary. Our application to nonlinear contracts nests his set-up as a special case. In particular, with a discrete type space, only efficient and fully separating outcomes can be stable. As this is true for all discrete type spaces and not only binary ones, our result can be seen as a stable-set extension of his result. Moreover, our approach lets us demonstrate that there is a conceptual difference between discrete and continuous type spaces. With a continuous type space, the optimal outcomes are pooling and thus inefficient. As Malcomson (2016) and Malcomson (2021) point out, in many applications agents of different types are pooled in groups, with the membership of each group persistent over time, and all group members treated the same despite the differences between them. For example, employees may be placed in pay grades, with those in a grade all paid the same; buyers may place their suppliers into a small number of categories that receive differential treatment. Malcomson (2016) provides a rationale for such pooling based on dynamics in relational contracts. Our results can be seen as providing a different rationale, based on stability.

Doval and Skreta (2022) derive a revelation principle for an infinite sequence of mechanisms. Our approach is complementary, as we abstract away from a specific mechanism-design game or renegotiation procedure. Moreover, Doval and Skreta (2022) assume that the mechanism designer has access to a mediator, that is, a device that privately collects messages from the agent and produces public signals conditional on the messages. The mediator allows them to focus on truth-telling mechanisms, with the message space being equal to the type space. By randomizing over signals, the mediator prevents the truth-telling messages from fully revealing the agent's private information to the mechanism designer. In contrast, we assume that communication is direct; that is, the principal directly observes the agent's message. However, full revelation of the agent's information is prevented by the principal's choice of the message space and by pooling or mixing on the part of the agent.

2. Model

Preferences and beliefs. A principal (she) and an agent (he) intend to implement a contract w from a set of contracts $\mathcal{W}$. If a contract w is implemented, the principal's utility amounts to $V(w)$, where $V : \mathcal{W} \to \mathbb{R}$ is a von Neumann–Morgenstern utility function. To simplify the use of probability measures on $\mathcal{W}$, we assume that $\mathcal{W}$ 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 $U : \mathcal{W} \times \Theta \to \mathbb{R}$ is a von Neumann–Morgenstern utility function and depends on the agent's type θ. The agent's type is private information to the agent and is drawn from $\Theta \subset \mathbb{R}$. We assume that Θ is a Polish space and endow Θ with the Borel σ-algebra; we denote by $\Delta(\Theta)$ the set of all probability measures on the Borel σ-algebra. We denote by θ_ϵ the ϵ-neighborhood of θ—that is, the set of all $\theta' \in \Theta$ such that $|\theta - \theta'| < \epsilon$. The principal's prior about the agent's type is $\mu_0(\cdot) \in \Delta(\Theta)$. The prior is common knowledge between the agent and the principal. For $\mu \in \Delta(\Theta)$, we denote by $\mathrm{supp}(\mu)$ the support of μ. The support is the closure of the set $\{\theta \in \Theta : \mu(\theta_\epsilon) > 0 \ \forall \epsilon > 0\}$. If no contract is implemented, both parties receive the outside option contract, denoted by $w_0 \in \mathcal{W}$.

Mechanisms. A mechanism is a pair $(\mathcal{Z}, \varphi(\cdot))$ consisting of a Polish space of messages $\mathcal{Z}$ endowed with the Borel σ-algebra and a measurable function $\varphi : \mathcal{Z} \to \mathcal{W}$.[3] The agent (possibly randomizing) chooses messages from $\mathcal{Z}$. His choice is encoded by a choice measure σ, which is a joint measure on $\Theta \times \mathcal{Z}$ endowed with the Borel σ-algebra such that the marginal of σ on Θ is the one defined by the prior μ_0. That is, $\sigma(A, \mathcal{Z}) = \mu(A)$ for any subset A in the Borel σ-algebra on Θ. When a message z is determined by σ, $\varphi(z)$ assigns a contract. The belief of the principal is updated to the regular conditional probability $\mu_0(\cdot \mid z)$ on Θ.[4]

[3] As we consider the possibility of renegotiation, the revelation principle does not necessarily apply to our set-up. Thus, we need to consider message sets that may differ from Θ. To be able to work with measurable functions and define probability measures on $\mathcal{Z}$, we restrict $\mathcal{Z}$ to be a metric space.

[4] As $\mathcal{Z}$ and Θ are Polish spaces, $\mu_0(A \mid z)$ exists.

Example 1. *To fix ideas, consider the following specification of the Coase conjecture, which we will analyze in more detail in Section 5. The principal is a seller who sells a good to the agent, a buyer. A contract consists of a pair* (p, q)*, where* $p \in \mathbb{R}^+$ *specifies the price of the good and* $q \in \{0, 1\}$ *specifies whether the good is traded* $(q = 1)$ *or not* $(q = 0)$*. The principal incurs a cost of* $c = 0$ *when selling the good. Her payoff from a contract* (p, q) *is* $p - qc = p$*. A buyer of type* $\theta \in \Theta = [0, 1]$*, with* θ *uniformly distributed, enjoys a utility of* θ *when consuming the good. We denote by* $\mu^{uni}(a, b)$ *the uniform measure on the interval* (a, b)*. The buyer's utility from a contract* (p, q) *is* $q\theta - p$*. A potential mechanism is* $\mathcal{Z} = \{0, 1\}$ *with* $\varphi(1) = (p, 1)$ *and* $\varphi(0) = (0, 0)$*. That is, the good is sold at a price of* p *if the message is* 1 *and not sold if the message is* 0*.*

The problem. Suppose a mechanism is played and produces a contract $w = \varphi(z)$. The principal then updates her belief to $\mu = \mu_0(\cdot \mid z)$. We call (w, μ) an outcome of the mechanism. The problem arises because the agent and the principal can decide to change (renegotiate) the outcome. This possibility may become relevant if they expect that w can be improved for both parties under the belief μ. The possibility that outcomes can be changed complicates the analysis of the mechanism design problem. If the agent expects that outcomes of a mechanism will be changed, he will take this into account when choosing a message, which makes his choice measure hard to pin down.

Example 2. *Consider the mechanism in Example 1 with* $p = 0.5$ *and the agent's choice measure choosing* 1 *if* $\theta > 0.5$ *and choosing* 0 *otherwise. After observing a message of* 1 *the buyer updates her belief to* $\mu^{uni}(0.5, 1]$*, and after observing a message of* 0 *she updates her belief to* $\mu^{uni}[0, 0.5]$*. The contract* $(0, 0)$ *with updated belief* $\mu^{uni}[0, 0.5]$ *can be improved for both parties with the contracts* $(0.25, 1)$ *and* $(0, 0)$*.*

However, if the agent expects that the outcome $((0,0),\mu^{uni}[0,0.5])$ *will be changed, it is initially not optimal for him to choose message* 1 *to receive the contract* $(0.5,1)$*, as he expects to be offered* $(0.25,1)$*—a contract with a lower price—later on if he chooses message* 0 *instead. Thus,* $((0.5,1),\mu^{uni}[0.5,1])$ *does not seem to be a plausible outcome of the proposed mechanism once renegotiation is considered.*

3. Solution concept: stable sets

Rather than focusing on specific renegotiation games, we focus on outcomes, that is, contract–belief pairs arising from a mechanism. In particular, we will define conditions characterizing outcomes that will not be changed. We call such outcomes *stable outcomes*. Suppose that all outcomes that satisfy the (yet to be defined) conditions form a set Ω, and suppose a mechanism yields outcomes that are all in Ω, given the choices of the agent. If the agent expects that the outcomes in Ω will not be changed, his choice measure in this mechanism will maximize his utility and thus can be pinned down. The principal in turn maximizes over mechanisms that yield outcomes in Ω, given that the agent will maximize his utility.

A natural condition to impose is that stable outcomes should be unimprovable. This is essentially the approach taken by Neeman and Pavlov (2013). However, their approach does not take into account that potential improvements may themselves be subject to change. In what follows, we explore a solution concept that takes this issue into account. In our approach, to determine whether an outcome is stable, we evaluate it only against other (stable) outcomes that will not themselves be changed. In what follows we explore a solution concept in the spirit of von Neumann and Morgenstern (1944) that takes this issue into account.

Example 3. *A mechanism that yields unimprovable outcomes in Example 2 is to implement the contract* $(0, 1)$*—that is, to sell the good at a price of* 0 *if the message is* 1 *and not to sell the good if the message is* 0*. The agent chooses message* 1 *irrespective of his type, and the only outcome of this mechanism is* $((0, 1), \mu^{uni}[0, 1])$*. This outcome cannot be improved for both parties, as a higher price or no trade would make the agent strictly worse off, and a lower price would make the principal strictly worse off.*

The question becomes whether this is the only stable outcome. The reason we have ruled out the outcomes $((0.5, 1), \mu^{uni}(0.5, 1])$ *and* $((0, 0), \mu^{uni}[0, 0.5])$ *is that* $((0, 0), \mu^{uni}[0, 0.5])$ *can be improved further if the principal, for example, offers to sell the good at a price of* 0.25*. However, by the same token, it would be an improvement for the principal to lower the price further, for example, to* 0.125*, in case there is no sale at the price* 0.25*. But if the agent expects further improvements, he might (rationally) reject the price of* 0.25 *even if his valuation is above* 0.25*. But if the agent will not accept a price of* 0.25*, is the offer at that price an improvement after all?*

Our solution concept adapts the stable sets of von Neumann and Morgenstern (1944) to incorporate private information. Without private information the concept works as follows. Suppose the principal and the agent bargain to implement a contract $w \in \mathcal{W}$. A contract w' *improves* the contract w if both the agent and the principal weakly prefer w' over w and at least one of the parties strictly gains. The main idea of the von Neumann–Morgenstern stable set is the following: just because w' improves w, it does not follow that w will be changed to w' through renegotiation. This is because there may be yet another contract w'' that improves w'; then, by the same logic, w' would be changed to w'', and thus would not serve as an improvement of w. To determine whether w can be implemented, we need to compare it to other contracts that can be implemented. For example, if w'' can be implemented and it improves on w, then w cannot be implemented. The von Neumann–Morgenstern stable set resolves the circularity introduced here. It defines all potential results of the bargaining simultaneously, as a stable set of contracts satisfying the following two properties:

(i) *Internal stability:* If w is in the stable set, it cannot be improved by any w' in the stable set.

(ii) *External stability:* If w is not in the stable set, there exists w' in the stable set that improves w.

The elements of the stable set are those contracts that are not dominated by any other contracts in the stable set. Elements of the stable set are called stable contracts. No stable contract can be improved by another stable contract (internal stability), and all unstable contracts can be improved by stable contracts (external stability).

Dominance. We adapt the concept of von Neumann–Morgenstern stability to private information. There are two main steps in this adaptation. First, as compared to the case without private information, we have an additional object to take care of: the beliefs of the principal about the agent's type. We will therefore define a dominance relation on outcomes $\hat{o} = (\hat{w}, \hat{\mu})$ consisting of contracts and beliefs rather than contracts only. Second, as different agent types may receive different contracts in an improvement, the dominance relation must compare outcomes with sets of outcomes. A potential improvement is therefore summarized by a probability measure $\gamma_{\hat{o}}$ on $\Theta \times \mathcal{W}$. The marginal of $\gamma_{\hat{o}}$ on Θ is given by $\hat{\mu}$. The principal's beliefs $\mu(A \mid w)$ in the outcomes of the potential improvement are implicitly given by $\gamma_{\hat{o}}$ through the regular conditional probability $\mu(\cdot \mid w)$. As $\mathcal{W}$ and Θ are Polish spaces, $\mu(\cdot \mid w)$ exists.

Let W be the support of $\gamma_{\hat{o}}$ on $\mathcal{W}$. We call $\mathcal{O}(\gamma_{\hat{o}}) = \{(w, \mu(\cdot \mid w)) : w \in W\}$ the set of outcomes and refer to $\gamma_{\hat{o}}$ as an outcome measure.

Let Ω be a stable set (a concept we have yet to define formally); roughly, Ω is a set of outcomes that will not be changed. A stable set helps us to define outcome measures with outcomes that dominate unstable outcomes and pin down beliefs in stable outcomes. Consider an outcome measure $\gamma_{\hat{o}}$ with marginal $\hat{\mu}$ whose set of outcomes $\mathcal{O}(\gamma_{\hat{o}})$ is a subset of Ω—that is, the outcome measure $\gamma_{\hat{o}}$ produces only stable outcomes. Since these outcomes (by definition) will not be changed, the principal's beliefs should reflect that in these outcomes, each type of agent receives a contract maximizing his utility. This is because the agent would fail to receive his optimal contract only if there were a further change yielding a better contract, which cannot occur for outcomes in Ω. This idea leads to the following definition of dominance.

Definition 1 (dominance)**.** *Let $\hat{o} = (\hat{w}, \hat{\mu})$ be an outcome, let $\gamma_{\hat{o}}$ be an outcome measure with marginal $\hat{\mu}$ on Θ, let $W = supp(\gamma(\Theta, \cdot))$, and let $\mathcal{O}(\gamma_{\hat{o}}) = \{(w, \mu(\cdot \mid w)) : w \in W\}$ be its set of outcomes. We say $(\gamma_{\hat{o}}, \mathcal{O}(\gamma_{\hat{o}}))$* dominates *$\hat{o} = (\hat{w}, \hat{\mu})$ if the following three properties hold:*

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

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

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

$$\int_{\mathcal{W}} V(w) \mathrm{d}\gamma_{\hat{o}}(\Theta, w) \geq V(\hat{w}). \tag{1}$$

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

We say $(\gamma_{\hat{o}}, \mathcal{O}(\gamma_{\hat{o}}))$ strictly dominates *$\hat{o}$ if inequality (1) is a strict inequality.*

Definition 1 captures the idea that if the principal and the agent change the outcome $\hat{o}$, they change it in such a way that each agent type is weakly better off with one of the new contracts (requirement (i)) and the principal is (weakly) better off in expectation (requirement (ii)). Requirement (iii) accounts for the fact that if the outcomes resulting from $\gamma_{\hat{o}}$ will not be changed, the principal should believe that the agent receives his optimal contract. In particular, after observing that a contract w has been implemented, she should not believe that she faces an agent of type θ if there exists an outcome with a contract w' that would make an agent of type θ strictly better off.

Note that while the formal definition of dominance in Definition 1 makes no reference to the yet-to-be-defined notion of a stable set, the main idea of the definition, namely that each agent's type receives an optimal contract, relies on the implicit assumption that the outcomes of a dominating outcome measure are stable. We have written the definition in this way to avoid referring to objects that are not yet defined. However, the definition of a stable set in the next paragraph will make the connection between dominance and stable sets precise, and will refer to dominance only with respect to outcome measures with stable outcomes. We discuss outcome measures with unstable outcomes at the end of this section.

Stable sets. We are now in a position to define stable sets. As with von Neumann–Morgenstern stable sets, stable outcomes are those that are not strictly dominated by outcomes of any outcome measure with stable outcomes; unstable outcomes are those that are dominated by outcomes of some outcome measure with stable outcomes.

Definition 2 (stable set)**.** *A set* $\Omega \subset \{(w, \mu) : w \in \mathcal{W} \wedge \mu \in \Delta(\Theta)\}$ *of outcomes is a* stable set *if the following hold:*

(i) Internal stability: *If o is in Ω, there exists no $(\gamma_o, \mathcal{O}(\gamma_o))$ with $\mathcal{O}(\gamma_o) \subset \Omega$ that strictly dominates o.*

(ii) External stability: *If o is not in Ω, there exists $(\gamma_o, \mathcal{O}(\gamma_o))$ with $\mathcal{O}(\gamma_o) \subset \Omega$ that dominates o.*

We refer to outcome measures γ_o with $\mathcal{O}(\gamma_o) \subset \Omega$ for a stable set Ω as stable outcome measures*.*

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. No stable outcome can be improved by other stable outcomes making the principal strictly better off. On the other hand, unstable outcomes do not block stable outcomes, as any unstable outcome will be improved to a set of stable outcomes.

Example 4. *Consider the set-up from Example 1. One stable set Ω for this example consists of the following types of outcomes:*

(1) All efficient outcomes where the good is traded—that is, outcomes $((p, 1), \mu)$ with $\mu \in \Delta([0, 1])$.

(2) The single outcome $((0, 0), \mu^{uni}[0, 0.5])$—that is, the outcome where the good is not traded and the principal believes that she faces only types θ below 0.5.

This set is internally stable: first, outcomes where the good is traded are efficient and cannot be improved for the agent and the principal at the same time. Second, the "no-trade" outcome is also internally stable. This is because, starting from $((0, 0), \mu^{uni}[0, 0.5])$, the only stable outcomes in Ω that can be reached by an outcome measure with marginal $\mu^{uni}[0, 0.5]$ are $((0, 0), \mu^{uni}[0, 0.5])$ and $((0, 1), \mu^{uni}[0, 0.5])$—that is, the same outcome with no trade, and another outcome with trade at a price of 0, neither of which makes the principal strictly better off. In other words, in a situation where the agent does not buy at a *price of* 0.5*, the principal will not lower her price, as the only feasible stable outcome is to trade at a price of* 0*, which would not make her better off.*

The set Ω *is also externally stable: as the principal's cost is* 0*, all outcomes* $((p,0),\mu)$ *where the good is not traded can be improved by an outcome measure yielding a single stable outcome* $((p,1),\mu)$*.*

We are now in a position to state the principal's problem.

The principal's problem. The principal chooses a stable set Ω and a mechanism. She is restricted to mechanisms that yield stable outcomes. The agent chooses messages to optimize the contract he receives from the mechanism. Given that stable outcomes will not be changed, it is indeed optimal for the agent to optimize within the mechanism.

More formally, the principal chooses a stable set Ω and a mechanism to maximize her utility. She designs a message space $\mathcal{Z}$ and a contract function $\varphi : \mathcal{Z} \to \mathcal{W}$. The agent chooses messages according to a choice measure σ with the property that (θ, z) is in the support of σ if and only if $z \in \arg\max_{z \in \mathcal{Z}} u(\varphi(z), \theta)$. That is, the agent maximizes his utility within the mechanism. The outcomes of the mechanism $\{(\varphi(z), \mu_0(\cdot \mid z) : z \in \mathcal{Z}\}$, given the agent's choice measure, have to be a subset of Ω. If there is more than one such choice measure, we choose the one that maximizes the principal's utility, which is given by

$$\int_{\mathcal{Z}} V(\varphi(z)) \mathrm{d}\sigma(\Theta, z). \tag{2}$$

We call the resulting mechanism the *optimal stable mechanism.*

Example 5. *In the set-up from Example 1, suppose the principal chooses the stable set from Example 4 and the mechanism from Example 1 with* $p = 0.5$*, while the agent chooses message* 1 *if* $\theta > 0.5$ *and message* 0 *otherwise. This mechanism and choice function result in stable outcomes as outlined in*

Example 4. As the principal receives the full-commitment monopoly price of 0.5*, this is the solution to the principal's problem.*

Discussion of the solution concept. Our solution concept allows us to reason about mechanisms and improvements without assuming a specific renegotiation procedure. In principle, improvements can come about in many ways: for example, the principal can propose a new mechanism or the agent and the principal can engage in an alternating-offers bargaining game, etc. The outcome measure captures different ways of generating improvements.

Outcome measures with unstable outcomes. The concepts of internal and external stability allow us to reason about proposed changes to unstable outcomes, or (more formally) about outcome measures with unstable outcomes. For outcome measures with unstable outcomes, the underlying assumption in the definition of dominance—that the agent receives an optimal contract—no longer applies, as unstable outcomes will be changed. However, by external stability, any outcome outside the stable set will be dominated by (and therefore changed to) outcomes in the stable set. Thus, external stability implies a recursive restriction on beliefs for outcome measures with unstable outcomes.

To see the logic of the recursive restriction, consider an outcome $(\hat{w}, \hat{\mu})$ and an outcome measure γ with marginal $\hat{\mu}$ and set of outcomes $\mathcal{O}(\gamma)$, some of which are not in Ω. Let $\{(y, \mu(\cdot \mid y)) : y \in \mathcal{W}\} \subset \mathcal{O}(\gamma)$ be the set of outcomes that are not in Ω, and let $\{(x, \mu(\cdot \mid x)) : x \in \mathcal{W}\} \subset \mathcal{O}(\gamma)$ be the set of outcomes that are in Ω.

For outcomes $o^y = (y, \mu(\cdot \mid y)) \in \mathcal{O}(\gamma)$ that are not in Ω, by external stability there exists an outcome measure γ_{o^y} with marginal $\mu(\cdot \mid y)$ and outcomes $\mathcal{O}(\gamma_{o^y})$ in Ω that dominate o^y. We can therefore replace each unstable outcome o^y in $\mathcal{O}(\gamma)$ with $\mathcal{O}(\gamma_{o^y})$ and construct an outcome measure $\hat{\gamma}$ with marginal $\hat{\mu}$ that yields this new set of outcomes, which are all in the stable set.

As unstable outcomes will be replaced by sets of stable outcomes, the beliefs in $\{(y, \mu(\cdot \mid y)) : y \in \mathcal{W}\} \subset \mathcal{O}(\gamma)$ and $\{(x, \mu(\cdot \mid x)) : x \in \mathcal{W}\} \subset \mathcal{O}(\gamma)$ should reflect this, restricting the outcome measures γ we can consider in the first place. In particular, consider two unstable outcomes of γ, o^y and $o^{y'}$. If there exists an outcome $(w', \mu(\cdot \mid w'))$ in $\mathcal{O}(\gamma_{o^{y'}})$ such that $u(w', \theta)$ is strictly larger than $u(w, \theta)$ for some $\theta \in \Theta$ and for all outcomes $(w, \mu(\cdot \mid w))$ in $\mathcal{O}(\gamma_{o^y})$, then the principal should not believe that she is facing an agent of type θ in outcome o^y. The same argument can be made comparing stable outcomes of γ with stable and with unstable outcomes.

Put differently, if the principal and the agent anticipate that unstable outcomes will be changed to stable outcomes, the principal's beliefs should take into account that each type of agent will end up with his optimal contract after the final change. Thus, the constructed outcome measure $\hat{\gamma}$ has property (iii) of Definition 1.[5] In particular, $(\hat{\gamma}, \mathcal{O}(\hat{\gamma}))$ dominates $(\hat{w}, \hat{\mu})$ in the sense of Definition 1, and all its outcomes are in Ω.

Summing up, whenever we consider an outcome measure with unstable outcomes, external stability ensures that the unstable outcomes can be further improved to stable outcomes. The principal's beliefs should take such changes into account; that is, the principal should assume that after the final change the agent will end up with an optimal contract for his type. Thus, it is sufficient to evaluate improvements in relation to stable outcome measures.

In particular, if the principal and the agent agree on a stable outcome, any further improvements will ultimately result in stable outcomes. Either

[5] To see this more formally, suppose that there are two outcomes $o = (w, \mu)$ and $o' = (w', \mu')$ in $\mathcal{O}(\hat{\gamma})$ such that $u(w', \theta) > u(w, \theta)$ but θ is in the support of μ. This cannot be the case if o and o' are in the same $\mathcal{O}(\gamma_y)$. Thus, suppose that o is in $\mathcal{O}(\gamma_y)$ and o' is in $\mathcal{O}(\gamma_{y'})$ for $y \neq y'$. As the agent of type θ receives an optimal contract in $\mathcal{O}(\gamma_y)$, $u(w', \theta) > u(w, \theta)$ implies that $u(w', \theta)$ is strictly larger than $u(w, \theta)$ for all outcomes $(w, \mu(\cdot \mid w))$ in $\mathcal{O}(\gamma_{o^y})$. Thus, θ is not in the support of $\mu(\cdot \mid y)$. As the marginal of γ_{o^y} is $\mu(\cdot \mid y)$, it cannot be in the support of μ. The same argument can be made if o or o' is a stable outcome of γ.

these outcomes will make the principal worse off, or there is a stable outcome measure that dominates the original stable outcome. In the latter case, internal stability guarantees that the principal is not strictly better off. We illustrate this recursive restriction with an example.

Example 6. *Consider the stable set from Example 4. Suppose the principal offers the two contracts* $(0.25, 1)$ *and* $(0, 0)$*. We demonstrate that no type of agent should choose the first contract. To see this, observe that any outcome* $((0.25, 1), \mu(\cdot \mid (0.25, 1)))$ *is stable, and the no-trade outcome* $((0, 0), \mu(\cdot \mid (0, 0))$ *is stable only if all agent types in* $[0, 0.5]$ *end up with the contract* $(0, 0)$*. However, this cannot be the case, as then both outcomes would be stable and the types in* $(0.25, 0.5)$ *would be strictly better off with the contract* $(0.25, 1)$*. Thus, assume that* $\mu(\cdot \mid (0, 0)) \neq \mu^{uni}[0, 0.5]$*. In this case, the no-trade outcome is by definition not in* Ω *and is therefore unstable. The outcome measure producing the sole stable outcome* $((0, 1), \mu(\cdot \mid (0, 0))$ *dominates* $((0, 0), \mu(\cdot \mid (0, 0))$*. We can then construct an outcome measure* $\hat{\gamma}$ *by replacing* $((0, 0), \mu(\cdot \mid (0, 0))$ *by* $((0, 1), \mu(\cdot \mid (0, 0))$*.*

Irrespective of his type, the agent is strictly better off with the contract $(1, 0)$ *than with any outcome* $((0.25, 1), \mu(\cdot \mid (0.25, 1))$*, since the latter is stable and will not be changed regardless of the principal's belief. Thus, the principal should believe that* $\mu(\cdot \mid (0, 0)) = \mu^{uni}[0, 1]$*.*

Indifferences reflecting the principal's bargaining power. We have defined strict dominance solely with respect to the principal's payoff. This captures the idea that even though outcomes can be changed, the principal retains some of her bargaining power throughout the renegotiation process. That is, if new outcomes are only weakly but not strictly better from the principal's point of view, she can unilaterally block or propose changes.

This is reflected in the definition of internal and external stability. Internal stability is defined with respect to the principal's payoff: if a given change would not make the principal strictly better off, she can unilaterally block the change. Such an assumption makes the definition of a stable set more permissive: a set of outcomes may satisfy internal stability even if it permits a change that would make a positive measure of agent types strictly better off, provided that such a change would make the principal only weakly better off.

For external stability, it is sufficient that any unstable outcome is *weakly* dominated by stable outcomes. Again, this assumption makes the definition of a stable set more permissive: the principal can exclude outcomes from a stable set that would otherwise undermine internal stability, even if such outcomes would not make her worse off.

Multiplicity and existence of stable sets. As we show in the application below, stable sets are not unique. We therefore assume that the principal chooses the stable set. This is in line with the assumption that the principal retains some bargaining power during renegotiation. Moreover, it allows us to compare the results with those of the standard case, in which the principal can commit to one mechanism and chooses the optimal mechanism. The assumption that the principal chooses the stable set provides a lower bound for the gap between the optimal outcomes with and without commitment.

The existence of stable sets can be established under some mild assumptions. For example, we can establish existence for quasi-linear utility as follows. Let the set of contracts be $\mathcal{W} = (Y, \mathbb{R})$; a contract is given by a pair (y, p) with $y \in Y$ being an allocation and $p \in \mathbb{R}$ a transfer. The agent's utility is $u(y, \theta) - p$, and the principal's utility is $v(y) + p$. Assume that $y^*(\theta) = \max_{y \in Y}\{u(y, \theta) + v(y)\}$ exists for all $\theta \in \Theta$. In this case, we claim that the set of all ex-post efficient outcomes $\Omega = \{((y, p), \mu^{y^*}) \mid \theta \in \text{supp}(\mu^{y^*}) \Rightarrow y = y^*(\theta)\}$ is a stable set. The set Ω satisfies internal stability because any outcome in Ω is ex-post efficient for all types in the support of the belief. Thus, there is no outcome measure that would make the principal strictly better off without leaving some agent types worse off. For external stability, take any outcome $((y, p), \mu)$ that is not in Ω. Consider the outcomes of the form $\{(y^*(\theta), p - v(y^*(\theta)) + v(y), \mu^{y^*}) \mid \theta \in \operatorname{supp}(\mu)\}$. As

$$\begin{aligned} & u(y, \theta) - p \leq u(y^*(\theta), \theta) - p - v(y) + v(y^*(\theta)) \\ \Leftrightarrow\ & u(y, \theta) + v(y) \leq u(y^*(\theta), \theta) + v(y^*(\theta)), \end{aligned}$$

each agent type θ (weakly) prefers the contract $(y^*(\theta), p - v(y^*(\theta)) + v(y))$ to (y, p). The principal is indifferent between all contracts $(y^*(\theta), p - v(y^*(\theta)) + v(y))$ and (y, p). Thus, there exists an outcome measure γ with $\mathcal{O}(\gamma) = \{(y^*(\theta), p - v(y^*(\theta)) + v(y), \mu^{y^*}) \mid \theta \in \operatorname{supp}(\mu)\}$ that (weakly) dominates $((y, p), \mu)$. This shows that Ω is also externally stable.

Communication and information revelation. We assume that communication is direct in the mechanism; that is, the principal directly observes the message chosen by the agent. Direct communication does not imply, however, that the principal always learns the agent's type by observing the message. Agent types can pool on messages; that is, multiple agent types can choose the same message (potentially mixing between messages). Thus, by designing the message space, the principal can control how much she learns from a mechanism. The information structure arising from the message space and the agent's strategy can be replicated in a setting with a mediator by assuming that the agent truthfully reports his type to the mediator and that the mediator produces signals and allocations that match the messages chosen by the agent in the mechanism with direct communication.

As we show in Section 4 for the case of nonlinear contracts, the principal chooses a message space with strictly fewer messages than types, enforcing pooling of agent types.

Some useful results. Before we turn our attention to specific applications of the solution concept, we state two results that will facilitate the construction of stable sets.

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

This result follows directly from external stability

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 argued in Example 4 and will 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 $o = (\hat{w}, \hat{\mu})$, if there exists an outcome $o' = (w', \hat{\mu})$ such that $V(w') > V(w)$ and $U(w') \geq U(\hat{w})$ for all θ in the support of μ, then o is not in Ω.*

In the following all proofs are relegated to the appendix. Lemma 2 has an intuitive interpretation: given any potential outcome, if there exists a single contract that would be accepted by the agent independent of his type and that would make the principal strictly better off, then the initial outcome cannot be stable. If such a contract exists, the agent and the principal can change to that contract without the agent's revealing any additional private information, and this change will make the principal strictly better off. If the resulting outcome is stable, this already demonstrates that the original outcome could not have been stable. If the resulting outcome is not stable, by external stability there exists a stable outcome measure that dominates it. However, the same outcome measure then strictly dominates the original outcome.

4. Nonlinear contracts

We now turn to our main application of stable sets: the analysis of nonlinear contracts. We proceed as follows. Firstly, we set up the model. Secondly, we introduce a simplified example to illustrate the intuition. Thirdly, we state the main result. Finally, we prove the main result.

Set-up. Consider a principal who wants to implement a two-dimensional contract $w = (p, q)$ with $q \in \mathbb{R}_+$ and $p \in \mathbb{R}$. If a contract (p, q) is implemented, the utility of the principal is given by

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

Denote by $c_q(\cdot)$ the derivative of $c(\cdot)$ with respect to q and by $c_{qq}(\cdot)$ the second derivative of $c(\cdot)$ with respect to q. Assume that $c_q(\cdot) > 0$ and $c_{qq}(\cdot) > 0$.

The type θ of the agent is taken from $\Theta = \left[\underline{\theta}, \overline{\theta}\right]$. The utility of the agent is given by

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

Denote by u_q the derivative of u with respect to q and by u_{qq} the second derivative of u with respect to q. Similarly, denote by u_θ the derivative of u with respect to θ and by u_{q_θ} the cross-derivative of u with respect to q and θ. Assume that $u_q > 0$ and $u_{qq} \leq 0$, and that u satisfies single crossing. That is, $u_\theta > 0$ and $u_{q\theta} > 0$: a larger type receives larger utility and larger marginal utility from a given q. The principal's prior on the agent's type is given by $\mu_0 \in \Delta(\Theta)$. Assume that μ_0 has full support (that is, $\text{supp}(\mu_0) = \Theta$) and that

μ_0 is atomless (that is, $\mu_0(\theta) = 0$ for all $\theta \in \Theta$).[6] The initial contract w_0 is $(0, 0)$.

While this model fits many applications, we adopt an interpretation known as "selling when price and quality matter".[7] The principal is a seller who sells a good to the agent, a buyer. The contract (p, q) specifies the price p of and the quality q of the good. The seller incurs a cost of $c(q)$ when producing a good of quality q. A buyer of type θ enjoys a utility of $u(q, \theta)$ when consuming a good of quality q. Buyers of higher types enjoy a higher utility and a higher marginal utility from consuming the good.

Useful properties of the model. Denote by $q^*(\theta)$ the efficient quality for a given type θ. The efficient quality is implicitly given by

$$-c_q(q^*(\theta)) = u_q(q^*(\theta), \theta). \tag{3}$$

Given our assumptions about v and u, $q^*(\theta)$ is unique and satisfies

$$q^*_\theta(\theta) = \frac{u_{q\theta}(q^*(\theta), \theta)}{V_{qq}(q^*(\theta)) - u_{qq}(q^*(\theta), \theta)} > 0.$$

Definition 3. *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 $\mu_0(\text{supp}(\mu)) > 0$, with μ_0 being the principal's initial belief.*

[6] In particular, these assumptions rule out discrete distributions.

[7] Other interpretations include procurement from a monopolistic supplier, labor contracts, regulation of a monopolist, and financial contracts. An overview can be found in Chapter 2.15 of Laffont and Martimort (2009).

Lemma 3. *There exists a price function $p(\theta)$ such that for all types θ,*

$$u(q^*(\theta),\theta) - p(\theta) \geq u(q^*(\hat{\theta}),\theta) - p(\hat{\theta})$$

for all $\hat{\theta} \in \Theta$. This holds true for all $p(\theta)$ such that

$$p_\theta(\theta) = u_q(q^*(\theta),\theta)q^*_\theta(\theta) > 0. \tag{4}$$

See Theorem 7.3 in Fudenberg and Tirole (1991) for a proof.

Lemma 4. *If every type obtains his efficient quality and the prices satisfy (4), then the principal is indifferent between all contracts; that is, $p(\theta) - c(q^*(\theta)) = k$ for some constant k.*

Lemma 4 is a direct consequence of the fact that

$$p_\theta(\theta) - c_q(q^*(\theta))q^*_\theta(\theta) = (u_q q^*(\theta),\theta) - c_q(q^*(\theta)))q^*_\theta(\theta) = 0,$$

by equations (3) and (4).

A simple example and intuition for the main result. In this section, we outline our main result using a simplified example and provide informal arguments to build intuition. All claims will be formalized after this section.

Let $V(p,q) = p - 0.5q^2$ and $U(p,q,\theta) = \theta q - p$ with $u(q,\theta) = \theta q$. Let $\Theta = [1,2]$, and let μ_0 denote the uniform measure on $[1,2]$. In this case, the efficient quality for each type θ is given by $q^*(\theta) = \theta$.

Optimal mechanism without renegotiation. Consider the well-known optimal mechanism without renegotiation. The principal uses a direct mechanism with $\mathcal{Z} = \Theta$ and $\varphi(\theta) = (\theta^2 - 1, 2(\theta - 1))$. Without renegotiation, it is optimal for the agent to report his true type. Call this mechanism the *second-best mechanism* and the resulting outcomes the *second-best outcomes*. In the second-best mechanism, the agent of type $\theta = 2$ obtains a contract with his efficient quality. Every other type of agent $\theta \in [1,2)$ obtains a contract with a quality lower than his efficient quality. The distortion for the lower types reduces the information rent for the higher types. This result is often called "no distortion at the top". The utility of the principal is given by

$$\int_1^2 \theta^2 - 1 - 2(\theta - 1)^2 \,\mathrm{d}\theta = \frac{2}{3}.$$

The agent reports his type truthfully, and the outcomes of this mechanism are fully separating but inefficient outcomes of the form $((\theta^2 - 1, 2(\theta - 1)), \mu^\theta)$.

Second-best outcomes are not stable. There is no stable set that contains separating but inefficient outcomes. For any type $\theta < 2$, the separating but inefficient second-best outcome can be improved by a single outcome yielding the efficient quality for type θ at a higher price. For example, the outcome $((\theta^2 - 1 + u(\theta, \theta) - u(2(\theta - 1), \theta)), \mu^\theta)$ transfers all the surplus from the improved quality to the principal, and the agent is indifferent between this improvement and the second-best outcome. By Lemma 2, $((\theta^2 - 1, 2(\theta - 1)), \mu^\theta)$ cannot be stable.

Pooling outcomes can be stable. By Lemma 1, all ex-post efficient and separating outcomes are in every stable set. Lemma 2 implies that inefficient and separating outcomes cannot be stable. Thus, the only outcomes that could improve the principal's utility relative to an efficient and separating outcome are pooling outcomes, in which a positive measure of agent types end up with the same contract. Pooling outcomes can be stable if they are not dominated by any efficient and separating outcomes (all of which, as argued above, are in all stable sets). In particular, this implies that if the principal does not want to implement efficient and separating outcomes, she has to restrict information revelation by the agent.

Stable pooling outcomes have two properties. First, the support of the principal's belief in a stable pooling outcome needs to be connected. Otherwise, the principal would be better off with a set of efficient and separating outcomes violating internal stability. To see this, consider an outcome $o = ((p, q), \mu)$ such that the support of μ is given by two disjoint intervals $[1, 1.25]$ and $[1.75, 2]$. We will construct a set of separating and efficient outcomes that make the principal strictly better off than with the outcome o. Using Lemma 3 we can find a price function that makes it optimal for each agent type to end up in an outcome with his efficient quality. However, as we do not have to worry about providing the right prices for types in the interval $(1.25, 1.75)$, this price function can have a jump, and the principal can charge a price premium to the types in $[1.75, 2]$ equal to the difference between $u(1.25, 1.75)$ and $u(1.75, 1.75)$. This premium makes the principal strictly better off.

Second, one of the types in the support of the principal's belief needs to receive his efficient quality. If, instead, all types received a contract with a lower (higher) quality than the efficient one, then slightly increasing (decreasing) the quality and charging a higher (lower) price would constitute an improvement both for the principal and for all types in the pool. Lemma 2 implies that such an outcome cannot be stable.

Optimal stable outcomes. Using the two properties stated above, we can derive the pooling outcomes that maximize the principal's profits. We call the resulting outcomes *third-best outcomes*. Figure 1 depicts the optimal pooling outcomes for our example.

Observe first that it is the lowest type in each pool who receives his efficient quality, a property we call "no distortion at the bottom". Note that the binding participation or incentive constraint in any pool is that of the lowest type. Thus, to maximize profits within a pool, it is optimal to offer the efficient quality to the lowest type. Moreover, if the lowest type in each pool receives his efficient quality, then lower pools are less attractive to agents in higher pools than they would be if a higher type received his efficient quality. Thus,

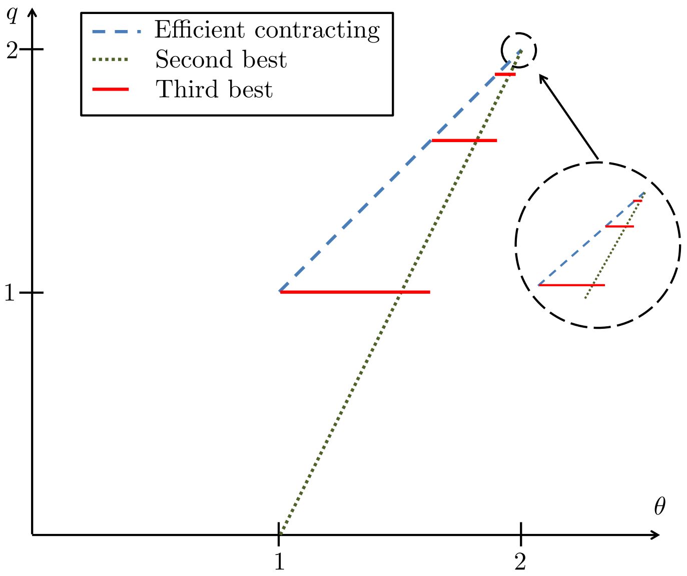

Figure 1. Comparison of the efficient quality levels (blue) with the quality levels under the second-best solution with commitment (green) and the third-best solution without commitment (red).

the incentive constraints of types in higher pools are relaxed. Put differently, if any type other than the lowest type in the pool received his efficient quality, then lower types would be distorted in the wrong direction, and the principal would need to increase the information rent to the types in higher pools.

Next, observe that the number of pooling outcomes is countably infinite. To see this, suppose that there is a "final" pool. By construction, the lowest type in this pool receives his efficient quality. In such a situation, the principal could benefit by splitting this outcome into two pooling outcomes: one with the quality from the original "final" pooling outcome, and one with a higher quality. Then the lower types would still receive the original quality, while the higher types would prefer the quality in the additional outcome, which means the principal could collect higher payments from them. As there are no higher types to consider, such a contract does not distort the incentives of other types.

In the section below we solve the principal's optimization problem by finding the optimal pooling outcomes. For the current example, approximating the solution with six pooling outcomes yields a utility of 0.5773 for the principal. The pools resulting from the approximation are shown in Figure 1.

The main result. To solve the principal's problem, we need to construct a stable set Ω and a mechanism $(\mathcal{Z}, \varphi)$ that yields outcomes that are in Ω. We then need to prove that the stable set and mechanism that we have constructed maximize the principal's utility among all stable sets and all mechanisms that produce outcomes in those stable sets.

Our main result is that there exist such a stable set Ω and a corresponding optimal stable mechanism. The utility-maximizing stable set Ω consists of all efficient and separating outcomes together with a countably infinite set of pooling outcomes. The beliefs in the pooling outcomes partition the type space into infinitely many connected intervals $[\theta^*_n, \theta^*_{n+1})$ with $\theta^*_{n+1} > \theta^*_n$ for all $n \in \mathbb{N}$. In each of the pooling outcomes, the lowest type receives his efficient quality $q^*_n = q^*(\theta^*_n)$. The prices p^*_n are given by incentive compatibility constraints such that type θ^*_n is indifferent between the contracts (p^*_n, q^*_n) and (p^*_{n-1}, q^*_{n-1}). We denote the set of optimal pooling outcomes as $\{o^*(n) = ((p^*_n, q^*_n), \mu^*_n) \mid n \in \mathbb{N}\}$.

The optimal mechanism is to offer $\mathcal{Z} = \mathbb{N}$ and $\varphi(n) = (p^*_n, q^*_n)$. The strategy of the agent of type θ is to choose message n if $\theta \in [\theta^*_n, \theta^*_{n+1})$. The mechanism and the choices of the agent lead to stable outcomes with respect to Ω.

Proposition 1. *The optimal stable mechanism exists. The profit-maximizing stable set Ω is given by the union of the set of all efficient and separating outcomes $\left\{((p, q^*(\theta)), \mu^\theta) : \theta \in \Theta, p \in \mathbb{R}\right\}$ and a countably infinite set of outcomes $\{o^*(n) = ((p^*_n, q^*_n), \mu^*_n) \mid n \in \mathbb{N}\}$ with the following properties:*

(1) Pooling: *All $o^*(n) = ((p^*_n, q^*_n), \mu^*_n)$ are pooling outcomes with $\mu^*_n(\cdot) = \mu_0(\cdot \mid [\theta^*_n, \theta^*_{n+1}))$.*

(2) No distortion at the bottom: *All $o^*(n)$ provide the efficient quality to the lowest type in the pool; that is, $q_n^* = q^*(\theta_n^*)$.*

(3) Incentive compatibility: *The prices are given by*

$$p_n^* = \sum_{m=1}^n u(q^*(\theta_m^*), \theta_m^*) - u(q^*(\theta_{m-1}^*), \theta_m^*) + \underline{\theta},\ p_0^* = u(q^*(\underline{\theta}), \underline{\theta}).$$

The optimal $\{\theta_n^*\}_{n\in\mathbb{N}}$ are the solution to the following maximization problem:

$$\begin{aligned}\max_{\{\theta_n\}_{n\in\mathbb{N}}} \quad & \sum_{n\in\mathbb{N}} \Bigg[\left(\sum_{m=1}^n u(q^*(\theta_m), \theta_m) - u(q^*(\theta_{m-1}), \theta_m) + \underline{\theta} \right) \\ & \qquad - c(q^*(\theta_n))\mu_0([\theta_n, \theta_{n+1})) \Bigg] \\ s.t. \quad & \theta_{n+1} > \theta_n,\ \theta_0 = \underline{\theta},\ \text{and}\ \theta_n < \bar{\theta}.\end{aligned} \tag{5}$$

The optimal stable mechanism is given by $\mathcal{Z} = \mathbb{N}$ and $\varphi(n) = (p_n^*, q_n^*)$.

The proof of Proposition 1 proceeds by a series of lemmata. Before we turn to the formal proofs we provide an outline of the argument.

We first show that any outcome is weakly dominated by an outcome measure yielding only efficient (and separating) outcomes (Lemma 5). Then we demonstrate that the following outcomes *cannot* be part of any stable Ω:

(1) Inefficient and separating outcomes (Lemma 6).

(2) Outcomes in which the support of the principal's belief has a gap (Lemma 7).

(3) Outcomes such that none of the types in the support of the principal's belief receives his efficient quality (Lemma 8).

Thus, every stable set must consist of a combination of (1) efficient outcomes and (2) pooling outcomes with connected support, such that one of the types in the pool receives his efficient quality.

Next, we prove that among these outcomes the principal's profit-maximizing set of outcomes consists solely of pooling outcomes. This set is countably infinite, and in each outcome, the lowest type in the support of the principal's belief receives his efficient quality (Lemma 9).

We then demonstrate that the $\{\theta_n^*\}_{n\in\mathbb{N}}$ that define the optimal contracts are the solution to the maximization problem (5) and that this solution exists (Lemma 10). Finally, we show that the set resulting from the union of all efficient outcomes and the profit-maximizing outcomes is internally and externally stable (Lemma 11).

Lemma 5. *For any outcome $o = ((p, q), \mu)$ there exists an outcome measure γ_o with only efficient and separating outcomes $\mathcal{O}(\gamma_o) = \{(p(\theta), q(\theta)), \mu^\theta) : \theta \in supp(\mu)\}$, such that $(\gamma_o, \mathcal{O}(\gamma_o))$ dominates o.*

Lemma 5 together with Lemma 1 implies that the set of all separating and efficient outcomes is internally and externally stable. As a consequence, to construct an optimal stable set, we can add outcomes to the set of efficient and separating outcomes while considering only internal stability. Put differently, we can add outcomes o to the stable set such that the principal is indifferent between o and changing o to an efficient and separating outcome. The following set of lemmata establishes which outcomes we may consider adding.

Lemma 6. *If $q \neq q^*(\theta)$, a separating outcome $((p, q), \mu^\theta)$ cannot be an element of any stable set Ω.*

Lemma 6 implies that to find an optimal stable set we need to consider pooling outcomes. The next two lemmata establish which pooling outcomes can be part of a stable set in general.

Lemma 7. *Let $o = ((p, q), \mu)$ be a pooling outcome. If* $\text{supp}(\mu)$ *is not connected, o cannot be an element of any stable set.*

Lemma 5, Lemma 6, and Lemma 7 taken together illustrate why models with discrete type spaces, as in Strulovici (2017), lead to efficient outcomes:

in a discrete type space the support of μ cannot be connected. Thus, stable outcomes need to be separating, and this is only possible if they are efficient. In the following, we refer to outcomes such that $\operatorname{supp}(\mu)$ is connected as *connected pooling outcomes.*

Lemma 8. *Let $o = ((p, q), \mu)$ be a connected pooling outcome. If $q \neq q^*(\theta)$ for all $\theta \in \operatorname{supp}(\mu)$, o cannot be an element of any stable set.*

At this point, we have shown that stable outcomes must be either efficient and separating, or pooling and connected, with the property that one of the agent types in the support obtains his efficient quality. We now show that among these outcomes, the profit-maximizing set of outcomes takes the form described in Proposition 1.

Lemma 9. *Among outcomes that are either efficient and separating, or pooling and connected with the property that one of the agent types in the support obtains his efficient quality, the profit-maximizing set of outcomes $\{o^*(n)\}_{n\in\mathbb{N}}$ is countably infinite and has the following properties:*

(1) Pooling: *All $o^*(n) = ((p^*_n, q^*_n), \mu^*_n)$ are pooling outcomes.*
(2) No distortion at the bottom: *All $o^*(n)$ provide the efficient quality to the lowest type in the pool.*

The proof of Lemma 9 proceeds in five steps. We first show that in the set of optimal outcomes, higher types receive higher quality. We then demonstrate that the principal optimally makes a higher profit from higher agent types. Next we argue that efficient and separating outcomes are not optimal for the principal. We show that in the profit-maximizing set of outcomes, the lowest type in a pooling and connected outcome will receive his efficient quality. Finally, we demonstrate that the set of profit-maximizing outcomes must be countably infinite.

All of the previous results taken together imply that the principal optimally partitions the support of her initial belief into countably many intervals such that the lowest type in each interval receives his efficient quality. She then chooses prices that maximize her utility, providing each type the incentive to select into the right outcome.

Lemma 10. *The solution $\{\theta_n^*\}_{n\in\mathbb{N}}$ to the maximization problem*

$$\begin{aligned}\max_{\{\theta_n\}_{n\in\mathbb{N}}} \quad & \sum_{n\in\mathbb{N}} \Bigg[\left(\sum_{m=1}^{n} u(q^*(\theta_m),\theta_m) - u(q^*(\theta_{m-1}),\theta_m) + \underline{\theta} \right) \\ & \qquad\qquad - c(q^*(\theta_n))\mu_0([\theta_n,\theta_{n+1})) \Bigg] \\ \text{s.t.} \quad & \theta_{n+1} > \theta_n, \;\; \theta_0 = \underline{\theta}, \;\; \text{and } \theta_n < \bar{\theta}\end{aligned} \tag{6}$$

yields the optimal set of outcomes through $o^*(n) = ((p_n^*, q^*(\theta_n^*)), \mu_0(\cdot \mid [\theta_n^*, \theta_{n+1^*}))$ and $p_n^* = \sum_{m=1}^{n} u(q^*(\theta_m^*),\theta_m^*) - u(q^*(\theta_{m-1}^*),\theta_m^*) + \underline{\theta}$, $p_0^* = u(q^*(\underline{\theta}),\underline{\theta})$. In particular, the solution exists.

Finally, we show that the set comprising all of the efficient and separating outcomes and the optimal pooling outcomes is stable.

Lemma 11. *Define $\{o^*(n) = ((p_n^*, q_n^*), \mu_n^*) \mid n \in \mathbb{N}\}$ as in Proposition 1. The union $\Omega = \big\{((p, q^*(\theta)), \mu^\theta) : \theta \in \Theta, p \in \mathbb{R}\big\} \bigcup \{o^*(n) = ((p_n^*, q_n^*), \mu_n^*) \mid n \in \mathbb{N}\}$ is a stable set.*

The proof of Lemma 11 follows directly from the construction. The set of all separating and efficient outcomes is externally stable, and we have constructed the optimal pooling outcomes in such a way that the principal is indifferent between those outcomes and the efficient and separating outcomes. Thus, the union is a stable set.

5. The Coase conjecture

In this section we consider the Coase conjecture, which is a special instance of our set-up. The Coase conjecture argues that if a seller is not able to commit to not selling a durable good, she can charge at most a price equal to 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 whenever the seller's cost is greater than or equal to the buyer's lowest valuation (the no-gap case), the seller is able to charge the monopoly price even without commitment. Our approach, using stable sets, allows us to rederive this result in a simple manner.

Set-up. Consider a monopolistic seller who is selling one good to a buyer. A contract is a pair $w = (p, q)$ with $p \in \mathbb{R}$ specifying the price and $q \in \{0, 1\}$ specifying whether the good is exchanged $(q = 1)$ or not $(q = 0)$. The principal incurs a cost of c to produce the good. Thus, the seller's utility function is given by

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

The buyer has a valuation of $\theta \in \Theta = \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}_0^+$ for consuming the good, which is private information to the buyer. Thus, the buyer's utility is 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.

The initial contract w_0 is $(0, 0)$. In the following we study two cases: the gap case, in which $c < \underline{\theta}$, and the no-gap case, in which $c \geq \underline{\theta}$. In the first case, there is a gap between the seller's cost and the buyer's lowest valuation, so it is common knowledge that there will be gains from trade. In the second case, the buyer can have a valuation that makes trade inefficient.

The gap case. In what follows we show that if $c < \underline{\theta}$, the unique stable set contains only outcomes in which the good is traded. This implies that the seller will optimally charge a price equal to the buyer's lowest valuation.

Proposition 2. *Let $c < \underline{\theta}$; then the unique stable set is $\Omega = \{((p, q), \mu) : p \in \mathbb{R}$ and $q = 1\}$. An optimal stable mechanism for the seller is given by $\mathcal{Z} = 1$ and $\varphi(1) = (\underline{\theta}, 1)$.*

Only outcomes in which the good is traded are stable. Any no-trade outcome can be improved by a single outcome in which trade occurs, if the seller increases the price by an amount between her cost and the lowest valuation of the buyer in the support of the no-trade outcome. As in any stable mechanism, the good needs to be traded with probability one, so the optimal price for the seller to charge is the buyer's lowest valuation. This result is in agreement with the literature on the Coase conjecture: the monopolist competes herself down to the lowest valuation.[8]

The no-gap case. We now consider $c \geq \underline{\theta}$. We construct an internally and externally stable Ω such that in the optimal set of outcomes within this Ω, the seller can charge the monopoly price. As the seller's problem of finding an optimal stable mechanism is a more constrained version of the problem of a seller who is not restricted by stability, charging the monopoly price must be the optimal stable mechanism.

Proposition 3. *Let $p^M = \arg\max_p(\mu_0([p, \bar{\theta}]))(p - c) > c$ denote the monopoly price. Let Ω consist of the following outcomes:*

(1) All outcomes $((p, 1), \mu)$ with $\min \operatorname{supp}(\mu) \geq c$—that is, all outcomes where the good is traded and all buyer types in the support of μ are above c.

[8] See for example Fudenberg et al. (1985).

(2) All outcomes $((p, 0), \mu)$ *with* $\max \operatorname{supp}(\mu) < c$*—that is, all outcomes where the good is not traded and all buyer types in the support of* μ *are below* c*.*

(3) The outcome $((0, 0), \mu^M)$ *with* $\operatorname{supp}(\mu^M) = [\underline{\theta}, p^M)$*—that is, the outcome where the good is not traded, the price is* 0*, and the support of* μ^M *is* $[\underline{\theta}, p^M)$*.*

Then Ω *is a stable set. An optimal stable mechanism for the seller is given by* $\mathcal{Z} = \{0, 1\}$*,* $\varphi(0) = (0, 0)$*, and* $\varphi(1) = (p^M, 1)$*.*

The stable set Ω consists of all efficient outcomes and one additional outcome in which the good is not traded. The efficient outcomes are either (1) outcomes in which the good is traded and all valuations are above costs, or (2) outcomes in which the good is not traded and all valuations are below costs. In the additional outcome (3), the good is not traded and all valuations are below the monopoly price.

Observe that this defines a stable set. Firstly, all efficient outcomes are in Ω. Secondly, the only way for the seller to change the additional outcome (3) to other stable outcomes is to offer a price equal to the cost, which yields her a profit of 0. The set Ω is therefore internally stable. It is also externally stable, because for any outcome there exists a stable outcome measure with efficient outcomes that makes both the principal and the corresponding buyer types better off.

To summarize, the set of stable outcomes in Proposition 3 is constructed in such a way that the sller can propose a mechanism that leads to one of two outcomes: either the good is traded at a price p^M and the seller believes that all buyers have a valuation above p^M, or the good is not traded and the seller believes that all buyers have a valuation below p^M. If the buyer does not buy at the price p^M, the seller does not lower the price further. This is because the only stable outcomes that dominate an outcome in which the good is not traded are those in which the good is either traded at a price equal to c or not traded at all. Thus, the seller cannot profit from renegotiation.

This result is in agreement with the non-cooperative bargaining literature. Ausubel and Deneckere (1989) show that in the no-gap case, in an infinite-horizon bargaining game, the seller can sustain the monopoly price in equilibrium if the frictions go to zero.

References

Asheim, G. and T. Nilssen (1997): "Insurance monopoly and renegotiation," *Economic Theory*, 354, 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. Strausz (2001): "Contracting with imperfect commitment and the revelation principle: The single agent case," *Econometrica*, 69, 1077–1098.

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.

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.

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.

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

Gretschko, V. and A. Wambach (2017): "Contract (re-) negotiation with private and common values," *ZEW-Centre for European Economic Research Discussion Paper*.

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, 2625–66.

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.

——— (2021): "Grouping agents with persistent types," *Available at SSRN 4291728*.

Mussa, M. and S. Rosen (1978): "Monopoly and product quality," *Journal of Economic Theory*, 18, 301–317.

Myerson, R. B. (1979): "Incentive compatibility and the bargaining problem," *Econometrica*, 47, 61–74.

Neeman, Z. and G. Pavlov (2013): "Ex-post renegotiation-proof mechanism design," *Journal of Economic Theory*, 148, 473–501.

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.

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*, 807–816.

Appendix A. Omitted proofs

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

Proof of Lemma 5. Three cases are relevant. Either there exists a type $\tilde{\theta}$ in $\operatorname{supp}(\mu)$ such that $q = q^*(\tilde{\theta})$, or for $\theta' = \min \operatorname{supp}(\mu)$ it holds that $q < q^*(\theta')$, or for $\theta'' = \max \operatorname{supp}(\mu)$ it holds that $q > q^*(\theta'')$. We focus on the first case only; the other cases can be proven analogously.

Let $q = q^*(\tilde{\theta})$ for some $\tilde{\theta} \in \operatorname{supp}(\mu)$. Consider the following set of efficient and separating outcomes: $\{(p(\theta), q^*(\theta)), \mu^\theta) : \theta \in \operatorname{supp}(\mu)\}$ such that $p(\tilde{\theta}) = p$ and $p_\theta(\theta) = u_q(q^*(\theta), \theta)q^*(\theta)$. Under this set of outcomes, an agent of type $\tilde{\theta}$ will receive the initial contract, and all agent types $\theta \neq \tilde{\theta}$ will receive a contract with his efficient quality that makes him strictly better off. Thus, the agent is better off with this set of outcomes (condition (i) of Definition 1). Moreover, $p(\theta)$ satisfies equation (4). Thus, each type of agent receives his optimal contract in the set of outcomes, and this is reflected in the beliefs (condition (iii) of Definition 1). Now consider an outcome measure γ_o that assigns each type θ in the support of μ the contract $(p(\theta), q^*(\theta))$ with probability one. In this case, the marginal of γ_o on Θ is μ and $\mathcal{O}(\gamma_o) = \{(p(\theta), q(\theta)), \mu^\theta) : \theta \in \operatorname{supp}(\mu)\}$. As a consequence of

$$p_\theta(\theta) - c_q(q^*(\theta))q^*_\theta(\theta) = (u_q q^*(\theta), \theta) - c_q(q^*(\theta))q^*_\theta(\theta) = 0,$$

the principal is indifferent between all outcomes in $\mathcal{O}(\gamma_o)$. In particular, she is then indifferent between the original outcome o and any outcome in

$$\{((p(\theta), q(\theta)), \mu^\theta) : \theta \in \operatorname{supp}(\mu)\}.$$

Thus, the outcome measure γ_o makes the principal weakly better off (condition (iii) of Definition 1). Therefore, $(\gamma_o, \mathcal{O}(\gamma_o))$ dominates o.

Proof of Lemma 6. The single outcome $(p+(u(q^*(\theta), \theta)-(u(q, \theta)), q^*(\theta)), \mu^\theta)$ makes the principal strictly better off and the agent indifferent. Lemma 2 implies the result.

Proof of Lemma 7. That $\operatorname{supp}(\mu)$ is not connected implies that there exist θ' and θ'' in $\operatorname{supp}(\mu)$ with $\theta'' > \theta'$ such that $\mu((\theta', \theta'')) = 0$. Suppose that $q = q^*(\hat{\theta})$ for some $\hat{\theta} \leq \theta'$. (The case $\hat{\theta} > \theta'$ works analogously.) We show that there exists $(\gamma_o, \mathcal{O}(\gamma_o))$ such that $\mathcal{O}(\gamma_o)$ is a set of efficient and separating outcomes that strictly dominates o. Consider the menu of contracts $\{(p(\theta), q^*(\theta)) : \theta \in \operatorname{supp}(\mu)\}$ with $p_\theta = u_q q^*_\theta$. For $\theta \leq \theta'$ set the initial condition to $p(\theta') = p$. For $\theta \geq \theta''$ set the initial condition to $p(\theta'') = p + u(q^*(\theta''), \theta'') - u(q^*(\theta'), \theta'')$. Now consider an outcome measure γ_o that assigns each type θ in the support of μ the contract $(p(\theta), q^*(\theta))$. In this case, the marginal of γ_o on Θ is μ and $\mathcal{O}(\gamma_o) = \{((p(\theta), q(\theta)), \mu^\theta) : \theta \in \operatorname{supp}(\mu)\}$.

From Lemma 3 it follows that each type of agent is better off with the contract offering his efficient quality. That is, an agent of type θ is better off with $(p(\theta), q^*(\theta))$. As a consequence of

$$\begin{aligned}&[p(\theta'') - c(q^*(\theta''))] - [p(\theta') - c(q^*(\theta'))] = \\ &\qquad [u(q^*(\theta''), \theta'') - c(q^*(\theta''))] - [u(q^*(\theta'), \theta'') - c(q^*(\theta'))] > 0,\end{aligned}$$

the principal is strictly better off: she makes the same profit from all types $\theta \leq \theta'$ and strictly more profit from types $\theta \geq \theta''$.

Thus, we have constructed $(\gamma_o, \mathcal{O}(\gamma_o))$ that strictly dominates o. By Lemma 5, all efficient and separating outcomes are in every stable set. Thus, by internal stability, the initial outcome o could not have been part of any stable set.

Proof of Lemma 8. Let $\min \operatorname{supp}(\mu) = \tilde{\theta}$ and suppose $q < q^*(\tilde{\theta})$. (The case with $\max \operatorname{supp}(\mu) = \theta'$ and $q > q^*(\theta')$ works analogously.) We show that there exists a single outcome $\tilde{o}$ that makes the principal and all types in the support of μ better off. Consider the contract $(\tilde{p}, q^*(\tilde{\theta}))$ with $\tilde{p} = p + \left[u(q^*(\tilde{\theta}), \tilde{\theta}) - u(q, \tilde{\theta})\right] > p + \left[c(q^*(\tilde{\theta})) - c(q)\right]$. The principal is strictly better off with this new contract. Type $\tilde{\theta}$ is indifferent between the two contracts, and all other types in the support of μ are strictly better off. Thus, the outcome $\tilde{o} = ((\tilde{p}, q^*(\theta')), \mu)$ meets the conditions of Lemma 2 and thus o cannot be stable.

Proof of Lemma 9. The proof is divided into five steps. We show the following:

Step 1. *In the set of optimal outcomes, higher types obtain a weakly higher quality.*

Step 2. *In the set of optimal outcomes, the principal achieves a weakly higher profit from higher types.*

Step 3. *Efficient and separating outcomes are not in the set of optimal outcomes.*

Step 4. *For every outcome in the set of optimal outcomes, the lowest type in the support of the principal's belief receives his efficient quality.*

Step 5. *The set of optimal outcomes is countably infinite.*

Let $N \subset \mathbb{R}$ be an index set. Denote by $\{o^*(n) = ((p^*_n, q^*_n), \mu^*_n) : n \in N\}$ the set of profit-maximizing outcomes, among all efficient and separating outcomes and all connected outcomes in which one type receives his efficient quality; denote by $\{w^*(n) = ((p^*_n, q^*_n)) : n \in N\}$ the corresponding set of contracts. Denote by γ^* the outcome measure with marginal μ_0 and $\mathcal{O}(\gamma^*) = \{o^*(n)\}_{n \in N}$. For any outcome measure γ' with marginal μ_0 such that the set $\mathcal{O}(\gamma')$ consists of efficient and separating outcomes or connected outcomes with one type receiving his efficient quality, it holds that

$$\int_{\mathcal{W}} V(w) \mathrm{d}\gamma^*(\Theta, w) \geq \int_{\mathcal{W}} V(w) \mathrm{d}\gamma'(\Theta, w).$$

Step 1: In the set of optimal outcomes, higher types obtain higher quality. The principal maximizes among outcomes that are either efficient and separating, or pooling and connected with one of the types receiving his efficient quality. Let $\theta_2 > \theta_1$, and let q_2 and q_1 denote the quality obtained by types θ_2 and θ_1 respectively. Then three cases are relevant. The first case is that each type obtains his efficient quality: $q_2 = q^*(\theta_2)$ and $q_1 = q^*(\theta_1)$. In this case, as q^* is an increasing function, $q_2 > q_1$. The second case is that the two types obtain different pooling outcomes.[9] In this case, we again have $q_2 > q_1$, because one of the types in each pool receives his efficient quality, the pools are connected, and q^* is an increasing function. The third case is that both types obtain the same pooling outcome. In this case, $q_2 = q_1$. Summing up, it follows that if $\theta_2 > \theta_1$, $q_2 \geq q_1$; that is, higher types receive higher quality.

Step 2: In the set of optimal outcomes, the principal achieves a weakly higher profit from higher types. We will show that for any type θ there exists an $\epsilon > 0$ such that for all types $\theta' \in (\theta, \theta+\epsilon)$, the principal realizes a profit greater than or equal to her profit with type θ. Let $o = ((p, q), \mu)$ be such that $\theta \in \text{supp}(\mu)$. If $\theta < \max \text{supp}(\mu)$—that is, if o is a pooling and connected outcome and θ is not the largest type in the pool—there exists an $\epsilon > 0$ such that all types $\theta' \in (\theta, \theta + \epsilon)$ receive the same contract. Thus, the principal makes the same profit with all these types.

Assume that $\theta = \max \text{supp}(\mu)$—that is, assume that θ is the largest type in the pool or that o is an efficient and separating outcome. Then no type $\theta' > \theta$ obtains the same contract as type θ. If there exists an $\epsilon > 0$ such that almost all types $\theta' \in (\theta, \theta + \epsilon)$ receive their efficient contract, then the principal makes the same profit from all of these types (Lemma 4). If type θ also obtains his efficient quality, the principal makes the same profit from θ and any $\theta' \in (\theta, \theta + \epsilon)$. If type θ does not obtain his efficient quality, then the principal makes a strictly larger profit from any type $\theta' \in (\theta, \theta + \epsilon)$.

Thus, assume that for every $\epsilon > 0$ the mass of types in $(\theta, \theta + \epsilon)$ who do not receive their efficient contract is of positive measure. In this case there exists an $\epsilon > 0$ such that all types in $(\theta, \theta + \epsilon)$ are in the same pooling outcome, with a contract $\hat{w} = (\hat{p}, q^*(\hat{\theta}))$ for some $\hat{\theta}$. Call this outcome $\hat{o} = (\hat{w}, \hat{\mu})$. If

[9] Or one of the types obtains his efficient quality and the other is in a pooling outcome.

the principal's profit with $(\hat{p}, q^*(\hat{\theta}))$ is greater than or equal to her profit with (p, q), we are done.

Thus, assume that the principal makes less profit with $(\hat{p}, q^*(\hat{\theta}))$. We will show that there exists a set of outcomes that makes the principal strictly better off. This is done by adding an additional contract and splitting $\hat{o}$ in two outcomes. There exists a type θ' with $\theta < \theta' < \hat{\theta}$ and a contract $w' = (p', q^*(\theta'))$ such that $U((p, q), \theta) = U((p', q^*(\theta')), \theta)$, $U((\hat{p}, q^*(\hat{\theta})), \hat{\theta}) > U((p', q^*(\theta')), \hat{\theta})$, and $V((p', q^*(\theta')) > V(\hat{p}, q^*(\hat{\theta}))$. That is, we construct a contract giving some type $\theta' \in (\theta, \hat{\theta})$ his efficient quality with type θ being indifferent between his original contract w and the new contract w'. Type $\hat{\theta}$ strictly prefers his original contract $\hat{w}$, and the principal makes a higher profit from the new contract w'.

By construction, there exists a type $\theta'' \in (\theta, \hat{\theta})$ such that all types between θ and θ'' prefer w' to w and all types between θ'' and $\max \operatorname{supp}(\hat{\mu}) = \tilde{\theta}$ prefer $\hat{w}$. Set $\mu'(\cdot) = \mu_0(\cdot \mid (\theta, \theta''])$ and $\mu''(\cdot) = \mu_0(\cdot \mid (\theta'', \tilde{\theta}))$. Denote by $o' = (w', \mu')$ and and by $o'' = (\hat{w}, \mu'')$. Thus, there exists an outcome measure γ' with marginal μ_0 and $\mathcal{O}(\gamma') = (\{o^*(n)\}_{n \in N} \setminus \{\hat{o}\}) \cup \{o', o''\}$. It follows that

$$\int_{\mathcal{W}} V(w) \mathrm{d}\gamma'(\Theta, w) > \int_{\mathcal{W}} V(w) \mathrm{d}\gamma^*(\Theta, w).$$

This contradicts the assumption that $\{o^*(n)\}_{n \in N}$ is the optimal set of outcomes.

Step 3: Efficient and separating outcomes are not in the set of optimal outcomes. The idea of this proof is the following. If the principal gives each type of agent in some interval his efficient quality, she does not earn any additional rent from higher types. If the principal instead offers a pooling outcome for some of the types in the interval, the information rent to all higher types is reduced. We will show that there is no interval $[\theta', \theta''] \subset \Theta$ with $\mu_0 ([\theta', \theta'']) > 0$

such that all types $\theta \in [\theta', \theta'']$ obtain their efficient quality $q^*(\theta)$. To obtain a contradiction, suppose such an interval exists. We will find a different set of outcomes and a corresponding outcome measure under which the principal is strictly better off, yielding a contradiction. To this end, we pool the types at the lower end of the interval into a pooling outcome and thereby can increase the price for all higher types holding the quality constant.

Take some type $\hat{\theta}$ strictly in the interior of $[\theta', \theta'']$. That is, pick $\hat{\theta}$ such that there exists an $\epsilon > 0$ such that $\theta' + \epsilon < \hat{\theta} < \theta'' - \epsilon$. We construct a new pooling outcome for types in $[\theta', \hat{\theta})$ as follows: denote by $\{o(\theta)\}_{\theta \in [\theta', \hat{\theta})} \subseteq \{o^*(n)\}$ the set of outcomes that contains all types $\theta \in [\theta', \hat{\theta})$. Define $\hat{p}$ through $u(q^*(\theta'), \hat{\theta}) - p(\theta') = u(q^*(\hat{\theta}), \hat{\theta}) - \hat{p}$. As the outcomes in $\{o(\theta)\}_{\theta \in [\theta', \hat{\theta})}$ are separating and $\hat{p}$ ignores the incentive constraints of all types between θ' and $\hat{\theta}$, it holds that $\hat{p} > p(\hat{\theta})$. Let $\Delta p = \hat{p} - p(\hat{\theta}) > 0$. Take $\hat{N} \subset N$ such that if $n \in \hat{N}$, then $\theta \geq \hat{\theta}$ for all θ in the support of μ_n^*. For $n \in \hat{N}$, consider the contracts $\hat{w}(n) = (p_n^* + \Delta p, q^*(n))$ and the set of outcomes $(\{o^*(n)\}_{n \in N} \setminus \{o(\theta)\}_{\theta \in [\theta', \hat{\theta})} \setminus \{o^*(n)\}_{n \in \hat{N}}) \cup (w^*(\theta'), \mu_1) \bigcup_{n \in \hat{N}} (\hat{w}(n), \mu^*(n))$ with $\text{supp}(\mu_1) = [\theta', \hat{\theta})$. That is, the set comprises the original outcomes for types below θ', the pooling outcome $(w^*(\theta'), \mu_1)$ for types in $[\theta', \hat{\theta})$, and the original outcome with the price increased by the constant Δp for all other types. By construction, each type prefers the contract assigned to him in this new set of outcomes. The principal is at least as well off with the pooling outcome $(w^*(\theta'), \mu_1)$ as with the outcomes in $\{o(\theta)\}_{\theta \in [\theta', \hat{\theta})}$ and strictly better off in all outcomes $(\hat{w}(n), \mu^*(n))$.

Thus, there exists an outcome measure γ' with marginal μ_0 and $\mathcal{O}(\gamma') = (\{o^*(n)\}_{n \in N} \setminus \{o(\theta)\}_{\theta \in [\theta', \hat{\theta})} \setminus \{o^*(n)\}_{n \in \hat{N}}) \cup (w^*(\theta'), \mu_1) \bigcup_{n \in \hat{N}} (\hat{w}(n), \mu^*(n))$. It follows that

$$\int_{\mathcal{W}} V(w) \mathrm{d}\gamma'(\Theta, w) > \int_{\mathcal{W}} V(w) \mathrm{d}\gamma^*(\Theta, w).$$

This contradicts the assumption that $\{o^*(n)\}_{n\in N}$ is the optimal set of outcomes.

Step 4: For every outcome in the set of optimal outcomes, the lowest type in the support of the principal's belief receives his efficient quality. The idea of the proof is the following. If a pooling outcome contains a contract where q is optimal for some intermediate type, the lower types are distorted in the wrong direction: they obtain too much quality. The principal thus prefers to give some of those types a contract with a lower quality. Suppose for the sake of contradiction that there exists $o^*(\hat{n}) = (w^*(\hat{n}), \mu^*_{\hat{n}}) = ((p^*_n, q^*_n), \mu^*_{\hat{n}}) \in \{o^*(n)\}$ such that $\min \operatorname{supp}(\mu^*_{\hat{n}}) = \theta$ but $q^*_n > q^*(\theta)$. By step 3, all outcomes in $\{o^*(n)\}$ are connected pooling outcomes.

Consider the contract $(p', q^*(\theta))$ with $p' = u(q^*(\theta), \theta) - u(q, \theta) + p$ and the set of outcomes $(\{o(n)^*\}_{n\in N} \setminus o^*(\hat{n}) \setminus o^*(\hat{n}+1)) \cup o' \cup o''$ with $o' = ((p', q^*(\theta)), \mu')$ and $o'' = (w^*(\hat{n}+1), \mu'')$. That is, construct a set of outcomes such that all contracts which are in the new set of outcomes are the same as in $\{o^*(n)\}$, except that (p^*_n, q^*_n) is swapped for $(p', q^*(\theta))$. Note that the principal makes a higher profit with this contract. By construction, all types below θ still prefer their old contract. Moreover, there exists a $\theta' > \theta$ such that all types in $[\theta, \theta') \subseteq \operatorname{supp}(\mu^*_{\hat{n}})$ prefer $(p', q^*(\theta))$ and all types in $[\theta', \max \operatorname{supp}(\mu^*_{\hat{n}})]$ prefer the contract $w^*(\hat{n}) + 1$. All types larger than $\max \operatorname{supp}(\mu^*)$ prefer their original contract. Thus, o'' is a connected pooling outcome. The principal makes a strictly higher profit from types in $[\theta, \theta')$. Steps 1 and 2 imply that the principal makes at least as much profit from types in $[\theta', \max \operatorname{supp}(\mu^*_n)]$ as if the set of outcomes were $\{o^*(n)\}$. Thus, there exists an outcome measure γ' with marginal μ_0 and $\mathcal{O}(\gamma') = (\{o(n)^*\}_{n\in N} \setminus o^*(\hat{n}) \setminus o^*(\hat{n}+1)) \cup o' \cup o''$. It follows that

$$\int_{\mathcal{W}} V(w) \mathrm{d}\gamma'(\Theta, w) > \int_{\mathcal{W}} V(w) \mathrm{d}\gamma^*(\Theta, w).$$

This contradicts the assumption that $\{o^*(n)\}_{n\in N}$ is the optimal set of outcomes.

Step 5: The set of optimal outcomes is countably infinite. So far we have shown that the optimal set of outcomes partitions $\left[\underline{\theta},\overline{\theta}\right]$ into connected intervals of strictly positive measure. This implies that if the set of optimal outcomes is infinite it must be countable ($N \subset \mathbb{N}$). Thus, we merely need to show that the set of optimal outcomes is not finite. Suppose to the contrary that there exist a $\theta \in \Theta = [\underline{\theta},\bar{\theta}]$ and $o = ((p_n^*, q^*(\theta)), \mu_n^*) \in \{o^*(n)\}_{n\in N}$ such that $\text{supp}(\mu_n^*) = [\theta,\bar{\theta}]$. Now take any $\theta' \in (\theta,\bar{\theta})$, consider the contract $w' = (p', q^*(\theta'))$ with $p' = u(q^*(\theta'),\theta') - u(q^*(\theta),\theta') + p$, and construct a new set of outcomes $(\{o^*(n)\}_{n\in N} \setminus\{o\}) \cup \{o_1, o_2\}$ with $o_1 = ((p, q^*(\theta)), \mu_1)$ and $o_2 = ((p', q^*(\theta')), \mu_2)$. Types in $[\theta,\theta')$ prefer $(p, q^*(\theta))$ and types in $[\theta',\bar{\theta}]$ prefer $(p', q^*(\theta'))$. This defines μ_1 and μ_2. The principal obtains a strictly higher profit from $(p', q^*(\theta'))$. Thus, there exists an outcome measure γ' with marginal μ_0 and $\mathcal{O}(\gamma') = (\{o^*(n)\}_{n\in N} \setminus\{o\}) \cup \{o_1, o_2\}$. It follows that

$$\int_{\mathcal{W}} V(w)\mathrm{d}\gamma'(\Theta,w) > \int_{\mathcal{W}} V(w)\mathrm{d}\gamma^*(\Theta,w).$$

This contradicts the assumption that $\{o^*(n)\}_{n\in N}$ is the optimal set of outcomes.

Proof of Lemma 10. Given the previous results, the principal's optimization problem becomes

$$\begin{aligned}
&\max_{\{\theta_n,p_n\}_{n\in\mathbb{N}}} && \sum_{n\in\mathbb{N}}(p_n - c(q^*(\theta_n))\mu_0([\theta_n,\theta_{n+1}))\\
&\text{s.t.} && \theta_{n+1} > \theta_n,\ \theta_0 = \underline{\theta},\ \theta_n < \bar{\theta},\ \text{and} \qquad (7)\\
& && u(q^*(\theta_{n+1}),\theta_{n+1}) - p_{n+1} \geq u(q^*(\theta_n),\theta_{n+1}) - p_n.
\end{aligned}$$

Optimally, the incentive constraints $u(q^*(\theta_{n+1}), \theta_{n+1}) - p_{n+1} \geq u(q^*(\theta_n), \theta_{n+1}) - p_n$ are binding. Thus, we can use the incentive constraints to back out the optimal prices. That is, setting $p_0 = u(q^*(\underline{\theta}), \underline{\theta})$ and solving recursively yields $p_n^* = \sum_{m=1}^n u(q^*(\theta_m^*), \theta_m^*) - u(q^*(\theta_{m-1}^*), \theta_m^*) + \underline{\theta}$. The principal's optimization problem is then the one stated in Lemma 10. Consider an auxiliary problem replacing the restrictions $\theta_{n+1} > \theta_n$ and $\theta_n < \bar{\theta}$ by $\theta_{n+1} \geq \theta_n$ and $\theta_n \leq \bar{\theta}$

$$\begin{aligned} \max_{\{\theta_n\}_{n\in\mathbb{N}}} \quad & \sum_{n\in\mathbb{N}} \Bigg[\left(\sum_{m=1}^{n} u(q^*(\theta_m), \theta_m) - u(q^*(\theta_{m-1}), \theta_m) + \underline{\theta} \right) \\ & \qquad - c(q^*(\theta_n))\mu_0([\theta_n, \theta_{n+1})) \Bigg] \\ \text{s.t.} \quad & \theta_{n+1} \geq \theta_n, \ \theta_0 = \underline{\theta}, \ \text{and } \theta_n \leq \bar{\theta}. \end{aligned} \tag{8}$$

Denote by $\theta^\infty \in \mathbb{R}^\infty$ a vector $(\theta_1, \theta_2, \ldots)$, let

$$\mathcal{K} = \left\{ \theta^\infty \in \mathbb{R}^\infty : \theta_{n+1} \geq \theta_n, \theta_0 = \underline{\theta}, \text{ and } \theta_z \leq \bar{\theta} \right\},$$

and let

$$\hat{V}(\theta^\infty) = \begin{cases} \sum_{n\in\mathbb{N}} [(\sum_{m=1}^n u(q^*(\theta_m), \theta_m) & \\ -u(q^*(\theta_{m-1}), \theta_m) + \underline{\theta}) - c(q^*(\theta_n))\mu_0([\theta_n, \theta_{n+1}))] & \text{if } \theta^\infty \in \mathcal{K}, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $\hat{V}$ is bounded, continuous, and positive on $\mathcal{K}$. Because $\mathcal{K}$ is compact, for any sequence θ_k^∞ that converges to some $\theta^\infty \in \mathbb{R}^\infty$ it holds that $\limsup_{k\to\infty} \hat{V}(\theta_k^\infty) \leq \hat{V}(\theta^\infty)$. Let $\alpha = \sup_{\theta^\infty \in \mathbb{R}^\infty} \hat{V}(\theta^\infty)$. Then there exists a sequence θ_k^∞ such that $\limsup_{k\to\infty} \hat{V}(\theta_k^\infty) = \alpha$. As $\hat{V}$ is continuous and positive on $\mathcal{K}$, there exists an $M > 0$ such that the level set $\left\{ \theta^\infty \in \mathbb{R}^\infty : \hat{V}(\theta^\infty) \geq M \right\} \subset \mathcal{K}$ is compact. Thus, for k sufficiently large we must have $\hat{V}(\theta_k^\infty) \geq M$, and therefore θ_k^∞ is contained in a compact set. Consequently, there exists a convergent subsequence $\theta_{k_l}^\infty$ such that $\theta_{k_l}^\infty \to \tilde{\theta}^\infty$ for some

$\tilde{\theta}^{\infty} \in \mathcal{K}$. Thus, we get $\alpha \geq \hat{V}(\tilde{\theta}^{\infty}) \geq \limsup_{k\to\infty} \hat{V}(\theta_k^{\infty}) = \alpha$. Therefore, $\tilde{\theta}^{\infty}$ is a global maximizer of the auxiliary maximization problem. However, Lemma 9 demonstrates that this solution necessarily needs to be in the interior of $\mathcal{K}$. Thus, a solution of the principal's problem exists.

Proof of Lemma 11. Firstly, we consider external stability. By Lemma 5, for any outcome o there exists an outcome measure γ_o with only efficient and separating outcomes that (weakly) dominates o. Thus, Ω is externally stable.

Secondly, we consider internal stability. For internal stability we only need to consider outcomes $o^*(n)$. However, all outcome measures $\gamma_{o^*(n)}$ with outcomes in Ω that dominate $o^*(n)$ yield efficient and separating outcomes such that θ_n^* obtains the same contract as before. From Lemma 4, it follows that the principal makes the same profit as before. Thus, internal stability is not violated.

Proof of Proposition 2. We first show that for a given Ω that is internally and externally stable, $((p,q),\mu)$ is in Ω if and only if $q = 1$.
If: For any outcome $((p,1),\mu)$, there exist no outcomes that would make the principal and all agent types in the support of μ weakly better off. It follows from Lemma 1 that $((p,1),\mu)$ must be in Ω.
Only if: Let $((p,0),\mu)$ be an outcome with $q = 0$. Consider the contract $(p + c + 0.5(\underline{\theta} - c), 1)$. The buyer strictly prefers this contract to $(p,0)$, independent of his type. Moreover the seller is strictly better off than with the contract $(p,0)$. Thus, the single outcome $\{((p + c + 1/2(\underline{\theta} - c), 1), \mu)\}$ makes both the seller and the buyer strictly better off. Lemma 2 yields the desired result.

Given that Ω is unique, the optimal stable mechanism for the seller is then to sell the good at a price equal to $\underline{\theta}$.

Proof of Proposition 3. We start by showing that the proposed Ω is externally stable. Let $o = ((p, q), \mu)$ be an outcome not in Ω. Consider the two outcomes $o^1 = ((p+(1-q)c, 1), \mu(\cdot \mid (c, \bar{\theta}]))$ and $o^2 = ((p-qc, 0), \mu(\cdot \mid [\underline{\theta}, c]))$—that is, an outcome in which the good is traded at a price of $p + (1-q)c$, and an outcome in which the good is not traded and the price is $p - qc$. All buyer types $\theta > c$ prefer the trade outcome, and all types $\theta \leq c$ prefer the no-trade outcome. The seller is indifferent between all three outcomes o, o^1, and o^2. Thus, there exists an outcome measure γ_o with marginal μ and with $\mathcal{O}(\gamma_o) = \{o^1, o^2\} \subset \Omega$ such that $(\gamma_o, \mathcal{O}(\gamma_o))$ weakly dominates o.

Having shown that Ω is externally stable, we turn our attention to internal stability. We only need to consider the outcome $o^M = ((0, 0), \mu^M)$ with $\text{supp}(\mu^M) = [\underline{\theta}, p^M]$. We show that there exists no outcome measure γ_{o^M} with outcomes $\mathcal{O}(\gamma_{o^M}) \subset \Omega$ such that $(\gamma_{o^M}, \mathcal{O}(\gamma_{o^M}))$ strictly dominates o^M.

By property (iii) of Definition 1 (which says that the buyer does not receive an suboptimal contract in $\mathcal{O}(\gamma_{o^M})$), $\mathcal{O}(\gamma_{o^M})$ consists of at most two outcomes: one where the good is traded and one where the good is not traded. The no-trade outcome cannot have a price above 0, as this would make all types of buyers strictly worse off than they are in the outcome o^M. The price in the trade outcome has to be less than or equal to c, since in all stable trade outcomes the minimum of the support of buyer types is above c. These observations imply that the seller cannot be strictly better off and that o^M is therefore stable.

When the seller offers the mechanism $\mathcal{Z} = \{0, 1\}$, $\varphi(0) = (0, 0)$, and $\varphi(1) = (p^M, 1)$, all buyer types $\theta < p^M$ choose message 0 and all other types choose message 1. This results in two stable outcomes: $((0, 0), \mu^M)$ and $((p^M, 1), \mu_0(\cdot \mid [p^M, \underline{\theta})]))$. The seller makes the same profit as in the optimal unconstrained mechanism; thus the proposed mechanism is also the optimal stable mechanism.