Optimally Coarse Contracts

Roberto Corrao* MIT

Joel P. Flynn† Yale

Karthik A. Sastry‡ Princeton

November 4, 2023

Abstract

We study a principal-agent model in which actions are imperfectly contractible and the principal chooses the extent of contractibility at a cost. If contractibility costs satisfy a monotonicity property—which is implied by costs that come from difficulties in distinguishing actions when writing the contract—then optimal contracts are necessarily *coarse*: they specify finitely many actions out of a continuum of possibilities. This result holds even if contractibility costs are arbitrarily small. Applying our results to a nonlinear pricing model, we study how changes in consumer demand, production costs, and informational asymmetries affect the optimally coarse set of quality options.

*MIT Department of Economics, 50 Memorial Drive, Cambridge, MA, 02142. Email: rcorrao@mit.edu.

†Yale University Department of Economics, 30 Hillhouse Avenue, New Haven, CT, 06511. Email: joel.flynn@yale.edu.

‡Princeton University Department of Economics, Julis Romo Rabinowitz Building, Princeton, NJ, 08544, Email: ksastry@princeton.edu.

We are grateful to Daron Acemoglu, Alessandro Bonatti, Oğuzhan Çelebi, Daniel Chen, Piotr Dworczak, Drew Fudenberg, Robert Gibbons, Zi Yang Kang, Giacomo Lanzani, Stephen Morris, Ellen Muir, Jean Tirole, Jonathan Weinstein, Alexander Wolitzky, and Kai Hao Yang for helpful comments. First posted version: November 2, 2023.

1 Introduction

Few contracts completely specify obligations for all observable, payoff-relevant outcomes. At least since Coase (1960), economists have argued that this *incompleteness* of contracts arises from costs inherent to their writing. Hart and Moore (2008) describe incompleteness of contracts as the difference between contractible actions in the "letter" of the contract and non-contractible actions in the "spirit" of the contract. In this paper, we study how costs of determining the "letter" of the contract affect optimal contract design.

To do this, we study a principal-agent model with privately informed agents in which agents' actions are imperfectly contractible and contractibility is costly for the principal. We model contractibility via a correspondence that translates a recommended action from the principal into a set of allowable actions for the agent—that is, a relationship between the "spirit" and the "letter" of the contract. Contractibility costs formalize the difficulty in distinguishing what is allowable under the "letter" of the contract from what is not.

We then analyze optimal contracts in two steps. First, we characterize implementable and optimal mechanisms for a fixed extent of contractibility. Second, we leverage this characterization to derive our main result: if contractibility has marginal costs that decline sufficiently slowly, then the principal chooses a *coarse contract* that specifies finitely many recommendations. This property of marginal costs is satisfied by a large family of costs that is based on distinguishing what is in the "letter" of the contract from what is not. Importantly, other cost functions that are motivated by costly enforcement of the contract *ex post*, as opposed to costly writing of the contract *ex ante*, do not generate the prediction of coarse contracts. Thus, in our analysis, it is the *ex ante* cost of determining the "letter" of the contract that yields coarseness.

In further results, we derive an upper bound on the optimal number of contractible outcomes as a function of the principal's payoffs, the agent's payoffs, the distribution of agents' types, and the cost of contractibility. Finally, we derive necessary conditions that describe not only how many outcomes, but also which outcomes, are optimally contractible.

We apply the model to study when and why incomplete contracts emerge in product markets, manifested as optimally designed coarse quality grades for a differentiated good or service. To do this, we study a variant of the nonlinear pricing model of Mussa and Rosen (1978) in which contracting on quality is costly. We analytically characterize the optimal qualities offered by the monopolist and show that lower production costs and greater consumer demand both lead to menus that feature fewer quality options. We also find that contracts are endogenously coarser under incomplete information about buyers' willingness to pay than under complete information.

Model. A principal contracts with an agent of an unknown type. The agent can take actions that influence the payoff of both the principal and the agent. Higher types value higher actions relatively more and all types have monotone increasing preferences over the action (*i.e.,* it is a "good").[1] The principal writes a contract that specifies payments associated with *recommendations*. Agents select a recommendation and then take a realized action which we call the *outcome*. The scope of contracts to discipline outcomes is specified by a *contractibility correspondence*, which describes all possible actions from which the agent can choose after receiving a given recommendation. Thus, the contractibility correspondence relates the spirit of the contract—the set of recommendations—to the letter of the contract—the set of actions that agents can legally take.

We impose four economic axioms on the relationship between spirit and letter, which translate to restrictions on the contractibility correspondence. The first is *reflexivity*: if the agent is called upon to do y, then y is within the letter of the contract. The second is *transitivity*: if the contract calls upon the agent to do y and they can, within the letter of y, also do x, then the set of actions consistent with doing x is a subset of the set of actions consistent with doing y. The third is *monotonicity*: if the contract recommends a higher action, then the consistent actions in the letter of the contract are also higher. The fourth is *excludability*, which allows the principal to not transact with the agent.[2] These axioms translate into natural patterns of incomplete contracting, in which the outcome space is composed of regions with perfectly contractible actions, regions that permit deviations up or down, and regions that are fully indistinguishable.

We allow the principal to select the contractibility correspondence at some cost. The cost reflects the principal's efforts in writing the contract. As a leading example, we define a class of *costs of distinguishing outcomes*. In this class, the cost of a given correspondence is the total cost, over all possible outcomes, of the inverse distance between what is within the letter of the contract and what is outside of it.

Main Results. To begin, we fix the contractibility correspondence and study how the principal optimally designs the contract. We first show that the principal can implement an outcome function, a mapping from agents' types to outcomes, if and only if it is monotone increasing and supported on a given set that depends on the contractibility correspondence. This set is the image of the action space under the maximum selection from the contractibility correspondence. Intuitively, agents prefer to take the highest possible action within the letter of the contract. The optimal outcome function maximizes virtual surplus (*i.e.,* total surplus

[1] The case in which preferences are monotone decreasing and the action is a "bad" is symmetric and our results apply.

[2] We also impose technical axioms that the correspondence is closed-valued and lower hemicontinuous.

net of information rents) subject to being supported on the given set.[3] We show that this takes a simple form: pick the best contractible action that is "close" to what the principal would pick with full contractibility.

We leverage this result to study optimal contractibility. Using our implementation result, we re-express our costs of contractibility correspondences in terms of the closed set of implementable outcomes that they induce. Under the technical condition that the cost is lower semicontinuous, the problem of optimal contractibility is well-posed: there exists a solution set, which is nonempty and compact.

To study the form of optimal contractibility, we place one additional assumption on the cost that we call *strong monotonicity*. This property is most easily understood in the context of contracting upon intervals of the action space. In this case, strong monotonicity implies that the marginal cost of introducing perfect contracting in an interval of the action space is (at most) second-order in the length of the interval. Strong monotonicity in its full form disciplines the marginal cost of not only adding intervals but also adding countably infinite sets and uncountably infinite and nowhere dense sets (*e.g.,* the Cantor set). The implicit requirement is the same: adding a small such set induces a marginal cost that converges to zero sufficiently slowly in the size of the set. While these requirements of strong monotonicity may seem specific, we show that any aforementioned cost of distinguishing outcomes is strongly monotone. Intuitively, distinguishing an interval of length t from all other outcomes moves measure t outcomes into the letter of the contract. Thus, the principal must distinguish the outcomes in this new contract, which are of total measure t, from the nearby outcomes that are now outside of the contract, which are also of measure t; yielding a second-order cost that is proportional to $t \times t = t^2$.

Our main result is that, if costs are strongly monotone, then optimal contractibility specifies a finite number of contractible actions. By implication, optimal contracts are *coarse*, or supported on a finite menu. These contracts are incomplete in a particularly strong way—they not only fail to specify *some* potentially verifiable outcomes, they in fact fail to specify *almost all of them* and leave a bounded-size gap between any two adjacent items. This result holds even when the cost of implementing the complete contract is arbitrarily low.

We prove this result by using variational arguments in the space of the closed sets of implementable outcomes that are induced by contractibility correspondences. For example, to rule out intervals of perfect contractibility, we construct a payoff-improving alternative contractibility correspondence that introduces "local incompleteness," or replaces a subset of such an interval with its two boundary points. The principal's surplus loss under the

[3] Formally, we make the standard assumptions that virtual surplus is strictly quasi-concave and strictly supermodular.

optimal contract that we previously characterized is *third-order* in the length of the interval. Intuitively, for each type that is allocated an outcome in the interior of this interval, the principal was originally maximizing the virtual surplus function—that is, for this type, the principal was unconstrained by incompleteness. Thus, there is no *first-order* cost in slightly moving the allocation, and any losses can be described by a *second-order term.* To obtain the total loss in surplus, we integrate these second-order losses over the interval of types whose allocation changes, which is also proportional to the width of the interval—thus obtaining a *third-order loss.* For a small enough interval, this will always be lower than the second-order savings in costs of contractibility, which are guaranteed under strong monotonicity. This argument rules out intervals of perfect contractibility. More technical arguments based on estimates of the value of other set-valued perturbations of the contract space rule out all other infinite sets, including the uncountable and nowhere dense sets.

Finally, we derive results that inform how coarse optimal contracts can be and which outcomes will be optimally contractible. First, we derive an analytical upper bound on the number of items on the menu or, more informally, a lower bound for the "extent of incompleteness." This bound increases in the maximum concavity of the virtual surplus function because this scales the principal's loss from moving agents' allocations; it increases in the maximum density of types and decreases in the minimum complementarity of types with actions because this scales how tightly packed the principal's preferred allocations can be in small intervals; and it increases in a parameter scaling the costs, for obvious reasons. Combined with the structure of payoffs and information rents, which themselves determine the virtual surplus function, we can use this result to gauge when contracts are "more or less incomplete." Second, we show how to determine optimal coarse contracts using simple first-order conditions that equate the marginal benefits of changing allocations on virtual surplus with the marginal costs of writing this into the contract.

Importantly, the coarseness of contracts does not stem from the presence of costly contractibility *per se.* Instead, the prediction of coarse contracts hinges on the notion that the *ex ante* writing of contracts is costly. We demonstrate this claim by showing that costs of contractibility which are natural but do not stem from a foundation of costly *ex ante* writing of contracts are not strongly monotone and do not yield coarse contracts. Concretely, we consider a setting in which writing contracts is free *ex ante* but has *ex post* enforcement costs. We can capture such a situation with an *ex post* variant of a cost of distinguishing what is allowed from what is not: instead of paying for each action described, the principal instead pays in proportion to how likely it is that a given action will be taken *ex post.* We show that such costs never yield coarse contracts alone: while these costs distort allocations, they do not affect the choice of contractibility.

Application: Monopoly Pricing with Coarse Contracts. We apply our model of optimal contractibility in the Mussa and Rosen (1978) nonlinear pricing model. This model describes a monopolist selling a service (*e.g.,* a vacation rental) that may differ in quality. The monopolist chooses both a menu of utilization levels (abstractly, "qualities") and prices, as in the standard nonlinear pricing problem. Moreover, they must write a contract that describes what levels of utilization by the buyers are acceptable. Contractibility is costly because the monopolist has to describe the acceptable levels of utilization of the good—for example, what constitutes a unit in "good" versus "bad" condition.

First, we show that the optimal contract features *uniformly* spaced qualities. Intuitively, in this quadratic-uniform setting of the Mussa and Rosen (1978) model, the monopolist's losses from coarse contracting are the same at all points in the menu. Thus, the monopolist has no incentive to make contracts more or less precise for high *vs.* low quality levels.

Second, we give a formula for the number of points in the menu (up to integer rounding) in terms of the parameters that control production costs (*i.e.,* concavity), differentiation in preferences (*i.e.,* supermodularity), and costs of contractibility. These parameters enter this formula exactly as they did in the general analysis' bound: contracts are more complete in environments with higher concavity, lower supermodularity, and lower costs of contractibility.

Finally, we study the impact of incomplete information on the optimally incomplete contract. We show that contracts are always "more complete," or contain more menu items under complete information than under incomplete information. Intuitively, adverse selection reduces the size of the pie available to the monopolist and dulls their incentives to contract more precisely. Thus, incomplete information begets more incomplete contracts in this setting.

Related Literature. Our approach to modeling incomplete contracts is inspired by the dichotomy between perfunctory performance (the letter) and consummate performance (the spirit) introduced by Hart and Moore (2008).[4] Under complete information, Hart and Moore (2008) adopt a behavioral approach to modeling contracting, in which contracts act as reference points. We retain their dichotomy between the letter and the spirit of a contract for understanding that some actions cannot be contracted upon, but follow the standard mechanism design literature in studying implementable and optimal contracts when the principal does not know the type of the agent. In general, this strand of the literature on incomplete contracts relies on the possibility of the parties renegotiating ex-post a previously specified and potentially optimally incomplete contract. For example, Segal (1999) provides a foundation of optimally incomplete contracts based on the classical renegotiation approach.

[4] In turn, this language choice is inspired by Williamson (1975).

Another strand of literature on incomplete contracts, closer to the analysis in this paper, explicitly models the complexity and the cost of writing and enforcing contracts by studying the derived trade-off for the principal between the benefits of more complete contracts and the costs of writing more complete contracts. Two notable examples are Bajari and Tadelis (2001) and Battigalli and Maggi (2002). By working in finite state and action settings, neither speaks to the issue of the endogenous coarseness of contracts. Moreover, neither of these papers considers ex-ante asymmetric information between the principal and the agent.

By incorporating incomplete contracts into principal-agent problems, our results fit into the theoretical literature on mechanism design with *ex post* moral hazard (*e.g.,* Laffont and Tirole, 1986; Carbajal and Ely, 2013; Strausz, 2017; Gershkov, Moldovanu, Strack, and Zhang, 2021; Yang, 2022). Within this literature, the most related analysis is by Grubb (2009) and Corrao, Flynn, and Sastry (2023), who study how fully non-contractible utilization (the possibility of free disposal) matters for optimal nonlinear pricing of goods. Our analysis significantly generalizes the scope of contractibility away from this fully non-contractible case. An important contrast between our approach and the standard one is that we model imperfect contractibility, while most analyses of moral hazard concern imperfect observability (with perfect contractibility). As we show, this difference in perspective leads to qualitatively different optimal mechanisms.

Finally, our work is related to models of optimal design where a continuous variable is optimally discretized as a result of a trade-off between the benefit of higher flexibility and its exogenous or endogenous costs. For example, in models of rational inattention as in Jung, Kim, Matějka, and Sims (2019) or optimal categorization as in Mohlin (2014) the designer faces an exogenously given cost of respectively refining information or labelings. In a setting closer to ours, Bergemann, Heumann, and Morris (2022) study a variant of a standard Mussa and Rosen (1978) nonlinear pricing model and show that if the monopolist can simultaneously choose the selling mechanism and the buyer's information, then both can be optimally chosen to be discrete. Differently from the previous two papers, here, the "cost" of finer information and contract is given by the information rents that the monopolist needs to guarantee to the buyer. In particular, Bergemann, Heumann, and Morris (2022) generalize results in Wilson (1989) showing that, under perfect information, coarsening the domain of contractibility into uniform cells is second-order in the length of the grid.

Outline. Section 2 introduces the model. Section 3 characterizes optimal contracts for a fixed contractibility correspondence. Section 4 studies optimal contractibility. Section 5 applies our results to study optimal contractibility in a nonlinear pricing model. Section 6 studies optimal contractibility under alternative assumptions on costs. Section 7 concludes.

2 Model

2.1 The Agent and the Principal

There is a single agent with privately known type $\theta \in \Theta = [0, 1]$. The type distribution $F \in \Delta(\Theta)$ admits a density f that is bounded away from zero on Θ. Each agent can take an action x in the interval $X = [0, \overline{x}] \subset \mathbb{R}$.

The agent's preferences are represented by a twice continuously differentiable utility function $u : X \times \Theta \to \mathbb{R}$. We assume that higher types value higher actions more and that all types have monotone preferences over actions with the following three conditions: (i) $u(\cdot)$ satisfies strict single-crossing in (x, θ); (ii) for each $x \in X$, $u(x, \cdot)$ is monotone increasing over Θ; and (iii) for each $\theta \in \Theta$, $u(\cdot, \theta)$ is strictly monotone increasing over X. The case with strictly decreasing preferences over X is analogous. All agent types value the zero action the same as their outside option payoff, which we normalize to zero, or $u(0, \theta) = 0$ for all types $\theta \in \Theta$. Agents have quasilinear preferences over actions and money $t \in \mathbb{R}$, so their transfer-inclusive payoff is $u(x, \theta) - t$.

The principal's payoff derives from three sources. The first is the sum of monetary payments $t \in \mathbb{R}$ from agents to the seller. The second is a (potentially type-dependent) payoff that derives from agents' actions, represented by a continuously differentiable $\pi : X \times \Theta \to \mathbb{R}$. We normalize $\pi(0, \theta) = 0$ for all $\theta \in \Theta$. The third is a cost of contractibility, which we will introduce in due course.

2.2 Partial Contractibility

To model the principal's inability to contract perfectly on outcomes, we define a *contractibility correspondence* $C : X \rightrightarrows X$ that maps every recommendation $y \in X$ to a feasible set of final actions that the agents can take $x \in C(y)$. In our interpretation, y embodies "the spirit of the contract" and the collection of outcomes $C(y)$ consistent with y according to C embodies "the letter of the contract." This terminology is also consistent with the following terminology from Williamson (1975) and Hart and Moore (1988): y is "consummate performance" and x is "perfunctory performance."

Regular Contractibility Correspondences. We discipline the relationship between the spirit and letter of a contract by imposing six axioms. The first four are economic in nature:

Axiom 1 (Reflexivity)**.** *For every $y \in X$, $y \in C(y)$.*

Reflexivity requires that the agent *can* undertake action y when they are called upon to take action y by the contract.

Axiom 2 (Transitivity)**.** *For every $x, y \in X$, if $x \in C(y)$, then $C(x) \subseteq C(y)$.*

Transitivity requires that, if an agent can reach action x by deviating from y and z by deviating from x, then they can reach z by deviating from y.

Axiom 3 (Monotonicity)**.** *For every $x, y \in X$, if $x \leq y$, then $C(x) \leq_{SSO} C(y)$, where $\leq_{SSO}$ denotes the strong set order.*

Monotonicity requires that, if an agent starts from being called upon to do $z \leq y$, then the set of things they can do after z is also lesser than the set of things they can do after y.

Axiom 4 (Excludability)**.** $C(0) = \{0\}$.

Excludability imposes that the principal can always exclude the agent from the contract by giving them their outside option.

The final two axioms are technical:

Axiom 5 (Closed-valuedness)**.** *For all $y \in X$, $C(y)$ is closed.*

Axiom 6 (Lower hemicontinuity)**.** *The correspondence C is lower hemicontinuous.*

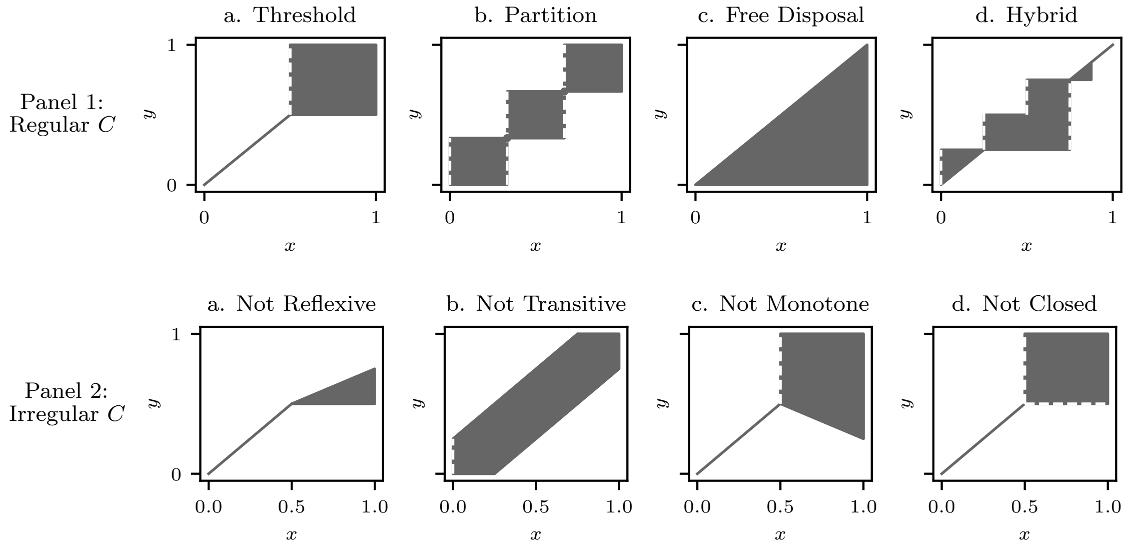
Closed-valuedness and *Lower hemicontinuity* ensure the existence of an optimal contract given any contractibility correspondence that satisfies the axioms above.

Throughout our analysis, we will study contractibility correspondences that satisfy all six axioms. We will refer to such contractibility correspondences as *regular*. We let $\mathcal{C}$ denote the set of regular contractibility correspondences.

Examples. We plot four examples of regular correspondences in Panel 1 of Figure 1. In the first regular example (1a), all $x \leq 1/2$ can be specified "perfectly" in the contract, while all $x > 1/2$ are indistinguishable: an agent recommended any action in this region can pick any other action in the region. In (1b), the action space is coarsened into four partitions of indistingushable actions. In (1c), agents have access to unrestricted *free disposal* as studied by Grubb (2009) and Corrao, Flynn, and Sastry (2023). In (1d), we combine these basic patterns into a "'hybrid."

We also show four irregular examples in the second row to better illustrate what our axioms rule out. Example (2a) is not reflexive, since the correspondence does not include the 45 degree line; (2b) is not transitive, since there are "chains" whereby an agent can reach x from y and z from x but not z from y; (2c) is not monotone, for $x > 1/2$; and (2d) is not closed, since the boundary of $C(x)$ is open for $x > 1/2$.

Figure 1: Illustrating Regular Contractibility Correspondences



Note: Each graph illustrates a contractibility correspondence for $X = [0, 1]$, with dark shading denoting the graph. The examples in Panel 1 (top row) are regular, with informative names. The examples in Panel 2 (bottom row) are not regular, respectively failing Axioms 1-3 and 5.

Representing Regular Contractibility. We now provide two characterizations of regular contractibility correspondences that clarify their economic properties. In our later analysis, these representations also turn out to be mathematically convenient.

Lemma 1 (Representations of Contractibility)**.** *Fix a contractibility correspondence C. The following statements are equivalent:*

1. *C is regular*
2. *There exist an upper semi-continuous increasing function $\underline{\delta} : X \to X$ and a lower semi-continuous increasing function $\overline{\delta} : X \to X$ such that for all $y \in X$: (i) $C(y) = [\underline{\delta}(y), \overline{\delta}(y)]$, (ii) $\underline{\delta}(y) \leq y \leq \overline{\delta}(y)$, (iii) $\underline{\delta}(x) = \underline{\delta}(y)$ for all $x \in [\underline{\delta}(y), y)$, (iv) $\overline{\delta}(x) = \overline{\delta}(y)$ for all $x \in (y, \overline{\delta}(y)]$, and (v) $\overline{\delta}(0) = 0$.*
3. *There exist two closed sets $\underline{D} \subseteq X$ and $\overline{D} \subseteq X$ such that: (i) $0 \in \underline{D}$ and $0, \overline{x} \in \overline{D}$, (ii) For all $x \in X$, we have*

$$C(x) = \left[\max_{z \leq x : z \in \underline{D}} z, \min_{z \geq x : z \in \overline{D}} z \right] \tag{1}$$

In this case, we have $\overline{D} = \overline{\delta}(X)$, $\underline{D} = \underline{\delta}(X)$. Moreover, given C, $(\underline{\delta}, \overline{\delta})$ and $(\underline{D}, \overline{D})$ are unique, and vice versa.

Proof. See Appendix A.1. □

The first characterization (Part 2) is in terms of the upper and lower envelope of the correspondence, $\overline{\delta}(y) = \max\{x \in C(y)\}$ and $\underline{\delta}(y) = \min\{x \in C(y)\}$. The first two properties of its definition ensure that $C(y)$ is a closed and convex interval including y. Properties three and four are most easily understood via the graphical illustrations of Figure 1: upper and lower boundaries of the graph $C(X)$, if they deviate from the identity line, must be flat. In this sense, imperfect contractibility in our model always presents as "disposal" ("lower triangles"), "creation" ("upper triangles"), or complete indistiguishability ("boxes").

The second alternate characterization (Part 3) is in terms of the images of these functions, which are equal to the sets of fixed points of these functions: $\underline{D} = \underline{\delta}(X) \subseteq X$ and $\overline{D} = \overline{\delta}(X) \subseteq X$. These correspond to the recommendations that an agent with monotone decreasing or increasing preferences (respectively) would follow.

2.3 Costly Contractibility

To achieve a specific level of contractibility, the principal pays a cost. This cost formalizes the difficulty that the principal faces in writing a contract with more elaborate contingencies. Our primary interpretation is that the cost is borne *ex ante*, for instance in the process of writing a contract with more descriptive language or even understanding how to express the relevant outcomes. However, the cost may reflect the expectation of a cost borne *ex post*, for instance in litigation. We express these costs via a function $\Gamma : \mathcal{C} \to [0, \infty]$. For now, we place no economic restrictions on this cost. Later, restrictions on the cost will be key for our main result about optimally incomplete contracts.

To make these costs concrete and to make clear the core economics that we wish to study, we now introduce a class of cost functionals based on the idea that writing contracts is costly because the principal must distinguish what is within the letter of the contract and what is outside of it. Consider a principal writing a contract that describes rights and obligations under a variety of "scenarios." In our formalism, each scenario is labeled by a recommendation x, the obligations by a monetary transfer, and a description of the rights embodied by $C(x)$. An important challenge for the principal is to differentiate the rights under x, $C(x)$, from the actions *outside* of the agent's rights in the same scenario, $X \setminus C(x)$. We embody this idea by assuming that the cost of distinguishing $C(x)$ from $X \setminus C(x)$ is equal to some decreasing function of the distance between $C(x)$ and $X \setminus C(x)$. Formally, we define such a cost of distinguishing as follows:

Definition 1 (Costs of Distinguishing Outcomes)**.** *Define the inverse distance between $C(x)$ and $X \setminus C(x)$ as:*

$$\hat{d}(C(x), X \setminus C(x)) = \int_{X \setminus C(x)} \min_{z \in C(x)} \tilde{d}(z, y) \, \mathrm{d}y \tag{2}$$

where $\tilde{d} = h \circ d$, $h : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuously differentiable, strictly decreasing function that is strictly positive on $\mathbb{R}_{++}$, and $d : X \times X \to \mathbb{R}_+$ is a continuously differentiable distance function, except potentially on the set $\{(x,x) : x \in X\}$. The cost of distinguishing outcomes is given by the total inverse distance over all possible outcomes:

$$\Gamma(C) = \int_X \hat{d}(C(x), X \setminus C(x))\,\mathrm{d}x \tag{3}$$

As this definition is somewhat abstract, we give some specific examples:.

Example 1 (Discrete and $p-$Distance Costs of Distinguishing)**.** The special case of the discrete metric, $d(z,y) = \mathbb{I}[y \neq z]$, is particularly natural. This cost is equal to the total Lebesgue measure over $x \in X$ of all points $y \in X \setminus C(x)$ that are distinguished from $C(x)$:

$$\Gamma(C) = \int_X \mu(X \setminus C(x))\,\mathrm{d}x = \int_0^{\overline{x}} \left(\overline{x} - \overline{\delta}(x)\right)\mathrm{d}x + \int_0^{\overline{x}} \underline{\delta}(x)\,\mathrm{d}x \tag{4}$$

where μ is the Lebesgue measure. Geometrically, in this case, the cost equals the area lying above the graph of $\overline{\delta}$ and below the graph of $\underline{\delta}$. Observe that this cost is 0 for the zero-contractibility correspondence $C(x) = [0, \overline{x}]$ and it is equal to its maximum of $\overline{x}^2$ for the perfect contractibility correspondence $C(x) = \{x\}$. Alternative distances, such as the family of $p-$distances, $\tilde{d}(z,y) = (z-y)^{-\frac{1}{1+p}}$ for $p \in (0,\infty)$, allow for the cost to depend on how many nearby actions are distinguished from each other. The discrete cost is nested in the family of $p-$distances as the $p \to \infty$ limit. △

In Section 6, we will give several other classes of cost functional based on notions of costly enforcement, costly clauses, and menu costs.

2.4 The Principal's Problem

We now state the principal's mechanism and contractibility design problem. Given the revelation principle, we consider direct and truthful mechanisms and restrict attention to deterministic mechanisms. Thus, a mechanism is a triple (ϕ, ξ, T) comprising a recommendation $\xi : \Theta \to X$, a final action or outcome $\phi : \Theta \to X$, and a tariff $T : X \to \overline{\mathbb{R}}$. The tariff and the recommendation jointly determine the transfer between the principal and the agent $T(\xi(\theta))$. The final action is then taken by the agent and must lie within the contractibility correspondence $\phi(\theta) \in C(\xi(\theta))$. Principal and agent payoffs both depend on the final action $\phi(\theta)$ and the monetary transfer $T(\xi(\theta))$. We now define what it means for a mechanism to be implementable:

Definition 2 (Implementable Mechanism)**.** *A mechanism* (ϕ, ξ, T) *is implementable given contractibility* C *if and only if the following three conditions are satisfied:*

1. *Obedience:*
$$\phi(\theta) \in \arg\max_{x \in C(\xi(\theta))} u(x, \theta) \quad \textit{for all} \;\; \theta \in \Theta \quad \textit{(O)} \tag{5}$$

2. *Incentive Compatibility:*
$$\xi(\theta) \in \arg\max_{y \in X} \left\{ \max_{x \in C(y)} u(x, \theta) - T(y) \right\} \quad \textit{for all} \;\; \theta \in \Theta \quad \textit{(IC)} \tag{6}$$

3. *Individual Rationality:*
$$u(\phi(\theta), \theta) - T(\xi(\theta)) \geq 0 \quad \textit{for all} \;\; \theta \in \Theta \quad \textit{(IR)} \tag{7}$$

We let $\mathcal{I}(C)$ *denote the set of implementable mechanisms under* C*.*

Obedience requires that each agent θ chooses an optimal final action $\phi(\theta)$ by optimally exploiting what is possible under the contract given the initial recommendation $\xi(\theta)$, *i.e.,* they choose a favorite element from $C(\xi(\theta))$.[5] Incentive Compatibility ensures that the agent wishes to actually perform the initial action $\xi(\theta)$ required by the mechanism, taking into account both the transfer they pay and their subsequent ability to optimize their final action within the scope described by the contract. Individual Rationality ensures that all agents are willing to participate in the mechanism.

Conditional on a level of contractibility C, the principal maximizes the sum of transfers and payoffs arising from agents' final actions or solves

$$\mathcal{J}(C) := \sup_{(\phi, \xi, T) \in \mathcal{I}(C)} \int_{\Theta} \left(\pi(\phi(\theta), \theta) + T(\xi(\theta)) \right) \mathrm{d}F(\theta) \tag{8}$$

We refer to a maximizer (ϕ, ξ, T), if it exists, as an *optimal contract* given C.

The principal's full problem encompasses the aforementioned inner problem and the choice of contractibility. The principal chooses contractibility $C \in \mathcal{C}$ to maximize expected surplus net of costs, or

$$\sup_{C \in \mathcal{C}} \quad \mathcal{J}(C) - \Gamma(C) \tag{9}$$

As this representation makes clear, designing "contractibility" and designing "the contract" are tightly linked, since the former determines what is implementable in the latter problem.

[5] We use the word "obedience" in the sense of Myerson (1982).

3 Optimal Contracts

We begin by studying the mechanism design problem with a fixed extent of contractibility. We characterize implementable and optimal contracts, and illustrate the optimal contract when partial contractibility induces a coarse menu.

3.1 The Optimal Contract

In principle, partial contractibility affects the problem in complex ways due to the interactions between obedience and incentive compatibility: when deciding what type to report, the agent takes into account their ability to later ignore the spirit of the contract (recommendation y) and instead take a different action within the letter of the contract (a different $x \in C(y)$). Put differently, allowing for imperfect contractibility ($C(y) \neq \{y\}$) widens the scope for deviations for each agent θ—they can now pretend to be type θ' while also taking an action that differs from the recommendation or action of θ'. Such double deviations place additional global constraints on what the principal can implement.

Despite this complication, we show that optimal mechanisms can be fully characterized. To do this, we first define the virtual surplus function $J : X \times \Theta \to \mathbb{R}$ as:

$$J(x, \theta) = \pi(x, \theta) + u(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(x, \theta) \tag{10}$$

which is the total surplus from θ taking action x, net of any payments that must be made to the agent to ensure local incentive compatibility. As is standard, we assume that J is strictly supermodular in (x, θ) and strictly quasiconcave in x. We define the principal's favorite final outcome function $\phi^P : \Theta \to X$ as:

$$\phi^P(\theta) = \arg\max_{x \in X} J(x, \theta) \tag{11}$$

Moreover, we define the lowest implementable final action greater than $\phi^P(\theta)$ and the greatest implementable final action less than $\phi^P(\theta)$ as:

$$\overline{\phi}(\theta) = \min\{x \in \overline{D} : x \geq \phi^P(\theta)\} \quad \text{and} \quad \underline{\phi}(\theta) = \max\{x \in \overline{D} : x \leq \phi^P(\theta)\} \tag{12}$$

Given that $\overline{D}$ is closed, these minimum and maximum values are attained. We finally define the difference in the virtual surplus between these two allocations as:

$$\Delta J(\theta) = J(\overline{\phi}(\theta), \theta) - J(\underline{\phi}(\theta), \theta) \tag{13}$$

With these objects in hand, we can now describe optimal contracts:

Theorem 1 (Optimal Contract)**.** *Fix a regular contractibility correspondence C with upper image set $\overline{D}$. Any optimal final outcome function is almost everywhere equal to:*

$$\phi^*(\theta) = \begin{cases} \overline{\phi}(\theta), & \Delta J(\theta) > 0, \\ \underline{\phi}(\theta), & \Delta J(\theta) \leq 0. \end{cases} \tag{14}$$

Moreover, ϕ^* is supported by $\xi^* = \phi^*$ and tariff:

$$T^*(x) = u(x, (\phi^*)^{-1}(x)) - \int_0^{(\phi^*)^{-1}(x)} u_\theta(\phi^*(s), s)\, \mathrm{d}s \tag{15}$$

Proof. See Appendix A.2. □

We prove this result in three parts in the appendix. In the first part, we characterize implementable allocations: a final outcome function ϕ is implementable if it is monotone increasing in θ and its image lies in $\phi(\Theta) \subseteq \overline{D}$. Intuitively, after being given any $y \in X$, the agent's favorite point is $\overline{\delta}(y)$. Thus, if $y < \overline{\delta}(y)$, Obedience fails and the contract is not implementable. The substantive part of the proof establishes sufficiency by ruling out double deviations: if $\phi(\theta) \in \overline{D}$, and ϕ is monotone, then transfers can be designed so that Obedience and Incentive Compatibility hold. This characterization of implementation also implies yields our formula for the tariff (Equation 15), which follows from application of the envelope theorem to the standard reporting problem of single deviations.

In the second part of the result, we combine our novel characterization of implementation with standard mechanism design arguments to reduce the principal's problem to an optimal control problem for the final action function.

The final part of the result characterizes the optimal final outcome function by solving this control problem. Intuitively, the optimal contract implements the "next best" thing to $\phi^P(\theta)$ that is actually conctractible, in an incentive-compatible way. This is $\overline{\phi}(\theta)$ when $\Delta J(\theta) > 0$ and $\underline{\phi}(\theta)$ when $\Delta J(\theta) < 0$. Our assumption that J is supermodular guarantees that this pointwise optimal policy is monotone and therefore globally optimal. As this result shows that ξ can be taken equal to ϕ; we henceforth focus on (ϕ, T) as the key objects of the contract.

3.2 Coarse Contracts

We finally specialize and illustrate Theorem 1 in a case that will become important later: when $\overline{D}$ can be written as a sequence of ordered isolated points, or $\overline{D} = \{x_1, \ldots, x_K\}$ with

$x_1 = 0$ and $x_K = \overline{x}$. In this case, the contract has the following structure:

Proposition 1 (Coarse Contracts)**.** *If $\overline{D} = \{x_1, \ldots, x_K\}$, any optimal final outcome function is almost everywhere equal to:*

$$\phi^*(\theta) = \sum_{k=1}^{K} x_k \mathbb{I}[\theta \in (\hat{\theta}_k, \hat{\theta}_{k+1}]] \tag{16}$$

where for $k \in \{2, \ldots, K\}$, $\hat{\theta}_k$ is defined as the unique solution to $J(x_k, \hat{\theta}_k) = J(x_{k-1}, \hat{\theta}_k)$ if one exists, one if $J(x_k, \theta) < J(x_{k-1}, \theta)$ for all $\theta \in \Theta$, and zero if $J(x_k, \theta) > J(x_{k-1}, \theta)$ for all $\theta \in \Theta$, with the normalization that $\hat{\theta}_1 = 0$ and $\hat{\theta}_{K+1} = 1$. The optimal on-menu tariff, $T : \overline{D} \to \mathbb{R}$, is given by

$$T^*(x_k) = \mathbb{I}[k \geq 2] \sum_{j=2}^{k} \left[u(x_j, \hat{\theta}_j) - u(x_{j-1}, \hat{\theta}_j) \right] \tag{17}$$

Proof. See Appendix A.3. □

In an optimal *coarse contract* with K contractible actions, the principal offers a K-item menu. The items are priced such that the types separate into a K-interval partition and the types in interval k purchase item k. The boundary types separating these intervals, $\{\hat{\theta}_k\}_{k=1}^{K}$, are such that the *principal* is indifferent between their purchasing adjacent items, taking into account the marginal effect of that type's choices on the required information rents. The profit-maximizing pricing has prices jump by exactly the willingness-to-pay of the threshold type for moving from the previous allocation to the next.

We now illustrate the coarse contract in an example of monopoly pricing à la Mussa and Rosen (1978) in Section 5.

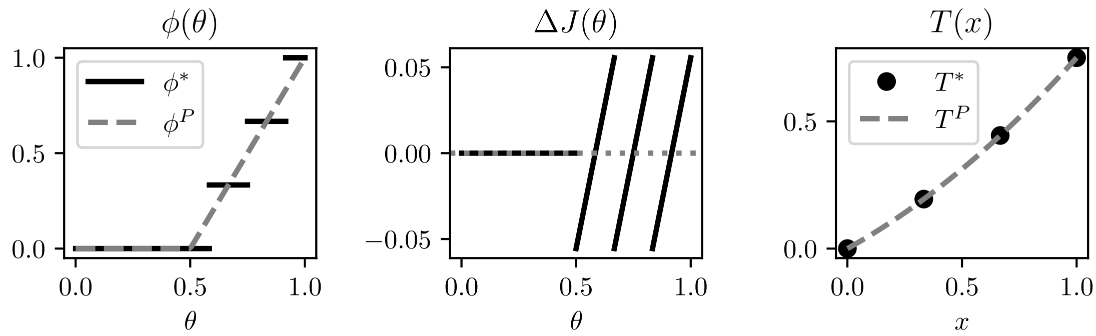
Example 2. We study a case with linear utility for the agent, quadratic costs for the principal, and uniformly distributed types:

$$u(x, \theta) = x\theta \qquad \pi(x, \theta) = -\frac{1}{2}x^2 \qquad \theta \sim U[0, 1] \tag{18}$$

We allow for contractibility on a four-point, evenly spaced partition of the action space $X = [0, 1]$: $\overline{D} = \{0, 1/3, 2/3, 1\}$. One contractibility correspondence that induces such an $\overline{D}$ is the "Partition" example of Figure 1, Panel 1b. Moreover, as implied by Theorem 1, the specification of the lower image set $\underline{D}$ is not relevant for the the optimal contract.

We remind that the optimal contract under *full contractibility*, as studied by Mussa and Rosen (1978) *inter alia*, assigns $\phi^P(\theta) = 0$ for $\theta \in [0, 1/2]$ and $\phi^P(\theta) = 2\theta - 1$ for $\theta \in (1/2, 1]$.

Figure 2: An Optimal Coarse Contract



Note: The optimal coarse contract in a setting with $u(x,\theta) = x\theta$, $\pi(x,\theta) = -\frac{x^2}{2}$, $\theta \sim U[0,1]$, and $\overline{D} = \{0, 1/3, 2/3, 1\}$. The first panel shows the assignment ϕ; the second panel shows the function $\Delta J(\theta)$ defined in Equation 13 and Theorem 1; and the third panel shows the tariff T. In the first and third panel, we graph both the optimal coarse contract (ϕ^*, T^*) and the contract under perfect contractibility (ϕ^P, T^P).

The optimal contract under full contractibility charges tariff $T(x) = \frac{x^2}{4} + \frac{x}{2}$.

The optimal contract in this quadratic case "coarsens" the familiar contract (ϕ^P, T^P) as illustrated in Figure 2. As described in the discussion of Theorem 1 and Corollary 1, the principal partitions the types into intervals receiving each item (first panel) and determines the boundaries of these intervals based on their indifference, or when ΔJ crosses zero (second panel). That the partition of the type space also features even intervals and that the optimal tariff connects points on T^P (third panel) are special features of this model, which features quadratic u and J. We discuss these special features in more depth when we study optimal contractibility in the same model in Section 5. $\triangle$

4 Optimal Contractibility

We now study the principal's optimal choice of contractibility. We show our main result: if costs of contractibility satisfy a *strong monotonicity* property defined below, then optimal contracts are *coarse*, *i.e,* they are supported on finitely many outcomes.

4.1 Existence of Solution

We first use the results of Section 3 to restate the principal's optimal contractibility problem and show that it is well-posed. As shown in Theorem 1, the set $\overline{D}$ summarizes the effects of imperfect contractibility on the optimal contract. We let $\mathcal{D}$ denote the set of possible

$\overline{D}$, or closed subsets of X that contain $\overline{x}$ and 0, and endow it with the topology induced by the Hausdorff distance between closed sets (see Lemma 1).[6] With an abuse of notation, let $\mathcal{J} : \mathcal{D} \to \mathbb{R}$ define the value induced by solving the non-linear pricing problem given a particular contractibility support $\overline{D} \in \mathcal{D}$. This is formally defined in Lemma 8 in Appendix A.2. The same lemma implies that the value induced by the optimal contract does not depend on $\underline{D}$. For this reason, here we fix $\underline{D} = \{0\}$, that is, complete absence of contractibility for deviations below the recommended outcome.[7] With this, and with some abuse of notation, for every $\overline{D} \in \mathcal{D}$, we let $\Gamma(\overline{D})$ denote the cost of the regular contractibility correspondence represented by $\overline{D}$ and $\{0\}$. We assume henceforth that $\Gamma : \mathcal{D} \to \overline{\mathbb{R}}$ is lower semi-continuous. For example, this is satisfied by costs of distinguishing (Definition 1). Using this we can rewrite the program of Equation 9 as the following choice of $\overline{D}$:

$$\sup_{\overline{D} \in \mathcal{D}} \quad \mathcal{J}(\overline{D}) - \Gamma(\overline{D}) \tag{19}$$

Our results in Section 3 moreover imply that $\mathcal{J}$ is continuous, allowing us to show the following:

Proposition 2. *The set of optimal contractibility supports $\mathcal{D}^*(\Gamma)$ solving Problem 19 is nonempty and compact.*

Proof. See Appendix A.4. □

Theorem 1 and Proposition 2 together imply that the joint design problem of optimally choosing a contractibility correspondence and then a contract has well-defined solutions, despite its high dimensionality.

4.2 Key Property: Strongly Monotone Costs

We next introduce a property of contractibility costs that will be crucial for our coarseness result. The property concerns the cost of differentiating a given action x from others with arbitrarily high "precision." Formally, we consider an $x \in \overline{D}$ that is an *accumulation point*, or a point around which any small neighborhood contains another point in $\overline{D}$. Economically, the principal can differentiate such an action x from many arbitrarily close actions. We consider the thought experiment of removing contractibility in a small region around x, or eliminating these fine distinctions between actions. The *strong monotonicity* property, stated

[6] Recall that the Hausdorff distance between sets in the real line is defined as $d_H(A, B) = \max\{\sup_{a\in A} \inf_{b\in B} |a - b|, \sup_{b\in B} \inf_{a\in A} |a - b|\}$.

[7] Observe that, whenever adding any contractibility from below involves strictly positive costs, setting $\underline{D} = \{0\}$ is part of any solution of the principal's overall problem.

below, disciplines the rate at which this cost of precise contracting declines to zero as we focus on an arbitrarily small part of the action space around x:

Definition 3. *A cost function* Γ *is strongly monotone if there exists* $\epsilon > 0$ *such that:*

$$\liminf_{m} \frac{\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (a_m, b_m))}{(x_m - a_m)(b_m - x_m)} \geq \epsilon \tag{20}$$

for all $\overline{D} \in \mathcal{D}$*, accumulation points* $x \in \overline{D}$*, and sequences* $\{a_m, x_m, b_m\}_{m=1}^{\infty} \subseteq \overline{D}$ *such that* $x_m \in (a_m, b_m)$ *and* $\overline{D} \cap (a_m, b_m) \to \{x\}$*, where the limit is in the topological sense.*[8]

Note that this property allows the costs of "precise contracting" to go to zero, as we can take $x_m - a_m$ and $b_m - x_m$ each to zero. The content of the property is to restrict how *quickly* these costs reach zero.

One important and illustrative implication of strong monotonicity is that there are *second order costs of perfect contractibility* in the following sense. Consider an x and $\overline{D}$ such that there is perfect contractibility in a neighborhood around x, or $B_t(x) \in \overline{D}$ for all sufficiently small $t > 0$.[9] In this construction, x is an (interior) accumulation point that the principal can precisely differentiate from all of its neighbors. Applying Definition 3, we can take a sequence $\{t_m\}_{m=0}^{\infty}$ such that $t_m \to 0$ and construct sequences $a_m = x - t_m$ and $b_m = x + t_m$. In this case, the operation in Definition 3 is to remove a sequence of shrinking balls centered around x. A cost Γ is strong monotone only if, in such a scenario, the cost of removing these balls is asymptotically bounded by a constant times their radius squared, or ϵt_m^2.

Definition 3 generalizes this idea to also discipline the cost of precise contracting around non-interior accumulation points. For example, the set $\overline{D} = \{1 - 2^{-k}\}_{k=0}^{\infty} \cup \{1\}$ has an empty interior, but 1 is an accumulation point which the principal can distinguish from any close action $1 - 2^{-k}$, for arbitrarily large k. Similarly, if $\overline{D}$ were the Cantor set, then *all* of its elements are non-interior accumulation points. The full form of Definition 3 is required to consider set-valued perturbations that allow for countably infinite sets and irregular sets, such as the Cantor set.

We argue that strong monotonicity is a natural property to possess because is *any* cost of distinguishing outcomes (recall Definition 1) satisfies it:

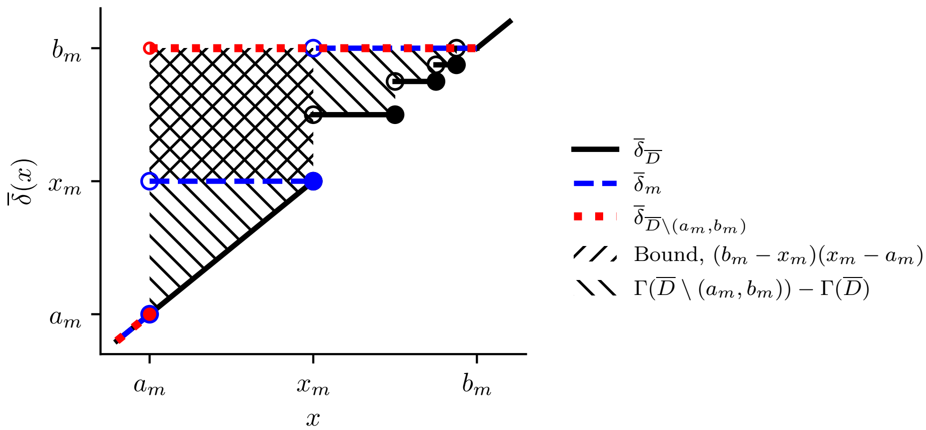
Proposition 3. *Any cost of distinguishing outcomes is strongly monotone with* $\epsilon = \tilde{d}(0, \overline{x})$.

Proof. See Appendix A.5. □

[8] The upper topological limit of a sequence of sets $\{A_m\}_{m=1}^{\infty} \subseteq X$ is the set of points $x \in X$ such that every neighborhood intersects infinitely many sets A_m. The lower topological limit is the set of points such that every neighborhood contains intersects almost all sets A_m. The topological limit exists if the upper and lower topological limits are equal.

[9] Here $B_t(x)$ denotes the open ball centered at x and with radius t.

Figure 3: Strong Monotonicity for Costs of Distinguishing Outcomes



Note: An illustration of strong monotonicity for discrete costs of distinguishing (Example 1). The function δ_m is constructed in the proof of Proposition 3. Note that, in this example, the bound is not tight.

We can give a simple geometric intuition why costs of distinguishing are strongly monotone. For simplicity, suppose that $\tilde{d}$ is the discrete metric (recall Example 1), in which case $\epsilon = 1$ and the cost coincides with the area above $\overline{\delta}$ (Figure 3). We first observe that any $\overline{D}$ which induces an upper envelope $\overline{\delta}_{\overline{D}}$ (black solid line, illustrating perfect contractibility), is "greater" than a variant set of contractibility that includes $\{a_m, x_m, b_m\}$ but no other points in the interval (a_m, b_m), represented by some upper envelope $\overline{\delta}_m$ (blue dashed line). This is itself "greater" than $\overline{\delta}_{\overline{D}\setminus(a_m,b_m)}$ (red dotted line). The cost savings of moving between the dashed line and the dotted line is the right-hatched rectangle, with side lengths $b_m - x_m$ and $x_m - a_m$. These cost savings are a lower bound for the cost savings of moving from $\overline{D}$ to $\overline{D} \setminus (a_m, b_m)$, which are indicated with left-hatched shading. Thus, $\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (a_m, b_m)) \geq (x_m - a_m)(b_m - x_m)$ and strong monotonicity is satisfied. Beyond this case, we show by the mean value theorem that any cost of distinguishing outcomes is bounded below by $\tilde{d}(0, \overline{x})$ times the cost of distinguishing outcomes under the discrete metric. For example, when the cost of distinguishing is induced by a $p-$distance (from Example 1), we have that $\epsilon = \overline{x}^{-\frac{1}{1+p}}$.

4.3 Optimal Coarse Contracts

We now state our main theoretical result on the optimality of coarse contracts and the extent of their coarseness. To do this, we define of maximum concavity $\bar{J}_{xx} = \max_{x,\theta} |J_{xx}(x, \theta)|$, minimum complementarity $\underline{J}_{x\theta} = \min_{x,\theta} J_{x\theta}(x, \theta)$, and maximum density $\bar{f} = \max_\theta f(\theta)$.

Note that, under our maintained assumptions, $0 < \bar{J}_{xx}, \underline{J}_{x\theta}, \bar{f} < \infty$. With these objects in hand, we have that:

Theorem 2 (Optimally Coarse Contractibility)**.** *If* Γ *is strongly monotone, then every optimal contractibility support* $\overline{D}^*$ *is finite with* $|\overline{D}^*| \leq \left\lfloor 2\left(\frac{3\overline{x}\bar{J}_{xx}^2\bar{f}}{\epsilon \underline{J}_{x\theta}} + 1\right)\right\rfloor$.

Before proving this result, we remark on what these properties for optimal contractibility imply for optimal contracts. We say that a final outcome function ϕ is supported on a set $\overline{D} \subseteq [0, \overline{x}]$ if there exists a tariff T with proper domain $\overline{D}$ that induces ϕ.

Corollary 1 (Optimally Coarse Contracts)**.** *If* Γ *is strongly monotone, every optimal final outcome function* ϕ^* *is supported on a finite menu with at most* $\left\lfloor 2\left(\frac{3\overline{x}\bar{J}_{xx}^2\bar{f}}{\epsilon \underline{J}_{x\theta}} + 1\right)\right\rfloor$ *items.*

This combination of Theorem 2 and Corollary 1 provides a foundation for endogenous incomplete contracts under the presence of contractibility costs. This incompleteness takes a strong form under a coarse contract because *almost all* actions are left unspecified. Moreover, this result holds for any arbitrarily small degree of cost of writing contracts, since the ϵ in Definition 3 can be made arbitrarily small.

We now describe the proof of Theorem 2 in three parts: i) finding estimates of the loss in value from set-valued perturbations of contractibility, ii) combining these estimates with strong monotonicity to rule out infinite sets, and iii) constructing an explicit bound for the extent of contractibility.

Part I: The Opportunity Cost of Coarsening a Contract. We first give an intermediate result that bounds the loss to the principal from removing contractibility:

Lemma 2. *Consider any* $\overline{D} \in \mathcal{D}$ *and any* $a, b \in \overline{D}$ *such that* $a < b$. *Then,*

$$\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (a, b)) \leq \frac{3}{2}\frac{\bar{J}_{xx}^2\bar{f}}{\underline{J}_{x\theta}}(b-a)^3 \tag{21}$$

Moreover, if $(a, b) \cap \overline{D} \neq \emptyset$, *then there exists* $c \in (a, b) \cap \overline{D}$ *such that:*

$$\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (a, b)) \leq \frac{3}{2}\frac{\bar{J}_{xx}^2\bar{f}}{\underline{J}_{x\theta}}(b-a)\left[(c-a)^2 + (b-c)^2\right] \tag{22}$$

Furthermore, if $\{a, b, c\}$ *are sequential, or* $\overline{D} \cap (a, b) = \emptyset$ *and* $\overline{D} \cap (b, c) = \emptyset$, *then*

$$\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (a, b)) \leq 3\frac{\bar{J}_{xx}^2\bar{f}}{\underline{J}_{x\theta}}(b-a)(c-a)(b-c) \tag{23}$$

Proof. See Appendix A.6. □

The first statement says that the opportunity cost of removing all points of contractibility within an interval (a, b) is *third-order* in the length of that interval. The next two statements refine this bound when there is a known point of contractibility $a < c < b$ and when the three points of interest are isolated. All three bounds share the following basic comparative statics: they loosen when J has higher concavity, when J has lower supermodularity, and when the type density is more concentrated.

We omit the full proof because it involves detailed calculations. But, to provide intuition for the form of these bounds, we sketch the proof of the first statement (Equation 21). We first observe, exploiting our results from Section 3.1, that optimal allocations conditional on any level of contractibility solve a pointwise program (see Lemma 8). Thus, we can re-express $\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (a,b))$ as

$$\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (a,b)) = \int_{\Theta} (J(\phi^*(\theta), \theta) - J(\phi^{*'}(\theta), \theta)) \, \mathrm{d}F(\theta) \tag{24}$$

where ϕ^* and $\phi^{*'}$ respectively denote the optimal assignments under each level of contractibility. We next observe, using our characterization of the optimal contract (Theorem 1), that $\phi^* \neq \phi^{*'}$ only for types such that the actions $\overline{\phi}(\theta)$ or $\underline{\phi}(\theta)$, defined relative to $\overline{D}$, were within (a, b). The third-order bound derives from two steps: showing that this set of affected types has measure proportional to $b - a$ and showing that the payoff losses for each such type are bounded by something proportional to $(b - a)^2$.

For the first step, we observe that a necessary condition for a type θ to be affected by the removal of the interval (a, b) is that $\phi^P(\theta) \in (a, b)$: in words, that the principal would prefer (absent imperfect contractibility) to allocate these types something between a and b. We can define this set of types as the pre-image of (a, b) via ϕ^P; intuitively, it has large mass if the ϕ^P mapping is very flat (*i.e.,* nearby types map to similar actions) or if the type density is very large in this region. We bound the (inverse) slope of the type distribution by $\frac{\bar{J}_{xx}}{\underline{J}_{x\theta}}$ and the maximum type distribution by $\bar{f}$. Together, this contributes a term $(b-a)\frac{\bar{J}_{xx}}{\underline{J}_{x\theta}}\bar{f}$ to the bound.

For the second step, we exactly express $J(\cdot, \theta)$ to second order around $\phi^*(\theta)$ using Taylor's remainder theorem. We next express the first-order effects as also *second-order*, using the fact that $\phi^*(\theta)$ and $\phi^{*'}(\theta)$ are close to $\phi^P(\theta)$, and the fact that $J_x(\phi^P(\theta), \theta) = 0$ due to that allocation's pointwise optimality. This contributes a term $\frac{3}{2}\bar{J}_{xx}(b-a)^2$, where we use the uniform bound on concavity. Putting steps one and two together gives the bound in Equation 21.

Part II: Establishing Finite Contractibility. We now establish that there exists some $K^* \in \mathbb{N}$ such that every optimal contractibility support is finite with $|\overline{D}^*| \leq K^*$. We prove this by contradiction. Suppose instead that the that an optimal contractibility support $\overline{D}^*$ is an infinite set. As $\overline{D}^*$ is compact, this implies that $\overline{D}^*$ contains an accumulation point x.

We now consider the closed set $\overline{B}_t(x) \cap \overline{D}$, which is the neighborhood around x in $\overline{D}$ and is infinite as x is an accumulation point. There are four exhaustive possibilities for the properties of this set:

1. $\overline{B}_t(x) \cap \overline{D}$ is a perfect set: that is, all of its members are accumulation points.
   (a) Moreover, the set is somewhere dense. In this case, the set necessarily contains an interval.
   (b) Moreover, the set is nowhere dense. For example, the set could be the Cantor set.
2. $\overline{B}_t(x) \cap \overline{D}$ is not a perfect set.
   (a) Moreover, the set is uncountably infinite. In this case, by application of the Cantor-Bendixson Theorem, it contains a perfect set (see, *e.g.,* p. 67 of Apostol, 1974).
   (b) Moreover, the set is countably infinite. In this case, the set contains an isolated point. If it did not, then all points in the set would be accumulation points, and the set would be a perfect set.

We proceed to show that each of these cases contradicts optimality. In each case, our argument will be that, given strong monotonicity (Definition 3), the marginal costs of precise contracting near an accumulation point x go to zero more slowly than the marginal benefits. In each case, we will rely on a different "costs" implication of strong monotonicity and a different "benefits" implication of Lemma 2.

Lemma 3. *If Γ is strongly monotone, then the following statements are true:*

1. *If $\overline{D} \in \mathcal{D}$ contains an interval, then $\overline{D}$ is not optimal*
2. *If $\overline{D} \in \mathcal{D}$ contains an accumulation point x such that $\overline{B}_t(x) \cap \overline{D}$ is a perfect and nowhere dense set for some $t > 0$, then $\overline{D}$ is not optimal*
3. *If $\overline{D} \in \mathcal{D}$ is countably infinite, then $\overline{D}$ is not optimal.*

Proof. See Appendix A.7 □

Thus, strong monotonicity rules out intervals, nowhere dense perfect sets (*e.g.,* the Cantor set), and countably infinite sets. We finally put these steps together to complete the proof of finiteness, referring back to our exhaustive list of cases. Under case 1(a), claim 1. of Lemma 3 contradicts optimality. Under case 1(b), claim 2. of Lemma 3 contradicts optimality. Under case 2(a), the problem reduces to either 1(a) or 1(b) and the previous arguments apply. Under case 2(b), claim 3. of Lemma 3 contradicts optimality. Thus, we have shown that $\overline{D}^*$ cannot contain an accumulation point. As the set is also compact, it must be finite.

Part III: Deriving the Bound. We now derive an explicit bound on the number of elements in $\overline{D}^*$.

Lemma 4. *If* Γ *is strongly monotone, then* $|\overline{D}^*| \leq \left\lfloor 2 \left(\frac{3\overline{x}\bar{J}_{xx}^2 \bar{f}}{\epsilon \underline{J}_{x\theta}} + 1 \right) \right\rfloor$.

Proof. See Appendix A.8. □

We prove this by using our explicit bound on the payoff gains from more complete contracts from Lemma 2. Concretely, if more than this many actions were contractible, we can show directly that eliminating at least one action would be payoff improving. The bound on the completeness of the contract inherits the comparative statics of our payoff bound in Lemma 2. That is, contracts are finer-grained when the losses from coarseness are higher, and those losses are higher with high concavity, low supermodularity, and high concentration of types. In Section 5, we will explore these predictions further in our application.

4.4 Designing Coarse Contracts

Having established that strong monotonicity implies coarse contracts and derived an explicit bound on the contract's "size," we now study how the principal chooses which outcomes are contractible. That is, how does a principal *design* a coarse contract to best suit their needs?

We first revisit our analysis from Section 3 to write the principal's payoffs when contractibility is finite. As observed in Proposition 1, the optimal contract given a coarse contractibility correspondence allocates action x_k to types $\theta \in [\hat{\theta}_k, \hat{\theta}_{k+1})$ (recall that $\hat{\theta}_k$ is defined as the solution to $J(x_k, \hat{\theta}_k) = J(x_{k-1}, \hat{\theta}_k)$ when one exists for $k \in \{2, \ldots, K\}$, with the normalization that $\hat{\theta}_1 = 0$ and $\hat{\theta}_{K+1} = 1$). Given this, we have that the principal's total profit is given by:

$$\mathcal{J}\left(\{x_k\}_{k=1}^K\right) = \sum_{k=1}^{K} \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J(x_k, \theta)\, \mathrm{d}F(\theta) \tag{25}$$

Let $\mathcal{D}_K$ be the set of all $\overline{D} \in \mathcal{D}$ such that $|\overline{D}| = K$. Observe that each set $\overline{D} = \{x_1, ..., x_K\} \in \mathcal{D}_K$ is uniquely identified by the vector $(x_1, ..., x_K) \in X^K$. Therefore, with a slight abuse of notation, we identify $\mathcal{D}_K$ with the finite-dimensional set X^K. Given any Γ and $K \in \mathbb{N}$, define the family of restricted cost functions $\Gamma_K : \mathcal{D}_K \to \bar{\mathbb{R}}$ with $\Gamma_K(\overline{D}) = \Gamma(\overline{D})$ for all $\overline{D} \in \mathcal{D}_K$. We now define the differentiability notion that we employ:

Definition 4 (Finite Differentiability)**.** Γ *is finitely differentiable if* Γ_K *is a continuously differentiable function for all* $K \in \mathbb{N}$.[10]

[10] As standard, here we mean that each Γ_K admits a continuously differentiable extension to an open set that contains X^K.

When a cost function is finitely differentiable, its derivatives coincide with a more traditional notion in Euclidean space. We write these derivatives in some abuse of notation for $k \in \{2, \ldots, K-1\}$ as:

$$\Gamma_K^{(k)}(\overline{D}) = \lim_{\epsilon \downarrow 0} \frac{\Gamma(\{x_1, \ldots, x_k + \epsilon, \ldots, x_K\}) - \Gamma(\{x_1, \ldots, x_k, \ldots, x_K\})}{\epsilon} \tag{26}$$

We observe that any cost of distinguishing satisfies this property:

Proposition 4. *Any cost of distinguishing outcomes is finitely differentiable.*

Proof. See Appendix A.9. □

We now state a necessary condition for an optimally designed coarse contract, which intuitively requires that "marginal benefits equal marginal costs" for adjusting any contractible outcome x_k:

Proposition 5. *If Γ is strongly monotone and finitely differentiable, then any optimal contractibility support $\overline{D}^* = \{x_1, \ldots, x_{K^*}\}$ satisfies:*

$$\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k, \theta) \, \mathrm{d}F(\theta) = \Gamma_{K^*}^{(k)}(\overline{D}^*) \qquad \text{for } k \in \{2, \ldots, K^* - 1\} \tag{27}$$

where $\hat{\theta}_k$ is as defined in Proposition 1.

Proof. See Appendix A.10 □

The left-hand-side of Equation 27 says that the marginal benefit of changing a grid point x_k is the average increase in virtual surplus over all types allocated to that action. Note that these marginal changes in *virtual* surplus take into account the direct effects on revenues and costs (holding fixed agents' purchases) as well as the indirect effects on the rest of the contract via information rents. A second effect of changing x_k, the change in the marginal types $\hat{\theta}_k$ and $\hat{\theta}_{k+1}$, is only second order since the principal is indifferent between allocating those types either of two adjacent actions in the grid.

4.5 Efficient Contracts and Contractibility

We have so far considered optimal contracts. However, our analysis also applies to efficient contracts that maximize total surplus, rather than virtual surplus. To be concrete, define total surplus as $S(x, \theta) = \pi(x, \theta) + u(x, \theta)$ and assume that this is strictly supermodular in (x, θ) and strictly quasi-concave in x. The efficient mechanism design and contractibility problems are respectively given by:

$$\mathcal{S}(C) := \sup_{(\phi,\xi,T)\in\mathcal{I}(C)} \int_{\Theta} S(\phi(\theta), \theta)\mathrm{d}F(\theta) \tag{28}$$

and:

$$\sup_{C\in\mathcal{C}} \quad \mathcal{S}(C) - \Gamma(C) \tag{29}$$

Understanding efficient contractibility is interesting for three reasons. First, it is directly useful for understanding the welfare effects of incomplete contracts. Second, it allows us to understand how incomplete information affects incomplete contracts. This is because the principal's problem under complete information reduces to the efficient problem.[11] Third, it allows us to study settings in which the agents have the bargaining power and choose a contract to maximize their expected utility subject to the principal's participation.[12]

All of our results apply to this problem, where J in our earlier results must simply be substituted with S. This observation opens up the door to comparative statics results on the extent of optimal contractibility across the revenue-maximization cases and the efficient cases. For example, the new bound on the optimal extent of contractibility in the efficient case is $|\overline{D}_e^*| \leq \left\lfloor 2\left(\frac{3\overline{x}\bar{S}_{xx}^2\bar{f}}{\epsilon \underline{S}_{x\theta}} + 1\right)\right\rfloor$, where $\overline{D}_e^*$ is any efficient contractibility support and $\bar{S}_{xx} = \max_{x,\theta} |S_{xx}(x,\theta)|$ and $\underline{S}_{x\theta} = \min_{x,\theta} S_{x\theta}(x,\theta)$. Thus, changes in concavity and supermodularity induced by information rents can be seen to directly impact the difference between efficient and revenue-maximizing contractibility. In Section 5.3, we exploit this to derive exact comparative statics in our leading application.

5 Application: Optimally Coarse Monopoly Pricing

In this section, we apply our results to study monopoly pricing with endogenous and costly contractibility. We show that optimal pricing takes the form of discrete quality tiers, as the principal forgoes the opportunity for finer-grained price discrimination to economize on the costs of designing the contract. We derive comparative statics for optimal coarseness, *i.e.,*

[11] This is because the participation constraint of each type θ must bind under complete information and so the principal extracts full surplus from each type. Although Problem 28 is defined to include the incentive compatibility constraint implied by incomplete information, strict supermodularity of S implies that the global incentive compatibility constraint would be slack.

[12] Formally, this corresponds to the constraint that the principal's expected payoff is no less than their outside option (normalized to 0): $\int_{\Theta} (\pi(\phi(\theta), \theta) + T(\xi(\theta)))\,\mathrm{d}F(\theta) \geq 0$. It is then standard to show that this participation constraint must bind at the agent's optimal contract which in turn must solve Problem Problem 28. Therefore, the extent of optimal contractibility must again solve Problem 29.

the number of quality tiers, as a function of differentiation in consumers' tastes, production costs, and costs of contractibility. We find that the presence of asymmetric information leads to endogenously coarser contracts, or fewer quality tiers, by restricting the principal's potential gains from introducing a more fine-grained menu.

5.1 Set-up

We study the canonical linear-quadratic-uniform model of monopoly screening introduced by Mussa and Rosen (1978). A monopolist (the principal) is selling a good of potentially variable quality $x \in X = [0, 1]$. A continuum of consumers (the agents) have privately known taste $\theta \sim U[0, 1]$ and preferences

$$u(x, \theta) = \alpha\theta x \tag{30}$$

where $\alpha > 0$ scales the extent of differentiation in preferences. The monopolist has production or service cost

$$\pi(x, \theta) = -\beta\frac{x^2}{2} \tag{31}$$

where $\beta \in (0, \alpha]$ scales the extent of these costs.[13]

In the model of Mussa and Rosen (1978), and the broader literature on nonlinear pricing (Wilson, 1993), the principal has access to contracts that specify a mapping from continuous levels of quality $x \in [0, 1]$ to prices $T(x)$. This is nested in our setting by eliminating costs of contractibility and observing that the principal's problem has the fewest constraints, and hence the highest payoff, under perfect contractibility, $C(x) = \{x\}$ (see Lemma 8).

We instead assume that the principal faces costs when writing the contract. In particular, these take the form of the *costs of distinguishing actions* under the discrete metric introduced in Example 1:

$$\Gamma(C) = \gamma \int_X \mu(X \setminus C(x))\, \mathrm{d}x = \gamma \int_0^1 (1 - \bar{\delta}(x))\, \mathrm{d}x \tag{32}$$

where μ is the Lebesgue measure, $\overline{\delta}(x) = \max C(x)$, $\gamma > 0$ is a scaling parameter, and where we ignore the additive term corresponding to $\underline{\delta}(x) = \min C(x)$ due to its irrelevance for the problem with increasing preferences. As described in Example 1, these costs represent the monopolist's difficulty in describing the difference between levels of quality *ex ante*.

To sharpen this interpretation, consider an application of the model to monopoly pricing of *rentals*—for instance, of hotel rooms or cars. In this example, x is the consumer's intensity of use (the "quality" of their experience). The type represents consumers' differential taste to spend time in the room or drive. The production cost represents the monopolist's need

[13] We introduce the simplifying assumption that $\alpha \geq \beta$, so under all optimal contracts the highest types are allocated the maximum quality $x = 1$.

to offset damage and/or depreciation. The cost of contractibility is the cost of specifying the boundaries between different levels of utilization (when is a car's interior damaged?). To act in the spirit of the contract is to check out of a pristine hotel room or return a perfectly clean car; to act in the letter is to skirt the boundary of acceptable condition.

5.2 Optimal Pricing Features Uniform Quality Tiers

We now study the monopolist's optimal pricing policy when they jointly design contractibility and the optimal contract. We first apply our general theoretical results to significantly simplify the problem. First, since the cost faced is a cost of distinguishing outcomes (as per Definition 1), Proposition 3 establishes that it is strongly monotone. Thus, Theorem 2 implies that any optimal contractibility correspondence is finite. As a consequence, we can treat the monopolist as optimizing jointly over a number $K \in \mathbb{N}$ of distinct quality levels and a vector $\{x_k\}_{k=1}^{K}$ specifying those levels. Moreover, Proposition 5 implies that optimal quality levels necessarily solve a first-order condition which, applied to our monopoly pricing problem, embodies the trade-off between the cost of specifying the contract *ex ante* and the benefits from price discrimination *ex post*. To proceed further, we exploit the specific structure of production costs and consumer demand. Specifically, the first-order condition reduces to a second-order nonlinear difference equation which we can solve directly. Using this, we can calculate the firm's payoff conditional on optimally designing a contract with any number K of contractible quality levels and then optimize analytically over K.

We find that the optimal contract takes the specific form of uniformly spaced quality levels. Moreover, we can characterize the optimal number of qualities in closed form and describe its comparative statics.

Proposition 6 (Optimal Nonlinear Pricing Contract)**.** *The seller offers the menu*

$$x_k = \frac{k-1}{K^*-1} \qquad T(x_k) = \frac{1}{2}\frac{k-1}{K^*-1}\left(\frac{\beta}{2}\frac{k-1}{K^*-1} + \alpha\right) \qquad k \in \{1, \dots, K^*\} \tag{33}$$

where the optimal number of qualities, K^*, satisfies $|K^* - \tilde{K}| < 1$ and

$$\tilde{K} = 1 + \frac{\beta^2}{12\alpha\gamma} \tag{34}$$

Moreover, K^* decreases in α, increases in β, and decreases in γ. If $\gamma < \frac{\beta^2}{16\alpha}$, then $K^* \geq 3$.

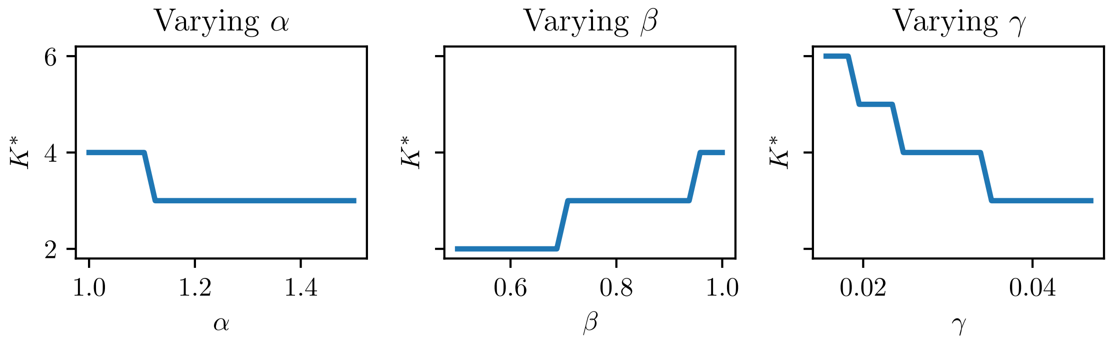
Proof. See Appendix A.11. □

Uniform Quality Tiers. The uniform spacing of qualities arises due to symmetries in the benefits of more precise price discrimination, irrespective of quality x or type θ, and the symmetry of the cost function. To understand the first property (symmetric benefits), we observe that the second derivative J_{xx} is constant as a function of (x, θ) and that the principal's optimal assignment absent contracting frictions induces a uniform distribution over actions. The following informal, constructive argument suggests the form of the solution. Starting from perfect contractibility, the opportunity cost of removing perfect contractibility in some interval of the action space is the same *regardless* of where that interval is located. This for two reasons. First, as virtual surplus is quadratic in this model, the seller has an equal opportunity cost of forgoing quality differentiation for high qualities. Second, because the optimal assignment function is linear (which is itself because of the uniformity of the distribution, the constant concavity of virtual surplus, and the constant supermodularity of virtual surplus), the same measure of types is affected. This is a specialization of the argument that underpins Theorem 2 in the general model, but *without* needing uniformity of concavity or the measure of affected types. In economic language, the seller has an equal opportunity cost of forgoing quality differentiation for high qualities (high-demand customers) or low qualities (low-demand customers). This is true even though the seller makes more money from the high-demand segment of the market. The corresponding symmetry in costs arises from our argument about distinguishing actions. In particular, this cost function implies that the difficulty in distinguishing actions does not vary over the action space—that is, nearby low qualities are not easier or harder to distinguish than nearby high qualities.

The Optimal Number of Tiers and Comparative Statics. The parameter $\tilde{K}$ is the unique maximum of the "smooth" (*i.e.*, non-integer) optimization problem. The comparative statics follow from applying the supermodularity of the objective function to the true, integer-domain problem. Economically, the comparative statics reinforce the lessons of our general bound of Theorem 2: contracts are more fine-grained or less incomplete when complementarity α is low, concavity β is high, and costs of contracting γ are low. In the monopoly-pricing contract, as described above, this corresponds to low consumer heterogeneity, high service costs, and high costs of distinguishing actions (*e.g.*, levels of utilization).

A Numerical Example. We have, in fact, already illustrated such a contract in the example of Section 3.1 shown in Figure 2. This example featured $K = 4$ and $\alpha = \beta = 1$; moreover, a four-quality contract is *optimal* for a range of cost scalings including $\gamma = \frac{1}{32}$. In Figure 4, we numerically illustrate the comparative statics of Proposition 6 near these parameter values.

Figure 4: Comparative Statics for Contract Coarseness



Note: In each panel, we illustrate comparative statics of the optimal level of contractibility K^* in the example of Section 5 with $\alpha = \beta = 1$ and $\gamma = \frac{1}{32}$. These results correspond to the analytical predictions of Proposition 6.

5.3 Incomplete Information Begets (More) Incomplete Contracts

We finally explore the interaction of incomplete information (*i.e.*, adverse selection) and incomplete contracts in the monopoly pricing setting. We do this by comparing the optimal monopoly pricing menu with the efficient allocation defined in Section 4.5.

Here, our preferred interpretation of that problem is that the monopolist can perfectly segment the market and propose an allocation that depends on the actual type θ of each consumer (*i.e.*, perfect third-degree price discrimination).[14] However, the monopolist must use the same extent of contractibility for all consumer types, for example because the choice of contractibility must be carried out before the monopolist learns the market segmentation.[15]

Under perfect contractibility, the monopolist would implement an "efficient" outcome that maximizes expected total surplus $S = \pi + u$ and perfectly extracts each consumer's willingness to pay. Under costly contractibility, however, the principal may prefer to *imperfectly* price discriminate and economize on the costs of writing a complex contract. We find that the efficient allocation also features uniform quality tiers, and that there are more tiers than in the monopoly allocation:

Proposition 7. *In the efficient contract, the optimal contractibility support is* $\overline{D}_e^* = \left\{\frac{k-1}{K^{*C}-1}\right\}_{k=1}^{K^{*C}}$ *where* $K^{*C} \geq K^*$*. Moreover,* K^{*C} *satisfies* $|K^{*C} - \tilde{K}^C| < 1$*, where* $\tilde{K}^C = 2\tilde{K} - 1$*.*

Proof. See Appendix A.12. □

[14] In other words, we consider the complete-information setting where the feasible direct mechanisms satisfy Obedience and Individual Rationality, but not Incentive Compatibility necessarily.

[15] As mentioned in Section 4.5, there is an alternative interpretation in which the consumer rather than the producer has bargaining power (*i.e.*, monopsony rather than monopoly).

The first part of the result has the same intuition as Proposition 6, relying on the symmetry of the benefits and cost functions. The second part follows by observing that

$$S(\theta, x) = J\left(\frac{1+\theta}{2}, x\right) \tag{35}$$

because $\hat{\theta} = 2\theta - 1$ is the "virtual type" of consumers, taking into account their effect on information rents. Thus, the complete-information monopolist faces the same trade-offs as the incomplete-information monopolist, but serves twice as large of a market (types in $[0, 1]$ rather than types in $[1/2, 1]$). Leveraging this observation, we show that the complete-information monopolist has exactly twice as much incentive to contract more precisely or employ more tiers.

Practically, this result implies that monopoly with adverse selection implies not just under-provision of quality—a classic result of Mussa and Rosen (1978)—but also under-differentiation of *qualities*. This arises in our environment because more incomplete information dulls the monopolist's incentives to price discriminate, which in turn dulls the monopolist's incentive to contractually differentiate different quality levels.

5.4 Additional Application: Optimal Quality Certification

To demonstrate the broad applicability of our framework, we apply our results to a model of optimal quality certification in Appendix B. Building on Albano and Lizzeri (2001) and Zapechelnyuk (2020), we consider a seller who is privately informed about how efficient they are in producing a good of a given quality. The quality actually produced by the seller is also their private information and, conditional on the realized quality, the seller offers a price to the market. The market is composed of a continuum of buyers who are privately informed about an outside option they forego when buying the seller's good. Therefore, each buyer purchases the good if and only if the expected quality of the good, minus the offered price, is no less than their outside option. We assume that the realized quality is not verifiable by the buyer and the sender cannot commit ex ante to any information disclosure policy. However, we consider a third-party certifier (*i.e.,* the designer) that, in exchange for payments from the seller, can commit ex-ante to an information policy disclosing information about the quality produced to the buyers. As described by Zapechelnyuk (2020), this setting captures a number of markets, such as crash safety testing in the car industry, food hygiene certifications for restaurants and factories, and educational inspections for schools and universities.

Differently from Albano and Lizzeri (2001) and Zapechelnyuk (2020), we assume that the certifier is also uninformed about how efficient the seller is. Moreover, each disclosure policy comes with a verification cost that the certifier has to pay. This captures the idea that the certifier has to invest resources in designing inspections and technology to ensure that the validity of their certification. Finally, we assume that the certifier maximizes profit.

We analyze this problem through a mechanism-design approach and show that it is mathematically equivalent to the problem analyzed in our main analysis. Therefore, when the verification costs for the certifier satisfy our strong monotonicity property, Theorem 2 implies that every optimal quality-certification policy must involve finitely many grades, a prediction that is in line with most certification policies that we practically observe. Concretely, in the context of the earlier examples: the European New Car Assessment Program gives a discrete star rating out of five for the crash safety of new vehicles; the New York City Health department gives grades of A, B, and C for restaurants' food hygiene; and the United Kingdom Office for Standards in Education operates a four-point grading system after school inspections.

6 Beyond Strongly Monotone Costs

While we have shown that the coarse contracting prediction holds for many reasonable costs, we have not yet demonstrated that the conclusion is non-trivial in general. That is, we have not shown that there exist reasonable costs of contractibility that do *not* deliver the prediction of coarse contracts. In this final section, we discuss the boundaries of the coarse contracting prediction under alternative costs. We show that: (i) costs motivated solely by enforcing contracts *ex post* do not deliver coarse contracts, (ii) some costs motivated by writing clauses deliver coarse contracts while some do not, and (iii) menu costs do not necessarily deliver coarse contracts.

6.1 Costly Enforcement

We have interpreted costly contractibility as something borne *ex ante*, or before the agent takes an action. As we argued above, this could capture the principal's difficulties in describing different outcomes in a legally precise way. A different model would instead focus on costs borne *ex post*, or after the agent takes (or attempts to take) an action. This could capture the expected cost of detecting a deviation from the contract or litigating a deviation from the contract, more reminiscent of the classic literature studying costly verification.

To shed light on the difference between these models, we show how an *ex post* variant of our costs of distinguishing outcomes (Definition 1) leads to optimally *complete* contracts. The reason turns out to be simple: *ex post* costs are equivalent to additional production costs for the principal, which do not by themselves induce coarseness. We use this observation to discuss the applicability of our coarse-contracts prediction to scenarios in which one might expect more costs to be borne *ex ante* vs. *ex post*.

To describe this scenario mathematically, we let $\Phi = \{\phi : \Theta \to X\}$ be the set of increasing assignment rules, define the generalized inverse $\phi^{-1}(x) = \inf\{\theta \in \Theta : \phi(\theta) \geq x\}$, and define an action-dependent cost as one that can be expressed by a function $\Gamma : \mathcal{C} \times \Phi \to \mathbb{R}$. In this context, we define *ex post* costs of distinguishing:

Definition 5 (*Ex Post* Costs of Distinguishing Outcomes)**.** *Fix $\phi \in \Phi$ and define the push-forward measure of F to X as $F_\phi(x) = F(\phi^{-1}(x))$. The ex post cost of distinguishing is:*

$$\Gamma(C, \phi) = \int_X \hat{d}(C(x), X \setminus C(x)) \mathrm{d}F_\phi(x) \tag{36}$$

where $\hat{d}$ is as in Definition 1.

This differs from the *ex ante* cost of distinguishing as the total cost is evaluated under the distribution of x that obtains *ex post*, which is F_ϕ, rather than under the uniform measure, which is relevant when costs are borne *ex ante*. We now give an example of such a cost that builds on Example 1, but differs critically in the timing of events:

Example 3 (Discrete *Ex Post* Costs of Distinguishing Outcomes)**.** Consider the discrete metric introduced by Example 1. The *ex post* cost of distinguishing is given by:

$$\begin{aligned} \Gamma(C, \phi) &= \int_X \mu(X \setminus C(x)) \mathrm{d}F_\phi(x) = \int_\Theta \mu(X \setminus C(\phi(\theta))) \mathrm{d}F(\theta) \\ &= \int_\Theta \left(\overline{x} - \overline{\delta}(\phi(\theta)) + \underline{\delta}(\phi(\theta))\right) \mathrm{d}F(\theta) \end{aligned} \tag{37}$$

Which has the same integrand as Example 1, but instead integrates over the space of types with respect to the distribution of types rather than the space of allocations with respect to the uniform measure over actions. △

This example hints at a fundamental difference between *ex ante* and *ex post* costs of distinguishing outcomes: *ex post* costs are linearly separable over types while *ex ante* costs are not. The only thing that ties different types together is $\overline{\delta}$, as this is common to all types. However, under any Obedient mechanism, we know that $\phi(\theta) = \overline{\delta}(\phi(\theta))$. Thus, fixing ϕ, we have pinned down $\overline{\delta}$, and the induced cost function is linearly separable over types in their final actions. Hence, it is as if *ex post* costs of distinguishing actions are a production cost. This logic yields the following result, which implies that optimal contracts are never coarse under *ex post* costs:

Proposition 8 (*Ex Post* Costs Do Not Yield Coarse Contracts)**.** *Under ex post costs of distinguishing outcomes, free disposal, $C(x) = [0, x]$ for all $x \in X$, is optimal.*

Proof. See Appendix A.13. □

Thus, the optimal contract makes additional usage impossible $\overline{\delta}(x)$ but allows for the possibility of free disposal; this generates no loss in value for the principal but economizes on the costs of monitoring for disposal, which they know will never actually happen as the agent has a positive marginal value for all units of the good.

Realistic scenarios might be described as a combination of both *ex ante* and *ex post* costs of distiguishing. That is, a principal may both have to write a contract that precisely distinguishes actions and enforce it. We might model such scenarios by allowing the "true" cost faced by the principal to be a weighted sum of *ex ante* and *ex post* costs. For instance, in the context of the aforementioned examples, we could have:

$$\Gamma(C, \phi) = \nu \Gamma^{EA}(C) + \Gamma^{EP}(C, \phi) \tag{38}$$

for some $\nu \in \mathbb{R}_+$, where Γ^{EA} is some cost of distinguishing outcomes and Γ^{EP} is some *ex post* cost of distinguishing outcomes. Provided that $\nu > 0$, Theorem 2 holds and optimal contracts are coarse. Moreover, the bound in Theorem 2 decreases in ν.

Thus, our theory predicts coarser contracts in scenarios in which defining outcomes *ex ante* is particularly difficult compared to scenarios in which outcomes are very well defined but merely difficult to detect, punish, or enforce. The first category might include variable quality services like hotel stays, vehicle rentals, or management consulting. What these scenarios have in common is that "success," "quality," and/or "damage" are inherently difficult to define. While there are surely issues also with enforcement, at least some meaningful fraction of costs comes from designing the contract in the first place ($\nu > 0$). The second category might include metered utilities, in which the sole difficulty is the precise measurement of *ex post* usage. This may include cases like the electrical service contracts which motivate Wilson's (1989) analysis.

6.2 Clause-Based Costs

One natural source for costly contractibility is a fixed cost for enumerating each relevant outcome. We call any cost that depends on the contractible set only via its cardinality a *clause-based cost*. These costs do not satisfy strong monotonicity, because they are insensitive to the structure of contractibility. Nevertheless, it is possible to recover the spirit of strong monotonicity and derive a sufficient condition for optimally coarse contracts in this class. This will highlight that the prediction of incompleteness is sensitive to the parametric structure of clause-based costs: while coarseness is guaranteed for any distance-based cost, not all clause-based costs will deliver incomplete contracts.

Definition 6 (Clause-Based Costs)**.** *A contractibility cost is clause-based if, for any* $\overline{D} \in \mathcal{D}$*, we can write* $\Gamma(\overline{D}) = \hat{\Gamma}(n(\overline{D}))$*, where* $n(\cdot)$ *denotes the cardinality of a set and* $\hat{\Gamma} : \overline{\mathbb{N}} \to \overline{\mathbb{R}}$ *is a* strictly increasing *cost defined on this cardinality with the normalization that* $\hat{\Gamma}(2) = 0$ *(as our axioms imply that all* $\overline{D}$ *contain* $\{0, \overline{x}\}$*).*

For such clause-based costs, we will discipline the rate at which marginal costs of adding a clause decline to zero with the following definition:

Definition 7 (Clause Strong Monotonicity)**.** *We say that* Γ*, with induced* $\hat{\Gamma}$*, is* $\beta-$*clause strongly monotone if there exist* β *and* $\epsilon > 0$ *such that:*

$$\liminf_{n \to \infty} \; (\hat{\Gamma}(K+1) - \hat{\Gamma}(K))K^{\beta} \geq \epsilon \tag{39}$$

We illustrate clause-based costs and $\beta-$clause strong monotonicity in the following examples:

Example 4. Consider first the linear cost $\hat{\Gamma}(K) = K - 2$, studied by Battigalli and Maggi (2002) in their analysis of optimally incomplete contracts. This cost is $\beta-$clause strongly monotone if and only if $\beta \geq 0$. As another example, the cost $\hat{\Gamma}(K) = \frac{1}{2} - \frac{1}{K}$, which is bounded and converges to $\frac{1}{2}$ as the number of clauses become infinite. This cost is $\beta-$clause strongly monotone if and only if $\beta \geq 1$. Finally, a cost with increments that are some power of the number of clauses written so far, *i.e.,* $\hat{\Gamma}(K) - \hat{\Gamma}(K-1) = (K-2)^{\alpha}$ for some $\alpha \in \mathbb{R}$, yields a cost $\hat{\Gamma}(K) = \sum_{k=1}^{K-2} k^{-\alpha}$. This cost is $\beta-$clause strongly monotone if and only if $\beta \geq \alpha$. △

It is obvious that any unbounded clause-based cost, such as the linear cost, implies a coarse contract. It is less obvious when coarseness will be obtained for bounded clause-based costs, such as $\hat{\Gamma}(K) = \frac{1}{2} - \frac{1}{K}$. The next proposition ties the optimality of coarse contracts to $\beta-$clause strong monotonicity.

Proposition 9. *If* Γ *is clause-based and* $\beta-$*clause strongly monotone for* $\beta < 3$*, then every optimal contractibility support is finite with* $|\overline{D}^*| \leq 2 + \left\lfloor \left(\frac{6\bar{J}_{xx}^2 \bar{f}}{\epsilon \underline{J}_{x\theta}} \right)^{\frac{1}{3-\beta}} \right\rfloor$.

Proof. See Appendix A.14. □

The proof of this result follows from three steps. We first observe that if any infinite-support contract is optimal, so too is perfect contractibility—this has the same cost, but higher benefits. We next show that the benefits of perfect contractibility relative to an evenly-spaced grid of sparse contracting points is second-order in the width of the grid. This is exactly consistent with integrating the third-order bound of Lemma 2's "individual grid cells" over the entire domain X. This step in the proof of Proposition 9 has precedents in the literature. In particular, Wilson (1989) shows under perfect information that coarsening the domain of contractibility into uniform cells is second-order in the length of the grid. Extending these ideas, Bergemann, Yeh, and Zhang (2021) show that this remains true with private information. By contrast, our earlier arguments away from clause-based costs that must consider set-valued perturbations are without precedent to our knowledge. The third step shows that, when costs are $\beta-$clause strongly monotone for $\beta < 3$, there is a fine enough grid that beats perfect contractibility, thereby contradicting that any infinite-support contractibility is optimal. Finally, the bound follows from using a similar argument to contradict the optimality of points spaced too close together.

To illustrate this result, let us return to the example $\hat{\Gamma}(K) = \frac{1}{2} - \frac{1}{K}$. As this cost is β-clause strongly monotone for $\beta = 1 < 3$, we have that the optimal contract is necessarily coarse. Moreover, we have a bound on the number of elements which is given by $2 + \left\lfloor \sqrt{\frac{6\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}} \right\rfloor$. Thus, despite the fact that the marginal cost of additional clauses converges to zero, there is nevertheless a finite bound on the number of clauses.

When a cost function is not β-clause strongly monotone for $\beta < 3$, it is possible that an optimal contract will be complete. Indeed, in our application from Section 5, it is easy to verify that a cost of the form $\hat{\Gamma}(K) = \sum_{k=1}^{K-2} k^{-\alpha}$ for $\alpha > 3$ would yield an optimally complete contract. This highlights that certain costs of contractibility could yield a prediction of complete contracts. Thus, the issue of whether contracts are complete hinges on the cost function and its economic properties.

We finally observe that the characterization of the optimally chosen actions in the clause-based case is much the same as the characterization in Proposition 5. The only difference is that the marginal cost term in the right of Equation 27 is zero, as there is no contractibility cost of changing the value of any x_k. Bergemann, Shen, Xu, and Yeh (2012) have previously studied this problem of optimally spacing grid points given an exogenous constraint in the setting with linear-quadratic preferences and found the same first-order condition that we have in this case. Relative to this work, we have shown how to optimally choose such points in the presence of costs and, more substantively, how many points the principal should elect to choose.

6.3 Menu Costs

Another natural source of non-production costs for the principal are *menu costs* of various forms: that is, costs of putting products up for sale rather than costs of delivering the final product *per se*. A rich class of menu costs can be described by the expanded class of costs $\Gamma(C, \phi)$. For example, our baseline costs of distinguishing actions can be re-interpreted as a type of menu cost that leads to coarse contract. Clause-based costs, which depend on the cardinality of the menu, can be interpreted as a menu cost that may or may not induce coarse contracts (Section 6). In general, however, not all reasonable menu costs induce coarse contracts, as we argue in the following eample.

Example 5 (Menu Costs from Maximum Quality)**.** Consider the cost function studied by Sartori (2021), in which the indirect cost of a menu corresponds to the cost of the most expansive quality to be produced. Formally, fix a continuous and increasing baseline cost function $c : X \to \overline{\mathbb{R}}$ and define

$$\Gamma(C, \phi) = \max_{x \in \phi(\Theta)} c(x) \tag{40}$$

The interpretation of this cost function is that the monopolist invests ex-ante in a maximum level of quality x of the good and then they are able to freely garble this quality by offering any smaller level $y \leq x$. It is easy to see that Γ_c does not satisfy the strong monotonicity properties of Section 4, since it depends only on the largest (relevant) item on the menu. In fact, the analysis in Sartori (2021) shows that, in general, the optimal menu offered by the monopolist is not coarse and involves a continuum of differentiated qualities. △

7 Conclusion

In this paper, we introduced a model of when and why incomplete contracts arise in an environment with costly contractibility. First, we studied contracting with fixed restrictions on what actions are contractible and we characterized implementable and optimal mechanisms. Second, we studied the problem of a principal that chooses the extent of contractibility subject to a cost. The cost, as we illustrated via examples, models the principal's difficulty in specifying and describing what outcomes are contractible. We then showed our main result: if the costs of contracting on outcomes are *strongly monotone* in a way that we formalized, then optimal contracts are coarse. Moreover, we derive a bound on the number of items in the optimal menu and derived necessary conditions that discipline which actions are contractible. Finally, we applied this model to study when and why incomplete contracts would arise in a monopoly pricing problem à la Mussa and Rosen (1978) that features costly contractibility. We showed that optimal menus feature uniformly spaced quality tiers and provided a formula for the number of tiers that featured the same comparative statics as our general bound. In this context, incomplete information induces more incomplete contracts relative to the complete information benchmark.

A Proofs

A.1 Proof of Lemma 1

(1) $\implies$ (2). Let $C : X \rightrightarrows X$ be a regular contractibility correspondence and define $\underline{\delta}(y) = \min C(y)$ and $\overline{\delta}(y) = \max C(y)$ for all $y \in X$. By Axiom 5, $\overline{\delta}$ and $\underline{\delta}$ exist. By Axiom 3, we have that $\underline{\delta}$ and $\overline{\delta}$ are increasing functions. By Axiom 1, we know that $y \geq \underline{\delta}(y)$ and $y \leq \overline{\delta}(y)$ for all y (part (ii) of 2). Moreover, by Lemma 17.29 in Aliprantis and Border (2006), $\overline{\delta}$ is lower semicontinuous and $\underline{\delta}$ is upper semicontinuous.

We now show part (i) of 2, that $C(y) = [\underline{\delta}(y), \overline{\delta}(y)]$. Assume by contradiction there exists some $y \in X$ and $x \in [\underline{\delta}(y), \overline{\delta}(y)]$ such that $x \notin C(y)$. Consider first the case where $x < y$. By the definition of $\underline{\delta}$, $\underline{\delta}(y) \in C(y)$ and $\underline{\delta}(y) < x$. As $x < y$, by Axiom 3, we have that $C(x) \leq_{SSO} C(y)$. Thus, as $x \in C(x)$ and $\underline{\delta}(y) \in C(y)$, we know that $\max\{x, \underline{\delta}(y)\} = x \in C(y)$. This is a contradiction. Consider now the case where $y < x$. Again, $\overline{\delta}(y) \in C(y)$ and $x < \overline{\delta}(y)$. By Axiom 3, we have that $\min\{x, \overline{\delta}(y)\} = x \in C(y)$. This is a contradiction.

We next show parts (iii), (iv), and (v) of 2. Fix $x, y \in X$ and assume that $x \in [\underline{\delta}(y), \overline{\delta}(y))$, which implies $x \in C(y)$. We start with part (iii), and mirror the argument for part (iv). Suppose $x < y$. As C is monotone, we know that $\underline{\delta}(x) \leq \underline{\delta}(y)$. Suppose by contradiction that $\underline{\delta}(x) < \underline{\delta}(y)$. But then, given the other properties of δ, for all $z \in (\underline{\delta}(x), \underline{\delta}(y))$ we would have that $z \in C(x)$ but $z \notin C(y)$, which contradicts Axiom 2. For part (iv), consider the same scenario but reversed. Suppose $x > y$. As C is monotone, we know that $\overline{\delta}(x) \geq \overline{\delta}(y)$. Imagine this held at strict inequality. Then there would exist $z \in (\overline{\delta}(y), \overline{\delta}(x))$ such that $z \in C(y)$ and $z \notin C(x)$, while $y \in C(x)$. This violates Axiom 2. It is immediate that $\overline{\delta}(0) = 0$ by Axiom 4 as $C(0) = \{0\}$.

(2) $\implies$ (3). We start with an ancillary lemma.

Lemma 5 (Fixed Point Lemma)**.** *Consider two functions $\underline{\delta}(x)$ and $\overline{\delta}(x)$ as in point (2) of Lemma 1. Then for all $\underline{z} \in \underline{\delta}(X)$ and $\overline{z} \in \overline{\delta}(X)$, it holds $\underline{\delta}(\underline{z}) = \underline{z}$ and $\overline{\delta}(\overline{z}) = \overline{z}$.*

Proof. Let $\underline{z} = \underline{\delta}(x)$ for some $x \in X$. It follows that $\underline{z} \in [\underline{\delta}(x), x]$. If $\underline{z} = x$, then we have that $\underline{\delta}(\underline{z}) = \underline{\delta}(x) = \underline{z}$. Alternatively, if $\underline{z} < x$, given property (iii) in part (2) of Lemma 1, we must have $\underline{\delta}(\underline{z}) = \underline{\delta}(x) = \underline{z}$. The proof for $\overline{z} \in \overline{\delta}(X)$ is symmetric, using property (iv) in part (2) of Lemma 1. □

Let $\underline{\delta}$ and $\overline{\delta}$ be as in (2) and define $\underline{D} = \underline{\delta}(X)$ and $\overline{D} = \overline{\delta}(X)$. First, observe that

$$\max_{z \leq x: z \in \underline{D}} z = \max_{z \leq x: z \in \underline{\delta}(X)} z \geq \underline{\delta}(x) \tag{41}$$

by construction. Let $\underline{z} = \max_{z \leq x: z \in \underline{D}} z$ and assume by contradiction that $\underline{z} > \underline{\delta}(x)$. If $\underline{z} = x$, then $x \in \underline{\delta}(X)$ and by Lemma 5 we have that $x = \underline{\delta}(x) < \underline{z}$, yielding a contradiction. If instead $\underline{z} < x$, then by Lemma 5 and the property (iii) of $\underline{\delta}$, we have $\underline{z} = \underline{\delta}(\underline{z}) = \underline{\delta}(x)$, yileding a contradiction. With this, we conclude that $\underline{z} = \underline{\delta}(x)$. With symmetric steps, we can show that $\min_{z \geq x: z \in \overline{D}} z = \overline{\delta}(x)$. Next, observe that necessarily we have $\underline{\delta}(0) = 0$, $\overline{\delta}(\overline{x}) = \overline{x}$, and $\overline{\delta}(0) = 0$ proving that $0 \in \underline{D}$ and $0, \overline{x} \in \overline{D}$. Finally, we need to show that $\underline{D}$ and $\overline{D}$ are closed. Take a sequence $z_n \in \underline{D}$ such that $z_n \to z$. Given that X is closed, we have that $z \in X$ and therefore $\underline{\delta}(z) \leq z$. Given that every z_n is in $\underline{D}$, Lemma 5 implies that $\underline{\delta}(z_n) = z_n$ for all n. Given that $\underline{\delta}$ is upper semicontinuous, it follows that

$$z = \lim_{n \to \infty} z_n = \lim_{n \to \infty} \underline{\delta}(z_n) \leq \underline{\delta}(z)$$

which implies that $z = \underline{\delta}(z)$ (as $z \geq \underline{\delta}(z)$) and therefore that $z \in \underline{D}$. This shows that $\underline{D}$ is closed. A symmetric argument shows that $\overline{D}$ is closed.

(3) $\Longrightarrow$ (2). Let $\underline{D}$ and $\overline{D}$ be as in (3) and define C as in equation 1. We want to show that C is a regular contractibility correspondence. Toward this goal define $\underline{\delta}(x) = \max_{z \leq x: z \in \underline{D}} z$ and $\overline{\delta}(x) = \min_{z \geq x: z \in \overline{D}} z$ and observe that $C(x) = [\underline{\delta}(x), \overline{\delta}(x)]$. It is immediate to see that both these functions are monotone increasing, such that $\underline{\delta}(x) \leq x \leq \overline{\delta}(x)$, and respectively upper semicontinuous and lower semicontinuous by Lemma 17.30 in Aliprantis and Border (2006). To see this, observe that the correspondences $x \rightrightarrows \{z \in \underline{D} : z \leq x\}$ and $x \rightrightarrows \left\{z \in \overline{D} : z \geq x\right\}$ are both upper hemicontinuous. Next, assume that $y \in [\underline{\delta}(x), x)$ and let $z = \underline{\delta}(x)$. We have $\underline{\delta}(y) \leq z$ by monotonicity. Moreover, by assumption $z \leq y$ and $z \in \underline{D}$, so that $z \leq \underline{\delta}(y)$ by definition. We then must have $z = \underline{\delta}(y)$. Symmetrically, assume that $y \in (x, \overline{\delta}(x)]$ and let $z = \overline{\delta}(x)$. We have $\overline{\delta}(y) \leq z$ by monotonicity. Moreover, by assumption $z \geq y$ and $z \in \overline{D}$, so that $z \geq \overline{\delta}(y)$ by definition. We then must have $z = \overline{\delta}(y)$. Finally, as $0 \in \overline{D}$, we have that $\overline{\delta}(0) = 0$.

(2) $\Longrightarrow$ (1). Fix $\underline{\delta}$ and $\overline{\delta}$ that satisfy (2). $C(y) = [\underline{\delta}(y), \overline{\delta}(y)]$ is regular. C is reflexive since because of (ii), closed because the intervals of the construction are closed, and monotone because $\underline{\delta}, \overline{\delta}$ are monotone. To show transitivity, consider $x \in C(y)$ and, first, the case $x < y$. From (iii), we have $\underline{\delta}(x) = \underline{\delta}(y)$. Moreover, from monotonicity, $\overline{\delta}(x) \leq \overline{\delta}(y)$. Therefore, $C(x) \subseteq C(y)$. Next, consider the case where $x > y$. From (iv), we have $\overline{\delta}(x) = \overline{\delta}(y)$. Moreover, from monotonicity, $\underline{\delta}(x) \geq \underline{\delta}(y)$. Therefore, $C(x) \subseteq C(y)$. Moreover, if $x = y$, clearly $C(x) \subseteq C(y)$. Given that these arguments hold for any x, this shows transitivity. Finally, as $\overline{\delta}(0) = \underline{\delta}(0)$, we have that $C(0) = \{0\}$, which establishes excludability. These arguments together establish that C is regular.

A.2 Proof of Theorem 1

We prove the result in three parts. First, we present a characterization of implementable allocations. Second, we use this characterization to derive the principal's control problem. Third, we solve this control problem for the optimal contract.

Part 1: Implementation

We begin by establishing a general taxation principle with partial contractibility. Given a regular contracting correspondence C, we say that $T : X \to \bar{\mathbb{R}}$ is monotone with respect to C if $T(x) \geq T(y)$ for all $x, y \in X$ such that $y \in C(x)$. We now show monotonicity of the tariff with respect to C is necessary and sufficient for implementability (Definition 2).

Lemma 6 (C-Monotone Taxation Principle)**.** *Fix a regular contractibility correspondence C. A final outcome function ϕ is implementable given C if and only if there exists a tariff $T : X \to \bar{\mathbb{R}}$ that is monotone with respect to C and such that:*

$$\phi(\theta) \in \arg\max_{x \in X} \{u(x, \theta) - T(x)\} \tag{42}$$

and $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$. In this case, ϕ is supported by $\xi = \phi$ and T.

Proof. **(Only if)** We begin by proving the necessity of the existence of a monotone tariff with respect to C. Suppose that ϕ is implementable. It follows that there exists (ξ, T) that support ϕ. In particular, observe that (O) implies that $\phi(\theta) \in C(\xi(\theta))$ for all $\theta \in \Theta$. Next define $\hat{T} : X \to \bar{\mathbb{R}}$ as:

$$\hat{T}(x) = \inf_{y \in X} \{T(y) : x \in C(y)\} \tag{43}$$

We next show that ϕ is also supported by $(\phi, \hat{T})$. By (O) of (ϕ, ξ, T), we have

$$u(\phi(\theta), \theta) \geq u(x, \theta) \tag{44}$$

for all $x \in C(\phi(\theta)) \subseteq C(\xi(\theta))$ (by transitivity) and for all $\theta \in \Theta$, yielding (O) of $(\phi, \phi, \hat{T})$. By (IR) of (ϕ, ξ, T) and the definition of $\hat{T}$, we have

$$u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \geq u(\phi(\theta), \theta) - T(\xi(\theta)) \geq 0 \tag{45}$$

for all $\theta \in \Theta$, yielding (IR) of $(\phi, \phi, \hat{T})$. Next, assume toward a contradiction that $(\phi, \phi, \hat{T})$ does not satisfy (IC), that is, there exists $\theta \in \Theta$ and $y \in X$ such that

$$\max_{x \in C(y)} u(x, \theta) - \hat{T}(y) > u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \tag{46}$$

Thus:

$$\begin{aligned} &\max_{x \in C(y)} u(x, \theta) - T(y) > u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \\ &\geq u(\phi(\theta), \theta) - T(\xi(\theta)) = \max_{x \in C(\xi(\theta))} u(x, \theta) - T(\xi(\theta)) \end{aligned} \tag{47}$$

The second inequality follows from the construction of $\hat{T}$. The final equality follows as (ϕ, ξ, T) satisfies (O). However, the previous inequality yields a contradiction of (IC) of (ϕ, ξ, T), proving that $(\phi, \phi, \hat{T})$ satisfies (IC). This shows that $(\phi, \phi, \hat{T})$ is implementable, hence that Equation 42 holds and that $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$.

Finally, we argue that $\hat{T}$ is monotone with respect to C. Fix $x, y \in X$ such that $y \in C(x)$. By Transitivity of C we have

$$\{\hat{x} \in X : x \in C(\hat{x})\} \subseteq \{\hat{x} \in X : y \in C(\hat{x})\} \tag{48}$$

yielding that $\hat{T}(y) \leq \hat{T}(x)$, as desired.

(If) We now establish sufficiency. Suppose that there exists a tarrif $T : X \to \bar{\mathbb{R}}$ that is monotone with respect to C and such that Equation 42 holds and $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$. We will show that (ϕ, ϕ, T) is implementable. (IR) is immediately satisfied. Next, we show that (IC) is satisfied. Suppose, toward a contradiction, that it were not. That is, there exist $\theta \in \Theta$, $y \in X$, and $x \in C(y)$ such that

$$u(x, \theta) - T(y) > \max_{\hat{x} \in C(\phi(\theta))} u(\hat{x}, \theta) - T(\phi(\theta)) \geq u(\phi(\theta), \theta) - T(\phi(\theta)) \tag{49}$$

But then, we have the following contradiction of monotonicity of T in C:

$$u(x, \theta) - T(y) > u(\phi(\theta), \theta) - T(\phi(\theta)) \geq u(x, \theta) - T(x) \tag{50}$$

where the second inequality uses the fact that $\phi(\theta)$ solves the program in Equation 42. Finally, we show that (O) is satisfied. Toward a contradiction, assume that it were not.That is, there exists $\theta \in \Theta$ and $x \in C(\phi(\theta))$ such that:

$$u(x, \theta) > u(\phi(\theta), \theta) \tag{51}$$

However, by monotonicity of T in C, we know that $T(\phi(\theta)) \geq T(x)$. Thus,

$$u(x,\theta) - T(x) > u(\phi(\theta),\theta) - T(\phi(\theta)) \tag{52}$$

yielding a contradiction to IC, which we just showed. This proves sufficiency.

Finally, the fact that any implementable final outcome function can be implemented as part of an allocation (ϕ, ϕ, T) follows by the construction in the necessity part of our proof. □

With this taxation principle in hand, we now characterize implementation:

Lemma 7 (Implementation)**.** *A final outcome function ϕ is implementable under C, associated with upper and lower image sets $(\overline{D}, \underline{D})$, if and only if it is monotone increasing and such that: (i) if agent preferences are monotone increasing, then $\phi(\Theta) \subseteq \overline{D}$, (ii) if preferences are monotone decreasing, then $\phi(\Theta) \subseteq \underline{D}$. Moreover, ϕ is supported by $\xi = \phi$ and tariff:*

$$T(x) = T(0) + u(x, \phi^{-1}(x)) - \int_0^{\phi^{-1}(x)} u_\theta(\phi(s), s)\,\mathrm{d}s \tag{53}$$

where $\phi^{-1}(s) = \inf\{\theta \in \Theta : \phi(\theta) \geq s\}$.

Proof. **(Only If for First Part)** If ϕ is implementable, then there exists (ξ, T) that support ϕ. By Lemma 6, we may take that $\xi = \phi$. By (IC) and Lemma 6, there exists a transfer function $t : \Theta \to \mathbb{R}$ such that $u(\phi(\theta), \theta) - t(\theta) \geq u(\phi(\theta'), \theta) - t(\theta')$ for all $\theta, \theta' \in \Theta$. As u is strictly single-crossing, Proposition 1 in Rochet (1987) then implies that ϕ is monotone. Without loss of generality, consider the case with monotone increasing preferences and toward a contradiction suppose that $\phi(\theta) \notin \overline{D}$. Deviating to $\overline{\delta}(\phi(\theta)) > \phi(\theta)$ is a strict improvement for the agent. Thus, if ϕ is implementable, then it is monotone, and $\phi(\Theta) \in \overline{D}$ (or $\phi(\Theta) \in \underline{D}$ with montone decreasing preferences) holds.

(If For First Part) Without loss of generality, we gain prove this part for he case with monotone increasing preferences. Now suppose that $\phi(\theta) \in \overline{D}$ holds for all $\theta \in \Theta$ and ϕ is monotone increasing. Define the function $t : \Theta \to \mathbb{R}$ as

$$t(\theta) = K + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s \tag{54}$$

for some $K \leq 0$, and the tariff $T : X \to \overline{\mathbb{R}}$ as

$$T(x) = \inf_{\theta' \in \Theta} \{t(\theta') : x \in C(\phi(\theta'))\} \tag{55}$$

Fix $x, y \in X$ such that $y \in C(x)$. By Transitivity, for all $\theta \in \Theta$, if $x \in C(\phi(\theta))$, then $y \in C(\phi(\theta))$. This shows that

$$\{\theta \in \Theta : x \in C(\phi(\theta))\} \subseteq \{\theta \in \Theta : y \in C(\phi(\theta))\} \tag{56}$$

Therefore, applying the construction of T, $T(x) \geq T(y)$. Thus, T is monotone with respect to C.

As T is monotone with respect to C, if we can show that $\phi(\theta) \in \arg\max_{x \in X}\{u(x, \theta) - T(x)\}$ and $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$, then we have shown by Lemma 6 that ϕ is implementable.

We start with the second condition. For every $\theta \in \Theta$, we have

$$u(\phi(\theta), \theta) - T(\phi(\theta)) \geq u(\phi(\theta), \theta) - t(\theta) = \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s - Z \tag{57}$$

Note that the right-hand side of this last equation is monotone increasing in θ since it is continuously differentiable with derivative $u_\theta(\phi(\theta), \theta) \geq 0$ for all $\theta \in \Theta$, owing to the fact that u is monotone increasing over Θ. Given that $Z \leq 0$, we have that $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$.

We are left to prove that (ϕ, T) satisfy Equation 42. We first prove that, for all $\theta, \theta' \in \Theta$:

$$u(\phi(\theta), \theta) - t(\theta) \geq \max_{x \in C(\phi(\theta'))} u(x, \theta) - t(\theta') \tag{58}$$

This is a variation of the standard reporting problem under consumption function ϕ and transfers t, where each agent, on top of misreporting their type, can also consume everything allowed by C. Violations of this condition can take two forms. First, an agent of type θ could report type θ' and consume $x = \phi(\theta')$. We call this a single deviation. Second, an agent of type θ could report type θ' and consume $x \in C(\phi(\theta')) \setminus \{\phi(\theta')\}$. We call this a double deviation. Under our construction of transfers t and monotonicity of ϕ, by a standard mechanism-design argument (*e.g.,* Nöldeke and Samuelson, 2007), there is no strict gain to any agent of reporting θ' and consuming $x = \phi(\theta')$. Thus, there are no profitable single deviations under (ϕ, t).

We now must rule out double deviations. Suppose that θ imitates θ' and plans to take final action $x \neq \phi(\theta')$. As $\phi(\theta') \in \overline{D}$ (in the monotone increasing case), $x < \phi(\theta')$. But in that case, simply taking action $\phi(\theta')$ is better. But then this is a single deviation, which we have ruled out. The same logic applies in the monotone decreasing case.

To derive the tariff, we can simply set $T(x) = t(\phi^{-1}(x))$. This yields the claimed formula.

□

Part 2: Control Problem

We now use this characterization of implementation to turn the principal's problem into an optimal control problem:

Lemma 8. *When agents have monotone increasing preferences, any optimal final outcome function solves:*

$$\begin{aligned} \mathcal{J}(\overline{D}) := \max_{\phi} \quad & \int_{\Theta} J(\phi(\theta), \theta)\, \mathrm{d}F(\theta) \\ s.t. \quad & \phi(\theta') \geq \phi(\theta), \phi(\theta) \in \overline{D}, \quad \theta, \theta' \in \Theta : \theta' \geq \theta \end{aligned} \tag{59}$$

When agents have monotone decreasing preferences, replace $\overline{D}$ with $\underline{D}$.

Proof. We begin by eliminating the proposed allocation and transfers from the objective function of the seller. From the proof of Lemma 7, we have that transfers for any incentive compatible triple (ξ, ϕ, t) are given by:

$$t(\theta) = Z + u(\phi(\theta), \theta) - \int_0^{\theta} u_{\theta}(\phi(s), s)\, \mathrm{d}s \tag{60}$$

for some constant $Z \in \mathbb{R}$. Thus, any ξ that supports ϕ leads to the same seller payoff and can therefore be made equal to ϕ without loss of optimality. Moreover, we know that ϕ being incentive compatible is equivalent to ϕ being monotone increasing and $\phi(\theta) \in \overline{D}$.

Plugging in the expression (60), we can simplify the expression for the seller's total transfer revenue as the following:

$$\begin{aligned} \int_{\Theta} t(\theta)\, \mathrm{d}F(\theta) &= \int_{\Theta} \left(Z + u(\phi(\theta), \theta) - \int_0^{\theta} u_{\theta}(\phi(s), s)\, \mathrm{d}s \right) \mathrm{d}F(\theta) \\ &= \int_{\Theta} (Z + u(\phi(\theta), \theta))\, \mathrm{d}F(\theta) - \int_0^1 \int_0^{\theta} u_{\theta}(\phi(s), s)\, \mathrm{d}s\, \mathrm{d}F(\theta) \end{aligned} \tag{61}$$

Using this expression for total transfer revenue, and the characterization of implementation from Lemma 7, we write the seller's problem as

$$\begin{aligned} \max_{\phi, Z} \quad & \int_{\Theta} \left(\pi(\phi(\theta), \theta) + Z + u(\phi(\theta), \theta) - \int_0^{\theta} u_{\theta}(\phi(s), s)\, \mathrm{d}s \right) \mathrm{d}F(\theta) \\ \text{s.t.} \quad & \phi(\theta') \geq \phi(\theta), \ \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \\ & u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^{\theta} u_{\theta}(\phi(s), s)\, \mathrm{d}s \right) \geq 0 \quad \forall \theta \in \Theta \end{aligned} \tag{62}$$

We further simplify this by applying integration by parts on the double integral of $u_{\theta}(\phi(s), s)$

over θ and s:

$$\begin{aligned}\int_0^1 \int_0^\theta u_\theta(\phi(s), s) \, \mathrm{d}s \, \mathrm{d}F(\theta) &= \left[F(\theta) \int_0^\theta u_\theta(\phi(s); s) \, \mathrm{d}s \right]_0^1 - \int_0^1 F(\theta) u_\theta(\phi(\theta), \theta) \, \mathrm{d}\theta \\ &= \int_0^1 (1 - F(\theta)) u_\theta(\phi(\theta), \theta) \, \mathrm{d}\theta \\ &= \int_0^1 \frac{(1 - F(\theta))}{f(\theta)} u_\theta(\phi(\theta), \theta) \, \mathrm{d}F(\theta)\end{aligned} \tag{63}$$

Plugging into the seller's objective, we find that the principal solves:

$$\begin{aligned}\max_{\phi, Z} \quad & \int_\Theta (J(\phi(\theta)) + Z) \, \mathrm{d}F(\theta) \\ \text{s.t.} \quad & \phi(\theta') \geq \phi(\theta), \ \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \\ & u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) \, \mathrm{d}s \right) \geq 0 \quad \forall \theta \in \Theta\end{aligned} \tag{64}$$

It follows that it is optimal to set $Z \in \mathbb{R}$ as large as possible such that:

$$V(\theta) = u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) \mathrm{d}s \right) \geq 0 \quad \forall \theta \in \Theta \tag{65}$$

We know that $V'(\theta) = u_\theta(\phi(\theta), \theta) \geq 0$ as $u(x, \cdot)$ is monotone over Θ. Thus, the tightest such constraint occurs when $\theta = 0$. Hence, the maximal Z must satisfy:

$$V(0) = -Z \geq 0 \tag{66}$$

This implies that Z is optimally 0 and ensures that the (IR) constraint holds for all types. Hence, the seller's program is:

$$\begin{aligned}\max_{\phi} \quad & \int_\Theta J(\phi(\theta), \theta) \, \mathrm{d}F(\theta) \\ \text{s.t.} \quad & \phi(\theta') \geq \phi(\theta), \ \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta\end{aligned} \tag{67}$$

This completes the proof. □

Part 3: The Optimal Contract

We first solve the pointwise problem in the control problem from Lemma 8 and then verify that this solution is monotone. The pointwise problem is $\max_{x \in \overline{D}} J(\phi(\theta), \theta)$, where the maximum exists as J is continuous and $\overline{D}$ is compact. As J is strictly quasi-concave, this maximum is either $\overline{\phi}(\theta)$ or $\underline{\phi}(\theta)$. When $\Delta J(\theta) > 0$, it is $\overline{\phi}(\theta)$. When $\Delta J(\theta) < 0$, it is $\underline{\phi}(\theta)$. When $\Delta J(\theta) = 0$, either is optimal. Thus, if it is monotone, the claimed solution is optimal (as it is supported on $\overline{D}$).

We next show that the claimed solution is monotone. Consider θ, θ' such that $\theta' > \theta$. If $\phi^*(\theta) = \underline{\phi}(\theta)$ and $\phi^*(\theta') = \underline{\phi}(\theta')$, then $\phi^*(\theta') \geq \phi^*(\theta)$ because $\underline{\phi}$ is increasing; similarly if $\phi^*(\theta) = \overline{\phi}(\theta)$ and $\phi^*(\theta') = \overline{\phi}(\theta')$. If $\phi^*(\theta) = \underline{\phi}(\theta)$ and $\phi^*(\theta') = \overline{\phi}(\theta')$, then $\phi^*(\theta') \geq \phi^*(\theta)$ because $\underline{\phi}$ is increasing and $\overline{\phi} \geq \underline{\phi}$. The only remaining case is if $\phi^*(\theta) = \overline{\phi}(\theta)$ and $\phi^*(\theta') = \underline{\phi}(\theta')$. Suppose toward a contradiction that $\overline{\phi}(\theta) > \underline{\phi}(\theta')$. We first observe that $\phi^P(\theta') < \overline{\phi}(\theta)$; otherwise $\underline{\phi}(\theta') = \max\{y \in \overline{D} : y \leq \phi^P(\theta')\} \geq \overline{\phi}(\theta)$. Moreover, since $\phi^P(\theta') \geq \phi^P(\theta)$, it must be the case that $\overline{\phi}(\theta') = \overline{\phi}(\theta)$. We next observe that $\phi^P(\theta) > \underline{\phi}(\theta')$; otherwise, $\overline{\phi}(\theta) = \min\{y \in \overline{D} : y \geq \phi^P(\theta)\} \leq \underline{\phi}(\theta')$. Again, since $\phi^P(\theta') \geq \phi^P(\theta)$, we must have $\underline{\phi}(\theta) = \underline{\phi}(\theta')$. But now we have the following contradiction: $J(\overline{\phi}(\theta), \theta) \geq J(\underline{\phi}(\theta), \theta)$ by optimality of $\overline{\phi}(\theta)$; $J(\overline{\phi}(\theta), \theta') > J(\underline{\phi}(\theta), \theta')$ by strict single crossing; $J(\overline{\phi}(\theta'), \theta') > J(\underline{\phi}(\theta'), \theta')$ because $\overline{\phi}(\theta) = \overline{\phi}(\theta')$ and $\underline{\phi}(\theta) = \underline{\phi}(\theta')$; but $J(\overline{\phi}(\theta'), \theta') \leq J(\underline{\phi}(\theta'), \theta')$ from the presumed optimality of $\underline{\phi}(\theta')$ for type θ'. This completes the argument that ϕ^* is monotone.

The claim that $\xi^* = \phi^*$ and the formula for the optimal tariff follow immediately from applying Lemma 7.

A.3 Proof of Proposition 1

We first derive the optimal allocation. As J is strictly single-crossing, $J(x_k, \theta) - J(x_{k-1}, \theta) = 0$ has no solution if and only if (i) $J(x_k, 0) - J(x_{k-1}, 0) > 0$ and (ii) $J(x_k, 1) - J(x_{k-1}, 1) < 0$. As J is strictly quasi-concave, if $J(x_k, 0) - J(x_{k-1}, 0) > 0$, then $J(\cdot, 0)$ is strictly increasing at x_{k-1}, and therefore at all x_j for $j \leq k-1$. Thus, if $J(x_k, 0) - J(x_{k-1}, 0) > 0$ holds for k, it holds for all $j \leq k$. Define $\underline{k} = \max\{k \in \{1, \ldots, K\} : J(x_k, 0) - J(x_{k-1}, 0) > 0\}$, with the convention that $\underline{k} = 1$ if this set is empty. Similarly, if $J(x_k, 1) - J(x_{k-1}, 1) < 0$, then $J(\cdot, 1)$ is strictly decreasing at x_k. Thus, if $J(x_k, 1) - J(x_{k-1}, 1) < 0$ holds for k, it holds for all $j \geq k$. Define $\overline{k} = \min\{k \in \{1, \ldots, K\} : J(x_k, 1) - J(x_{k-1}, 1) < 0\}$, with the convention that $\overline{k} = K$ if this set is empty. As J is strictly single crossing, $\overline{k} > \underline{k}$. We now have that $J(x_k, \theta) - J(x_{k-1}, \theta) = 0$ has a solution if and only if $k \in \{\underline{k}+1, \ldots, \overline{k}-1\}$ (if $\overline{k} = \underline{k} + 1$, then this set is empty). For all $k \geq \overline{k}$, we have that $\hat{\theta}_k = 1$. For all $k \leq \underline{k}$, we have that $\hat{\theta}_k = 0$. For all $k \in \{\underline{k}+1, \ldots, \overline{k}-1\}$, we have that $\hat{\theta}_k$ is the unique solution to $J(x_k, \hat{\theta}_k) = J(x_{k-1}, \hat{\theta}_k)$. As J is strictly quasi-concave, we know that $\phi^P(\hat{\theta}_k) \in (x_{k-1}, x_k)$, which implies that $\underline{\phi}(\hat{\theta}_k) = x_{k-1}$ and $\overline{\phi}(\hat{\theta}_k) = x_k$. Thus, by Theorem 1, we have that $\phi^*(\theta) = x_k$ for all $\theta \in (\hat{\theta}_k, \hat{\theta}_{k+1}]$.

We now derive the tariff that supports this allocation. Applying Equation 15 from Lemma

7, we have that:

$$\begin{aligned}T(x_k) &= u(x_k,\hat{\theta}_k) - \mathbb{I}[k\geq 2]\sum_{j=1}^{k-1}\int_{\hat{\theta}_j}^{\hat{\theta}_{j+1}} u_\theta(x_j,s)\,\mathrm{d}s \\ &= u(x_k,\hat{\theta}_k) - \mathbb{I}[k\geq 2]\sum_{j=1}^{k-1}\Big[u(x_j,\hat{\theta}_{j+1}) - u(x_j,\hat{\theta}_j)\Big] \\ &= u(x_1,0) + \mathbb{I}[k\geq 2]\sum_{j=2}^{k}\Big[u(x_j,\hat{\theta}_j) - u(x_{j-1},\hat{\theta}_j)\Big]\end{aligned} \tag{68}$$

where the second equality computes the integrals and the final equality telescopes the summation. Observing that $x_1 = 0$ and $u(0,0) = 0$ completes the proof.

A.4 Proof of Proposition 2

We first show that $\mathcal{D}$ is a compact set. The set of closed subsets of X is compact when endowed with the Hausdorff distance, so it is sufficient to show that $\mathcal{D}$ is closed. Take a sequence D_n inside $\mathcal{D}$ and assume that $D_n \to D$. We have that D is closed and given that $\overline{x} \in D_n$ for all n, it follows that $\overline{x} \in D$, yielding that $D \in \mathcal{D}$ and that the latter is closed.

By Lemma 8 and since $J(x,\theta)$ is strictly supermodular, we have

$$\mathcal{J}(\overline{D}) = \int_\Theta \mathcal{J}(\overline{D},\theta)\,\mathrm{d}F(\theta) \tag{69}$$

where

$$\mathcal{J}(\overline{D},\theta) := \max_{x\in\overline{D}} J(x,\theta) \tag{70}$$

for all $\theta \in \Theta$. By Berge's Maximum theorem, for every $\theta \in \Theta$, the map $\overline{D} \mapsto \mathcal{J}(\overline{D},\theta)$ is continuous in the Hausdorff topology. Given that Θ is compact and $\mathcal{J}(\overline{D},\theta)$ is bounded it follows that also the map $\overline{D} \mapsto \mathcal{J}(\overline{D})$ is continuous in the Hausdorff topology. With this, the result follows by Weierstrass Theorem applied to (19).

A.5 Proof of Proposition 3

Fix a sequence $\{a_m, x_m, b_m\}_{m=1}^\infty \subseteq \overline{D}$ such that $x_m \in (a_m, b_m)$ and $\overline{D} \cap (a_m, b_m) \to \{x\}$. For costs of distinguishing induced by $\tilde{d}$, using Lemma 1, we can re-express this cost in terms of the maximum and minimum selections from $C(x)$, $\overline{\delta}(x)$ and $\underline{\delta}(x)$:

$$\Gamma(C) = \int_0^{\overline{x}} \left[\int_{\overline{\delta}(x)}^{\overline{x}} \tilde{d}(\overline{\delta}(x), y) \,\mathrm{d}y + \int_0^{\underline{\delta}(x)} \tilde{d}(\underline{\delta}(x), y) \,\mathrm{d}y \right] \mathrm{d}x \tag{71}$$

Defining $I(w) = \int_w^{\overline{x}} \tilde{d}(w, y) \,\mathrm{d}y$, we have that:

$$\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (a_m, b_m)) = \int_{a_m}^{b_m} \left(I\left(\overline{\delta}_{\overline{D}}(z)\right) - I\left(b_m\right) \right) \mathrm{d}z \tag{72}$$

By the mean value theorem, we have for every $z \in [a_m, b_m]$ that there exists $w \in [\overline{\delta}_{\overline{D}}(z), b_m]$ such that:

$$\begin{aligned} I\left(\overline{\delta}_{\overline{D}}(z)\right) - I\left(b_m\right) &= -I'(w)\left(b_m - \overline{\delta}_{\overline{D}}(z)\right) \\ &= \left(\tilde{d}(w, w) - \int_w^{\overline{x}} \tilde{d}_w(w, y) \,\mathrm{d}y \right) \left(b_m - \overline{\delta}_{\overline{D}}(z)\right) \\ &\geq \tilde{d}(w, w)\left(b_m - \overline{\delta}_{\overline{D}}(z)\right) \\ &\geq \tilde{d}(0, \overline{x})\left(b_m - \overline{\delta}_{\overline{D}}(z)\right) \end{aligned} \tag{73}$$

where we obtain the derivative of I by applying Leibniz's rule, which is itself possible because $\tilde{d}(w, y)$ is continuously differentiable on $(w, \overline{x})$. The first inequality follows by noting that $\tilde{d}(\tilde{z}, y)$ is a decreasing function in its first argument when $y \geq \tilde{z}$, making $\int_w^{\overline{x}} \tilde{d}_w(w, y) \,\mathrm{d}y \leq 0$. The second inequality follows by noting that $\tilde{d}(x, y)$ is minimized by $\tilde{(}0, \overline{x})$. We therefore have that:

$$\begin{aligned} \Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (a_m, b_m)) &\geq \tilde{d}(0, \overline{x}) \int_{a_m}^{b_m} \left(b_m - \overline{\delta}_{\overline{D}}(z)\right) \mathrm{d}z \\ &= \tilde{d}(0, \overline{x}) b_m (b_m - a_m) - \tilde{d}(0, \overline{x}) \int_{a_m}^{b_m} \overline{\delta}_{\overline{D}}(z) \,\mathrm{d}z \end{aligned} \tag{74}$$

Set $\epsilon = \tilde{d}(0, \overline{x})$. As d is a distance, we have that $d(0, \overline{x}) > 0$. As h is strictly positive when evaluated on a strictly positive argument, $\tilde{d}(0, \overline{x}) > 0$. Thus, $\epsilon > 0$. Using this ϵ, costs of distinguishing are therefore strongly monotone if:

$$\begin{aligned} \int_{a_m}^{b_m} \overline{\delta}_{\overline{D}}(z) \,\mathrm{d}z &\leq b_m(b_m - a_m) - (x_m - a_m)(b_m - x_m) \\ &= b_m(b_m - x_m) + x_m(x_m - a_m) \\ &= \int_{a_m}^{b_m} \overline{\delta}_m(z) \,\mathrm{d}z \end{aligned} \tag{75}$$

where $\overline{\delta}_m : [a_m, b_m] \to [0,1]$ is given by:

$$\overline{\delta}_m(z) = \begin{cases} x_m, z \in [a_m, x_m], \\ b_m, z \in (x_m, b_m]. \end{cases} \tag{76}$$

As $a_m, x_m, b_m \in \overline{D}$, observe that $\overline{\delta}_{\overline{D}}(z) \leq \overline{\delta}_m(z)$ for all $z \in [a_m, b_m]$, completing the proof.

A.6 Proof of Lemma 2

Let ϕ^* denote the optimal allocation under $\overline{D}$ and $\phi^{*'}$ denote the optimal allocation under $\overline{D}' = \overline{D} \setminus (a,b)$, as defined in Theorem 1. By Lemma 8, the difference in values under these contractibility correspondences is

$$\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D}') = \int_0^1 (J(\phi^*(\theta), \theta) - J(\phi^{*'}(\theta), \theta))\, \mathrm{d}F(\theta) \tag{77}$$

First, we observe that $\phi^*(\theta) \neq \phi^{*'}(\theta)$ only if $\phi^*(\theta) \in (a,b)$. We denote the set of types who receive such allocations by $\Theta(a,b) = \{\theta \in \Theta : \phi^*(\theta) \in (a,b)\}$. As ϕ^* is monotone, this is an interval. If this interval is empty, then $\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D}') = 0$ and the proof is finished. If not, we construct the optimal $\phi^{*'}$. Define $\hat{\theta}(y,z)$ as the type for which the principal is indifferent between giving y or $z > y$, or the unique solution to $J(y, \hat{\theta}(y,z)) = J(z, \hat{\theta}(y,z))$. By Theorem 1, the following assignment function is optimal:

$$\phi^{*'}(\theta) = \begin{cases} a & \text{if } \theta \in [\inf \Theta(a,b), \hat{\theta}(a,b)], \\ b & \text{if } \theta \in (\hat{\theta}(a,b), \sup \Theta(a,b)], \\ \phi^*(\theta) & \text{otherwise.} \end{cases} \tag{78}$$

where we observe that $\sup \Theta(a,b) = (\phi^*)^{-1}(b)$. Defining the left generalized inverse as $\phi^\dagger(z) = \sup\{\theta \in \Theta : \phi(\theta) \leq z\}$, we also observe that $\inf \Theta(a,b) = (\phi^*)^\dagger(a)$. Because of this, we have that:

$$\inf \Theta(a,b) = \begin{cases} \min_{x \in \overline{D}: x > a} \hat{\theta}(a,x), \text{if it exists,} \\ \left(\phi^P\right)^{-1}(a), \text{otherwise.} \end{cases} \tag{79}$$

$$\sup \Theta(a,b) = \begin{cases} \max_{x \in \overline{D}: x < b} \hat{\theta}(b,x), \text{if it exists,} \\ \left(\phi^P\right)^{-1}(b), \text{otherwise.} \end{cases} \tag{80}$$

We can now bound the loss in value from the deletion of (a,b) from $\overline{D}$. By the previous arguments, we have that:

$$\begin{aligned}\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D}') = &\int_{\inf \Theta(a,b)}^{\hat{\theta}(a,b)} \left(J(\phi^*(\theta),\theta) - J(a,\theta)\right) \mathrm{d}F(\theta) \\ &+ \int_{\hat{\theta}(a,b)}^{\sup \Theta(a,b)} \left(J(\phi^*(\theta),\theta) - J(b,\theta)\right) \mathrm{d}F(\theta)\end{aligned} \tag{81}$$

We now proceed in three steps. We first bound the integrands, then bound the limits of integration, and finally put the two together.

Step 1: Bounding the Integrands. We first derive an upper bound for $J(\phi^*(\theta),\theta) - J(x,\theta)$. We expand $J(x,\theta)$ to the second order around $\phi^*(\theta)$. Using Taylor's remainder Theorem, and evaluating at $x = \phi^{*'}(\theta)$,

$$J(\phi^{*'}(\theta),\theta) = J(\phi^*(\theta),\theta) + J_x(\phi^*(\theta),\theta)(\phi^{*'}(\theta) - \phi^*(\theta)) + \frac{1}{2}J_{xx}(y(\theta),\theta)(\phi^{*'}(\theta) - \phi^*(\theta))^2 \tag{82}$$

for some $y(\theta) \in [\phi^*(\theta), \phi^{*'}(\theta)] \cup [\phi^{*'}(\theta), \phi^*(\theta)]$. We further apply Taylor's remainder theorem to take a first-order expansion of $J_x(x,\theta)$ around $x = \phi^P(\theta)$ and evaluate at $x = \phi^*(\theta)$:

$$\begin{aligned}J_x(\phi^*(\theta),\theta) &= J_x(\phi^P(\theta),\theta) + J_{xx}(z(\theta),\theta)(\phi^*(\theta) - \phi^P(\theta)) \\ &= J_{xx}(z(\theta),\theta)(\phi^*(\theta) - \phi^P(\theta))\end{aligned} \tag{83}$$

where the first equality defines the point $z(\theta) \in [\phi^*(\theta), \phi^P(\theta)] \cup [\phi^P(\theta), \phi^*(\theta)]$ and the second uses the fact that $J_x(\phi^P(\theta),\theta) = 0$ by definition, since ϕ^P maximizes J and J is strictly quasiconcave in its first argument. Combining these expansions, we have that:

$$\begin{aligned}|J(\phi^{*'}(\theta),\theta) - J(\phi^*(\theta),\theta)| &\leq |J_x(\phi^*(\theta),\theta)||\phi^{*'}(\theta) - \phi^*(\theta)| + \frac{1}{2}|J_{xx}(y(\theta),\theta)|(\phi^{*'}(\theta) - \phi^*(\theta))^2 \\ &\leq |J_{xx}(z(\theta),\theta)|(\phi^{*'}(\theta) - \phi^*(\theta))^2 + \frac{1}{2}|J_{xx}(y(\theta),\theta)|(\phi^{*'}(\theta) - \phi^*(\theta))^2 \\ &\leq \frac{3}{2}\bar{J}_{xx}(\phi^{*'}(\theta) - \phi^*(\theta))^2\end{aligned} \tag{84}$$

Thus, defining $c = \phi^*(\hat{\theta}(a,b))$, the integrand in the first line of Equation 81 is bounded above by $\frac{3}{2}\bar{J}_{xx}(c-a)^2$ and the integrand in the second line of Equation 81 is bounded above by $\frac{3}{2}\bar{J}_{xx}(b-c)^2$.

Step 2: Bounding the Limits of Integration. We first derive bounds for the limits of integration. There are two approaches to this that we use. The first approach yields Equation 21 and Equation 22. The second approach yields Equation 23.

In the first approach, we observe that $\hat{\theta}(a,b) - \inf \Theta(a,b), \sup \Theta(a,b) - \hat{\theta}(a,b) \leq \sup \Theta(a,b) -$

$\inf \Theta(a,b) \leq \left(\phi^P\right)^{-1}(b) - \left(\phi^P\right)^{-1}(a)$. Both ϕ^P and $(\phi^P)^{-1}$ are monotone and differentiable functions under our maintained assumption that J is twice continuously differentiable and strictly supermodular in (x,θ). In this case, the slope of the inverse function is $((\phi^P)^{-1})'(x) = \frac{1}{(\phi^P)'((\phi^P)^{-1}(x))}$. Moreover, by the implicit function theorem, $(\phi^P)'(\theta) = \frac{J_{x\theta}(\phi^P(\theta),\theta)}{J_{xx}(\phi^P(\theta),\theta)}$. Therefore, we can write the bound

$$((\phi^P)^{-1})'(x) = \frac{J_{xx}(x,(\phi^P)^{-1}(x))}{J_{x\theta}(x,(\phi^P)^{-1}(x))} \leq \frac{\sup_{y\in X,\theta\in\Theta} J_{xx}(y,\theta)}{\inf_{y\in X,\theta\in\Theta} J_{x\theta}(y,\theta)} = \frac{\bar{J}_{xx}}{\underline{J}_{x\theta}} < \infty \tag{85}$$

where penultimate inequality uses the definitions of $\bar{J}_{xx}$ and$\underline{J}_{x\theta}$; and the last inequality follows from the fact that J twice continuously differentiable and strictly supermodular over the compact set $X \times \Theta$. Thus, we have that:

$$\sup \Theta(a,b) - \inf \Theta(a,b) \leq \frac{\bar{J}_{xx}}{\underline{J}_{x\theta}}(b-a) \tag{86}$$

In the second approach, we suppose that $a < c < b$ are three sequential points in $\overline{D}$, *i.e*, c is isolated, and a and b are the closest elements to c in $\overline{D}$. In this case $\inf \Theta(a,b) = \hat{\theta}(a,c)$ and $\sup \Theta(a,b) = \hat{\theta}(c,b)$. We first bound $\hat{\theta}(a,b) - \hat{\theta}(a,c)$.

To do this, we define $\hat{\theta}(u) = \hat{\theta}(a, c+u)$ and note that $\hat{\theta}(b-c) = \hat{\theta}(a,b)$ and $\hat{\theta}(0) = \hat{\theta}(a,c)$. Under this reformulation, the definition of $\hat{\theta}(u)$ can be re-written as $J(c+u,\hat{\theta}(u)) = J(a,\hat{\theta}(u))$. We now implicitly differentiate this to obtain

$$\hat{\theta}'(u) = \frac{-J_x(c+u,\hat{\theta}(u))}{J_\theta(c+u,\hat{\theta}(u)) - J_\theta(a,\hat{\theta}(u))} \tag{87}$$

We now apply Taylor's remainder theorem to $\hat{\theta}(u)$ around $u=0$, evaluated at $u = b-c$, to obtain

$$\hat{\theta}(b-c) = \hat{\theta}(0) + \hat{\theta}'(\tilde{u})(b-c) \tag{88}$$

for some $\tilde{u} \in [0, b-c]$. Using our definitions, this implies

$$\hat{\theta}(a,b) - \hat{\theta}(a,c) = \hat{\theta}(b-c) - \hat{\theta}(0) = \frac{-J_x(c+\tilde{u},\hat{\theta}(\tilde{u}))}{J_\theta(c+\tilde{u}t,\hat{\theta}(\tilde{u})) - J_\theta(a,\hat{\theta}(\tilde{u}))}(b-c) \tag{89}$$

We now bound the numerator and denominator of the first fraction. For the numerator, we apply Taylor's remainder theorem to $J_x(\cdot,\hat{\theta}(\tilde{u}))$ around $x = \phi^P(\hat{\theta}(\tilde{u}))$ to write

$$\begin{aligned} J_x(c+\tilde{u},\hat{\theta}(\tilde{u})) =& J_x(\phi^P(\hat{\theta}(\tilde{u})),\hat{\theta}(\tilde{u})) + J_{xx}(z,\hat{\theta}(\tilde{u}))(c+\tilde{u}-\phi^P(\hat{\theta}(\tilde{u}))) \\ =& J_{xx}(z,\hat{\theta}(\tilde{u}))(c+\tilde{u}-\phi^P(\hat{\theta}(\tilde{u}))) \end{aligned} \tag{90}$$

for some $z \in [c+\tilde{u}, \phi^P(\hat{\theta}(\tilde{u}))]$, where we use $J_x(\phi^P(\theta), \theta) = 0$ in the second line. Moreover, we have that $(c+\tilde{u}-\phi^P(\hat{\theta}(\tilde{u}))) \leq b-a$. Therefore, we have that $|J_x(c+\tilde{u}, \hat{\theta}\tilde{u})| < \bar{J}_{xx}(b-a)$. For the denominator, we apply Taylor's remainder theorem to $J_\theta(\cdot, \hat{\theta}(\tilde{u}))$ around $x = a$ to write

$$J_\theta(c+\tilde{u}, \hat{\theta}(\tilde{u})) - J_\theta(a, \hat{\theta}(\tilde{u})) = J_{x\theta}(z, \hat{\theta}(\tilde{u}))(c+\tilde{u}-a) \tag{91}$$

for some $z \in [a, c+\tilde{u}]$. We observe that $c+\tilde{u}-a \geq c-a$. Therefore, $|J_\theta(c+\tilde{u}, \hat{\theta}(\tilde{u})) - J_\theta(a, \hat{\theta}(\tilde{u}))| \geq \underline{J}_{x\theta}(c-a)$. Combining these two bounds, we deduce that:

$$\hat{\theta}(a,b) - \hat{\theta}(a,c) \leq \frac{\bar{J}_{xx}(b-a)}{\underline{J}_{x\theta}(c-a)}(b-c) \tag{92}$$

To bound, $\hat{\theta}(c,b) - \hat{\theta}(a,b)$ we can apply analogous arguments. By doing this, we obtain:

$$\hat{\theta}(a,b) - \hat{\theta}(c,b) \leq \frac{\bar{J}_{xx}(b-a)}{\underline{J}_{x\theta}(b-c)}(c-a) \tag{93}$$

Step 3: Bounding the Value. Combining steps 1 and 2. We can now derive the payoff bound of Equation 22:

$$\begin{aligned}
\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D}') &= \int_{\inf \Theta(a,b)}^{\hat{\theta}(a,b)} (J(\phi^*(\theta), \theta) - J(a, \theta)) \, \mathrm{d}F(\theta) \\
&\quad + \int_{\hat{\theta}(a,b)}^{\sup \Theta(a,b)} (J(\phi^*(\theta), \theta) - J(b, \theta)) \, \mathrm{d}F(\theta) \\
&\leq \int_{\inf \Theta(a,b)}^{\hat{\theta}(a,b)} \frac{3}{2} \bar{J}_{xx}(c-a)^2 \, \mathrm{d}F(\theta) + \int_{\hat{\theta}(a,b)}^{\sup \Theta(a,b)} \frac{3}{2} \bar{J}_{xx}(b-c)^2 \, \mathrm{d}F(\theta) \\
&\leq \frac{3}{2} \bar{J}_{xx}[(c-a)^2 + (b-c)^2] \int_{\inf \Theta(a,b)}^{\sup \Theta(a,b)} \mathrm{d}F(\theta) \\
&\leq \frac{3}{2} \bar{J}_{xx}[(c-a)^2 + (b-c)^2] \frac{\bar{J}_{xx}}{\underline{J}_{x\theta}}(b-a)\bar{f} \\
&= \frac{3}{2} \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b-a)[(c-a)^2 + (b-c)^2]
\end{aligned} \tag{94}$$

Observing that $(c-a)^2 + (b-c)^2 \leq (b-a)^2$, we also obtain Equation 21.

Finally, we obtain Equation 23 by combining step 1 with the second approach to step 2.

Doing this, we obtain:

$$
\begin{aligned}
\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D}') &= \int_{\hat{\theta}(a,c)}^{\hat{\theta}(a,b)} \left(J(\phi^*(\theta),\theta) - J(a,\theta)\right) \mathrm{d}F(\theta) \\
&\quad + \int_{\hat{\theta}(a,b)}^{\sup \Theta(c,b)} \left(J(\phi^*(\theta),\theta) - J(b,\theta)\right) \mathrm{d}F(\theta) \\
&\leq \int_{\hat{\theta}(a,c)}^{\hat{\theta}(a,b)} \frac{3}{2}\bar{J}_{xx}(c-a)^2 \,\mathrm{d}F(\theta) + \int_{\hat{\theta}(a,b)}^{\hat{\theta}(c,b)} \frac{3}{2}\bar{J}_{xx}(b-c)^2 \,\mathrm{d}F(\theta) \\
&\leq 3\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b-a)(c-a)(b-c)
\end{aligned} \tag{95}
$$

Completing the proof.

A.7 Proof of Lemma 3

We prove the three claims in turn.

1. Intervals. Suppose that $\overline{D}$ contains an interval I. Let x be the midpoint of such an interval and consider a sequence of points $a_m = x - \frac{t}{m}$, $x_m = x$, and $b_m = x + \frac{t}{m}$, where $t > 0$ is small enough such that $(x-t, x+t)$ is contained in I. We use Equation 21 from Lemma 2. In particular, for every m, we have that:

$$
\mathcal{J}(\overline{D}) - \mathcal{J}\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) \leq 12\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} t^3 m^{-3} \tag{96}
$$

We observe that $\overline{D} \cap (x - \frac{t}{m}, x + \frac{t}{m}) = (x - \frac{t}{m}, x + \frac{t}{m})$ for all m by construction. Moreover, the topological limit of $(x - \frac{t}{m}, x + \frac{t}{m})$ is $\{x\}$. Thus, by strong monotonicity, there exists M such that for all $m \geq M$, we have that:

$$
\Gamma(\overline{D}) - \Gamma\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) \geq \epsilon t^2 m^{-2} \tag{97}
$$

Thus, for all $m > \max\left\{M, 12\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} \frac{t}{\varepsilon}\right\}$ we have that:

$$
\mathcal{J}(\overline{D}) - \Gamma(\overline{D}) < \mathcal{J}\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) - \Gamma\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) \tag{98}
$$

which contradicts the optimality of $\overline{D}$.

2. Perfect and Nowhere Dense Sets. As $\overline{B}_t(x) \cap \overline{D}$ is perfect for some $t > 0$, every element is an accumulation point. Moreover, as the set is nowhere dense, $\overline{B}_t(x) \cap \overline{D}$ must contain an accumulation point that is isolated from one side. We focus on the case in which the point is isolated from the left, *i.e.*, there exists $x^* \in \overline{B}_t(x) \cap \overline{D}$ such that $y = \max\{z \in \overline{D} : z < x^*\}$ exists; the argument is entirely symmetric if the point is isolated from the right. We now construct a sequence with $a_m = y$ and $\{b_m\}$ equal to a monotone decreasing sequence of points in $\overline{D}$ that converges to x^* (as x^* is a limit point, the Bolzano-Weierstrass theorem implies that this is always possible). Thus, we have from statement 2 of Lemma 2 (Equation 22) that there exists a sequence of points $z_m \in (x^*, b_m) \cap \overline{D}$ such that:

$$\begin{aligned} &\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (y, b_m)) = \mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus [x^*, b_m)) \\ &\leq \frac{3}{2}\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b_m - x^*)\left[(b_m - z_m)^2 + (z_m - x^*)^2\right] \leq \frac{3}{2}\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b_m - y)\left[(b_m - x^*)^2\right] \end{aligned} \tag{99}$$

We now fix the sequence $x_m = x^*$ and observe that the topological limit of $(y, b_m) \cap \overline{D}$ is $\{x^*\}$. By strong monotonicity, we have that there exists M such that for all $m \geq M$, we have that:

$$\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (y, b_m)) \geq \epsilon(x^* - y)(b_m - x^*) \tag{100}$$

As $b_m - x^*$ is common to both terms we have that for all $m \geq M$ that:

$$\begin{aligned} &\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (y, b_m)) - (\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (y, b_m))) \\ &\geq (b_m - x^*)\left[\epsilon(x^* - y) - \frac{3}{2}\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b_m - x)(b_m - y)\right] \end{aligned} \tag{101}$$

As $b_m \to x^*$, we have that there exists a $\hat{M}$ such that $\left[\epsilon(x^* - y) - \frac{3}{2}\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b_m - x)(b_m - y)\right] > 0$ for all $m \geq \hat{M}$, which implies that for all $m \geq \max\{M, \hat{M}\}$:

$$\mathcal{J}(\overline{D}) - \Gamma(\overline{D}) < \mathcal{J}(\overline{D} \setminus (y, b_m)) - \Gamma(\overline{D} \setminus (y, b_m)) \tag{102}$$

This contradicts the optimality of $\overline{D}$.

3. Countably Infinite Sets. If $\overline{D}$ is countably infinite it contains an accumulation point x. As $\overline{D}$ does not contain any perfect sets, we know that every neigborhood of x contains an isolated point. Let $\{x_m\} \subset \overline{D}$ be a monotone sequence of isolated points such that $x_m \to x$. As x_m is isolated, we may define $a_m = \max\{y \in \overline{D} : y < x_m\}$ and $b_m = \min\{y \in \overline{D} : y > x_m\}$. By statement 3. in Lemma 2 (Equation 23), we have that:

$$\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus \{x_m\}) \leq 3\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b_m - a_m)(x_m - a_m)(b_m - x_m) \tag{103}$$

By construction, we have that $x_m \in (a_m, b_m)$. Moreover, $\overline{D} \cap (a_m, b_m) = \{x_m\}$, the topological limit of which is $\{x\}$ as $x_m \to x$. Thus, by strong monotonicity, we have that there exists M such that for all $m \geq M$, we have that:

$$\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus \{x_m\}) \geq \epsilon(x_m - a_m)(b_m - x_m) \tag{104}$$

Factoring $(x_m - a_m)(b_m - x_m)$ from both expressions, we have that:

$$\begin{aligned} &\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus \{x_m\}) - (\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus \{x_m\})) \\ &\geq (x_m - a_m)(b_m - x_m) \left[\epsilon - 3\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b_m - a_m)\right] \end{aligned} \tag{105}$$

As $a_m, b_m \to x$, we have that there exists $\hat{M}$ such that $\epsilon - 3\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(b_m - a_m) > 0$ for all $m \geq \hat{M}$. This implies that for all $m \geq \max\{M, \hat{M}\}$ that:

$$\mathcal{J}(\overline{D}) - \Gamma(\overline{D}) < \mathcal{J}(\overline{D} \setminus \{x_m\}) - \Gamma(\overline{D} \setminus \{x_m\}) \tag{106}$$

which contradicts the optimality of $\overline{D}$.

A.8 Proof of Lemma 4

We have already show that $\overline{D}^*$ is finite under strong monotonicity. Thus, we can express it as a sequence of ordered points. Take any three sequential points $x_{m-1}, x_m, x_{m+1} \in \overline{D}^*$. We can apply statement 3 of Lemma 2 (Equation 23) to bound the loss from eliminating contractibility at x_m:

$$\mathcal{J}(\overline{D}^*) - \mathcal{J}(\overline{D}^* \setminus \{x_m\}) \leq 3\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(x_m - x_{m-1})(x_{m+1} - x_m)(x_{m+1} - x_{m-1}) \tag{107}$$

Moreover, we can take constant sequences $a_n = x_{m-1}$, $\tilde{x}_n = x_m$ $b_n = x_{m+1}$ for all $n \in \mathbb{N}$. $a_n, \tilde{x}_n, b_n \in \overline{D}^*$ for all $n \in \mathbb{N}$ and $\overline{D}^* \cap (a_n, b_n) = \{x_m\}$ for all $n \in \mathbb{N}$. Thus, strong monotonicity of Γ implies that:

$$\Gamma(\overline{D}^*) - \Gamma(\overline{D}^* \setminus \{x_m\}) \geq \epsilon(x_m - x_{m-1})(x_{m+1} - x_m) \tag{108}$$

Optimality of $\overline{D}^*$ requires that $\mathcal{J}(\overline{D}^*) - \mathcal{J}(\overline{D}^* \setminus \{x_m\}) \geq \Gamma(\overline{D}^*) - \Gamma(\overline{D}^* \setminus \{x_m\})$. Combining this with Inequalities 107 and 108, we have that:

$$3\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(x_m - x_{m-1})(x_{m+1} - x_m)(x_{m+1} - x_{m-1}) \geq \epsilon(x_m - x_{m-1})(x_{m+1} - x_m) \tag{109}$$

Dividing both sides by $(x_{m+1}-x_m)(x_m-x_{m-1})$ yields

$$x_{m+1}-x_{m-1} \geq \frac{\epsilon}{3}\frac{\underline{J}_{x\theta}}{\overline{J}_{xx}^2\overline{f}} \tag{110}$$

Thus, we have that:

$$\bar{x} \geq x_{K^*}-x_1 = \sum_{j=1}^{\lfloor K^*/2 \rfloor} x_{2j+1}-x_{2j-1} \geq K^*\frac{\epsilon}{6}\frac{\underline{J}_{x\theta}}{\overline{J}_{xx}^2\overline{f}} \tag{111}$$

Re-arranging this equation yields the desired bound.

A.9 Proof of Proposition 4

Using the representation we derived in Proposition 3, we have that costs of distinguishing satisfy:

$$\Gamma_K(\overline{D}) = \int_0^{\overline{x}} I(\overline{\delta}(x))\,\mathrm{d}x = \sum_{k=2}^{K} I(x_k)(x_k-x_{k-1}) \tag{112}$$

Thus, we have that:

$$\Gamma_K^{(k)}(\overline{D}) = I'(x_k)(x_k-x_{k-1}) + I(x_k) - I(x_{k+1}) \tag{113}$$

where $I'(x_k) = -\tilde{d}(x_k,x_k) + \int_{x_k}^{\overline{x}} \tilde{d}_w(x_k,y)\,\mathrm{d}y = -h(0) + \int_{x_k}^{\overline{x}} \tilde{d}_w(x_k,y)\,\mathrm{d}y$.

A.10 Proof of Proposition 5

We first introduce some preliminary notation. Given a vector $(x_2,...,x_{K^*-1}) \in \mathbb{R}^{K^*-2}$, we let $(x_k+\varepsilon, x_{-k}) \in \mathbb{R}^{K^*-2}$ the vector where we replace x_k with $x_k+\varepsilon$ for some $k \in \{2,...,K^*-1\}$. As Γ is strongly monotone, $\overline{D}$ is finite. Thus, for $\{x_k\}$ to be optimal, as Γ is finitely differentiable, it must be true that $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{J}(x_k+\varepsilon,x_{-k})|_{\varepsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Gamma(x_k+\varepsilon,x_{-k})|_{\varepsilon=0}$ for any $k \in \{2,\dots,K^*-1\}$. The left-hand-side is

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{J}(x_k+\varepsilon,x_{-k})|_{\varepsilon=0} = \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k,\theta)\,\mathrm{d}F(\theta) + \\
&\quad \frac{\partial}{\partial x_k}\hat{\theta}_k\left(J(x_k,\hat{\theta}_k)-J(x_{k-1},\hat{\theta}_k)\right)f(\hat{\theta}_k) + \frac{\partial}{\partial x_k}\hat{\theta}_{k+1}\left(J(x_{k+1},\hat{\theta}_{k+1})-J(x_k,\hat{\theta}_{k+1})\right)f(\hat{\theta}_{k+1}) \\
&= \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k,\theta)\,\mathrm{d}F(\theta)
\end{aligned} \tag{114}$$

where, in the second equality, we use the fact that $J(x_k, \hat{\theta}_k) = J(x_{k-1}, \hat{\theta}_k)$ by definition. By the definition that $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Gamma(x_k + \varepsilon, x_{-k})|_{\varepsilon=0} = \Gamma_k(\overline{D})$, we obtain Equation 27. Finally, again by definition, we have that $x_1 = 0$ and $x_{K^*} = 1$

A.11 Proof of Proposition 6

We split the argument in two parts. We first calculate the optimal contract for fixed K. We then solve for the optimal K^*.

Optimal Contract for Fixed K. We leverage our characterization of the optimal contract in Proposition 5 to set up the optimization problem in closed form. The virtual surplus function in this setting is $J(x, \theta) = \alpha(2\theta - 1)x - \beta\frac{x^2}{2}$. Equation 25 gives the principal's interim payoff under the optimal contract conditional on any set of K contractible actions $\{x_k\}_{k=1}^K$. Moreover, the K-interval partition of types is defined by the indifference condition of Corollary 1. We therefore define the following value function describing the monopolist's favorite K-item contract as the solution of a quadratic constrained optimization problem:

$$
\begin{aligned}
V(K) = \max_{(x_1,\ldots,x_K)\in X^K} & \left\{ \sum_{k=1}^{K} \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \left(\alpha(2\theta - 1)x_k - \beta\frac{x_k^2}{2} \right) \mathrm{d}\theta \right. \\
& \left. - \gamma\left(1 - x_1^2 - \sum_{k=2}^{K} x_k(x_k - x_{k-1})\right) \right\} \\
\text{s.t. } & 0 \le x_k \le x_{k+1}, \qquad \forall k \le K-1 \\
& x_1 = 0,\ x_K = 1 \\
& \hat{\theta}_k = \frac{\beta}{4\alpha}(x_k + x_{k-1}) + \frac{1}{2}, \qquad 2 \le k \le K \\
& \hat{\theta}_1 = 0, \hat{\theta}_{K+1} = 1
\end{aligned}
\tag{115}
$$

The first constraint requires that the x_k be an ordered sequence. The second constraint requires that $x_K = 1$, since this action is always contractible. The third constraint solves for the cut-off types $\hat{\theta}_k$, to whom the principal is indifferent in allocating x_k or x_{k-1}. The final constraint gives the boundary conditions for the type space.

Applying Proposition 5, the first-order condition for $k \in \{2, K-1\}$ is

$$
\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} (\alpha(2\theta - 1) - \beta x_k)\,\mathrm{d}\theta - \gamma(-2x_k + x_{k-1} + x_{k+1}) = 0 \tag{116}
$$

This reduces to:

$$\begin{aligned}\gamma(-2x_k + x_{k-1} + x_{k+1}) &= (\hat{\theta}_{k+1} - \hat{\theta}_k)\left[\alpha(\hat{\theta}_{k+1} + \hat{\theta}_k - 1) - \beta x_k\right] \\ &= \frac{\beta^2}{16\alpha}(x_{k+1} - x_{k-1})(x_{k+1} + x_{k-1} - 2x_k)\end{aligned} \tag{117}$$

where, in the second equality, we use the fact that $\hat{\theta}_k = \frac{\beta}{4\alpha}(x_k + x_{k-1}) + \frac{1}{2}$. This can in turn be written as:

$$(x_{k+1} + x_{k-1} - 2x_k)\left[\frac{\beta^2}{16\alpha}(x_{k+1} - x_{k-1}) - \gamma\right] = 0 \tag{118}$$

This equation has two solutions,

$$x_k = \frac{x_{k+1} + x_{k-1}}{2}, \qquad x_{k+1} = x_{k-1} + \Delta \tag{119}$$

where $\Delta = \frac{16\alpha\gamma}{\beta^2}$. We now separately consider each case.

Case 1: Uniform Grid. From the boundary conditions, we have that $x_1 = 0$ and $x_K = 1$. Thus, we have that:

$$x_k = \frac{x_{k+1} + x_{k-1}}{2} \implies x_k = \frac{k-1}{K-1} \tag{120}$$

We can verify that this is a local maximum by checking the Hessian is negative definite at this solution. We calculate that:

$$\begin{aligned}\frac{\partial^2 \mathcal{J}}{\partial x_k^2} &= H^{\mathcal{J}}_{k-1,k-1} = -\frac{\beta^2}{4\alpha(K-1)} + 2\gamma = \kappa \\ \frac{\partial^2 \mathcal{J}}{\partial x_k \partial x_{k+1}} &= H^{\mathcal{J}}_{k,k-1} = H^{\mathcal{J}}_{k-1,k} = \frac{\beta^2}{8\alpha(K-1)} - \gamma = -\frac{1}{2}\kappa\end{aligned} \tag{121}$$

where we note that row and column $k-1$ of $H^{\mathcal{J}}$ corresponds to the variable x_k. Thus, the Hessian is a tridiagonal Toeplitz matrix, which implies that the Eigenvalues are, by Theorem 2.2 of Kulkarni, Schmidt, and Tsui (1999), given by:

$$\lambda_k = \kappa\left(1 + \cos\left(\frac{k-1}{K}\pi\right)\right) \tag{122}$$

for $k \in \{2, \ldots, K-1\}$. As $\cos\left(\frac{k-1}{K}\pi\right) > -1$ for all such k, we have that $\text{sgn}(\lambda_k) = \text{sgn}(\kappa)$. Thus, the Hessian is negative definite if and only if:

$$K < \bar{K} = 1 + \frac{\beta^2}{8\alpha\gamma} \tag{123}$$

We will later verify that this holds whenever K is set optimally, confirming the optimality of the uniform grid solution.

Case 2: Alternating Grid. The first solution yields a uniform grid. Under the second solution, it must be the case that even points form a uniform grid with spacing $\Delta \equiv \frac{16\alpha\gamma}{\beta^2}$ and the odd points form a uniform grid with spacing $\Delta \equiv \frac{16\alpha\gamma}{\beta^2}$. When K is odd, given the boundary conditions that $x_1 = 0$ and $x_K = 1$, we have that this is possible only when $K = 2 + \frac{2}{\Delta}$, which is itself only possible when $\frac{\beta^2}{8\alpha\gamma}$ is an odd integer. When K is even, the solution must be $x_k = \frac{k-1}{2}\Delta$ for k odd, and $x_k = 1 - \frac{K-k}{2}\Delta$ for k even. This is possible for any even $K < 2 + \frac{2}{\Delta}$.

We next show that the alternating grid is *not* a local maximum of the objective function. For a local maximum, a necessary condition is that the Hessian is negative semidefinite. We will show the existence of a vector $x \in \mathbb{R}^{K-2}$ such that $v \neq 0$ an $v' H^{\mathcal{J}} v > 0$, which implies that $H^{\mathcal{J}}$ is not negative semidefinite. To do this, we first evaluate the second-order conditions at the conjectured alternating grid solution. These simplify to

$$\begin{aligned}
\frac{\partial^2 \mathcal{J}}{\partial x_k^2} &= H^{\mathcal{J}}_{k-1,k-1} = -\frac{\beta^2}{8\alpha}\Delta + 2\gamma = 0 \\
\frac{\partial^2 \mathcal{J}}{\partial x_k \partial x_{k+1}} &= H^{\mathcal{J}}_{k,k-1} = H^{\mathcal{J}}_{k-1,k} = \frac{\beta^2}{8\alpha}(x_{k+1} - x_k) - \gamma
\end{aligned} \tag{124}$$

Using this, we define $v_k = e_{k-1} - e_k$, where e_k denotes the unit vector in dimension k. This direction corresponds to increasing x_k and decreasing x_{k+1}. We calculate

$$v_k' H^{\mathcal{J}} v_k = 2\left(\gamma - \frac{\beta^2}{8\alpha}(x_{k+1} - x_k)\right) \tag{125}$$

We now split the proof into two cases. First, consider the case in which $K > 4$. In this case, there must exist some x_k, x_{k+1} such that $x_{k+1} - x_k < \frac{\Delta}{2}$, since the grid is not uniform. Then,

$$v_k' H^{\mathcal{J}} v_k > 2\left(\gamma - \frac{\Delta\beta^2}{16\alpha}\right) > 0 \tag{126}$$

and, as desired, we have shown that the Hessian is not negative definite. Next, we consider the case in which $K = 4$. In this case, we take two candidate vectors. The first is $u = e_1 + e_2$, and we observe

$$u' H^{\mathcal{J}} u = 2\left(\frac{\beta^2}{8\alpha}(x_3 - x_2) - \gamma\right) \tag{127}$$

The second is $v_1 = e_1 - e_2$, and we observe

$$v_1' H^{\mathcal{J}} v_1 = 2\left(\gamma - \frac{\beta^2}{8\alpha}(x_3 - x_2)\right) = -u' H^{\mathcal{J}} u \tag{128}$$

We have therefore shown the desired result but for the case in which $u'H^{\mathcal{J}}u = v_1'H^{\mathcal{J}}v_1 = 0$. Here, $x_3 - x_2 = \frac{8\alpha\gamma}{\beta^2} = \frac{\Delta}{2}$. But this is precisely the case of the uniform grid.

Optimal K^*. We first prove a Lemma computing the costs and benefits of having K tiers:

Lemma 9. *The value to the monopolist of a K-item contract, or the solution to the program in Equation 115, can be written as $V(K) = \hat{\Pi}(K) - \hat{\Gamma}(K)$ where*

$$\begin{aligned} \hat{\Pi}(K) &= \frac{\alpha - \beta}{4} + \frac{\beta^2}{48\alpha}\frac{(2K-3)(2K-1)}{(K-1)^2} \\ \hat{\Gamma}(K) &= \frac{\gamma}{2}\frac{K-2}{K-1} \end{aligned} \tag{129}$$

Proof. Using the representation in Equation 25, we write

$$\begin{aligned} \hat{\Pi}(K) &= \sum_{k=1}^{K} \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \left(\alpha(2\theta - 1)x_k - \beta\frac{x_k^2}{2}\right) \mathrm{d}\theta \\ &= \sum_{k=1}^{K} \left[\alpha x_k \theta^2 - x_k\left(\alpha + \frac{\beta}{2}x_k\right)\theta\right]_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \\ &= \sum_{k=1}^{K} \left(\alpha x_k(\hat{\theta}_{k+1} - \hat{\theta}_k)(\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k\left(\alpha + \frac{\beta}{2}x_k\right)(\hat{\theta}_{k+1} - \hat{\theta}_k)\right) \\ &= \frac{\beta}{2\alpha(K-1)} \sum_{k=2}^{K-1} \left(\alpha x_k(\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k\left(\alpha + \frac{\beta}{2}x_k\right)\right) + (1 - \hat{\theta}_K)\left(\alpha\hat{\theta}_K - \frac{\beta}{2}\right) \end{aligned} \tag{130}$$

where, in the fourth equality, we use that $\hat{\theta}_{k+1} - \hat{\theta}_k = \frac{\beta}{2\alpha(K-1)}$ for $k < K$ and that $\hat{\theta}_{K+1} = 1$

and $x_K = 1$. We simplify the summation term as

$$\begin{aligned}
&\sum_{k=2}^{K-1} \left(\alpha x_k (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\alpha + \frac{\beta}{2} x_k \right) \right) \\
&= \sum_{k=2}^{K-1} \left(\alpha x_k \left(1 + \frac{\beta}{\alpha} x_k \right) - x_k \left(\alpha + \frac{\beta}{2} x_k \right) \right) \\
&= \frac{\beta}{2} \sum_{k=2}^{K-1} x_k^2 \\
&= \frac{\beta}{2} \sum_{k=2}^{K-1} \left(\frac{k-1}{K-1} \right)^2 = \frac{\beta}{12(K-1)} (K-2)(2K-3)
\end{aligned} \tag{131}$$

where we use that $\hat{\theta}_k + \hat{\theta}_{k+1} = 1 + \frac{\beta}{\alpha} x_k$. To simplify the second term, we observe that

$$\begin{aligned}
\hat{\theta}_K &= \frac{1}{2} + \frac{\beta}{4\alpha} \left(1 + \frac{K-2}{K-1} \right) \\
&= \frac{2\alpha(K-1) + \beta(2K-3)}{4\alpha(K-1)} \\
1 - \hat{\theta}_K &= \frac{2\alpha(K-1) - \beta(2K-3)}{4\alpha(K-1)}
\end{aligned} \tag{132}$$

Putting this together, we write

$$\begin{aligned}
\hat{\Pi}(K) = &\frac{\beta^2}{24\alpha(K-1)^2} \Bigg((K-2)(2K-3) + \frac{3}{2\beta^2} \left(4\alpha^2 (K-1)^2 - \beta^2 (2K-3)^2 \right) - \\
&\frac{3}{\beta} \left(2\alpha (K-1)^2 - \beta(2K-3)(K-1) \right) \Bigg) \\
= &\frac{\alpha - \beta}{4} + \frac{\beta^2}{48\alpha} \frac{(2K-3)(2K-1)}{(K-1)^2}
\end{aligned} \tag{133}$$

We next show the desired representation of $\hat{\Gamma}$. This follows by direct calculation:

$$\begin{aligned}
\hat{\Gamma}(K) &= \gamma \left(1 - \left(\frac{1-1}{K-1} \right)^2 - \sum_{k=2}^{2} \frac{k-1}{K-1} \frac{1}{K-1} \right) \\
&= \frac{\gamma}{2} \left(1 - \frac{1}{K-1} \right) = \frac{\gamma}{2} \frac{K-2}{K-1}
\end{aligned} \tag{134}$$

Completing the proof. □

To derive $\tilde{K}$, we take the first derivative of V:

$$V'(K) = \frac{\beta^2}{24\alpha(K-1)^3} - \frac{\gamma}{2(K-1)^2} \tag{135}$$

We observe that $V'(K) > 0$ if and only if

$$K < \tilde{K} := \frac{\beta^2}{12\alpha\gamma} + 1 \tag{136}$$

We now prove that $|K^* - \tilde{K}| < 1$. If $K^* - \tilde{K} > 1$, then we know that $V(K^*-1) > V(K^*)$ as $V' < 0$ for all $K^* - 1 < K < K^*$; this contradicts optimality. Similarly, if $\tilde{K} - K^* > 1$, we know that $V(K^*+1) > V(K^*)$ as as $V' > 0$ for all $K^* < K < K^* + 1$; this contradicts optimality. Recall that we needed to check if the Hessian was negative definite. This is true so long as $K^* < \bar{K}$. As $\bar{K} = \frac{4}{3}\tilde{K}$, this holds whenever $\tilde{K} \geq 3$. It remains to check when $\tilde{K} \in (2,3)$ and $K^* = 3$. Direct calculation shows that indifference between $K = 2$ and $K = 3$ occurs when $\gamma = \frac{\beta^2}{16\alpha}$. At this point, $\tilde{K} = 7/3$. Thus, whenever $K^* > 2$ is strictly optimal (which is when $\gamma < \frac{\beta^2}{16\alpha}$), we have that $K^* < \bar{K}$. The comparative statics follow from standard monotone comparative statics arguments, after the observations that $V_{K\alpha} < 0$, $V_{K\beta} > 0$, and $V_{K\gamma} < 0$. Finally, $V(3) - V(2) = \frac{1}{4}\left(\frac{\beta^2}{16\alpha} - \gamma\right)$. Thus, whenever $\gamma < \frac{\beta^2}{16\alpha}$ we have that $V(3) > V(2)$, which implies that $K^* \geq 3$.

A.12 Proof of Proposition 7

We now consider the problem of maximizing total surplus subject to the implementability constraint, or in which

$$S(x,\theta) := u(x,\theta) + \pi(x,\theta) = \alpha x\theta - \beta\frac{x^2}{2} \tag{137}$$

We first derive the principal's expected surplus as a function of the number of contractibility points. Using Equation 25, we calculate:

$$\begin{aligned}\hat{\Pi}^C(K) &= \sum_{k=1}^{K} \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \left(\alpha\theta x_k - \beta\frac{x_k^2}{2}\right) \mathrm{d}\theta \\ &= \sum_{k=1}^{K} \left[\frac{\alpha}{2} x_k \theta^2 - x_k \left(\frac{\beta}{2} x_k\right) \theta\right]_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \\ &= \sum_{k=1}^{K} \left(\frac{\alpha}{2} x_k (\hat{\theta}_{k+1} - \hat{\theta}_k)(\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\frac{\beta}{2} x_k\right)(\hat{\theta}_{k+1} - \hat{\theta}_k)\right) \\ &= \frac{\beta}{\alpha(K-1)} \sum_{k=2}^{K-1} \left(\frac{\alpha}{2} x_k (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\frac{\beta}{2} x_k\right)\right) + (1 - \hat{\theta}_K)\left(\frac{\alpha}{2}(1 + \hat{\theta}_K) - \frac{\beta}{2}\right)\end{aligned} \tag{138}$$

We simplify the summation term as

$$\begin{aligned}&\sum_{k=2}^{K-1} \left(\frac{\alpha}{2} x_k (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\frac{\beta}{2} x_k\right)\right) \\ &= \sum_{k=2}^{K-1} \left(\frac{\alpha}{2} x_k \left(\frac{2\beta}{\alpha} x_k\right) - x_k \left(\frac{\beta}{2} x_k\right)\right) \\ &= \frac{\beta}{2} \sum_{k=2}^{K-1} x_k^2\end{aligned} \tag{139}$$

where we use that $\hat{\theta}_k + \hat{\theta}_{k+1} = \frac{2\beta}{\alpha} x_k$. Comparing to Equation 131 in the proof of Proposition 6, we observe that $\hat{\Pi}^C(K) = 2\hat{\Pi}(K)$.

Using Lemma 9, it follows that the optimal contract with complete information and cost scaling $\hat{\gamma}$ is the same as the optimal contract under a transformed problem with incomplete information and $\gamma = \frac{\hat{\gamma}}{2}$. Thus, $\tilde{K}^C = 2\tilde{K} - 1$.

A.13 Proof of Proposition 8

We start by using the change of variables formula for pushforward measures to rewrite the cost as:

$$\Gamma(C, \phi) = \int_X \hat{d}(C(x), X \setminus C(x))\mathrm{d}F_\phi(x) = \int_\Theta \hat{d}(C(\phi(\theta)), X \setminus C(\phi(\theta))\mathrm{d}F(\theta) \tag{140}$$

Using the representation of $\hat{d}$ derived in Proposition 3 and observing that $\underline{\delta} = 0$ is without loss of optimality, we can further simplify the cost as:

$$\Gamma(C, \phi) = \int_{\Theta} \int_{\overline{\delta}(\phi(\theta))}^{\overline{x}} \tilde{d}(\overline{\delta}(\phi(\theta)), y) \mathrm{d}F(\theta) = \int_{\Theta} I(\overline{\delta}(\phi(\theta))) \mathrm{d}F(\theta) \tag{141}$$

By Lemma 7, we have that $\overline{\delta}(\phi(\theta)) = \phi(\theta)$ for any implementable mechanism. Thus, conditional on ϕ, we have that the cost must satisfy:

$$\Gamma(C, \phi) = \tilde{\Gamma}(\phi) = \int_{\Theta} I(\phi(\theta)) \mathrm{d}F(\theta) \tag{142}$$

We can subsume this cost into the virtual surplus. Define $\tilde{\pi}(x, \theta) = \pi(x, \theta) - I(x)$ and define $\tilde{J} = \tilde{\pi} + u - \frac{1-F}{f} u_\theta$. By the arguments of Lemma 8, we then have that any optimal final action function solves:

$$\max_{\phi : \Theta \to X :\, \phi \text{ is increasing}} \int \tilde{J}(x, \theta) \mathrm{d}F(\theta) \tag{143}$$

Let $X^+ = \phi^*(\Theta)$ be the image of a solution to this problem and let $X^- = X \setminus X^+$. We have that $C(x) = [0, x]$ for every $x \in X^+$. As $F_\phi(X^-) = 0$, the choice of $C(x)$ for any $x \in X^-$ has no effect on costs or benefits. Thus, we can set $C(x) = [0, x]$ for every $x \in X^-$ without loss of optimality.

A.14 Proof of Proposition 9

We start with a preliminary lemma.

Lemma 10. *If Γ has a clause-based representation $\hat{\Gamma}$ then it is lower semicontinuous in the Hausdorff topology of closed sets.*

Proof. Given that $\hat{\Gamma} : \mathbb{N} \to \overline{\mathbb{R}}$ is strictly increasing we have as $K \to \infty$ either $\hat{\Gamma}(K)$ asymptotes to some value $\hat{\gamma} \in \overline{\mathbb{R}}$, potentially equal to ∞. In particular, it must be the case that $\hat{\Gamma}(D) = \hat{\gamma}$ for all sets such that $n(D) = \infty$. Consider a sequence of closed sets D_n such that $D_n \to D$ in the Hausdorff sense. There are four cases:

1. If eventually D_n and D have infinite many points, then $\hat{\Gamma}(D_n) = \hat{\gamma}$ for all n and $\hat{\Gamma}(D) = \hat{\gamma}$, as desired.
2. If eventually D_n has infinite many points, but $n(D) < \infty$, then we have $\liminf_n \hat{\Gamma}(D_n) = \hat{\gamma} > \hat{\Gamma}(D)$, as desired.
3. If every D_n has finitely many points, but $n(D) = \infty$, then by Hausdorff convergence we must have that $n(D_n) \to \infty$. Monotonicity then implies that $\liminf_n \hat{\Gamma}(D_n) = \hat{\gamma} = \hat{\Gamma}(D)$, as desired.

4. If every D_n and D have all finitely many points, then by Hausdorff convergence we must have that $n(D_n) \to n(D)$. Discrete convergence then implies that $\liminf_n \hat{\Gamma}(D_n) = \hat{\Gamma}(D)$, as desired.

□

We now first prove that $\overline{D}^*$ is finite. We first rule out the case in which the cardinality of $\overline{D}$ is infinite but $\overline{D} \neq X$, or contractibility is not perfect. Under clause-based costs, $\Gamma(\overline{D}) = \Gamma(X)$, or there is no increase in cost to consider perfect contractibility. However, $\mathcal{J}(X) \geq \mathcal{J}(\overline{D})$. Therefore, there must also be a solution with perfect contractibility. It will therefore suffice to show that perfect contractibility cannot be optimal.

To do this, we show that there is a strict payoff improvement from replacing perfect contractibility with a uniform grid of K points, evenly spaced with width $\overline{x}/K$. Recall that ϕ^P denotes the assignment under perfect contractibility, let ϕ^*_K denote the assignment under the grid, and let $G_K = \{\overline{x}i/K\}_{i=1}^K \in \mathcal{D}$ denote the grid. To derive the benefits of this contractibility correspondence, we apply a close variant of Lemma 2. Using the bound derived in the proof of that result for $|J(\phi^P(\theta), \theta) - J(x, \theta)|$ for any x, we derive

$$\begin{aligned} \mathcal{J}(X) - \mathcal{J}(G_K) &= \int_0^1 (J(\phi^P(\theta), \theta) - J(\phi^*_n(\theta), \theta))\, \mathrm{d}F(\theta) \\ &\leq \int_0^1 \frac{1}{2K^2} \bar{J}_{xx}\, \mathrm{d}F(\theta) = \frac{1}{2K^2} \bar{J}_{xx} \end{aligned} \tag{144}$$

We next observe that, if costs are clause strongly monotone, for sufficiently large n

$$\Gamma(X) - \Gamma(G_K) \geq \sum_{j=K}^{\infty} j^{-\beta} \epsilon \tag{145}$$

If $\beta \leq 1$, then $\Gamma(X) - \Gamma(G_K) = \infty$ and it is clearly preferred to set G_K. If $\beta > 1$, then we note that

$$\Gamma(X) - \Gamma(G_K) \geq \epsilon \sum_{j=K}^{\infty} j^{-\beta} \geq \epsilon \int_K^{\infty} s^{-\beta}\, \mathrm{d}s = \epsilon \left[-\frac{1}{\beta} s^{-\beta+1} \right]_K^{\infty} = \frac{\epsilon}{\beta} K^{-\beta+1} \tag{146}$$

where the first inequality uses the fact that $s^{-\beta}$ is a decreasing function for $s > 0$, and therefore the integral is smaller than its approximation via left end-point steps (*i.e.,* the sum). In this case, we have

$$\mathcal{J}(G_K) - \Gamma(G_K) \geq \mathcal{J}(X) - \Gamma(X) + \left(\frac{\epsilon}{\beta} K^{-\beta+1} - \frac{1}{2} \bar{J}_{xx} K^{-2} \right) \tag{147}$$

But, for $\beta < 3$, there is a contradiction to optimality. In particular,

$$K > \left(\frac{\beta}{2\epsilon}\bar{J}_{xx}\right)^{\frac{1}{3-\beta}} \to \mathcal{J}(G_K) - (\Gamma(G_K) - \mathcal{J}(X) - \Gamma(X)) \geq 0 \tag{148}$$

Thus, an optimal contracting support cannot be full contractibility. Finally, by Lemma 10 we can invoke Proposition 2 to establish that the solution set is compact. In turn, this yields the upper bound on the number of points of the optimal contracting supports.

We now derive the bound on the number of clauses. Our overall strategy will be to show that, if the number of clauses exceeded the claimed upper bound, then we could remove one clause and achieve a strict improvement. We first observe that, in a K clause contract, there must exist some ordered triple of points (x_{m-1}, x_m, x_{m+1}) such that $x_{m+1} - x_{m-1} < 2\overline{x}/(K-2)$. Otherwise, there would be a contradiction:

$$\begin{aligned} x_K - x_1 &= \sum_{j=1}^{\lfloor K/2 \rfloor} x_{2j+1} - x_{2j-1} \geq \lfloor K/2 \rfloor \frac{2\overline{x}}{K-2} \\ &> \left(\frac{K}{2} - 1\right) \frac{2\overline{x}}{\frac{K}{2} - 1} > \overline{x} \end{aligned} \tag{149}$$

We first apply the third statement of Lemma 2 to bound the loss from eliminating contractibility at some point x_m:

$$\begin{aligned} \mathcal{J}(\overline{D}^*) - \mathcal{J}(\overline{D}^* \setminus \{x_m\}) &\leq 3\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(x_m - x_{m-1})(x_{m+1} - x_m)(x_{m+1} - x_{m-1}) \\ &\leq \frac{3}{4}\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(x_{m+1} - x_{m-1})^3 \end{aligned} \tag{150}$$

where in the second inequality we use the fact that $\max_{w+y\leq z} wy = z^2/4$. Next, applying the clause strong monotonicity of $\Gamma(D) = \hat{\Gamma}(n(D))$ to a K-clause contract, we have

$$\hat{\Gamma}(K) - \hat{\Gamma}(K-1) \geq \epsilon(K-1)^{-\beta} > \epsilon(K-2)^{-\beta} \tag{151}$$

A sufficient condition for the principal to prefer to remove contractibility at point x_m is if the lower bound on cost reduction is larger than the upper bound on benefits loss, or

$$\epsilon(K-2)^{-\beta} > \frac{3}{4}\frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}}(x_{m+1} - x_{m-1})^3 \tag{152}$$

We now take $x_{m+1} - x_{m-1} < 2\overline{x}/(K-2)$ and re-arrange this to

$$K > 2 + \left(\frac{6\bar{J}_{xx}^2 \bar{f}}{\epsilon \underline{J}_{x\theta}} \right)^{\frac{1}{3-\beta}} \tag{153}$$

Thus, if K exceeds the right hand side, then we have found a contradiction to the optimality of the clause-based contract.

B Additional Application: Optimally Coarse Quality Certification

In this section, we apply our general results to a model of optimal quality certification provided by a third-party certifier that charges a price for certification to the producer. Our analysis combines and extends previous models of optimal certification provision by considering a certifier that is not informed about the producer's costs like in (as in Albano and Lizzeri, 2001), that potentially cares about the final consumer's utility (as in Zapechelnyuk, 2020), and for which testing is costly. This last feature is the main element of novelty of our analysis with respect to the previous literature. We argue that this feature is natural for the examples studied in this literature, such as optimal certification of bonds by rating agencies or optimal certification of safety (*e.g.,* for food, drugs, or cars) by a regulator. An adaptation of our main Theorem to this setting will reveal that, when testing costs are strongly monotone, every optimal certification policy entails a finite number of grades.

Our formalization of the basic economic environment closely follows the one in Zapechelnyuk (2020). Consider a producer choosing the price $p \geq 0$ and the quality $x \in X = [0, 1]$ of an indivisible good at cost $(1 - \theta)x^2/2$ where $\theta \in [0, 1]$ is the ability of the producer and is uniformly distributed. Consumers observe the price, receive some information about the quality produced by a certifier, and form an estimate $\hat{x}$ of the quality. They buy the good $a = 1$ if and only if $\hat{x} - p \geq b$ where $b \in [0, 1]$ is an outside option that the consumer forgoes in case they buy the producer's good. Consumers are heterogeneous in their outside option b, which is distributed according to $G(b) = b^\tau$ for some $\tau > 0$. With this, the revenue of the producer and the consumer's surplus given estimate $\hat{x}$ are respectively

$$r(\hat{x}) = \max_{p \geq 0} \{p(\hat{x} - p)^\tau\} = \left(\frac{\tau}{1+\tau}\right)^{1+\tau} \hat{x}^{1+\tau} \tag{154}$$

$$s(\hat{x}) = \frac{\tau}{1+\tau} r(\hat{x}) = \left(\frac{\tau}{1+\tau}\right)^{2+\tau} \hat{x}^{1+\tau} \tag{155}$$

where the unique optimal price is $p^*(\hat{x}) = \hat{x}/(1 + \tau)$.

The certifier can commit to some rating rule that reveals information about the quality x chosen by the producer. Formally, a rating rule is a right-continuous function $\zeta : X \to \mathbb{R}$ that assigns a grade to each chosen quality. This rule partitions X into sets of qualities x mapped to the same rating $\zeta(x) = z$. Given a rating z, the receiver learns that the quality of the producer's good must be in $\zeta^{-1}(z)$. Because higher qualities require a higher effort for the producer, the latter will always choose the lowest quality consistent with the desired rating, and therefore in equilibrium the estimated quality given rating z is $\hat{x}_\zeta(z) = \min \zeta^{-1}(z)$. With this, the set of qualities that can be chosen is equilibrium given ζ is $\underline{D}_\zeta = \hat{x}_\zeta(\zeta(X)) \subseteq X$, which by construction is a closed set always containing 0. It will be momentarily clear that this set corresponds to the set $\underline{D}$ in our general analysis, hence justifying our choice of notation.

Besides committing to a rating rule, the certifier commits to a price rule $T(z)$ that maps each rating to the price payed by the producer to the certifier. Given the raiting and price rules, the decision problem of a producer with ability θ is

$$\sup_{z \in \zeta(X)} \left\{ r(\hat{x}_\zeta(z)) - (1-\theta)\hat{x}_\zeta(z)^2 - T(z) \right\}$$

that is, the producer picks the rating by trading off the expected revenue induced in equilibrium with the minimum cost of effort consistent with that rating as well as the certifier fee.

Given fee t and quality estimate $\hat{x}$, the total payoff of the certifier is $(1-\beta)t + \beta s(\hat{x})$, that is the certifier potentially cares about both maximizing their profit and the consumers' surplus, with relative weight β. Therefore, the certifier chooses a pair of rating and pricing rules (ζ, T) as well as a recommendation rule $z : \Theta \to \zeta(X)$ to maximize

$$\int_\Theta (1-\beta)T(z(\theta)) + \beta s(\hat{x}_\zeta) dF(\theta) - \Gamma(\zeta) \tag{156}$$

under the constraint that

$$z(\theta) \in \arg\max_{z \in \zeta(X)} \left\{ r(\hat{x}_\zeta(z)) - (1-\theta)\hat{x}_\zeta(z(\theta))^2 - T(z) \right\} \tag{157}$$

Next, define

$$J(x,\theta) = \left(1 - \beta + \beta\frac{\tau}{\tau+1}\right)\left(\frac{\tau}{1+\tau}\right)^{1+\tau} x^{1+\tau} - (1-\beta)(2-\theta)x^2. \tag{158}$$

The certifier's problem can be simplified as follows.

Lemma 11. *The certifier's problem is equivalent to:*

$$\sup_{\underline{D}, \phi:\Theta\to X} \int_\Theta J(\phi(\theta),\theta)dF(\theta) - \Gamma(\underline{D}) \tag{159}$$

such that $\underline{D}$ *is closed, contains* 0 *and* ϕ *is nondecreasing and such that* $\phi(\Theta) \subseteq \underline{D}$.

Because J is strictly concave and supermodular, this problem falls under the umbrella of our main analysis. Thus, we can invoke Theorem 2 to establish that all the optimal $\underline{D}^*$

in the previous program are finite provided that the cost Γ is strongly monotone. In the certification setting, the assumption corresponds to a restriction on the costs of testing the difference between nearby quality grades.

In practice, the result implies that finite quality grades are optimal. This result is consistent, for instance, with the ubiquitous letter grading of bonds (*e.g.,* AAA *vs.* BAA) and restaurants (*e.g.,* sanitation grade A *vs.* B). Crucially, our result can rationalize grade systems other than a two-grade pass-fail, as studied by Zapechelnyuk (2020).

References

Albano, G. L., and A. Lizzeri (2001): "Strategic certification and provision of quality," *International Economic Review*, 42(1), 267–283.

Aliprantis, C. D., and K. Border (2006): *Infinite dimensional analysis: A Hitchhiker's Guide*. Springer.

Apostol, T. M. (1974): *Mathematical Analysis*. Pearson.

Bajari, P., and S. Tadelis (2001): "Incentives versus transaction costs: A theory of procurement contracts," *Rand journal of Economics*, pp. 387–407.

Battigalli, P., and G. Maggi (2002): "Rigidity, discretion, and the costs of writing contracts," *American Economic Review*, 92(4), 798–817.

Bergemann, D., T. Heumann, and S. Morris (2022): "Screening with persuasion," Pre-print 2212.03360, arXiv.

Bergemann, D., J. Shen, Y. Xu, and E. M. Yeh (2012): "Mechanism design with limited information: the case of nonlinear pricing," in *Game Theory for Networks: 2nd International ICST Conference, GAMENETS 2011, Shanghai, China, April 16-18, 2011, Revised Selected Papers 2*, pp. 1–10. Springer.

Bergemann, D., E. Yeh, and J. Zhang (2021): "Nonlinear pricing with finite information," *Games and Economic Behavior*, 130, 62–84.

Carbajal, J. C., and J. C. Ely (2013): "Mechanism design without revenue equivalence," *Journal of Economic Theory*, 148(1), 104–133.

Coase, R. H. (1960): "The problem of social cost," *The Journal of Law and Economics*, 3, 1–44.

Corrao, R., J. P. Flynn, and K. A. Sastry (2023): "Nonlinear Pricing with Underutilization: A Theory of Multi-part Tariffs," *American Economic Review*, 113(3), 836–860.

Gershkov, A., B. Moldovanu, P. Strack, and M. Zhang (2021): "A theory of auctions with endogenous valuations," *Journal of Political Economy*, 129(4), 1011–1051.

Grubb, M. D. (2009): "Selling to overconfident consumers," *American Economic Review*, 99(5), 1770–1807.

Hart, O., and J. Moore (1988): "Incomplete contracts and renegotiation," *Econometrica*, pp. 755–785.

——— (2008): "Contracts as reference points," *The Quarterly Journal of Economics*, 123(1), 1–48.

Jung, J., J. H. Kim, F. Matějka, and C. A. Sims (2019): "Discrete actions in information-constrained decision problems," *The Review of Economic Studies*, 86(6), 2643–2667.

Kulkarni, D., D. Schmidt, and S.-K. Tsui (1999): "Eigenvalues of tridiagonal pseudo-Toeplitz matrices," *Linear Algebra and its Applications*, 297, 63–80.

Laffont, J.-J., and J. Tirole (1986): "Using cost observation to regulate firms," *Journal of Political Economy*, 94(3, Part 1), 614–641.

Mohlin, E. (2014): "Optimal categorization," *Journal of Economic Theory*, 152, 356–381.

Mussa, M., and S. Rosen (1978): "Monopoly and product quality," *Journal of Economic Theory*, 18(2), 301–317.

Myerson, R. B. (1982): "Optimal coordination mechanisms in generalized principal–agent problems," *Journal of Mathematical Economics*, 10(1), 67–81.

Nöldeke, G., and L. Samuelson (2007): "Optimal bunching without optimal control," *Journal of Economic Theory*, 134(1), 405–420.

Rochet, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16(2), 191–200.

Sartori, E. (2021): "Competitive provision of digital goods," Working paper, Center for Studies in Economics and Finance.

Segal, I. (1999): "Complexity and renegotiation: A foundation for incomplete contracts," *The Review of Economic Studies*, 66(1), 57–82.

Strausz, R. (2017): "A theory of crowdfunding: A mechanism design approach with demand uncertainty and moral hazard," *American Economic Review*, 107(6), 1430–76.

Williamson, O. E. (1975): "Markets and hierarchies: analysis and antitrust implications: a study in the economics of internal organization," *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*.

Wilson, R. (1989): "Efficient and competitive rationing," *Econometrica*, pp. 1–40.

Wilson, R. B. (1993): *Nonlinear Pricing*. Oxford University Press, New York.

Yang, K. H. (2022): "Selling consumer data for profit: Optimal market-segmentation design and its consequences," *American Economic Review*, 112(4), 1364–93.

Zapechelnyuk, A. (2020): "Optimal quality certification," *American Economic Review: Insights*, 2(2), 161–76.