

Optimally Coarse Contracts

Roberto Corrao*
MIT

Joel P. Flynn†
Yale

Karthik A. Sastry‡
Princeton

October 17, 2023

Abstract

We study when a principal’s difficulty in describing payoff-relevant outcomes leads contracts to be optimally incomplete. Our setting is a general principal-agent model in which agents’ actions are partially contractible and the principal can choose the extent of contractibility at a cost. We first characterize implementable and optimal incomplete contracts for any fixed extent of contractibility. We next show that, if costs of contractibility satisfy a generalized notion of having a *strictly positive marginal cost*, then contracts are optimally *coarse*: they specify finitely many outcomes out of a continuum of possibilities. This provides a general foundation for incomplete contracts: even with arbitrarily small costs of contractibility, optimal contracts leave almost all outcomes unspecified. We apply these results in a standard nonlinear pricing setting to study when and why optimal price schedules have coarse tiers. In additional applications, we rationalize coarse product tiers in further assignment problems that feature menu costs, privacy concerns, and quality certification.

*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.

1 Introduction

Few contracts completely specify rights and obligations for all feasibly observable, payoff-relevant outcomes. Hart and Moore (1988) argue that this *incompleteness* of contracts stems from difficulties in specifying all relevant possibilities in a contract. Refining this argument, Hart and Moore (2008) differentiate the outcomes associated with a contract into two categories: verifiable and contractible outcomes in the “letter” of the contract and non-verifiable and non-contractible outcomes in the “spirit” of the contract. But they take as given the difference between the two types of outcomes, rather than relating them to costly effort by the principal to write or enforce a more complete contract.

We study how a principal optimally designs contracts in the face of costs of expanding the “letter” of the contract. To do this, we first introduce a model of optimal contracting with adverse selection and fixed constraints on the contractibility of a payoff-relevant action. These constraints are derived from axioms disciplining how the “spirit” of a contract relates to its “letter.” We show how our axioms accommodate rich forms of partial contractibility, and we characterize implementable and optimal mechanisms in this setting. We then allow the principal to choose the extent of contractibility itself at a cost. The cost formalizes the difficulty of describing outcomes and distinguishing them from one another, as we illustrate in examples. We show the following main result that formalizes Hart and Moore’s (1988) premise for studying incomplete contracts: if costs of contractibility have marginal costs that decline sufficiently slowly, in a specific sense that we define, then the principal chooses a *coarse contract* that has only finitely many items. In further results, we derive an upper bound for the optimal number of items in the menu and study how “contract incompleteness” depends on features of the environment. We apply the model to study when and why incomplete contracts would emerge in product markets, manifested as optimally designed coarse quality grades for a differentiated product or service.

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 preferences over the action (i.e., it is a “good” or a “bad”). 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 three economic axioms on that 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. We also impose a fourth, technical axiom, that the contractibility correspondence is closed-valued. We show that these axioms translate into natural patterns of incomplete contracting, in which the action space can be partitioned into regions with perfectly contractible actions, action regions that permit deviations up or down, and action regions that are fully indistinguishable.

We finally allow the principal to also select the contractibility correspondence, in a stage before the specific contract is written. Choosing a given level of contractibility has a cost. This cost reflects the principal's efforts to better monitor and verify actions, more precisely write the contract, or better justify the contract in a court of law.

This model is designed to capture two key ideas about incomplete contracts: that an incomplete contract imperfectly specifies the outcomes that are payoff relevant to both principal and agent and that the fundamental origin of incompleteness is an *ex ante* cost to specifying all contingencies. To focus on these elements, the model abstracts from other issues studied in this literature including planning for *ex post* renegotiation and manipulation of reference points (Hart and Moore, 2008).

Optimally Designed Incomplete Contracts. To begin, we fix the contractibility correspondence and study how the principal optimally designs an (incomplete) 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. If the agent's preferences are increasing in the action (i.e., it is a good), the set is given by the image of the action space of the maximum selection from the contractibility correspondence. Intuitively, with increasing preferences, agents prefer to take the highest possible action within the letter of the contract. If the agent's preferences are decreasing in the action (i.e., it is a bad), both the result and intuition are the opposite.

We next characterize optimal contracts. The optimal outcome function maximizes virtual surplus (i.e., total surplus net of information rents) subject to being supported on the given set. We show that this takes a simple form: to pick the best contractible action that is "close" to what the principal would pick with full contractibility.

Optimally Incomplete Contracts. We then return to the question of 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.¹ Under a 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.

We next place two substantial economic restrictions on these costs, which restrict the marginal cost of writing a “more complete” contract. We say that a cost is *interval strongly monotone* if the cost of introducing perfect contracting in an interval of the action space is (at most) second-order in the length of the interval. We say a cost is *tail strongly monotone* if a similar property holds near non-interior accumulation points (e.g., limits of sequences of isolated points). We show that both properties are obtained in a benchmark class of *costs of distinguishing*. We construct these costs as an *ex ante* cost for the principal, summing over all outcomes, of distinguishing against all other outcomes. In this class, the costs of removing contractibility as described are exactly second order for the following reason: the principal saves effort by not having to distinguish each outcome in the interval (i.e., measure μ) from each other outcome in the interval (i.e., measure μ), proportional to total cost $\mu \times \mu = \mu^2$.

Our main result is that, if costs are interval and tail 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* in the measure-theoretic sense and leave a bounded-size gap between any two adjacent items. Our result provides a foundation for incomplete contracts and sparse menus observed in practice based on formalizing the intuition that writing and enforcing more precise contracts is costly. This result holds even when the cost of implementing complete contracts is arbitrarily low.

The key step of our proof shows by contradiction that the optimal contractibility correspondence cannot contain intervals of perfect contractibility, even if they are arbitrarily small. To do so, we construct a payoff-improving alternative contractibility correspondence that introduces “local incompleteness,” or replaces these intervals with its two boundary points. The principal’s surplus loss under the optimal contract derived in the first part of the paper is *third-order* in the length of the interval. For each type that is allocated an outcome in the interior of this interval, the principal was originally maximizing their (quasi-concave) 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

¹We endow this space of closed subsets of the action space with the topology induced by the Hausdorff distance.

any losses can be described by a *second-order term*. To obtain the total loss in surplus, we integrate these second order losses over an interval of *types* that 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 interval strong monotonicity. This argument rules out intervals of perfect contractibility, and guarantees contracts supported on a set of isolated points. A very similar argument, this time using tail strong monotonicity, contradicts the existence of accumulation points and implies a finite support.

We next use the logic above to derive an upper bound for 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.” We return to this idea in the applications.

We finally derive alternative sufficient condition for finite contractability that holds under *clause-based costs*, or costs that depend solely on the number of outcomes that are perfectly contractible. These costs are natural, as the name suggests, if the principal’s primary difficulty is adding separate clauses to the contract which describe contractible actions. We derive a condition on these costs that guarantees optimal finiteness of contracts. We also derive and interpret a necessary condition for the planner’s choice of contractible actions, conditional on the number of optimally chosen contractible actions.

Application: Monopoly Pricing with Coarse Contracts. We apply our model to study monopoly pricing with endogenous and costly contractibility. We study a case with quadratic preferences and uniform type distribution, as in [Mussa and Rosen \(1978\)](#). We specialize the costs of contracting to the *costs of distinguishing*, introduced earlier as a leading example.

One interpretation of the model is that it 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, and a set of contractible utilization levels, which is new to our setting. Here, contractibility may be costly because the principal has to describe, *ex ante*, what are different levels of utilization of the good—for example, what constitutes a unit in “good” versus “bad” condition.

Our general results ensure that an optimal contract exists, that the optimal level of contractibility is supported on finite menu items, and that there is a closed-form upper bound for the number of items in that menu. Using the parametric structure of this example, we specialize these results in several ways.

First, we show that the optimal contract always features *uniformly* spaced quality levels. Intuitively, in this quadratic-uniform example, the monopolist’s losses from coarse contracting are the same at all points in the menu, and therefore the monopolist has no incentive to make contracts more or less precise for high *vs.* low quality levels.

Next, we give a formula for the number of points in the menu (up to integer rounding) up to parameters that control production costs (i.e., concavity), differentiation in preferences (i.e., supermodularity), and costs of contractibility. These parameters enter 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, to isolate the role of adverse selection in our results, we study a variant problem in which there is no asymmetric information about agents’ preference shifter. We show that contracts are always “more complete,” or contain more menu items, in this case. Intuitively, adverse selection reduces the size of the pie available to the monopolist and dulls their incentives to contract more precisely. Thus, in a formal sense, adverse selection begets more incomplete contracts in this setting.

Related Literature. Our approach to modeling incomplete contracts is inspired by the dichotomy between perfunctory (letter) and consummate (spirit) performance formally introduced by [Hart and Moore \(2008\)](#).² Under complete information, [Hart and Moore \(2008\)](#) take a behavioral approach to understand 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, the classical 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.

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 tradeoff for the principal. Two notable examples are [Bajari and Tadelis \(2001\)](#) and [Battigalli and Maggi \(2002\)](#). Neither of these papers considers ex-ante asymmetric information between the principal and the agent.

²In turn, this language choice is inspired by [Williamson \(1975\)](#).

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 analysis of moral hazard concerns imperfect observability (with perfect contractibility). As we show, this difference in perspective leads to qualitatively different optimal mechanisms.

Finally, our work is technically related to models of optimal design where a continuous variable is optimally discretized as a result of a tradeoff 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 much closer to the one of the present paper, [Bergemann, Heumann, and Morris \(2022\)](#) show that, in a standard [Mussa and Rosen \(1978\)](#) nonlinear pricing model, if the monopolist can simultaneously choose the selling mechanism and the buyer’s information, then the 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.

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 $X = [0, \bar{x}] \subset \mathbb{R}$.

The Agent’s preferences are represented by a twice continuously differentiable utility function $u : X \times \Theta \rightarrow \mathbb{R}$. 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 or decreasing) over X . 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 $\pi : X \times \Theta \rightarrow \mathbb{R}$. 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 *contracting 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 consistent with [Hart and Moore's \(1988\)](#) use of the terms *consummate performance* (our y) and *perfunctory performance* (our $C(y)$), borrowed from [Williamson \(1975\)](#).

Regular Contracting Correspondences. We economically discipline the relationship between the spirit and letter of a contract by imposing five axioms. The first three 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 .

The final two axioms are technical:

Axiom 4 (Closed-valuedness). *For all $y \in X$, $C(y)$ is closed.*

Axiom 5 (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 contracting correspondences that satisfy all five axioms. We will refer to such contracting 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 indistinguishable 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 $x > 1/2$.

Representing Regular Contractibility. The examples above suggested a specific structure for regular contracting correspondences. Below, we provide two characterizations of these correspondences that clarify these economic properties. In our later analysis, these representations also turn out to be mathematically convenient.

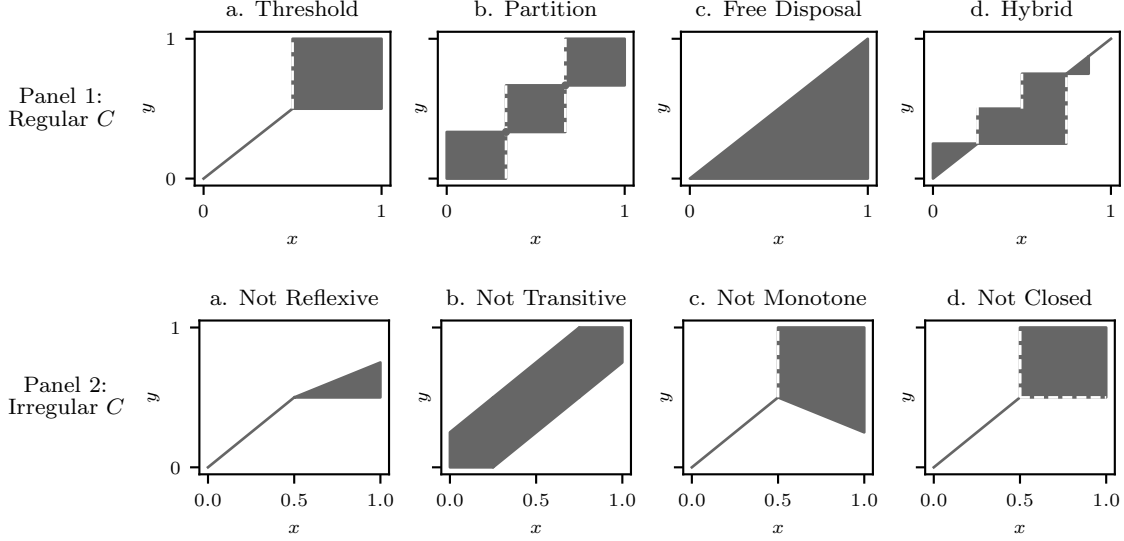
Lemma 1. *Fix a contractibility correspondence C . The following statements are equivalent:*

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

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

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

Figure 1: Illustrating Regular Contracting Correspondences



Note: Each graph illustrates a contracting 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 1-4.

Proof. See Appendix A.1. □

The first alternate characterization (Part 2) is in terms of the upper and lower envelope of the correspondence, $\bar{\delta}(y) = \max\{x \in C(y)\}$ 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 indistinguishability (“boxes”).

The second alternate characterization (Part 3) is in terms of the images of these functions, indeed we have $\underline{D} = \underline{\delta}(X) \subseteq X$ and $\overline{D} = \bar{\delta}(X) \subseteq X$. As will become more clear in our analysis, these correspond to the recommendations that an agent with monotone decreasing or increasing preferences (respectively) would follow. These are closed sets which respectively always include the minimum and maximum action.

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 preferred 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. But, mathematically, the cost may also be the expectation of a cost borne *ex post*, for instance in litigation. In both cases, we assume that the principal *commits* to a certain level of contractibility before proposing the actual contract to the agent.

We express these costs via a function $\Gamma : \mathcal{C} \rightarrow [0, \infty]$. 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.

We next provide two concrete examples that illustrate two leading motivations for costly contractibility: the difficulty of distinguishing different outcomes in the contract and a cost of adding clauses to a contract.

Example 1 (Costs of Distinguishing Outcomes). 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 y , the obligations by a monetary transfer (to be introduced to the model in due course), and a description of the rights embodied by $C(y)$. An important challenge for the principal is to differentiate the rights under y , $C(y)$, from the actions *outside* of the agent’s rights in the same scenario, $X \setminus C(y)$. We embody this idea by assuming that the cost of distinguishing $z \in C(x)$ from $y \in X$ is equal to $|z - y|^p$ for some $p > -1$. When $p = 0$, this cost is constant over all the pairs of actions and normalized to one $1_{[y \neq z]}$. When $p > 0$, this cost is proportional to L^p norm and makes *far away* actions harder to distinguish. When $p \in (0, 1)$, this cost is proportional to the inverse L^p norm and makes *nearby* actions harder to distinguish.

We therefore express these costs of a regular contractibility correspondence $C(x) = [\underline{\delta}(x), \bar{\delta}(x)]$ as the following:

$$\Gamma(C) = \int_0^{\bar{x}} \int_0^{\bar{x}} \min_{z \in [\underline{\delta}(x), \bar{\delta}(x)]} |z - y|^p dy dx \quad (2)$$

where the previous expression is obtained by summing over all the Hausdorff distances between any arbitrary action y and the set of feasible actions given an arbitrary initial assignment x .

With some simple algebra, we can simplify this cost function as follows:

$$\begin{aligned} \Gamma(C) &= \int_0^{\bar{x}} \int_{\bar{\delta}(x)}^{\bar{x}} (y - \bar{\delta}(x))^p dy dx + \int_0^{\bar{x}} \int_0^{\underline{\delta}(x)} (\underline{\delta}(x) - y)^p dy dx \\ &= \int_0^{\bar{x}} \frac{(\bar{x} - \bar{\delta}(x))^{1+p}}{1+p} dx + \int_0^{\bar{x}} \frac{\underline{\delta}(x)^{1+p}}{1+p} dx \end{aligned} \quad (3)$$

The special case of $p = 0$ is easiest to interpret. Geometrically, in this cost, the cost equals the area lying above the graph of $\bar{\delta}$ and below the graph of $\underline{\delta}$. Mathematically, this is the average of the Hausdorff distance between the points in X and the indistinguishable points in the contractibility correspondence, when we endow X with the discrete metric $1_{[y \neq z]}$. Observe that this cost is 0 for the zero-contractibility correspondence $C(x) = [0, \bar{x}]$ and it is equal to its maximum of $\bar{x}^2$ for the perfect contractibility correspondence $C(x) = \{x\}$. More generally, this cost is monotone in a natural way: costs increase when $\bar{D}$ or $\underline{D}$ increase in the set inclusion order. $\triangle$

Example 2 (Costs of Writing Clauses). Suppose that the principal faces an effort cost of specifying the number of clauses in the contract, where we define a clause as an outcome that can be exactly specified. Define $\bar{D} = \bar{\delta}(X)$ and $\underline{D} = \underline{\delta}(X)$. Let $n(\cdot)$ denote the cardinality of a set. The number of clauses in the contract is given by $n(\bar{D}) + n(\underline{D})$.

$$\Gamma(C) = \alpha \left(n(\bar{D}) + n(\underline{D}) \right)^\alpha \quad (4)$$

for some $\alpha \in \mathbb{R}$. For $\alpha \geq 1$, the cost is convex and satisfies $\lim_{n \rightarrow \infty} \hat{\Gamma}(n) = \infty$. For $\alpha \in (0, 1)$, the cost is concave and also infinite under infinite clauses. For $\alpha < 0$, the cost is concave and satisfies $\lim_{n \rightarrow \infty} \hat{\Gamma}(n) = 0$. $\triangle$

2.4 The Principal's Problem

We first consider the principal's mechanism design problem for a fixed level of contractibility. Given the revelation principle, we consider direct and truthful mechanisms, and in particular we restrict to the deterministic ones. Thus, a mechanism is a triple (ϕ, ξ, T) comprising a recommendation $\xi : \Theta \rightarrow X$, a final action or outcome $\phi : \Theta \rightarrow X$, and a tariff $T : X \rightarrow \mathbb{R}$. 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 contracting 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 1 (Implementable Mechanism). *A mechanism (ϕ, ξ, T) is implementable given contractibility C if and only if the following three conditions are satisfied:*

1. *Obedience:*

$$\phi(\theta) \in \arg \max_{x \in C(\xi(\theta))} u(x, \theta) \quad \text{for all } \theta \in \Theta \quad (O) \quad (5)$$

2. *Incentive Compatibility:*

$$\xi(\theta) \in \arg \max_{y \in X} \left\{ \max_{x \in C(y)} u(x, \theta) - T(y) \right\} \quad \text{for all } \theta \in \Theta \quad (IC) \quad (6)$$

3. *Individual Rationality:*

$$u(\phi(\theta), \theta) - T(\xi(\theta)) \geq 0 \quad \text{for all } \theta \in \Theta \quad (IR) \quad (7)$$

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 $C(\xi(\theta))$.³ 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, the principal maximizes the sum of transfers and payoffs arising from agents' final actions or solves

$$\mathcal{J}(C) = \sup_{(\phi, \xi, T) \in \mathcal{I}(C)} \int_{\Theta} (\pi(\phi(\theta), \theta) + T(\xi(\theta))) dF(\theta) \quad (8)$$

We refer to the maximum (ϕ, ξ, T) , if it exists, as the *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

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

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. Nonetheless, we will study the composite problem by first describing the optimal contract given contractibility (Problem 8) in Section 3 before studying optimal contractibility (Problem 9) in Section 4.

³We 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 Implementable Outcomes

In principle, partial contractibility affects the problem in complex ways due to the interactions of obedience and incentive compatibility. This interaction is “built into” Equation 6 of Definition 1 in the following way: 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' . This puts an additional *global* constraint on what the principal can implement. However, we show below that these constraints drastically simplify.

Proposition 1 (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:*

$$T(x) = T(0) + u(x, \phi^{-1}(x)) - \int_0^{\phi^{-1}(x)} u_\theta(\phi(s), s) \, ds \quad (10)$$

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

Proof. See Appendix A.2. □

The first part of the result says that actions are implementable if they are monotone increasing in θ and lie in $\overline{D}$ (the “goods” case) or $\underline{D}$ (the “bads” case), which are the natural representations of C in this context. We illustrate the intuition in the goods case. After being given any $y \in X$, the agent’s favorite point is $\bar{\delta}(y)$. Thus, if $y < \bar{\delta}(y)$, Obedience fails and the contract is not implementable. Moreover, if $\phi(\theta) \in \overline{D}$, and ϕ is monotone, then transfers can be designed so that that Obedience and Incentive Compatibility hold. The second part of the result gives an explicit formula for the tariff that supports ϕ . To derive it, we apply the envelope formula to the reporting problem of any type and then recast transfers as a tariff. Finally, the proof of the result shows how to construct this implementable allocation

such that recommendations ξ coincide with final outcomes ϕ ; we henceforth focus on (ϕ, T) as the key objects of the contract.

3.2 The Optimal Contract

We now describe optimal contracts. To do this, we first define the virtual surplus function $J : X \times \Theta \rightarrow \mathbb{R}$ as:

$$J(x, \theta) = \pi(x, \theta) + u(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(x, \theta) \quad (11)$$

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. We assume that J is strictly supermodular in (x, θ) and strictly quasiconcave in x . These assumptions are satisfied, for instance, in the *regular case* of the monopoly pricing model of [Mussa and Rosen \(1978\)](#). Applying standard arguments along with Proposition 1, we obtain that the optimal contract solves the following control problem:

Lemma 2. *When agents have monotone increasing preferences, any optimal final outcome function solves:*

$$\begin{aligned} \mathcal{J}(\overline{D}) &:= \max_{\phi} \int_{\Theta} J(\phi(\theta), \theta) dF(\theta) \\ \text{s.t. } &\phi(\theta') \geq \phi(\theta), \phi(\theta) \in \overline{D}, \quad \theta, \theta' \in \Theta : \theta' \geq \theta \end{aligned} \quad (12)$$

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

Proof. See Appendix A.3. □

Observe that this lemma implies that, for monotone increasing preferences, only $\overline{D}$ matters for solving for the optimal contract (and symmetrically $\underline{D}$ for decreasing preferences).⁴

For the rest of this section, we solve the monotone increasing case. The solution in the monotone decreasing case is analogous. We first define the principal's favorite final outcome function $\phi^P : \Theta \rightarrow X$ as:

$$\phi^P(\theta) = \arg \max_{x \in X} J(x, \theta) \quad (13)$$

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

$$\overline{\phi}(\theta) = \min\{x \in \overline{D} : x \geq \phi^P(\theta)\} \quad \text{and} \quad \underline{\phi}(\theta) = \max\{x \in \overline{D} : x \leq \phi^P(\theta)\} \quad (14)$$

⁴For this reason, we denote the induced value for the principal by $\mathcal{J}(\overline{D})$ as opposed to $\mathcal{J}(C)$.

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:

$$\Delta J(\theta) = J(\overline{\phi}(\theta), \theta) - J(\underline{\phi}(\theta), \theta) \quad (15)$$

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

Theorem 1 (Optimal Contract). *Fix a regular contractibility correspondence with upper image set $\overline{D}$ and assume that u is monotone increasing in x . Any optimal final outcome function is almost everywhere equal to:*

$$\phi^*(\theta) = \begin{cases} \overline{\phi}(\theta), & \Delta J(\theta) > 0, \\ \underline{\phi}(\theta), & \Delta J(\theta) \leq 0. \end{cases} \quad (16)$$

Proof. See Appendix A.4. □

Intuitively, the optimal contract implements the “next best” thing to $\phi^P(\theta)$ that is actually contractible, 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.⁵

We finally specialize and illustrate this result in a case that will become important later: when $\overline{D}$ can be written as a sequence of ordered isolated points, or $\overline{D} = \{x_1, \dots, x_K\}$. An implication of Theorem 1 is that the contract has the following structure:

Proposition 2 (Coarse Contracts). *If $\overline{D} = \{x_1, \dots, x_K\}$, any optimal final outcome function is almost everywhere equal to:*

$$\phi(\theta) = \sum_{k=1}^K x_k 1_{\theta \in (\hat{\theta}_k, \hat{\theta}_{k+1}]} \quad (17)$$

where for $k \in \{2, \dots, K\}$, $\hat{\theta}_k$ 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, $T : \overline{D} \rightarrow \mathbb{R}$, is given by

$$T(x_k) = u(x_1, 0) + \mathbb{I}[k \geq 2] \sum_{j=2}^k \left[u(x_j, \hat{\theta}_j) - u(x_{j-1}, \hat{\theta}_j) \right] \quad (18)$$

⁵The assumption of lower hemicontinuity of C is made purely for technical reasons to ensure that $\overline{D}$ is closed. When C is not lower hemicontinuous, an optimal contract can fail to exist, but it is arbitrarily well approximated by small deviations from the contract in Theorem 1.

Proof. See Appendix A.5. □

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

$$u(x, \theta) = x\theta \quad \pi(x, \theta) = -\frac{1}{2}x^2 \quad \theta \sim U[0, 1] \quad (19)$$

We allow for contractibility on a four-point, evenly spaced partition of the action space $X = [0, 1]$: $\bar{D} = \{0, 1/3, 2/3, 1\}$. One contractibility correspondence that induces such an $\bar{D}$ 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.

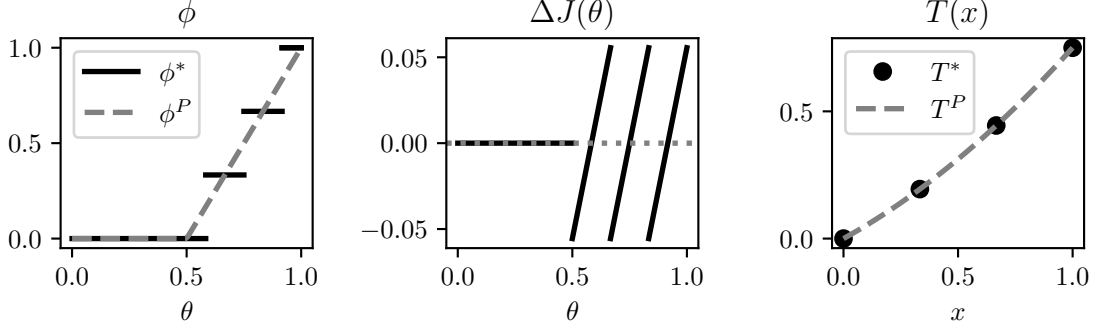
We remind that the optimal contract under *full contractibility*, as studied by [Mussa and Rosen \(1978\)](#) *inter alia*, assigns $\phi^P(\theta) = 0$ for $\theta \in [0, 1/2]$ and $\phi^P(\theta) = 2\theta - 1$ for $\theta \in (1/2, 1]$. The optimal contract under full contractibility charges tariff $T(x) = \frac{x^2}{4} + \frac{x}{2}$.

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 2, 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. △

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

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 $\bar{D} = \{0, 1/3, 2/3, 1\}$. The first panel shows the assignment ϕ ; the second panel shows the function $\Delta J(\theta)$ defined in Equation 15 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) .

contracts are *coarse* or supported on finitely many outcomes. As in the previous section, we focus on the case of monotone increasing preferences and note that all arguments are analogous in the decreasing case.

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 $\bar{D}$ summarizes the effects of imperfect contractibility on the optimal contract. We let $\mathcal{D}$ denote the set of possible $\bar{D}$, or closed subsets of X that contain $\bar{x}$ and 0, and endow it with the topology induced by the Hausdorff distance between closed sets (see Lemma 1).⁶ We defined $\mathcal{J} : \mathcal{D} \rightarrow \mathbb{R}$, the value induced by solving the non-linear pricing problem given a particular extent of contracting, in Lemma 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 $\bar{D} \in \mathcal{D}$, we let $\Gamma(\bar{D})$ denote the cost of the regular contractibility correspondence represented by $\bar{D}$ and $\{0\}$. We assume henceforth that $\Gamma : \mathcal{D} \rightarrow \mathbb{R}$ is lower semi-continuous.

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

⁷Observe 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.

This is satisfied by the cost functions introduced in Examples 1 and 2.

Using this we can rewrite the program of Equation 9 as the following choice of $\overline{D}$:

$$\sup_{\overline{D} \in \mathcal{D}} \mathcal{J}(\overline{D}) - \Gamma(\overline{D}) \quad (20)$$

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

Proposition 3. *The set of optimal contractibility supports $\mathcal{D}^*(\Gamma)$ solving (20) is nonempty and compact.*

Proof. See Appendix A.6. □

Putting together Theorem 1 and Proposition 3, we have that the joint design problem of optimally choosing a contractibility correspondence and then a contract has well-defined solutions. This is despite the high dimensionality of the initially posed problem of choosing a regular contractibility correspondence (Equation 9).

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 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 2. *A cost function Γ is strongly monotone if for all $\overline{D} \in \mathcal{D}$ and accumulation points $x \in \overline{D}$, there exists $\epsilon > 0$ such that:*

$$\liminf_m \frac{\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (a_m, b_m))}{(x_m - a_m)(b_m - x_m)} \geq \epsilon \quad (21)$$

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

⁸The 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

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

Definition 2 generalizes this idea to also discipline the cost of precise contracting around non-interior accumulation points. For example, the set $\bar{D} = \{1 - 2^{-k}\}_{k=0}^\infty \cup \{1\}$ 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 $\bar{D}$ were the Cantor set, then *all* of its elements are non-interior accumulation points. The more flexible form of Definition 2 is required to rule out the optimality of such sets.

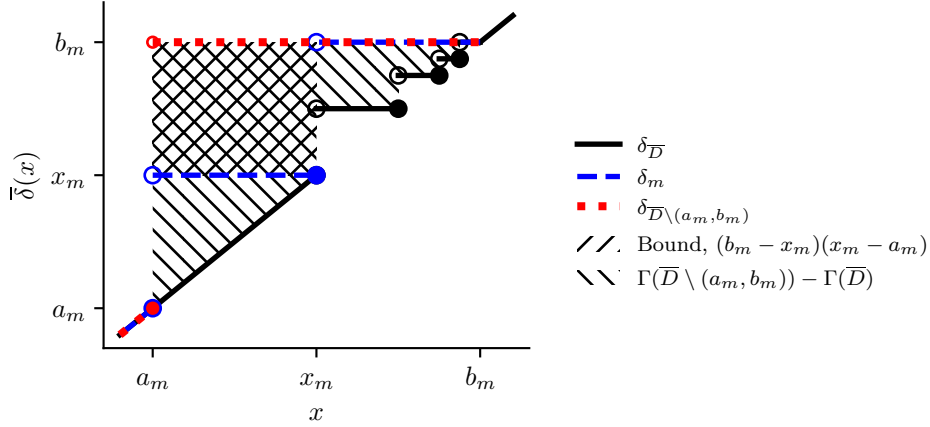
We finally observe that costs of distinguishing outcomes (Example 1) is strongly monotone, when $p \in (-1, 0]$:

Lemma 3. *If $p \in (-1, 0]$, the cost of distinguishing outcomes is strongly monotone with $\epsilon = \bar{x}^p$.*

We can give a simple geometric intuition for the proof when $p = 0$ and the cost coincides with the area above $\bar{\delta}$ (Figure 3). We first observe that any $\bar{D}$, which induces an upper envelope $\delta_{\bar{D}}$ (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 δ_m (blue dashed line). This is itself “greater” than $\bar{D} \setminus (a_m, b_m)$ (red dotted line). The cost savings moving between the dashed line and the dotted line is the right-hatched rectangle, with side lengths $b_m - x_m$ and $a_m - x_m$. These costs savings are a lower bound for the cost savings of moving from $\bar{D}$ to $\bar{D} \setminus (a_m, b_m)$, which are indicated with left-hatched shading.

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.

Figure 3: Strong Monotonicity for Costs of Distinguishing



Note: An illustration of strong monotonicity for costs of distinguishing (Example 1) with $p = 0$. The function δ_m is constructed in the proof of Lemma 3. Note that, in this example, the bound is not tight.

4.3 Optimal Coarse Contracts

We now state our main result for this section:

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

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 $\bar{D} \subseteq [0, \bar{x}]$ if there exists a tariff T with proper domain $\bar{D}$ that induces ϕ .

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

This combination of Theorem 2 and Corollary 1 provides a foundation for endogenous incomplete contracts under the presence of contractibility costs. The contracts are incomplete because they do not need to specify rights (e.g., allowable actions) or responsibilities (e.g., payments) for certain payoff-relevant outcomes in X . This incompleteness takes a very strong form under a coarse contract, because *almost all* actions (in the measure-theoretic sense) are left unspecified. Moreover, this result holds for any arbitrarily small degree of cost of writing contracts, since the ϵ in Definition 2 can be made arbitrarily small (relative to the scale of the contracting surplus).

We now prove each part of Theorem 2, up to some more technical steps which are relegated to the Appendix.

Intermediate Step: The Opportunity Cost of Coarsening a Contract. We first give an intermediate result that bounds the loss to the principal from “removing contractibility” in an interval of the action space. In this result, we use the definitions 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, $0 < \bar{J}_{xx}, \underline{J}_{x\theta}, \bar{f} < \infty$.

Lemma 4. *Consider any $\bar{D} \in \mathcal{D}$ and any $a, b \in \bar{D}$ such that $a < b$. Then,*

$$\mathcal{J}(\bar{D}) - \mathcal{J}(\bar{D} \setminus (a, b)) \leq \frac{3}{2} \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} (b - a)^3 \quad (22)$$

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

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

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

$$\mathcal{J}(\bar{D}) - \mathcal{J}(\bar{D} \setminus (a, b)) \leq 3 \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} (b - a)(c - a)(b - c) \quad (24)$$

Proof. See Appendix A.7. □

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 22).

We first observe, exploiting our results from Section 3.2, that optimal allocations conditional on any level of contractibility solve a pointwise program (see Lemma 2). Thus, we can re-express $\mathcal{J}(\bar{D}) - \mathcal{J}(\bar{D} \setminus (a, b))$ as

$$\mathcal{J}(\bar{D}) - \mathcal{J}(\bar{D} \setminus (a, b)) = \int_0^1 (J(\phi^*(\theta), \theta) - J(\phi^{*'}(\theta), \theta)) dF(\theta) \quad (25)$$

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 $\bar{\phi}(\theta)$ or $\underline{\phi}(\theta)$, defined relative to $\bar{D}$, 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 $\frac{\bar{J}_{xx}}{\bar{J}_{x\theta}}$ and the maximum type distribution by $\bar{f}$. Together, this contributes a term $(b - a) \frac{\bar{J}_{xx}}{\bar{J}_{x\theta}} \bar{f}$ 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 22.

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

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

1. $\bar{B}_t(x) \cap \bar{D}$ 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 (homeomorphic to) the Cantor set.
2. $\bar{B}_t(x) \cap \bar{D}$ 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.

Note finally that $\overline{B}_t(x) \cap \overline{D}$ cannot be finite. In this case, there would exist a closest point to x within $\overline{B}_t(x) \cap \overline{D}$ (call it z), the neighborhood $\overline{B}_{|z-x|/2}(x) \cap \overline{D}$ would necessarily be empty, and x would not be an accumulation point.

We proceed to show, by contradiction, that each of these cases contradicts optimality. In each case, our argument will be that, given strong monotonicity (Definition 2), 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 4.

To illustrate this method, we fully present the argument that rules out intervals:

Lemma 5. *If Γ is strongly monotone and $\overline{D} \in \mathcal{D}$ contains an interval, then $\overline{D}$ is not optimal.*

Proof. 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 22 from Lemma 4. In particular, for every m , we have that:

$$\mathcal{J}(\overline{D}) - \mathcal{J}\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) \leq 12 \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} t^3 m^{-3} \quad (26)$$

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:

$$\Gamma(\overline{D}) - \Gamma\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) \geq \epsilon t^2 m^{-2} \quad (27)$$

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

$$\mathcal{J}(\overline{D}) - \Gamma(\overline{D}) < \mathcal{J}\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) - \Gamma\left(\overline{D} \setminus \left(x - \frac{t}{m}, x + \frac{t}{m}\right)\right) \quad (28)$$

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

We relegate to the Appendix the detailed arguments that rule out nowhere dense perfect sets (e.g., the Cantor set) and countably infinite sets. These arguments use Equations 23 and 24, respectively, along with more specialized implications of strong monotonicity. Thus, while these arguments have the same basic form of the proof of Lemma 5, they have meaningfully different economic content.

Lemma 6. *If Γ is strongly monotone and $\bar{D} \in \mathcal{D}$ contains an accumulation point x such that $\bar{B}_t(x) \cap \bar{D}$ is a perfect and nowhere dense set for some $t > 0$, then $\bar{D}$ is not optimal.*

Proof. See Appendix A.8. □

Lemma 7. *If Γ is strongly monotone and $\bar{D} \in \mathcal{D}$ is countably infinite, then $\bar{D}$ is not optimal.*

Proof. See Appendix A.9. □

We finally put these steps together to complete the proof of finiteness, referring back to our exhaustive list of cases. Under case 1(a), Lemma 5 contradicts optimality. Under case 1(b), Lemma 6 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), Lemma 7 contradicts optimality. Thus, we have shown that $\bar{D}^*$ cannot contain an accumulation point. As the set is also compact, it must be finite.

Deriving the Bound. We now derive an explicit bound on the number of elements in $\bar{D}^*$. As this set must be finite (from the first argument), we can express it as a sequence of ordered points. Take any three sequential points $x_{m-1}, x_m, x_{m+1} \in \bar{D}^*$. We can apply statement 3 of Lemma 4 (Equation 24) to bound the loss from eliminating contractibility at x_m :

$$\mathcal{J}(\bar{D}^*) - \mathcal{J}(\bar{D}^* \setminus \{x_m\}) \leq 3 \frac{\bar{J}_{xx}^2 \bar{f}}{J_{x\theta}} (x_m - x_{m-1})(x_{m+1} - x_m)(x_{m+1} - x_{m-1}) \quad (29)$$

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

$$\Gamma(\bar{D}^*) - \Gamma(\bar{D}^* \setminus \{x_m\}) \geq \epsilon(x_m - x_{m-1})(x_{m+1} - x_m) \quad (30)$$

Optimality of $\bar{D}^*$ requires that $\mathcal{J}(\bar{D}^*) - \mathcal{J}(\bar{D}^* \setminus \{x_m\}) \geq \Gamma(\bar{D}^*) - \Gamma(\bar{D}^* \setminus \{x_m\})$. Combining this with Inequalities 29 and 30, we have that:

$$3 \frac{\bar{J}_{xx}^2 \bar{f}}{J_{x\theta}} (x_m - x_{m-1})(x_{m+1} - x_m)(x_{m+1} - x_{m-1}) \geq \epsilon(x_m - x_{m-1})(x_{m+1} - x_m) \quad (31)$$

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

$$x_{m+1} - x_{m-1} \geq \frac{\epsilon}{3} \frac{J_{x\theta}}{\bar{J}_{xx}^2 \bar{f}} \quad (32)$$

Thus, we have that:

$$\bar{x} \geq x_{K^*} - x_1 = \sum_{j=1}^{\lfloor K^*/2 \rfloor} x_{2j+1} - x_{2j-1} \geq K^* \frac{\epsilon}{6} \frac{J_{x\theta}}{J_{xx}^2 f} \quad (33)$$

Re-arranging this equation yields the desired bound.

The bound inherits the comparative statics of our payoff bound in Lemma 4. 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 a wage-schedule example.

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 Corollary 2, the optimal contract given a coarse contractibility correspondence allocates action x_k to types $\theta \in [\hat{\theta}_k, \hat{\theta}_{k+1})$, where $\hat{\theta}_k$ is defined as the solution to $J(x_k, \hat{\theta}_k) = J(x_{k-1}, \hat{\theta}_k)$ for $k \in \{2, \dots, K\}$, 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:

$$\Pi(\{x_k\}) = \sum_{k=1}^K \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J(x_k, \theta) dF(\theta) \quad (34)$$

Let $\mathcal{D}_K$ be the set of all $\bar{D} \in \mathcal{D}$ such that $|\bar{D}| = K$. We observe that $\mathcal{D}_K \subseteq X^K$. Given any Γ and $K \in \mathbb{N}$, define the family of restricted cost functions $\Gamma_K : \mathcal{D}_K \rightarrow \bar{\mathbb{R}}$ with $\Gamma_K(\bar{D}) = \Gamma(\bar{D})$ for all $\bar{D} \in \mathcal{D}_K$. To proceed, we temporarily restrict attention to cost functionals that are suitably differentiable on this domain:

Definition 3 (Finite Differentiability). Γ is *finitely differentiable* if Γ_K is a *continuously differentiable function* for all $K \in \mathbb{N}$.

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 as

$$\Gamma_K^{(k)}(\bar{D}) = \lim_{\epsilon \downarrow 0} \frac{\Gamma(\{x_1, \dots, x_k + \epsilon, \dots, x_K\}) - \Gamma(\{x_1, \dots, x_k, \dots, x_K\})}{\epsilon} \quad (35)$$

for $0 < k < K$.

We observe that the cost of distinguishing satisfies this property:

Lemma 8. *If $p \in (-1, 0]$, the cost of distinguishing outcomes is finitely differentiable.*

Proof. Immediate from observing that the cost, on the domain of coarse contractibility, is

$$\Gamma(\overline{D}) = \sum_{k=2}^{K-1} (x_k - x_{k-1}) \frac{(\overline{x} - x_k)^p}{1+p} \quad (36)$$

□

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 4. *If Γ is strongly monotone and finitely differentiable, then any optimal contractibility support $\overline{D}^* = \{x_1, \dots, x_{K^*}\}$ with $0 = x_1 < \dots < x_k < \dots < x_{K^*} = 1$ satisfies:*

$$\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k, \theta) dF(\theta) = \Gamma_{K^*}^{(k)}(\overline{D}^*) \quad \text{for } k \in \{2, \dots, K^* - 1\} \quad (37)$$

where $\hat{\theta}_k$ is defined by $J(x_k, \hat{\theta}_k) = J(x_{k-1}, \hat{\theta}_k)$.

Proof. See Appendix A.10

□

The left-hand-side of Equation 37 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.

Using Proposition 4, we can solve for optimal contracts under specific cost functionals. In Section 5, we will do so in a wage-schedule example.

4.5 Coarse Contracts with Clause-Based Costs

Our proof of finiteness in Theorem 2 relied on contradicting the optimality of including accumulation points in $\overline{D}^*$, or very precisely describing a particular action. The strong monotonicity property, and its embodiment in costs of distinguishing (Example 1), were tightly linked to this idea.

In this subsection, we build a theory of coarse contracts predicated on a slightly different idea: that costs of contracting depend only on the number of items on the contract, and these items have a sufficiently slowly *declining* marginal cost as contracts become more dense.

We say that a contractibility cost is *clause-based* if, for any $\bar{D} \in \mathcal{D}$, we can write $\Gamma(\bar{D}) = \hat{\Gamma}(n(\bar{D}))$, where $n(\cdot)$ denotes the cardinality of a set and $\hat{\Gamma} : \bar{\mathbb{N}} \rightarrow \bar{\mathbb{R}}$ is a *montone increasing* cost defined on this cardinality. Example 2 in Section 2 is an example of a clause-based cost. Note that, for $\alpha < 1$, the cost is *concave*, and for $\alpha < 0$ the cost has a finite value for perfect contractibility. In such cases, the marginal cost of adding another contractible action to a sufficiently “rich” contract declines to zero.

If clause based costs are increasing and convex, like the $\alpha \geq 1$ case in Example 2, it is immediate that optimal contracts are finite. If they are increasing but concave, then it is possible that non-finite contracts are optimal. The key economic consideration, in analogy to Theorem 2, is whether marginal costs outpace the declining marginal benefits of contracting at many points. Below, we formalize such a condition:

Definition 4. We say that Γ is clause strongly monotone if it has a clause-based representation $\hat{\Gamma}$ such that

$$\liminf_{n \rightarrow \infty} (\hat{\Gamma}(n+1) - \hat{\Gamma}(n))n^\beta \geq \epsilon \quad (38)$$

for some $\epsilon > 0$ and $\beta < 3$.

In Example 2, the condition holds with $\beta = 1 - \alpha$. Therefore, power clause-based costs are clause strongly monotone if and only if $\alpha > -2$.

We now state the result:

Theorem 3. If Γ is clause-based and clause strongly monotone, then every optimal contractibility support is finite with $|\bar{D}^*| \leq 2 + \left(\frac{6J_{xx}^2 \bar{f}}{\epsilon J_{x\theta}} \right)^{\frac{1}{3-\beta}}$.

Proof. See Appendix A.11. □

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 4’s “individual grid cells” over the entire domain X . The third step shows that, when costs are clause strongly monotone, 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.

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 4. The only difference is that the marginal cost term in the right of Equation 37 is zero, as there is no contractibility cost of changing the value of any x_k .

5 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 costs of implementing the contract. In this example, we study exact comparative statics of optimal coarseness, or number of quality tiers, as a function of differentiation in consumers' tastes, production costs, and costs of contractibility. We show also how 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

In this section, 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

$$u(x, \theta) = \alpha \theta x \tag{39}$$

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

$$\pi(x, \theta) = -\beta \frac{x^2}{2} \tag{40}$$

where $\beta > 0$ scales the extent of these costs. We introduce the simplifying assumption that $\alpha \geq \beta$, so under all optimal contracts the highest types are allocated the maximum quality $x = 1$.

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 2).

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

$$\Gamma(\bar{D}) = \gamma \int_0^1 (1 - \bar{\delta}(x)) dx \quad (41)$$

where $\bar{\delta}$ is defined as in Lemma 1, $\gamma > 0$ is a scaling parameter, and where we ignore the additive term corresponding to $\underline{\delta}$ due to its irrelevance for the problem with increasing preferences.⁹ Abstractly, as introduced 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 calls. 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 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 Contracts Feature Quality Tiers

We now study the monopolist's optimal pricing policy when they jointly design contractibility and the optimal contract. Theorem 2 implies that the optimal level of contractibility in this setting is *finite*. Therefore, to characterize the highest-payoff contract, we write and solve the problem of maximizing over all possible vectors of quality levels.

We next leverage our characterization of the optimal contract in Proposition 4 to set up this 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 34 gives the principal's interim payoff (*i.e.*, tariff revenue net of service costs) 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 2. We therefore define the following value function describing the monopolist's favorite K -item contract as the solution of a quadratic constrained optimization

⁹See footnote 7.

problem:

$$\begin{aligned}
V(K) = \max_{\{x_1, \dots, x_K\}} & \left\{ \sum_{k=1}^K \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \left(\alpha(2\theta - 1)x_k - \beta \frac{x_k^2}{2} \right) d\theta - \gamma \left(1 - x_1^2 - \sum_{k=2}^K x_k(x_k - x_{k-1}) \right) \right\} \\
\text{s.t. } & 0 \leq x_k \leq x_{k+1}, \quad \forall k \leq K-1 \\
& x_K = 1 \\
& \hat{\theta}_k = \frac{\beta}{4\alpha}(x_k + x_{k-1}) + \frac{1}{2}, \quad 2 \leq k \leq K-1 \\
& \hat{\theta}_0 = 0, \hat{\theta}_1 = 1
\end{aligned} \tag{42}$$

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.

The optimization problem in Equation 42 demonstrates the linkages between the optimal contract, described in the first part of the objective, and the costs of contractibility, described in the second part of the objective. In particular, when choosing the values of each x_k , the principal must jointly consider the cost of specifying the contract *ex ante* and its ability to price discriminate *ex post*.

By deriving the first-order necessary condition for optimal $\{x_k\}_{k=1}^K$ in this setting (Proposition 4), we show that optimal contracts with endogenous contractibility must have a specific, *uniform partition* structure:

Proposition 5. *The monopolist optimally offers K^* quality options, $\{x_k\}_{k=1}^{K^*}$, where*

$$x_k = \frac{k-1}{K-1} \tag{43}$$

In this contract, items are priced by the tariff

$$T(x_k) = \frac{1}{2} \frac{k-1}{K-1} \left(\frac{\beta}{2} \frac{k-1}{K-1} + \alpha \right) \tag{44}$$

Any optimal level K^ solves the program*

$$K^* \in \arg \max_{k \in \mathbb{N}} V(K) \tag{45}$$

Proof. See Appendix A.12. □

We have, in fact, already illustrated such a contract in the example of Section 3.2 shown

in Figure 2. This example featured $K = 4$ and $\alpha = \beta = 1$.

Intuitively, 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 virtual surplus J 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—because the same number of types are affected (uniform distribution) and each type’s re-allocation contributes the same loss in surplus (constant concavity). This is the same argument underpinning Theorem 2 in the general model, but *without* needing uniform bounds for concavity or the measure of affected types. In economic language, the principal has an equal opportunity cost of forgoing quality differentiation for high qualities (high-demand customers) or low qualities (low-demand customers).

The corresponding symmetry in costs arises from our argument about distinguishing actions (Example 1). In particular, we assume that the difficulty in distinguishing actions does not vary over the action space—that is, in the units of the problem, that nearby low qualities are not easier or harder to distinguish than nearby high qualities.

To complete our description of the solution, we finally observe that the optimal number of quality tiers K^* maximizes the value function V . In the next section, we explore the comparative statics of K^* .

5.3 When is Pricing More or Less Coarse?

So far, we have established that optimal contracts are not just coarse but also have a specific structure. We now exploit this structure to derive comparative statics for the number of tiers in the contract. These translate the economic lessons from our earlier general bound on the number of menu items (Theorem 2) to this setting *exactly*.

As a first step, we use the optimality of the uniform partition contract to derive a closed-form expression for the value to the principal of a K -item contract. We separate this into expressions for the “benefits” and “costs.”

Lemma 9. *The value to the monopolist of a K -item contract, or the solution to the program*

in Equation 42, can be written as $V(K) = \hat{\Pi}(K) - \hat{\Gamma}(K)$ where

$$\begin{aligned}\hat{\Pi}(K) &= \frac{\alpha - \beta}{4} + \frac{\beta^2}{48\alpha} \frac{(2K - 3)(2K - 1)}{(K - 1)^2} \\ \hat{\Gamma}(K) &= \frac{\gamma}{2} \frac{K - 2}{K - 1}\end{aligned}\tag{46}$$

Proof. See Appendix A.13. □

The costs and benefits functions are both increasing. They respectively asymptote toward $\frac{\alpha - \beta}{4} + \frac{\beta^2}{6\alpha}$, the principal's value under perfect contractibility, and $\frac{\gamma}{2}$, the cost of perfect contractibility.

We moreover observe that $\hat{\Pi}_{K\alpha} < 0$, $\hat{\Pi}_{K\beta} > 0$, and $\hat{\Gamma}_{K\gamma} > 0$. That is, lower taste heterogeneity, higher production costs (by extension, concavity), and lower contracting costs raise the principal's incentives to choose a more fine-grained contract (*i.e.*, a higher K).

We now use these to describe the optimal extent of contractibility—in other words, the optimal number of quality tiers or the optimal “completeness” of the contract.

Proposition 6. *The principal's optimal level of contractibility K^* satisfies $|K^* - \tilde{K}| < 1$, where*

$$\tilde{K} = 1 + \frac{\beta^2}{12\alpha\gamma}\tag{47}$$

Moreover, K^* decreases in α , increases in β , and decreases in γ .

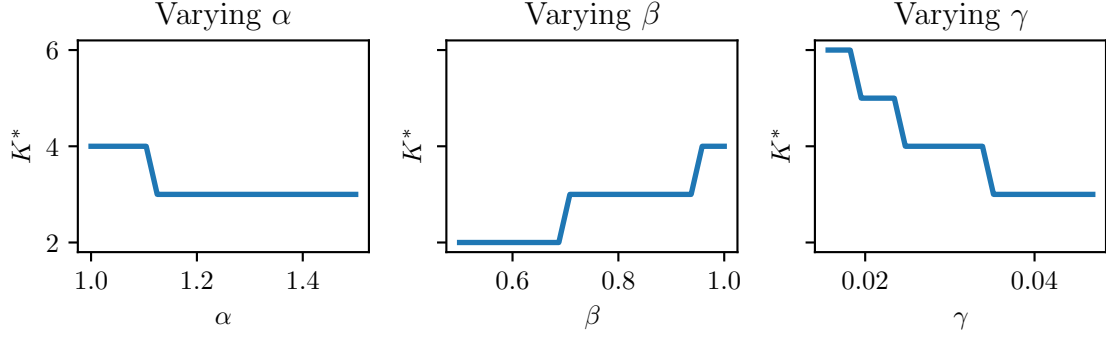
Proof. See Appendix A.14. □

The parameter $\tilde{K}$ is the unique maximum of the “smooth” (*i.e.*, non-integer) objective function $V(K)$. 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 is low, concavity is high, and costs of contracting 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).

We illustrate these comparative statics numerically in Figure 4. In the baseline calibration with $\alpha = \beta = 1$ and $\gamma = \frac{1}{32}$, the optimal contract has $K^* = 4$. This is exactly the contract shown in Figure 2. As we vary the parameters, K^* changes as predicted by Proposition 6.

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.

5.4 Incomplete Information Begets (More) Incomplete Contracts

We finally use the monopoly pricing example to further explore the interaction of incomplete information (*i.e.*, adverse selection) and incomplete contracts.

To do this, we now introduce a variant setting in which the monopolist observes the consumer's type θ .¹⁰ Under perfect contractibility, the monopolist would implement an “efficient” outcome that maximizes (the integral of) total surplus $S = \pi + u$ and perfectly extracts each consumer's willingness to pay. Under cost contractibility, however, the principal may prefer to *imperfectly* price discriminate and, in so doing, economize on the costs of writing a complex contract.

We model this scenario, in essence, by re-writing our studied monopolist's problem with $S(x, \theta) = \pi(x, \theta) + u(x, \theta)$ in place of $J(x, \theta)$. As $S(x, \theta)$ is concave in x and strictly super-modular in (x, θ) , our main results carry through to this setting. In particular, Theorem 2 guarantees that optimal contractibility in this scenario. We can therefore re-cast the problem of finding the optimal contract as an optimization over finite contractible actions, as in Section 5.2.

In the following Lemma, we show that the optimal contract has the same structure as the one previously studied as well a *exactly proportional and larger payoff*:

Lemma 10. *The optimal complete-information contract includes K^{*C} quality options, $\{x_k\}_{k=1}^{K^{*C}}$, where $x_k = \frac{k-1}{K-1}$. The level K^{*C} solves the program $K^{*C} \in \arg \max_{k \in \mathbb{N}} V^C(K)$ where $V^C(K) = 2\hat{P}i(K) - \hat{\Gamma}(K)$.*

¹⁰In other words, we consider the set of direct mechanisms that satisfy Obedience and Individual Rationality, but not Incentive Compatibility necessarily.

Proof. See Appendix A.15. □

The first part of the result has the same intuition as Lemma 5, relying on the symmetry of the benefits and cost functions. The second part follows from observing that

$$S(\theta, x) = J(2\theta - 1, x) \tag{48}$$

$\hat{\theta} = 2\theta - 1$ 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 $[1/2, 1]$).

From these observations, it is immediate that the complete-information monopolist has greater incentives to write a finer-grained contract—the benefits of price discrimination are twice as large. Thus,

Corollary 2. *Contracts are coarser (i.e., more incomplete) under incomplete information, that is, $K^* \leq K^{*C}$.*

6 Coarse Assignment: Additional Applications

In this section, we consider additional applications of our results in Section 4 on the optimality of coarse contracts.

6.1 Optimal Product Design

Consider a canonical monopoly environment similar to the Mussa and Rosen (1978) one analyzed in Section 5. Let $X = [0, 1]$ and, differently from before, assume that the buyer’s preferences have a bliss point and are given by $u(x, \theta) = \theta x - x^2/2$, with $\theta \sim U[0, 1]$. Moreover, assume that perfectly specified contracts are feasible for the monopolist, $C(x) = \{x\}$, that $\pi(x, \theta) = 0$, and that the production costs of the monopolist depend on the menu proposed to the buyer. Formally, if $\phi(\Theta) \subseteq X$ is the collection of qualities that the monopolist produces, then the cost they incur is given by $\Gamma(\phi(\Theta))$. In other words, we reinterpret the cost function Γ as the map assigning the production cost corresponding to a given menu of qualities. Also, observe that the total and virtual surplus functions $S(x, \theta)$ and $J(x, \theta)$ are exactly the same as the ones of Section 5 for the particular case $\alpha = \beta = 1$.

The monopolist commits to a direct mechanism (ϕ, T) that is implementable under perfect contractibility and pays the menu cost $\Gamma(\phi(\Theta))$. This problem can be split into two parts. First, the monopolist chooses a menu of qualities $\bar{D} \subseteq X$ to produce and that can

be sold to the buyer. Second, the monopolist proposes a direct implementable mechanism (ϕ, T) that is supported over $\overline{D}$, that is $\phi(\Theta) \subseteq \overline{D}$ and the effective domain of T is also contained in $\overline{D}$.

Given the menu $\overline{D} \subseteq X$, the problem of the monopolist is exactly the same as the control problem in equation 12. Thus, the overall problem of the monopolist is the same as the one in equation 20 and we can apply all of our results to this alternative production-cost interpretation. For example, if the menu-cost function Γ is the integral one of Example 1, then the optimal menu $\phi^*(\Theta)$ is uniform over X and we can solve for the optimal number of items as in Proposition 6. Similarly, if the menu cost only depends on the number of different qualities in the menu, that is it has the form of Example 2, then we can apply Theorem 3 to conclude that any optimal menu must be coarse.

In general, however, not all the reasonable cost functions representing menu cost satisfy the strong monotonicity properties introduced in Section 4. For example, one can consider the cost function of Sartori (2021) where 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 $c : X \rightarrow \overline{\mathbb{R}}$ and define

$$\Gamma_c(\overline{D}) = \max_{x \in \overline{D}} c(x) \quad (49)$$

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 any of the strong monotonicity properties of Section 4. In fact, the analysis in Sartori (2021) shows that, in general, the optimal menu offered by the monopolist is not fully coarse and involves a continuum of differentiated qualities.

6.2 Optimal 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 Albano and Lizzeri (2001)), that potentially cares about the final consumer's utility (like in Zapechelnyuk (2020)), and for which testing is costly. This latter feature is the main element of novelty of our analysis with respect to the previous literature and also the driver for our main result: under the conditions on the testing costs introduced in Section 2.2, 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. We assume that consumers are heterogeneous in their outside option b that is distributed according to $G(b) = b^t$ for some $t > 0$. With this, the revenue of the producer and the consumer's surplus given estimate $\hat{x}$ are respectively

$$r(\hat{x}) = \max_{p \geq 0} p(\hat{x} - p)^t = R_t \hat{x}^{1+t} \quad (50)$$

$$s(\hat{x}) = \frac{t}{1+t} r(\hat{x}) = \frac{t}{1+t} R_t \hat{x}^{1+t} \quad (51)$$

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

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 $\zeta : X \rightarrow \mathbb{R}$ 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 in equilibrium given ζ is $\underline{D}_\zeta = \hat{x}_\zeta(\zeta(X)) \subseteq X$, which by construction is a closed set always containing 0.¹¹

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

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

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'

¹¹It will be momentarily clear that this set corresponds to the set $\underline{D}$ in our general analysis, hence justifying our choice of notation.

surplus, with relative weight β . Therefore, the certifier chooses a pair of rating and pricing rules (ζ, T) as well as a recommendation rule $z : \Theta \rightarrow \zeta(X)$ to maximize

$$\int_{\Theta} (1 - \beta)T(z(\theta)) + \beta s(\hat{x}_{\zeta})dF(\theta) - \Gamma(\zeta) \quad (52)$$

under the constraint that

$$z(\theta) \in \arg \max_{z \in \zeta(X)} \{r(\hat{x}_{\zeta}(z)) - (1 - \theta)\hat{x}_{\zeta}(z(\theta))^2 - T(z)\} \quad (53)$$

Next, define

$$J(x, \theta) = \left(1 - \beta + \beta \frac{t}{t+1}\right) R_t x^{t+1} - (1 - \beta)(2 - \theta)x^2. \quad (54)$$

The certifier's problem can be simplified as follows.

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

$$\sup_{\underline{D}, \phi: \Theta \rightarrow X} \int_{\Theta} J(\phi(\theta), \theta) dF(\theta) - \Gamma(\underline{D}) \quad (55)$$

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

This allows us to invoke Theorem 2 to establish that all the optimal $\underline{D}^*$ in the previous program are finite, thereby inducing an optimal rating rule with finitely many grades.

7 Conclusion

In this paper, we introduced a model of when and why incomplete contracts arise in an environment with costly contractibility. First, we introduced a model of contracting with fixed restrictions on what actions are contractible and we characterized implementable and optimal mechanisms. Next, we studied the problem of a principal that chooses the extent of contractibility before the contracting stage 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 had at most finite support. Thus, contracts in this case are optimally incomplete and, moreover, coarse. Moreover, using the structure of our proof, we derive a bound on the number of items in the optimal menu that depended on model primitives. We applied this model to study when

and why incomplete contracts would arise in a monopoly pricing problem à la [Mussa and Rosen \(1978\)](#) but with costly contractibility. We showed that optimal finite contracts took the form of uniformly-spaced quality tiers and derived a formula for the number of tiers that mimicked our general bound. We finally showed in the example how adverse selection—equivalently, the presence of information rents that distort the allocation—induced coarser and more incomplete contracts by dulling incentives to precisely separate agents.

A Proofs

A.1 Proof of Lemma 1

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

We now show part (i) of 2, that $C(y) = [\underline{\delta}(y), \bar{\delta}(y)]$. Assume by contradiction there exists some $y \in X$ and $x \in [\underline{\delta}(y), \bar{\delta}(y)]$ 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, $\bar{\delta}(y) \in C(y)$ and $x < \bar{\delta}(y)$. By Axiom 3, we have that $\min\{x, \bar{\delta}(y)\} = x \in C(y)$. This is a contradiction.

We next show parts (iii) and (iv) of 2. Fix $x, y \in X$ and assume that $x \in [\underline{\delta}(y), \bar{\delta}(y)]$, 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 $\bar{\delta}(x) \geq \bar{\delta}(y)$. Imagine this held at strict inequality. Then there would exist $z \in (\bar{\delta}(y), \bar{\delta}(x))$ such that $z \in C(y)$ and $z \notin C(x)$, while $y \in C(x)$. This violates Axiom 2.

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

Lemma 12 (Fixed Point Lemma). *Consider two functions $\underline{\delta}(x)$ and $\bar{\delta}(x)$ as in point (2) of Lemma 1. Then for all $\underline{z} \in \underline{\delta}(X)$ and $\bar{z} \in \bar{\delta}(X)$, it holds $\underline{\delta}(\underline{z}) = \underline{z}$ and $\bar{\delta}(\bar{z}) = \bar{z}$.*

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 $\bar{z} \in \bar{\delta}(X)$ is symmetric, using property (iv) in part (2) of Lemma 1. $\square$

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

$$\max_{z \leq x: z \in \underline{D}} z = \max_{z \leq x: z \in \underline{\delta}(X)} z \geq \underline{\delta}(x) \quad (56)$$

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 12 we have that $x = \underline{\delta}(x) < \underline{z}$, yielding a contradiction. If instead $\underline{z} < x$, then by Lemma 12 and the property (iii) of $\underline{\delta}$, we have $\underline{z} = \underline{\delta}(\underline{z}) = \underline{\delta}(x)$, yielding a contradiction. With this, we conclude that $\underline{z} = \underline{\delta}(x)$. With symmetric steps, we can show that $\min_{z \geq x: z \in \bar{D}} z = \bar{\delta}(x)$. Next, observe that necessarily we have $\underline{\delta}(0) = 0$ and $\bar{\delta}(\bar{x}) = \bar{x}$ proving that $0 \in \underline{D}$ and $\bar{x} \in \bar{D}$. Finally, we need to show that $\underline{D}$ and $\bar{D}$ are closed. Take a sequence $z_n \in \underline{D}$ such that $z_n \rightarrow z$. 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 12 implies that $\underline{\delta}(z_n) = z_n$ for all n . Given that $\underline{\delta}$ is upper semicontinuous, it follows that

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

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 $\bar{D}$ is closed.

(3) $\implies$ (2). Let $\underline{D}$ and $\bar{D}$ 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 $\bar{\delta}(x) = \min_{z \geq x: z \in \bar{D}} z$ and observe that $C(x) = [\underline{\delta}(x), \bar{\delta}(x)]$. It is immediate to see that both these functions are monotone increasing, such that $\underline{\delta}(x) \leq x \leq \bar{\delta}(x)$, 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 $x \rightrightarrows \{z \in \bar{D} : z \geq x\}$ 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 $y \in (x, \bar{\delta}(x)]$ and let $z = \bar{\delta}(x)$. We have $\bar{\delta}(y) \leq z$ by monotonicity. Moreover, by assumption $z \geq y$ and $z \in \bar{D}$, so that $z \geq \bar{\delta}(y)$ by definition. We then must have $z = \bar{\delta}(y)$.

(2) $\implies$ (1). Fix $\underline{\delta}$ and $\bar{\delta}$ that satisfy (2). $C(y) = [\underline{\delta}(y), \bar{\delta}(y)]$ is regular. C is reflexive since because of (ii), closed because the intervals of the construction are closed, and monotone because $\underline{\delta}, \bar{\delta}$ 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, $\bar{\delta}(x) \leq \bar{\delta}(y)$. Therefore,

$C(x) \subseteq C(y)$. Next, consider the case where $x > y$. From (iv), we have $\bar{\delta}(x) = \bar{\delta}(y)$. Moreover, from monotonicity, $\underline{\delta}(x) \geq \underline{\delta}(y)$. Therefore, $C(x) \subseteq C(y)$. Finally, if $x = y$, clearly $C(x) \subseteq C(y)$. Given that these arguments hold for any x , this shows transitivity. These arguments together establish that C is regular.

A.2 Proof of Proposition 1

We begin by establishing a general taxation principle with partial contractibility.

Intermediate Result: A Monotone Taxation Principle. Given a regular contracting correspondence C , we say that $T : X \rightarrow \bar{\mathbb{R}}$ 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 1).

Lemma 13 (*C*-Monotone Taxation Principle). *Fix a regular contracting correspondence C . A final outcome function ϕ is implementable given C if and only if there exists a tariff $T : X \rightarrow \bar{\mathbb{R}}$ that is monotone with respect to C and such that:*

$$\phi(\theta) \in \arg \max_{x \in X} \{u(x, \theta) - T(x)\} \quad (57)$$

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 $\hat{T} : X \rightarrow \bar{\mathbb{R}}$ as:

$$\hat{T}(x) = \inf_{y \in X} \{T(y) : x \in C(y)\} \quad (58)$$

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

$$u(\phi(\theta), \theta) \geq u(x, \theta) \quad (59)$$

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

$$u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \geq u(\phi(\theta), \theta) - T(\xi(\theta)) \geq 0 \quad (60)$$

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

$$\max_{x \in C(y)} u(x, \theta) - \hat{T}(y) > u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \quad (61)$$

Thus:

$$\begin{aligned} \max_{x \in C(y)} u(x, \theta) - T(y) &> u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \\ &\geq u(\phi(\theta), \theta) - T(\xi(\theta)) = \max_{x \in C(\xi(\theta))} u(x, \theta) - T(\xi(\theta)) \end{aligned} \quad (62)$$

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 57 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

$$\{\hat{x} \in X : x \in C(\hat{x})\} \subseteq \{\hat{x} \in X : y \in C(\hat{x})\} \quad (63)$$

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

(If) We now establish sufficiency. Suppose that there exists a tariff $T : X \rightarrow \bar{\mathbb{R}}$ that is monotone with respect to C and such that Equation 57 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

$$u(x, \theta) - T(y) > \max_{\hat{x} \in C(\phi(\theta))} u(\hat{x}, \theta) - T(\phi(\theta)) \geq u(\phi(\theta), \theta) - T(\phi(\theta)) \quad (64)$$

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

$$u(x, \theta) - T(y) > u(\phi(\theta), \theta) - T(\phi(\theta)) \geq u(x, \theta) - T(x) \quad (65)$$

where the second inequality uses the fact that $\phi(\theta)$ solves the program in Equation 57. 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:

$$u(x, \theta) > u(\phi(\theta), \theta) \quad (66)$$

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

$$u(x, \theta) - T(x) > u(\phi(\theta), \theta) - T(\phi(\theta)) \quad (67)$$

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. $\square$

With this in hand, we now prove Proposition 1.

Proof. Only If for First Part. If ϕ is implementable, then there exists (ξ, T) that support ϕ . By Lemma 13, we may take that $\xi = \phi$. By (IC) and Lemma 13, there exists a transfer function $t : \Theta \rightarrow \mathbb{R}$ 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 $\bar{\delta}(\phi(\theta)) > \phi(\theta)$ 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 monotone decreasing preferences) holds.

If For First Part. Without loss of generality, we gain prove this part for the 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 $t : \Theta \rightarrow \mathbb{R}$ as

$$t(\theta) = K + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) ds \quad (68)$$

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

$$T(x) = \inf_{\theta' \in \Theta} \{t(\theta') : x \in C(\phi(\theta'))\} \quad (69)$$

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

$$\{\theta \in \Theta : x \in C(\phi(\theta))\} \subseteq \{\theta \in \Theta : y \in C(\phi(\theta))\} \quad (70)$$

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 13 that ϕ is implementable.

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

$$u(\phi(\theta), \theta) - T(\phi(\theta)) \geq u(\phi(\theta), \theta) - t(\theta) = \int_0^\theta u_\theta(\phi(s), s) ds - Z \quad (71)$$

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 57. We first prove that, for all $\theta, \theta' \in \Theta$:

$$u(\phi(\theta), \theta) - t(\theta) \geq \max_{x \in C(\phi(\theta'))} u(x, \theta) - t(\theta') \quad (72)$$

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.

Deriving the Tariff. We can simply set $T(x) = t(\phi^{-1}(x))$. This yields the claimed formula. $\square$

A.3 Proof of Lemma 2

We begin by eliminating the proposed allocation and transfers from the objective function of the seller. From the proof of Theorem 1, we have that transfers for any incentive compatible

triple (ξ, ϕ, t) are given by:

$$t(\theta) = Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) \, ds \quad (73)$$

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

$$\begin{aligned} \int_{\Theta} t(\theta) \, dF(\theta) &= \int_{\Theta} \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) \, ds \right) dF(\theta) \\ &= \int_{\Theta} (Z + u(\phi(\theta), \theta)) \, dF(\theta) - \int_0^1 \int_0^\theta u_\theta(\phi(s), s) \, ds \, dF(\theta) \end{aligned} \quad (74)$$

Using this expression for total transfer revenue, and the characterization of implementation from Proposition 1, we write the seller's problem as

$$\begin{aligned} \max_{\phi, Z} \quad & \int_{\Theta} \left(\pi(\phi(\theta), \theta) + Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) \, ds \right) dF(\theta) \\ \text{s.t.} \quad & \phi(\theta') \geq \phi(\theta), \quad \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \\ & u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) \, ds \right) \geq 0 \quad \forall \theta \in \Theta \end{aligned} \quad (75)$$

We further simplify this by applying integration by parts on the double integral of $u_\theta(\phi(s), s)$ over θ and s :

$$\begin{aligned} \int_0^1 \int_0^\theta u_\theta(\phi(s), s) \, ds \, dF(\theta) &= \left[F(\theta) \int_0^\theta u_\theta(\phi(s); s) \, ds \right]_0^1 - \int_0^1 F(\theta) u_\theta(\phi(\theta), \theta) \, d\theta \\ &= \int_0^1 (1 - F(\theta)) u_\theta(\phi(\theta), \theta) \, d\theta \\ &= \int_0^1 \frac{(1 - F(\theta))}{f(\theta)} u_\theta(\phi(\theta), \theta) \, dF(\theta) \end{aligned} \quad (76)$$

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

$$\begin{aligned}
& \max_{\phi, Z} \int_{\Theta} (J(\phi(\theta)) + Z) dF(\theta) \\
& \text{s.t. } \phi(\theta') \geq \phi(\theta), \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \\
& \quad u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) ds \right) \geq 0 \quad \forall \theta \in \Theta
\end{aligned} \tag{77}$$

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

$$V(\theta) = u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s) ds \right) \geq 0 \quad \forall \theta \in \Theta \tag{78}$$

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:

$$V(0) = -Z \geq 0 \tag{79}$$

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

$$\begin{aligned}
& \max_{\phi} \int_{\Theta} J(\phi(\theta), \theta) dF(\theta) \\
& \text{s.t. } \phi(\theta') \geq \phi(\theta), \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta
\end{aligned} \tag{80}$$

This completes the proof.

A.4 Proof of Theorem 1

We first solve the pointwise problem 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 $\bar{\phi}(\theta') = \bar{\phi}(\theta)$. We next observe that $\phi^P(\theta) > \underline{\phi}(\theta')$; otherwise, $\bar{\phi}(\theta) = \min\{y \in \bar{D} : y \geq \phi^P(\theta)\} \leq \underline{\phi}(\theta')$. 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: $J(\bar{\phi}(\theta), \theta) \geq J(\underline{\phi}(\theta), \theta)$ by optimality of $\bar{\phi}(\theta)$; $J(\bar{\phi}(\theta), \theta') > J(\underline{\phi}(\theta), \theta')$ by strict single crossing; $J(\bar{\phi}(\theta'), \theta') > J(\underline{\phi}(\theta'), \theta')$ because $\bar{\phi}(\theta) = \bar{\phi}(\theta')$ and $\underline{\phi}(\theta) = \underline{\phi}(\theta')$; but $J(\bar{\phi}(\theta'), \theta') \leq J(\underline{\phi}(\theta'), \theta')$ from the presumed optimality of $\underline{\phi}(\theta')$ for type θ' . This completes the argument that ϕ^* is monotone.

A.5 Proof of Proposition 2

Proof. 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 $\underline{k} = \max\{k \in \{1, \dots, K\} : J(x_k, 0) - J(x_{k-1}, 0) > 0\}$, 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 $\bar{k} = \min\{k \in \{1, \dots, K\} : J(x_k, 1) - J(x_{k-1}, 1) < 0\}$, with the convention that $\bar{k} = K$ if this set is empty. As J is strictly single crossing, $\bar{k} > \underline{k}$. We now have that $J(x_k, \theta) - J(x_{k-1}, \theta) = 0$ has a solution if and only if $k \in \{\underline{k} + 1, \dots, \bar{k} - 1\}$ (if $\bar{k} = \underline{k} + 1$, then this set is empty). For all $k \geq \bar{k}$, we have that $\hat{\theta}_k = 1$. For all $k \leq \underline{k}$, we have that $\hat{\theta}_k = 0$. For all $k \in \{\underline{k} + 1, \dots, \bar{k} - 1\}$, 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 $\bar{\phi}(\hat{\theta}_k) = x_k$. 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 10 from Proposition 1, we have that:

$$\begin{aligned}
T(x_k) &= u(x_k, \hat{\theta}_k) - \mathbb{I}[k \geq 2] \sum_{j=1}^{k-1} \int_{\hat{\theta}_j}^{\hat{\theta}_{j+1}} u_\theta(x_j, s) \, ds \\
&= u(x_k, \hat{\theta}_k) - \mathbb{I}[k \geq 2] \sum_{j=1}^{k-1} \left[u(x_j, \hat{\theta}_{j+1}) - u(x_j, \hat{\theta}_j) \right] \\
&= u(x_1, 0) + \mathbb{I}[k \geq 2] \sum_{j=2}^k \left[u(x_j, \hat{\theta}_j) - u(x_{j-1}, \hat{\theta}_j) \right]
\end{aligned} \tag{81}$$

where the second equality computes the integrals and the final equality telescopes the summation. $\square$

A.6 Proof of Proposition 3

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 $D_n \rightarrow D$. We have that D is closed and given that $\bar{x} \in D_n$ for all n , it follows that $\bar{x} \in D$, yielding that $D \in \mathcal{D}$ and that the latter is closed.

By Lemma 2 and since $J(x, \theta)$ is strictly supermodular, we have

$$\mathcal{J}(\bar{D}) = \int_{\Theta} \mathcal{J}(\bar{D}, \theta) dF(\theta) \quad (82)$$

where

$$\mathcal{J}(\bar{D}, \theta) := \max_{x \in \bar{D}} J(x, \theta) \quad (83)$$

for all $\theta \in \Theta$. By Berge's Maximum theorem, for every $\theta \in \Theta$, the map $\bar{D} \mapsto \mathcal{J}(\bar{D}, \theta)$ is continuous in the Hausdorff topology. Given that Θ is compact and $\mathcal{J}(\bar{D}, \theta)$ is bounded it follows that also the map $\bar{D} \mapsto \mathcal{J}(\bar{D})$ is continuous in the Hausdorff topology. With this, the result follows by Weierstrass Theorem applied to (20).

A.7 Proof of Lemma 4

Let ϕ^* denote the optimal allocation under $\bar{D}$ and $\phi^{*'} denote the optimal allocation under $\bar{D}' = \bar{D} \setminus (a, b)$, as defined in Theorem 1. By Lemma 2, the difference in values under these contractibility correspondences is$

$$\mathcal{J}(\bar{D}) - \mathcal{J}(\bar{D}') = \int_0^1 (J(\phi^*(\theta), \theta) - J(\phi^{*'}(\theta), \theta)) dF(\theta) \quad (84)$$

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 $\mathcal{J}(\bar{D}) - \mathcal{J}(\bar{D}') = 0$ 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:

$$\phi^{*'}(\theta) = \begin{cases} a & \text{if } \theta \in [\inf \Theta(a, b), \hat{\theta}(a, b)], \\ b & \text{if } \theta \in (\hat{\theta}(a, b), \sup \Theta(a, b)], \\ \phi^*(\theta) & \text{otherwise.} \end{cases} \quad (85)$$

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:

$$\inf \Theta(a, b) = \begin{cases} \min_{x \in \bar{D}: x > a} \hat{\theta}(a, x), & \text{if it exists,} \\ (\phi^P)^{-1}(a), & \text{otherwise.} \end{cases} \quad (86)$$

$$\sup \Theta(a, b) = \begin{cases} \max_{x \in \bar{D}: x < b} \hat{\theta}(b, x), & \text{if it exists,} \\ (\phi^P)^{-1}(b), & \text{otherwise.} \end{cases} \quad (87)$$

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

$$\begin{aligned} \mathcal{J}(\bar{D}) - \mathcal{J}(\bar{D}') &= \int_{\inf \Theta(a, b)}^{\hat{\theta}(a, b)} (J(\phi^*(\theta), \theta) - J(a, \theta)) dF(\theta) \\ &\quad + \int_{\hat{\theta}(a, b)}^{\sup \Theta(a, b)} (J(\phi^*(\theta), \theta) - J(b, \theta)) dF(\theta) \end{aligned} \quad (88)$$

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 $x = \phi^{*'}(\theta)$,

$$J(\phi^{*'}(\theta), \theta) = J(\phi^*(\theta), \theta) + J_x(\phi^*(\theta), \theta)(\phi^{*'}(\theta) - \phi^*(\theta)) + \frac{1}{2} J_{xx}(y(\theta), \theta)(\phi^{*'}(\theta) - \phi^*(\theta))^2 \quad (89)$$

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 $x = \phi^*(\theta)$:

$$\begin{aligned} J_x(\phi^*(\theta), \theta) &= J_x(\phi^P(\theta), \theta) + J_{xx}(z(\theta), \theta)(\phi^*(\theta) - \phi^P(\theta)) \\ &= J_{xx}(z(\theta), \theta)(\phi^*(\theta) - \phi^P(\theta)) \end{aligned} \quad (90)$$

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:

$$\begin{aligned}
|J(\phi^{*'}(\theta), \theta) - J(\phi^*(\theta), \theta)| &\leq |J_x(\phi^*(\theta), \theta)| |\phi^{*'}(\theta) - \phi^*(\theta)| + \frac{1}{2} |J_{xx}(y(\theta), \theta)| (\phi^{*'}(\theta) - \phi^*(\theta))^2 \\
&\leq |J_{xx}(z(\theta), \theta)| (\phi^{*'}(\theta) - \phi^*(\theta))^2 + \frac{1}{2} |J_{xx}(y(\theta), \theta)| (\phi^{*'}(\theta) - \phi^*(\theta))^2 \\
&\leq \frac{3}{2} \bar{J}_{xx} (\phi^{*'}(\theta) - \phi^*(\theta))^2
\end{aligned} \tag{91}$$

Thus, defining $c = \phi^*(\hat{\theta}(a, b))$, the integrand in the first line of Equation 88 is bounded above by $\frac{3}{2} \bar{J}_{xx} (c - a)^2$ and the integrand in the second line of Equation 88 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 22 and Equation 23. The second approach yields Equation 24.

In the first approach, we observe that $\hat{\theta}(a, b) - \inf \Theta(a, b), \sup \Theta(a, b) - \hat{\theta}(a, b) \leq \sup \Theta(a, b) - \inf \Theta(a, b) \leq (\phi^P)^{-1}(b) - (\phi^P)^{-1}(a)$. 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

$$((\phi^P)^{-1})'(x) = \frac{J_{xx}(x, (\phi^P)^{-1}(x))}{J_{x\theta}(x, (\phi^P)^{-1}(x))} \leq \frac{\sup_{y \in X, \theta \in \Theta} J_{xx}(y, \theta)}{\inf_{y \in X, \theta \in \Theta} J_{x\theta}(y, \theta)} = \frac{\bar{J}_{xx}}{\underline{J}_{x\theta}} < \infty \tag{92}$$

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:

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

In the second approach, we suppose that $a < c < b$ are three sequential points in $\bar{D}$, i.e., c is isolated, and a and b are the closest elements to c in $\bar{D}$. 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

$$\hat{\theta}'(u) = \frac{-J_x(c + u, \hat{\theta}(u))}{J_\theta(c + u, \hat{\theta}(u)) - J_\theta(a, \hat{\theta}(u))} \quad (94)$$

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

$$\hat{\theta}(b - c) = \hat{\theta}(0) + \hat{\theta}'(\tilde{u})(b - c) \quad (95)$$

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

$$\hat{\theta}(a, b) - \hat{\theta}(a, c) = \hat{\theta}(b - c) - \hat{\theta}(0) = \frac{-J_x(c + \tilde{u}, \hat{\theta}(\tilde{u}))}{J_\theta(c + \tilde{u}, \hat{\theta}(\tilde{u})) - J_\theta(a, \hat{\theta}(\tilde{u}))}(b - c) \quad (96)$$

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

$$\begin{aligned} J_x(c + \tilde{u}, \hat{\theta}(\tilde{u})) &= J_x(\phi^P(\hat{\theta}(\tilde{u})), \hat{\theta}(\tilde{u})) + J_{xx}(z, \hat{\theta}(\tilde{u}))(c + \tilde{u} - \phi^P(\hat{\theta}(\tilde{u}))) \\ &= J_{xx}(z, \hat{\theta}(\tilde{u}))(c + \tilde{u} - \phi^P(\hat{\theta}(\tilde{u}))) \end{aligned} \quad (97)$$

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

$$J_\theta(c + \tilde{u}, \hat{\theta}(\tilde{u})) - J_\theta(a, \hat{\theta}(\tilde{u})) = J_{x\theta}(z, \hat{\theta}(\tilde{u}))(c + \tilde{u} - a) \quad (98)$$

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:

$$\hat{\theta}(a, b) - \hat{\theta}(a, c) \leq \frac{\bar{J}_{xx}(b - a)}{\underline{J}_{x\theta}(c - a)}(b - c) \quad (99)$$

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

$$\hat{\theta}(a, b) - \hat{\theta}(c, b) \leq \frac{\bar{J}_{xx}(b - a)}{\underline{J}_{x\theta}(b - c)}(c - a) \quad (100)$$

Step 3: Bounding the Value. Combining steps 1 and 2. We can now derive the payoff bound of Equation 23:

$$\begin{aligned}
\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D}') &= \int_{\inf \Theta(a,b)}^{\hat{\theta}(a,b)} (J(\phi^*(\theta), \theta) - J(a, \theta)) dF(\theta) \\
&\quad + \int_{\hat{\theta}(a,b)}^{\sup \Theta(a,b)} (J(\phi^*(\theta), \theta) - J(b, \theta)) dF(\theta) \\
&\leq \int_{\inf \Theta(a,b)}^{\hat{\theta}(a,b)} \frac{3}{2} \bar{J}_{xx} (c - a)^2 dF(\theta) + \int_{\hat{\theta}(a,b)}^{\sup \Theta(a,b)} \frac{3}{2} \bar{J}_{xx} (b - c)^2 dF(\theta) \\
&\leq \frac{3}{2} \bar{J}_{xx} [(c - a)^2 + (b - c)^2] \int_{\inf \Theta(a,b)}^{\sup \Theta(a,b)} dF(\theta) \\
&\leq \frac{3}{2} \bar{J}_{xx} [(c - a)^2 + (b - c)^2] \frac{\bar{J}_{xx}}{\underline{J}_{x\theta}} (b - a) \bar{f} \\
&= \frac{3}{2} \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} (b - a) [(c - a)^2 + (b - c)^2]
\end{aligned} \tag{101}$$

Observing that $(c - a)^2 + (b - c)^2 \leq (b - a)^2$, we also obtain Equation 22.

Finally, we obtain Equation 24 by combining step 1 with the second approach to step 2. Doing this, we obtain:

$$\begin{aligned}
\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D}') &= \int_{\hat{\theta}(a,c)}^{\hat{\theta}(a,b)} (J(\phi^*(\theta), \theta) - J(a, \theta)) dF(\theta) \\
&\quad + \int_{\hat{\theta}(a,b)}^{\sup \Theta(c,b)} (J(\phi^*(\theta), \theta) - J(b, \theta)) dF(\theta) \\
&\leq \int_{\hat{\theta}(a,c)}^{\hat{\theta}(a,b)} \frac{3}{2} \bar{J}_{xx} (c - a)^2 dF(\theta) + \int_{\hat{\theta}(a,b)}^{\hat{\theta}(c,b)} \frac{3}{2} \bar{J}_{xx} (b - c)^2 dF(\theta) \\
&\leq 3 \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} (b - a) (c - a) (b - c)
\end{aligned} \tag{102}$$

Completing the proof.

A.8 Proof of Lemma 6

As is $\overline{B}_t(x) \cap \overline{D}$ 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 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. 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 4 (Equation 23) that there exists a sequence of points $z_m \in (x^*, b_m) \cap \overline{D}$ such that:

$$\begin{aligned} \mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus (y, b_m)) &= \mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus [x^*, b_m)) \\ &\leq \frac{3}{2} \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} (b_m - x^*) [(b_m - z_m)^2 + (z_m - x^*)^2] \leq \frac{3}{2} \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} (b_m - y) [(b_m - x^*)^2] \end{aligned} \quad (103)$$

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:

$$\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus (y, b_m)) \geq \epsilon(x^* - y)(b_m - x^*) \quad (104)$$

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

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

As $b_m \rightarrow x^*$, 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 $m \geq \max\{M, \hat{M}\}$:

$$\mathcal{J}(\overline{D}) - \Gamma(\overline{D}) < \mathcal{J}(\overline{D} \setminus (y, b_m)) - \Gamma(\overline{D} \setminus (y, b_m)) \quad (106)$$

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

A.9 Proof of Lemma 7

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 neighborhood of x contains an isolated point. Let $\{x_m\} \subset \overline{D}$ be a monotone sequence of isolated points such that $x_m \rightarrow x$. 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 4 (Equation 24), we have that:

$$\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus \{x_m\}) \leq 3 \frac{\bar{J}_{xx}^2 \bar{f}}{\underline{J}_{x\theta}} (b_m - a_m)(x_m - a_m)(b_m - x_m) \quad (107)$$

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 $x_m \rightarrow x$. Thus, by strong monotonicity, we have that there exists

M such that for all $m \geq M$, we have that:

$$\Gamma(\overline{D}) - \Gamma(\overline{D} \setminus \{x_m\}) \geq \epsilon(x_m - a_m)(b_m - x_m) \quad (108)$$

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

$$\begin{aligned} & \Gamma(\overline{D}) - \Gamma(\overline{D} \setminus \{x_m\}) - (\mathcal{J}(\overline{D}) - \mathcal{J}(\overline{D} \setminus \{x_m\})) \\ & \geq (x_m - a_m)(b_m - x_m) \left[\epsilon - 3 \frac{\bar{J}_{xx}^2 \bar{f}}{\bar{J}_{x\theta}} (b_m - a_m) \right] \end{aligned} \quad (109)$$

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

$$\mathcal{J}(\overline{D}) - \Gamma(\overline{D}) < \mathcal{J}(\overline{D} \setminus \{x_m\}) - \Gamma(\overline{D} \setminus \{x_m\}) \quad (110)$$

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

A.10 Proof of Proposition 4

We first introduce some preliminary notation. Given a vector $(x_1, \dots, x_{K-1}) \in \mathbb{R}^{K-1}$, we let $(x_k + \varepsilon, x_{-k}) \in \mathbb{R}^{K-1}$ the vector where we replace x_k with $x_k + \varepsilon$ for some $k \in \{2, \dots, K-1\}$. 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 $\frac{d}{d\varepsilon} \tilde{\Pi}(x_k + \varepsilon, x_{-k})|_{\varepsilon=0} = \frac{d}{d\varepsilon} \Gamma(x_k + \varepsilon, x_{-k})|_{\varepsilon=0}$ for any $k \in \{2, \dots, K-1\}$. The left-hand-side is

$$\begin{aligned} \frac{d}{d\varepsilon} \tilde{\Pi}(x_k + \varepsilon, x_{-k})|_{\varepsilon=0} &= \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k, \theta) dF(\theta) + \\ & \quad \frac{\partial}{\partial x_k} \hat{\theta}_k \left(J(x_k, \hat{\theta}_k) - J(x_{k-1}, \hat{\theta}_k) \right) f(\hat{\theta}_k) + \\ & \quad \frac{\partial}{\partial x_k} \hat{\theta}_{k+1} \left(J(x_{k+1}, \hat{\theta}_{k+1}) - J(x_k, \hat{\theta}_{k+1}) \right) f(\hat{\theta}_{k+1}) \\ &= \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k, \theta) dF(\theta) \end{aligned} \quad (111)$$

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 $\frac{d}{d\varepsilon} \Gamma(x_k + \varepsilon, x_{-k})|_{\varepsilon=0} = \Gamma_k(\overline{D})$, we obtain Equation 37. For $k = K$, by definition, we have that $x_K = 1$. For $k = 1$, we have that x_1 either solves the first-order condition of (37) or is equal to 0.

A.11 Proof of Theorem 3

We start with a preliminary lemma.

Lemma 14. *If Γ has a clause-based representation $\hat{\Gamma}$ then it is lower semicontinuous in the Hausdorff topology of closed sets.*

Proof. Given that $\hat{\Gamma} : \mathbb{N} \rightarrow \overline{\mathbb{R}}$ is strictly increasing we have as $K \rightarrow \infty$ either $\hat{\Gamma}(K)$ asymptots 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 $D_n \rightarrow D$ 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 conference we must have that $n(D_n) \rightarrow \infty$. 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 conference we must have that $n(D_n) \rightarrow n(D)$. 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 n points, evenly spaced with width $\bar{x}/n$. Recall that ϕ^P denotes the assignment under perfect contractibility, let ϕ_n^* denote the assignment under the grid, and let $G_n = \{\bar{x}i/n\}_{i=1}^n \in \mathcal{D}$ denote the grid. To derive the benefits of this contractibility correspondence, we apply a close variant of Lemma 4. Using the bound derived in the proof of that result for $|J(\phi^P(\theta), \theta) - J(x, \theta)|$ for any x , we derive

$$\begin{aligned} \mathcal{J}(X) - \mathcal{J}(G_n) &= \int_0^1 (J(\phi^P(\theta), \theta) - J(\phi_n^*(\theta), \theta)) dF(\theta) \\ &\leq \int_0^1 \frac{1}{2n^2} \bar{J}_{xx} dF(\theta) = \frac{1}{2n^2} \bar{J}_{xx} \end{aligned} \tag{112}$$

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

$$\Gamma(X) - \Gamma(G_n) \geq \sum_{j=n}^{\infty} j^{-\beta} \epsilon \quad (113)$$

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

$$\Gamma(X) - \Gamma(G_n) \geq \epsilon \sum_{j=n}^{\infty} j^{-\beta} \geq \epsilon \int_n^{\infty} s^{-\beta} ds = \epsilon \left[-\frac{1}{\beta} s^{-\beta+1} \right]_n^{\infty} = \frac{\epsilon}{\beta} n^{-\beta+1} \quad (114)$$

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

$$\mathcal{J}(G_n) - \Gamma(G_n) \geq \mathcal{J}(X) - \Gamma(X) + \left(\frac{\epsilon}{\beta} n^{-\beta+1} - \frac{1}{2} \bar{J}_{xx} n^{-2} \right) \quad (115)$$

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

$$n > \left(\frac{\beta}{2\epsilon} \bar{J}_{xx} \right)^{\frac{1}{3-\beta}} \rightarrow \mathcal{J}(G_n) - (\Gamma(G_n) - \mathcal{J}(X) - \Gamma(X)) \geq 0 \quad (116)$$

Thus, an optimal contracting support cannot be full contractibility. Finally, by Lemma 14 we can invoke Proposition 3 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 $x_{m+1} - x_{m-1} < 2\bar{x}/(K-2)$. Otherwise, there would be a contradiction:

$$\begin{aligned} x_K - x_1 &= \sum_{j=1}^{\lfloor K/2 \rfloor} x_{2j+1} - x_{2j-1} \geq \lfloor K/2 \rfloor \frac{2\bar{x}}{K-2} \\ &> \left(\frac{K}{2} - 1 \right) \frac{2\bar{x}}{\frac{K}{2} - 1} > \bar{x} \end{aligned} \quad (117)$$

We first apply the third statement of Lemma 4 to bound the loss from eliminating con-

tractibility at some point x_m :

$$\begin{aligned}\mathcal{J}(\overline{D}^*) - \mathcal{J}(\overline{D}^* \setminus \{x_m\}) &\leq 3 \frac{\bar{J}_{xx}^2 \bar{f}}{J_{x\theta}} (x_m - x_{m-1})(x_{m+1} - x_m)(x_{m+1} - x_{m-1}) \\ &\leq \frac{3}{4} \frac{\bar{J}_{xx}^2 \bar{f}}{J_{x\theta}} (x_{m+1} - x_{m-1})^3\end{aligned}\tag{118}$$

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

$$\hat{\Gamma}(K) - \hat{\Gamma}(K-1) \geq \epsilon(K-1)^{-\beta} > \epsilon(K-2)^{-\beta}\tag{119}$$

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

$$\epsilon(K-2)^{-\beta} > \frac{3}{4} \frac{\bar{J}_{xx}^2 \bar{f}}{J_{x\theta}} (x_{m+1} - x_{m-1})^3\tag{120}$$

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

$$K > 2 + \left(\frac{6\bar{J}_{xx}^2 \bar{f}}{\epsilon J_{x\theta}} \right)^{\frac{1}{3-\beta}}\tag{121}$$

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

A.12 Proof of Proposition 5

The second part of the result (the optimal tariff) follows directly from applying the tariff formula in Proposition 1. The third part (the program defining K^*) is trivial. Therefore, in the remainder of the proof, we show the optimality of the uniform grid. We do this by showing that this is the only solution that is compatible with the necessary first-order conditions derived in Proposition 4.

Applying Proposition 4, the first-order condition for $k \in \{2, K-1\}$ is

$$\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} (\alpha(2\theta - 1) - \beta x_k) d\theta - \gamma(-2x_k + x_{k-1} + x_{k+1}) = 0\tag{122}$$

This reduces to:

$$\begin{aligned}\gamma(-2x_k + x_{k-1} + x_{k+1}) &= (\hat{\theta}_{k+1} - \hat{\theta}_k) \left[\alpha(\hat{\theta}_{k+1} + \hat{\theta}_k - 1) - \beta x_k \right] \\ &= \frac{\beta^2}{16\alpha} (x_{k+1} - x_{k-1})(x_{k+1} + x_{k-1} - 2x_k)\end{aligned}\quad (123)$$

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:

$$(x_{k+1} + x_{k-1} - 2x_k) \left[\frac{\beta^2}{16\alpha} (x_{k+1} - x_{k-1}) - \gamma \right] = 0 \quad (124)$$

This equation has two solutions,

$$x_k = \frac{x_{k+1} + x_{k-1}}{2}, \quad x_{k+1} = x_{k-1} + \Delta \quad (125)$$

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

Case 1: Uniform Grid. For $k = K$, from the boundary condition, we have that $x_K = 1$. For $k = 1$, x_1 is equal to the solution to the corresponding first-order condition (when that solution is positive):

$$x_1 = \max \left\{ 0, \frac{1}{3}x_2 - \frac{2}{3}\frac{\alpha}{\beta} \right\} \quad (126)$$

Given any value for $x_1 \geq 0$, we therefore have that:

$$x_k = \frac{1 - x_1}{K - 1}k + \frac{Kx_1 - 1}{K - 1} \quad (127)$$

for all $k \geq 2$. Thus, when $x_1 = 0$, we have that:

$$x_k = \frac{k - 1}{K - 1} \quad (128)$$

We now show that x_1 cannot be strictly positive. Imagine it were so. The first-order condition implies that

$$x_1 = \frac{1 - 2\frac{\alpha}{\beta}(K - 1)}{2K - 1} \quad (129)$$

but this is negative under our maintained assumption that $\alpha/\beta \geq 1$, for any $K \geq 2$. Therefore, $x_1 = 0$.

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 $\Delta \equiv \frac{16\alpha\gamma}{\beta^2}$. To pin down both grids, as $x_K = 1$, it remains to consider x_1 . The first-order condition, which holds at equality if

$x_1 \geq 0$ and inequality if $x_1 = 0$, is:

$$\begin{aligned}
0 &\geq \hat{\theta}_2 \left[\alpha(\hat{\theta}_2 - 1) - \beta x_1 \right] - \gamma(-2x_1 + x_2) \\
&= \left(\frac{\beta}{4\alpha}(x_2 + x_1) + \frac{1}{2} \right) \left[\alpha \left(\frac{\beta}{4\alpha}(x_2 + x_1) - \frac{1}{2} \right) - \beta x_1 \right] - \gamma(-2x_1 + x_2) \\
&=: g(x_1, x_2)
\end{aligned} \tag{130}$$

where, to obtain the second equality, we substitute in the value of $\hat{\theta}_2$; and the third equality defines the function g .

We now show that $g(x, x_2) < 0$ for any $x < x_2$, thus establishing that $x_1 = 0$ is the only solution compatible with optimality. We first show that $g(0, x_2) < 0$ for any x_2 . To do this, we first calculate

$$g(0, x_2) = \alpha \left(\frac{\beta^2}{16\alpha^2} x_2^2 - \frac{1}{4} \right) - \gamma x_2 \tag{131}$$

From this, it is easy to see that the function $x_2 \mapsto g(0, x_2)$ is maximized at either $x_2 = 0$ or $x_2 = 1$. We have $g(0, 0) = -\alpha/4 < 0$, and $g(0, 1) = \alpha((\beta/\alpha)^2(1/16) - 1/4) - \gamma < 0$ under the maintained assumption that $\alpha/\beta \geq 1$ or $\beta/\alpha \leq 1$. We next observe that

$$g_{x_1}(x_1, x_2) = \underbrace{-\frac{\beta^2}{8\alpha}(x_1 + x_2)}_{\text{Term 1}} + \underbrace{2\gamma - \frac{\beta}{2} - \frac{\beta^2}{4}}_{\text{Term 2}} \tag{132}$$

Clearly, Term 1 is negative. Moreover, Term 2 is negative under the maintained assumptions that $\Delta < 1$ and $\beta/\alpha \leq 1$, which together imply that $\gamma < \frac{\beta}{16}$. Combining the observations that $g(0, x_2) < 0$ and $g_{x_1}(x_1, x_2) < 0$, for all x_1 and x_2 , we have as desired that $g(x, x_2) < 0$ for any $x < x_2$, thus establishing that $x_1 = 0$ is the only solution compatible with optimality.

Thus, the conjectured solution must have $x_1 = 0$; $x_k = \frac{k-1}{2}\Delta$ for k odd, and $x_k = 1 - \frac{K-k}{2}\Delta$ for k even. This is only possible if $K = 2\lfloor \frac{1}{\Delta} \rfloor$.

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 bordered Hessian, when projected into the null space of the Jacobian matrix of active constraints, is negative semidefinite. The bordered Hessian is the second derivative matrix of the Lagrangian,

$$\mathcal{L}(x_1, \dots, x_{K-1}) = F(x_1, \dots, x_{K-1}) - \lambda x_1 \tag{133}$$

where λ is the Lagrange multiplier on the constraint $x_1 \geq 0$ and F is the objective as a function of the grid points $x_1, \dots, x_{K-1}$ (substituting in $x_K = 1$). Since the active constraint is $x_1 = 0$, the null space of the Jacobian of active constraints is spanned by $[0, 0; 0, I_{K-2}]$,

i.e., the identity matrix for elements $x_2 \dots x_{K-1}$. Since the constraint is linear, the Hessian restricted to this space equals the second derivative matrix of F in $(x_2, \dots, x_{K-1})$, which we call H^F . Hence, to show that the alternating grid is *not* a local maximum, it will suffice to find a contradiction to this necessary condition. Specifically, we will show the existence of a vector $x \in \mathbb{R}^{K-2}$ such that $v \neq 0$ and $v'H^F v > 0$, which implies that H^F is not negative semidefinite. Equivalently, we could have argued that the existence of this vector contradicts the optimality of the alternating grid because it implies a direction in which we could perturb the grid, while satisfying the constraint (since x_1 is unchanged), such that payoffs increase.¹²

To do this, we first evaluate the second-order conditions at the conjectured alternating grid solution. These simplify to

$$\begin{aligned} \frac{\partial^2 F}{\partial x_k^2} &= H_{k-1,k-1}^F = -\frac{\beta^2}{8\alpha}\Delta + 2\gamma = 0 \\ \frac{\partial^2 F}{\partial x_k \partial x_{k+1}} &= H_{k,k-1}^F = H_{k-1,k}^F = \frac{\beta^2}{8\alpha}(x_{k+1} - x_k) - \gamma \end{aligned} \quad (134)$$

where we take care to observe that row and column $k-1$ of H^F corresponds to the variable x_k .

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

$$v_k' H^F v_k = 2 \left(\gamma - \frac{\beta^2}{8\alpha}(x_{k+1} - x_k) \right) \quad (135)$$

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,

$$v_k' H^F v_k > 2 \left(\gamma - \frac{\Delta\beta^2}{16\alpha} \right) > 0 \quad (136)$$

and, as desired, we have shown that the bordered 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

$$u' H^F u = 2 \left(\frac{\beta^2}{8\alpha}(x_3 - x_2) - \gamma \right) \quad (137)$$

¹²To see this, observe that the second-order Taylor expansion of F around the alternating-grid point has a linear term proportional to the gradient of F , which is zero; a quadratic term, which equals a quadratic form in H^F ; and a third-order remainder. Hence, if we find a direction in which the quadratic term is positive, then there exists a small enough perturbation in that direction that increases F .

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

$$v_1' H^F v_1 = 2 \left(\gamma - \frac{\beta^2}{8\alpha} (x_3 - x_2) \right) = -u' H^F u \quad (138)$$

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

A.13 Proof of Lemma 9

We first show the desired representation of Π . Using the representation in Equation 34, we write

$$\begin{aligned} \hat{\Pi}(K) &= \sum_{k=1}^K \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \left(\alpha(2\theta - 1)x_k - \beta \frac{x_k^2}{2} \right) d\theta \\ &= \sum_{k=1}^K \left[\alpha x_k \theta^2 - x_k \left(\alpha + \frac{\beta}{2} x_k \right) \theta \right]_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \\ &= \sum_{k=1}^K \left(\alpha x_k (\hat{\theta}_{k+1} - \hat{\theta}_k) (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\alpha + \frac{\beta}{2} x_k \right) (\hat{\theta}_{k+1} - \hat{\theta}_k) \right) \\ &= \frac{\beta}{2\alpha(K-1)} \sum_{k=2}^{K-1} \left(\alpha x_k (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\alpha + \frac{\beta}{2} x_k \right) \right) + (1 - \hat{\theta}_K) \left(\alpha \hat{\theta}_K - \frac{\beta}{2} \right) \end{aligned} \quad (139)$$

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

$$\begin{aligned} \sum_{k=2}^{K-1} \left(\alpha x_k (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\alpha + \frac{\beta}{2} x_k \right) \right) &= \sum_{k=2}^{K-1} \left(\alpha x_k \left(1 + \frac{\beta}{\alpha} x_k \right) - x_k \left(\alpha + \frac{\beta}{2} x_k \right) \right) \\ &= \frac{\beta}{2} \sum_{k=2}^{K-1} x_k^2 \\ &= \frac{\beta}{2} \sum_{k=2}^{K-1} \left(\frac{k-1}{K-1} \right)^2 = \frac{\beta}{12(K-1)} (K-2)(2K-3) \end{aligned} \quad (140)$$

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

$$\begin{aligned}\hat{\theta}_K &= \frac{1}{2} + \frac{\beta}{4\alpha} \left(1 + \frac{K-2}{K-1}\right) \\ &= \frac{2\alpha(K-1) + \beta(2K-3)}{4\alpha(K-1)} \\ 1 - \hat{\theta}_K &= \frac{2\alpha(K-1) - \beta(2K-3)}{4\alpha(K-1)}\end{aligned}\tag{141}$$

Putting this together, we write

$$\begin{aligned}\hat{\Pi}(K) &= \frac{\beta^2}{24\alpha(K-1)^2} \left((K-2)(2K-3) + \frac{3}{2\beta^2} (4\alpha^2(K-1)^2 - \beta^2(2K-3)^2) - \right. \\ &\quad \left. \frac{3}{\beta} (2\alpha(K-1)^2 - \beta(2K-3)(K-1)) \right)\end{aligned}\tag{142}$$

We now simplify further, collecting terms:

$$\begin{aligned}\hat{\Pi}(K) &= \frac{\beta^2}{24\alpha} \left(\frac{12\alpha^2}{2\beta^2} - \frac{6\alpha}{\beta} \right) + \frac{\beta^2}{24\alpha(K-1)^2} (2K-3) \left((K-2) - \frac{3}{2}(2K-3) + 3(K-1) \right) \\ &= \frac{\alpha - \beta}{4} + \frac{\beta^2}{48\alpha} \frac{(2K-3)(2K-1)}{(K-1)^2}\end{aligned}\tag{143}$$

We next show the desired representation of $\hat{\Gamma}$. This follows by direct calculation, starting from Equation 3:

$$\begin{aligned}\hat{\Gamma}(K) &= \gamma \left(1 - \left(\frac{1-1}{K-1} \right)^2 - \sum_{k=2}^2 \frac{k-1}{K-1} \frac{1}{K-1} \right) \\ &= \frac{\gamma}{2} \left(1 - \frac{1}{K-1} \right) = \frac{\gamma}{2} \frac{K-2}{K-1}\end{aligned}\tag{144}$$

A.14 Proof of Proposition 6

The comparative statics follow from standard monotone comparative statics arguments, after the observations that $\Pi_{K\alpha} < 0$, $\Pi_{K\beta} > 0$, and $\Pi_{K\gamma}$ (which themselves imply that Π satisfies single-crossing in the relevant dimensions).

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

$$V'(K) = \frac{\beta^2}{24\alpha(K-1)^2} - \frac{\gamma}{2(K-1)^2}\tag{145}$$

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

$$K > \tilde{K} := \frac{\beta^2}{12\alpha\gamma} + 1 \quad (146)$$

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 $V' > 0$ for all $K^* < K < K^* + 1$; this contradicts optimality.

A.15 Proof of Lemma 10

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

$$J(x, \theta) = u(x, \theta) + \pi(x, \theta) = \alpha x \theta - \beta \frac{x^2}{2} \quad (147)$$

Applying Proposition 4, the first-order condition for $k \in \{2, K - 1\}$ is

$$\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} (\alpha \theta - \beta x_k) d\theta - \gamma(-2x_k + x_{k-1} + x_{k+1}) = 0 \quad (148)$$

The FOC reduces to:

$$\gamma(-2x_k + x_{k-1} + x_{k+1}) = (\hat{\theta}_{k+1} - \hat{\theta}_k) \left[\frac{\alpha}{2}(\hat{\theta}_{k+1} + \hat{\theta}_k) - \beta x_k \right] \quad (149)$$

We next observe that $\hat{\theta}_k = \frac{\beta}{\alpha} \frac{x_k + x_{k-1}}{2}$. Therefore, $\hat{\theta}_{k+1} - \hat{\theta}_k = \frac{\beta}{\alpha} \frac{x_{k+1} - x_{k-1}}{2}$ and $\frac{\alpha}{2} (\hat{\theta}_{k+1} + \hat{\theta}_k) = \frac{\beta}{4} (x_{k+1} + x_{k-1} + 2x_k)$. Therefore, we can write

$$0 = \frac{\partial \hat{\Pi}}{\partial x_k} = \left(\frac{\beta}{2\alpha} (x_{k+1} - x_{k-1}) - \gamma \right) \left(\frac{\beta}{4} (x_{k+1} + x_{k-1} + 2x_k) \right) \quad (150)$$

We can also write the second-order condition

$$\frac{\partial^2 \hat{\Pi}}{\partial x_k^2} = -\frac{\beta}{2} \left(\frac{\beta}{2\alpha} (x_{k+1} - x_{k-1}) - \gamma \right) \quad (151)$$

Therefore, mirroring the logic above, the only solution compatible with the first-order condition is the even grid with $x_k = \frac{x_{k+1} + x_{k-1}}{2}$. The first-order condition for x_1 implies that:

$$\gamma(-2x_1 + x_2) \geq \hat{\theta}_2 \left[\frac{\alpha}{2} \hat{\theta}_2 - \beta x_1 \right] \Rightarrow (x_2 - 2x_1) \left(\frac{\beta^2}{2\alpha} x_2 - \gamma \right) \leq 0 \quad (152)$$

If the second term is positive, then the candidate solutions $x_1 \geq x_2/2$ are not local minima, since the second derivative is weakly positive. If the second term is negative, then the only possibility is that $x_1 = 0$.

We have therefore established that the optimal contract for maximizing total surplus is also a uniformly spaced grid.

We now consider the optimal principal's payoffs as a function of the number of contractibility points. We first write

$$\begin{aligned} \Pi(\hat{K}) &= \sum_{k=1}^K \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \left(\alpha \theta x_k - \beta \frac{x_k^2}{2} \right) d\theta \\ &= \sum_{k=1}^K \left[\frac{\alpha}{2} x_k \theta^2 - x_k \left(\frac{\beta}{2} x_k \right) \theta \right]_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \\ &= \sum_{k=1}^K \left(\frac{\alpha}{2} x_k (\hat{\theta}_{k+1} - \hat{\theta}_k) (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\frac{\beta}{2} x_k \right) (\hat{\theta}_{k+1} - \hat{\theta}_k) \right) \\ &= \frac{\beta}{\alpha(K-1)} \sum_{k=2}^{K-1} \left(\frac{\alpha}{2} x_k (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\frac{\beta}{2} x_k \right) \right) + (1 - \hat{\theta}_K) \left(\frac{\alpha}{2} (1 + \hat{\theta}_K) - \frac{\beta}{2} \right) \end{aligned} \quad (153)$$

We simplify the summation term as

$$\begin{aligned} \sum_{k=2}^{K-1} \left(\alpha x_k (\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k \left(\frac{\beta}{2} x_k \right) \right) &= \sum_{k=2}^{K-1} \left(\frac{\alpha}{2} x_k \left(\frac{2\beta}{\alpha} x_k \right) - x_k \left(\alpha + \frac{\beta}{2} x_k \right) \right) \\ &= \frac{\beta}{2} \sum_{k=2}^{K-1} x_k^2 \\ &= \frac{\beta}{2} \sum_{k=2}^{K-1} \left(\frac{k-1}{K-1} \right)^2 = \frac{\beta}{12(K-1)} (K-2)(2K-3) \end{aligned} \quad (154)$$

where we use that $\hat{\theta}_k + \hat{\theta}_{k+1} = \frac{2\beta}{\alpha} x_k$.

To simplify the second term, we observe that

$$\begin{aligned}
\hat{\theta}_K &= \frac{\beta}{2\alpha} \frac{2K-3}{K-1} \\
1 - \hat{\theta}_K &= \frac{2\alpha(K-1) - \beta(2K-3)}{2\alpha(K-1)} \\
1 + \hat{\theta}_K &= \frac{2\alpha(K-1) + \beta(2K-3)}{2\alpha(K-1)}
\end{aligned} \tag{155}$$

Putting this together, we write

$$\begin{aligned}
\Pi^T(K) &= \frac{\beta^2}{4\alpha(K-1)^2} \left(\frac{1}{3}(K-2)(2K-3) + \frac{1}{2\beta^2} (4\alpha^2(K-1)^2 - \beta^2(2K-3)^2) - \right. \\
&\quad \left. \frac{1}{\beta}(K-1)(2\alpha(K-1) - \beta(2K-3)) \right)
\end{aligned} \tag{156}$$

We observe, comparing this to Equation 142 in the proof of Lemma 9, that $\hat{\Pi}^T(K) = 2\hat{\Pi}(K)$. The program for defining K^{*C} follows immediately from this definition of Π^T .

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.
- CARBAJAL, J. C., AND J. C. ELY (2013): “Mechanism design without revenue equivalence,” *Journal of Economic Theory*, 148(1), 104–133.
- 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: Journal of the Econometric Society*, 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.
- 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. 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.