Contractibility Design

Roberto Corrao[*] Stanford

Joel P. Flynn[†] Yale

Karthik A. Sastry[‡] Princeton

October 10, 2025

Abstract

We introduce a model of incentive contracting in which the principal, in addition to writing contracts, must engage in contractibility design: creating an evidence structure that allows them to prove when the agent has breached the contract. Designing an evidence structure entails both (i) front-end costs borne *ex ante*, such as those of drafting contracts, and (ii) back-end costs borne *ex post*, such as those of generating evidence. We find that, under even small front-end costs, optimal contracts are coarse, specifying finitely many contingencies out of a continuum of possibilities. In contrast, under even large back-end costs, optimal contracts are complete. Applied to the design of procurement contracts, our results rationalize: (i) the discreteness of contracts, (ii) the presence of similarly vague contracts in low-stakes and high-stakes settings, and (iii) the discontinuous adjustment of contracts to changes in the economic environment.

[*]Stanford University Department of Economics, 579 Jane Stanford Way, Stanford, CA, 94305. Email: rcorrao@stanford.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.

This paper was formerly circulated under the title: "Optimally Coarse Contracts." We are grateful to Dirk Bergemann, Alessandro Bonatti, Oğuzhan Çelebi, Simone Cerreia-Vioglio, Tommaso Denti, Laura Doval, Piotr Dworczak (Discussant), Drew Fudenberg, Robert Gibbons, Marina Halac, Navin Kartik, Giacomo Lanzani, Elliot Lipnowski, Meg Meyer, Stephen Morris, Ellen Muir, Marco Ottaviani, Jacopo Perego, Andrea Prat, Ilya Segal, Ludvig Sinander, Francesco Squintani, Philipp Strack, Roland Strausz, Alexander Teytelboym, Jean Tirole, Rakesh Vohra, Alexander Wolitzky, Kai Hao Yang, and seminar participants at Columbia, Chicago, Collegio Carlo Alberto, PSE, Sciences Po, Oxford, Warwick, Nottingham, Yale, Arizona State, UBC, the North American Winter Meeting of the Econometric Society, SITE, Washington University in St. Louis, the NYU IO Day, UC Berkeley, and UT Austin for helpful comments. First posted version: November 2, 2023.

1 Introduction

Contracts are not enforceable by *fiat*. Contractual enforceability obtains only under careful writing and the generation of evidence that can be used to prove or disprove the legality of a given action. Accordingly, the literature on contract law emphasizes two classes of contracting costs. The first class comprises *front-end* costs of "foreseeing possible future contingencies, determining the efficient obligations that should be enforced in each contingency, [...] and drafting the contract language that communicates their intent to courts" (Scott and Triantis, 2005). The second class comprises *back-end* costs of "observing and proving the existence (or nonexistence) of any relevant fact after uncertainty has been resolved" (Scott and Triantis, 2005). The fact that lawyers spend up to 60% of their time drafting and reviewing documents (Thomson Reuters, 2024) underscores the importance of these costs for contract design.

In practice, contracting parties weigh such costs of complex contracts against their incentive gains. The contracts that result from this balancing act are often vague: even billion-dollar commercial contracts can be perplexingly imprecise and filled with phrases like "best efforts," "reasonable care," and "good faith."

The literature in microeconomic theory has approached this issue in two ways. On the one hand, the textbook mechanism design approach to optimal incentive contracting (see *e.g.,* Bolton and Dewatripont, 2004; Laffont and Martimort, 2009) abstracts away from costs of writing contracts and allows all contracts to be as precise as possible. On the other hand, a large and influential literature on incomplete contracts studies the optimal design of contracts given the constraint that they are vague, often with the justification that complete contracts are prohibitively costly (see *e.g.,* Anderlini and Felli, 1994; Battigalli and Maggi, 2002).

In this paper, we propose a new framework for studying contractibility design in principal-agent settings with incomplete information and where evidence is costly. Our goal is to understand how the nature of contracting costs affects the properties of optimal contracts. When are vaguely specified contracts optimal? What economic principles guide their design?

Formally, we consider a setting in which a principal contracts with an agent who has private information about a continuous type and may take a continuous action. The principal designs both a menu of contracts, which specify promised actions and payments, and an *evidence structure* that maps the action the agent takes to a set of evidence. The principal can use this evidence to prove whether the agent deviated from the promised action, imposing a fine upon the agent when they can prove such a breach of contract. We assume only that the feasible evidence structures for the principal satisfy natural continuity and monotonicity properties and always allow the principal to prove that they excluded the agent.

Evidence structures are costly. We consider two classes of costs: front-end and back-end costs. Both share the same basic form and capture the idea that the principal bears a cost whenever the evidence structure allows the distinction of one action from another. Front-end costs are borne *ex ante* and depend on the entire evidence structure, independently of the actions taken and the eventual generation and use of evidence. By contrast, back-end costs are borne *ex post* and the principal internalizes their expected value when designing the evidence structure. Thus, back-end costs depend on the likelihood that different sets of evidence are generated. Importantly, evidence structures that perfectly distinguish all pairs of actions have a finite and potentially arbitrarily small cost. The joint design of the evidence structure and menu involves a trade-off between the benefits of sharper incentives and the costs of creating and employing an evidence structure that guarantees these incentives.

In our main results, we establish that front-end and back-end costs generate qualitatively distinct optimal contracts. When the principal faces only front-end costs, every optimal evidence structure is *coarse*: it allows for the distinction of *finitely many* equivalence classes of actions. This, in turn, implies that there exists a finite optimal menu of contracts. Thus, front-end costs always lead to incomplete contracts: there are finitely many contingencies, but within each contingency many actions are legally permissible.[1] By contrast, when the principal faces only back-end costs, all optimal evidence structures and menus have the cardinality of the continuum. That is, despite transaction costs, it is optimal for the principal to specify a *complete* contract. Taken together, our results show that large transaction costs are neither necessary nor sufficient for optimal contracts to be coarse. Instead, coarseness hinges on the *nature* of these costs and, more specifically, their timing.

The intuition behind the coarseness of optimal contracts under front-end costs is that, when contracts are sufficiently precise, the marginal benefits of more precise contracting are an order of magnitude smaller than the corresponding marginal costs. More specifically, the principal allocates to each type of the agent the action that maximizes their virtual surplus among those actions that are distinguishable given the evidence structure. When there are already distinguishable actions that are close enough to the principal's second best, we show that adding another one even closer can only yield modest gains of *third-order*. By contrast, the front-end cost of adding that action in the middle of an interval of length t of indistinguishable actions leads to a *second-order* loss. Intuitively, to distinguish the t actions above from the t actions below, the number of costly comparisons that must be made is proportional to t^2.

[1] We adopt the definition of contractual incompleteness used, among others, by Spier (1992) and Scott and Triantis (2005). That is, while the contract specifies an obligation for each outcome, it *does not* specify a different obligation for each outcome, even when it is efficient to do so.

The contrasting result that back-end costs do not yield coarseness is subtle: for a fixed distribution of the agents' actions, the back-end cost is mathematically equivalent to a front-end cost. The difference in the two cases arises purely because the distribution of actions is determined endogenously and the principal internalizes its effects on this distribution. This renders the marginal gains and costs of more precise evidence of a comparable magnitude.

When both costs are present, contracts are always incomplete, but front-end and back-end costs interact to determine the structure of the optimal evidence structure and menu. Our approach, based on variational arguments, allows us to derive an upper bound on the optimal number of contractible actions. This bound is invariant to the slope of the modified virtual surplus in the agent's action, an object that we call the stakes of contracting. In this way, our analysis not only predicts the coarseness of contracts, but also the possibility that high- and low-value contracts could be equally coarse. This rationalizes the perhaps puzzling phenomenon that even billion-dollar contracts may remain vague. Moreover, the optimality of coarse contracts allows us to reduce the original infinite-dimensional problem to a finite-dimensional one, where simple first-order conditions describe optimal menus and evidence structures.

In an application, we use the model and our theoretical results to study contractibility design in optimal procurement. Agents are contractors who differ in their privately known productivity and can exert effort to produce output for a principal. The principal is a purchaser who uses incentive contracts to induce effort, but also bears the front-end and back-end costs of evidence to make these contracts enforceable.

The purchaser optimally offers discrete levels of payment corresponding to discrete tiers of effort. Such arrangements are common in practice, even in settings where piece rates or more complicated nonlinear contracts are feasible (see *e.g.,* Bewley, 1999). Applying our necessary optimality conditions, we solve analytically for the optimal menu. Front-end costs affect only the number of elements of an optimal menu, while back-end costs only distort the payments received for each menu item. Turning to comparative statics, we show that: (i) the vagueness of optimal contracts is invariant to type-neutral changes in productivity and (ii) contracts are rigid and unchanged in response to small changes to type-augmenting productivity, while large changes can induce a complete restructuring. Thus, our results rationalize the possibility that procurement contracts are vague in a way that is insensitive to economic stakes and adjust discontinuously to changes in the economic environment.

Related Literature. We build our theory on the formalization of evidence developed by Green and Laffont (1986) and Hart, Kremer, and Perry (2017). Two novel features of our evidentiary model are: (i) evidence is endogenously generated by the agent's action, rather than exogenously by their type (see Ball and Kattwinkel, 2019; Krähmer and Strausz, 2025, for recent contributions on mechanism design with evidence about exogenous types) and (ii) the evidence structure is optimally designed. We also formalize two general classes of costs of evidence structures (front-end and back-end) and show that their analysis is tractable, despite the richness of the space of evidence structures over which they are defined.

Endogenous evidence about the agent's actions leads to endogenous contractibility of the agent's actions. This links our analysis to the literature studying imperfect contractibility and the dichotomy between the letter of the contract and the spirit of the contract emphasized by Williamson (1975), Hart and Moore (2008), and Squintani (2019). We formalize this distinction via evidence structures, which nest as a particular and special case the type of imperfect contractibility (free disposal) studied by Grubb (2009) and Corrao, Flynn, and Sastry (2023). Differently from these papers, we study the *optimal* extent of contractibility. In addition, we derive the properties of contractibility from an evidentiary foundation.

Our work fits into a larger literature that provides foundations for incomplete contracts based on transaction costs (Simon, 1951; Coase, 1960). With respect to the Tirole (1999) classification of approaches, our analysis shows how costs of writing contracts (*i.e.*, front-end costs) lead to incompleteness while costs of enforcing contracts (*i.e.*, back-end costs) do not.[2]

Existing work on contracting with costly writing and/or enforcement studies problems with computational constraints on what events can be described (Anderlini and Felli, 1994, 1999; Al-Najjar, Anderlini, and Felli, 2006) and costs of writing that scale linearly with the number of clauses in a contract (Dye, 1985; Bajari and Tadelis, 2001; Battigalli and Maggi, 2002, 2008). Relative to this work, our analysis is different in two ways. First, we study an infinite (rather than finite) state and action space in which continuum contracts are feasible at finite cost. Thus, our analysis can directly speak to whether it is desirable to implement infinite contracts, like piece rates for workers (Holmström, 1979; Holmström and Milgrom, 1987) and nonlinear pricing with smooth quantity discounts (Wilson, 1993). This allows us to overcome the Tirole (1999) critique of the incomplete contracts literature that implicitly or explicitly assumes that such contracts are infinitely costly to write. Second, as mentioned above, our analysis proposes a new framework for justifying costly contractibility based on costly evidence. This microfoundation for the form of contractibility costs is especially important in light of our result that some, but not all, costs lead to optimal coarseness.

Finally, our methods are related to a literature on the finiteness of optimal contracts and menus. In Section 4.6, we elaborate on this connection in detail.

[2] There are, of course, other important perspectives on why contracts may be incomplete. One classical approach studies how parties can renegotiate previously specified incomplete contracts *ex post* (Segal, 1999; Hart and Moore, 1999; Che and Hausch, 1999). Another approach studies how *ex ante* costs of contracting serve a signaling role in the presence of private information for the principal (Spier, 1992). Our goal is not to subsume all such perspectives but instead to study how the nature of contracting costs affects incompleteness.

2 Model

2.1 The Agent and the Principal

The core of our model is a standard principal-agent setting. The agent's type $\theta \in \Theta = [0, 1]$ is drawn from a distribution F with strictly positive density f. The agent privately knows their type and can take an action x in the interval $X = [0, \overline{x}] \subset \mathbb{R}$. The agent has a twice continuously differentiable utility function $u : X \times \Theta \to \mathbb{R}$. We assume that higher types value higher actions more and that preferences are monotone increasing in the action: (i) u is strictly supermodular in (x, θ) and (ii) for each $\theta \in \Theta$, $u(\cdot, \theta)$ is strictly monotone increasing over X. The case with strictly decreasing preferences over X is analogous. All agent types value the zero action the same as their outside option payoff, which we normalize to zero, or $u(0, \theta) = 0$ for all $\theta \in \Theta$. The agent has quasilinear preferences over actions and money $\tau \in \mathbb{R}$, so their transfer-inclusive payoff is $u(x, \theta) - \tau$.

The principal's payoff derives from three sources. The first is a payoff that depends on the agent's action and type given by $\pi : X \times \Theta \to \mathbb{R}$, a twice continuously differentiable function such that $\pi(0, \theta) = 0$ for all $\theta \in \Theta$. The second is the monetary payment $\tau \in \mathbb{R}$ from the agent. The third is the cost of contractibility, which we will introduce in due course. We define the virtual surplus function $J : X \times \Theta \to \mathbb{R}$ as:

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

This is the total surplus generated when agent θ takes action x, net of the payments from the agent that ensure local incentive compatibility. We assume that J is twice continuously differentiable, strictly supermodular in (x, θ), and strictly quasiconcave in x.[3]

2.2 Evidence

In standard mechanism design, actions—or noisy signals thereof (*e.g.*, Holmström, 1977)—are assumed to be perfectly contractible. That is, upon seeing an action or a signal generated by an action, the contract completely specifies what the transfer between the principal and the agent will be. To go beyond this and model contractibility, we start from the premise that actions are contractible only to the extent that the principal can prove, with hard and

[3]Sufficient primitive conditions are that u and π are three times continuously differentiable with $u_{x\theta}, u_{xx\theta}, \pi_{x\theta} > 0$, $u_{xx}, u_{x\theta\theta}, \pi_{xx} < 0$, and that F has the increasing hazard rate property. We assume strict supermodularity of J for clarity of exposition, as this ensures there is no mechanical incentive for the principal to assign agents to finitely many allocations. As we later discuss, our analysis extends to settings with a strictly concave (but not supermodular) virtual surplus by leveraging the results of Toikka (2011).

legally verifiable evidence, that the agent breached the contract if they did indeed violate its terms. We formalize this by building a model of evidence regarding the agent's actions.

Our approach is motivated by existing frameworks for modeling evidence about *exogenous* types that we adapt to modeling *endogenous* actions (Green and Laffont, 1986; Hart, Kremer, and Perry, 2017). Formally, we posit a totally ordered space of evidence $(\Omega, \geq)$ with the cardinality of X. An evidence structure is a correspondence $\mathcal{E} : X \rightrightarrows \Omega$ that returns a set of evidence $\mathcal{E}(x)$ when the agent takes action $x \in X$. The principal can exhibit any subset of $\mathcal{E}(x)$ to an external arbitrator. We assume that the agent is "innocent until proven guilty." That is, the agent's action is found consistent with the contract unless the principal can prove that the agent did not act in accordance with the contract. Formally, the principal can prove that the agent did not do some promised action y if there exists a piece of evidence generated by their action $e \in \mathcal{E}(x)$ that could not have been generated by doing what they promised to do, *i.e.*, $e \notin \mathcal{E}(y)$. We say that the principal can *distinguish* action x from action y if and only if $\mathcal{E}(y)^c \cap \mathcal{E}(x) \neq \emptyset$.

We restrict attention to evidence structures with the following *regularity* properties:

Definition 1 (Regularity). *An evidence structure $\mathcal{E} : X \rightrightarrows \Omega$ is regular if it satisfies:*

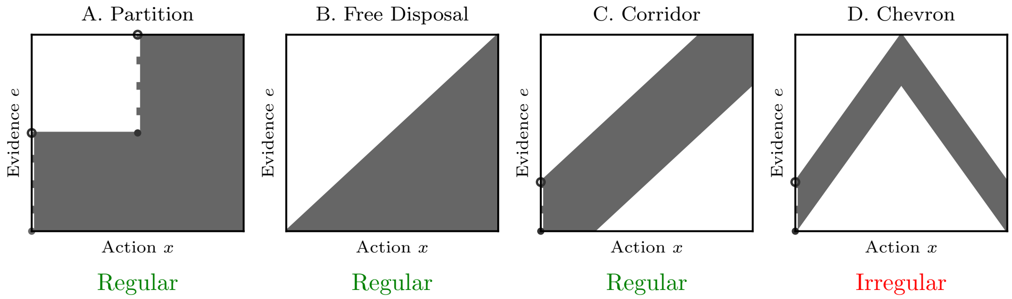
1. *(Continuity) Fix $x, y \in X$: (i) For every sequence $x_n \to x$ with $\mathcal{E}(x_n) \subseteq \mathcal{E}(y)$ for all $n \in \mathbb{N}$, it holds that $\mathcal{E}(x) \subseteq \mathcal{E}(y)$, (ii) Conversely, if $\mathcal{E}(x) \subseteq \mathcal{E}(y)$, then for every sequence $y_n \to y$, there exists a subsequence $y_{n_k} \to y$ and a sequence $x_k \to x$ such that $\mathcal{E}(x_k) \subseteq \mathcal{E}(y_{n_k})$ for all $k \in \mathbb{N}$.*
2. *(Definitive evidence of exclusion) For all $x \in X$, if $\mathcal{E}(x) \subseteq \mathcal{E}(0)$, then $x = 0$.*
3. *(Monotonicity) $\mathcal{E}$ is monotone in the strong set order.*

Continuity is a technical assumption, necessary to derive the existence of optimal contracts given an evidence structure. Definitive evidence of exclusion asserts that the principal can always prove that the agent was excluded if that is indeed the case. Monotonicity rules out situations in which high actions generate lower evidence than low actions and corresponds to a Boolean version of the monotone likelihood-ratio property.[4]

Figure 1 illustrates evidence structures and the content of regularity. Examples A, B, and C each correspond to different regular evidence structures. Example D corresponds to an evidence structure that fails monotonicity. In particular, taking $X = \Omega = [0, 1]$, we can observe that $x' = 1$ generates the piece of evidence $e = 0$ and $x = 1/2$ generates the piece of evidence $e' = 1$. However, it is *not* the case that the higher action generates the higher piece of evidence: $e' = 1 \notin \mathcal{E}(x') = [0, 1/4]$.

[4] To see this, observe that monotonicity of $\mathcal{E}$ in the strong set order is equivalent to the following property. For all $x, x' \in X$ and $e, e' \in \Omega$, if $e' \geq e$ and $x' \geq x$, then $\mathbb{I}[e' \in \mathcal{E}(x')]\mathbb{I}[e \in \mathcal{E}(x)] \geq \mathbb{I}[e' \in \mathcal{E}(x)]\mathbb{I}[e \in \mathcal{E}(x')]$.

Figure 1: Examples of Evidence Structures



Note: This figure shows examples of evidence structures. Examples A, B, and C are regular evidence structures (see Definition 1). Example D is an irregular evidence structure because it fails monotonicity.

An assumption implicit in our model is that the principal's payoff is not evidence *to an external arbitrator* in and of itself. This is motivated by practice in the contracting settings in which we are interested. Typically, one cannot establish breach of contract by merely claiming that you suffered a loss; one must provide concrete evidence of the underlying conduct and prove its inconsistency with a codified contract (see, *e.g.*, New York City Bar, 2014). In such settings, Maskin and Tirole's (1999) suggested approach to specify contracts directly over payoffs would not *by itself* overcome the necessity of evidence generation.

Extension: Noisy Evidence. In our baseline model, we assume that evidence and the arbitrator's finding of breach of contract are deterministic. However, it is natural to imagine that evidence is a noisy signal and arbitration is therefore random. As breach of contract is intrinsically binary, there are only two possible types of arbitration errors that may result from noise: type I errors, where the agent satisfied the terms of the contract but is nevertheless found to be in violation, and type II errors, where the agent violated the terms of the contract but is not found to be in violation. Because of this observation, in Appendix D we extend our model to allow for exogenous noise in evidence and arbitration and we show that our main results generalize to such a setting (see Proposition 4).

2.3 Costs of Evidence: Front-End *vs.* Back-End

We next describe the costs of evidence structures. Suppose the principal wishes to implement action $x \in X$. We assume that the evidence structure $\mathcal{E}$ gives rise to a cost every time it allows the distinction of any other action $z \in X$ from x. That is, when z can be distinguished from x, or $\mathcal{E}(x)^c \cap \mathcal{E}(z) \neq \emptyset$, the evidence structure yields a cost $g(x, z)$, where $g : X \times X \to \mathbb{R}_{++}$ is a continuously differentiable function and we let $\mathcal{G}$ denote the set of such functions.[5] Thus,

[5] Observe that this implies that $g(x, x) > 0$. This has no practical implications. As the diagonal is measure zero, nothing in our analysis changes if we consider $\hat{g}(x, y) = g(x, y)\mathbb{I}[x \neq y]$.

the cost of the evidence structure $\mathcal{E}$ for each possible action x is then given by:

$$\gamma_g(\mathcal{E}, x) = \int_{\{z \in X \,:\, \mathcal{E}(x)^c \cap \mathcal{E}(z) \neq \emptyset\}} g(x, z) \, \mathrm{d}z \tag{2}$$

The function g captures how hard it is to distinguish two actions. This may depend on the identities of those actions, how far apart those actions are, and/or whether those actions are distinguished *before* the agent acts or *after* the agent acts. The simplest case corresponds to paying a uniform cost of $\kappa > 0$ every time z is distinguished from x, or $g(x, z) = \kappa$.

Existing legal scholarship (Scott and Triantis, 2005) identifies two distinct stages of a contractual relationship at which actions must be distinguished from one another: before the agent takes an action and after the agent takes an action. We formalize this idea by considering two separate components for the cost of evidence structures.

The first component of costs corresponds to *front-end costs* that are borne before the agent acts. These include, for example, the costs of "foreseeing possible future contingencies, determining the efficient obligations that should be enforced in each contingency, [...] and drafting the contract language that communicates their intent to courts" (Scott and Triantis, 2005). The principal needs to pay these costs *regardless* of what the agent ultimately does. To model this, we specify the front-end cost as

$$\Gamma_f(\mathcal{E}) = \int_X \gamma_{g_f}(\mathcal{E}, x) \, \mathrm{d}x \tag{3}$$

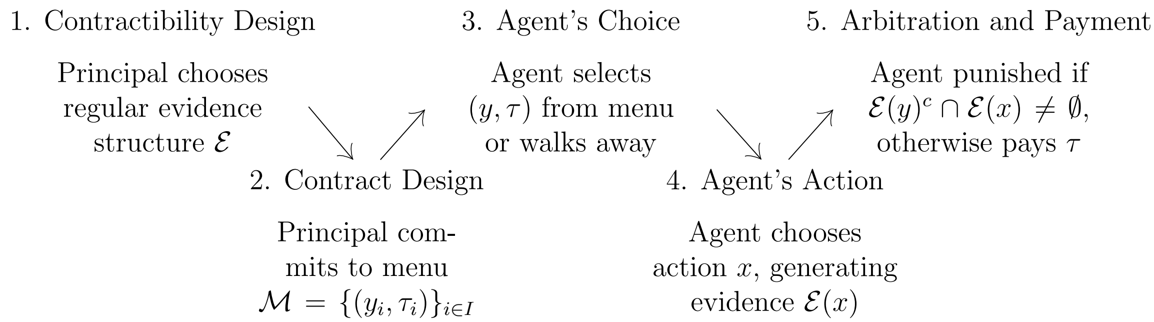
where $g_f \in \mathcal{G}$ measures the cost of distinguishing pairs of actions *ex ante*. Aggregating the action-specific costs under the uniform measure amounts to assuming that each action contributes equally to the total cost, consistent with our interpretation. Possible asymmetries in the costs of evidence about different actions can be encoded in the function g_f.

The second component of costs corresponds to *back-end costs* that are borne after the agent acts. These include costs of "observing and proving the existence (or nonexistence) of any relevant fact after uncertainty has been resolved" (Scott and Triantis, 2005). Importantly, these are *expected* costs and therefore depend on the likelihood that actions occur. We model this by specifying the back-end cost:

$$\Gamma_b(\mathcal{E}, Q) = \int_X \gamma_{g_b}(\mathcal{E}, x) \, \mathrm{d}Q(x) \tag{4}$$

where $g_b \in \mathcal{G}$ measures the cost of distinguishing pairs of actions *ex post* and $Q \in \Delta(X)$ is the principal's belief about the agent's action. Intuitively, every time that the agent takes action x, leading to the generation of evidence $\mathcal{E}(x)$, the principal incurs the cost $\gamma_{g_b}(\mathcal{E}, x)$.

Figure 2: Model Timeline



Thus, at the time of choosing the evidence structure and given beliefs about the distribution of the agent's potential actions, the principal's expectation of the costs that it will incur *ex post* corresponds to the back-end cost that we have specified in Equation 4.[6]

The key, but perhaps subtle, difference between Equations 3 and 4 is that the uniform weighting in the former one does not represent the likelihood of actions in *equilibrium*, as does Q in the latter equation, but rather the *fixed* relative contribution of each action to the total front-end costs. In fact, when Q is a fixed distribution, the back-end cost is *payoff* equivalent to a front-end cost with $g_f = g_b \times \mathrm{d}Q$. Yet, because Q is an equilibrium object, the analysis of the two costs will be significantly different. Finally, we observe that while $g_f, g_b \in \mathcal{G}$ can be distinct, they do not need to be.

Extension: More General Costs. Of course, parameterizing front-end and back-end costs by $g_f, g_b \in \mathcal{G}$ imposes a specific structure. In Appendix E, we consider a more general, non-parametric family of front-end and back-end costs and provide general sufficient conditions under which our main results hold (see Propositions 6 and 7).

2.4 Timeline and the Principal's Problem

We now describe the full timeline of events, which is also summarized in Figure 2. First, the principal chooses a regular evidence structure $\mathcal{E}$. Second, the principal commits to a menu $\mathcal{M} = \{(y_i, \tau_i)\}_{i \in I} \subseteq X \times \mathbb{R}$ of promised actions $y_i \in X$ and corresponding payments $\tau_i \in \mathbb{R}$ for some index set I. Third, the agent draws their type $\theta \in \Theta$ from the distribution F and either selects an element $(y, \tau) \in \mathcal{M}$ or walks away from the mechanism and takes their outside option.[7] Fourth, if the agent participates in the mechanism, they choose an action

[6] Observe that this formulation assumes that the principal must pay back-end costs to prove that it excluded agents. If this is not the case, our analysis differs only in determining a threshold type that participates in the mechanism and our main results hold as stated (see the arguments in Proposition 4 in Appendix D).

[7] As is standard, we assume that the agent, if indifferent, takes the principal's preferred action.

$x \in X$ and this generates a set of evidence $\mathcal{E}(x)$. Fifth, if x can be distinguished from y, then the agent is punished by an arbitrarily large fine,[8] and otherwise pays τ to the principal.

We next describe the principal's problem of contractibility design via evidence. We define $\phi : \Theta \to X$ and $t : \Theta \to \mathbb{R}$ as mappings from types to actions and transfers, respectively. We let $\mathcal{I}(\mathcal{M}, \mathcal{E})$ denote the pairs of measurable (ϕ, t) that arise in some pure-strategy subgame-perfect Nash equilibrium of the game between the principal and agent given $(\mathcal{M}, \mathcal{E})$. The principal jointly designs the evidence structure $\mathcal{E}$ and the menu $\mathcal{M}$ to maximize its expected payoffs net of front-end and back-end costs of evidence. For convenience, we break this problem into two parts. First, we maximize over the menu and the allocation to define the value of an evidence structure, including back-end costs but not front-end costs:

$$\mathcal{R}(\mathcal{E}) = \sup_{\mathcal{M},\, (\phi,t) \in \mathcal{I}(\mathcal{M},\mathcal{E})} \int_{\Theta} (\pi(\phi(\theta), \theta) + t(\theta))\, \mathrm{d}F(\theta) - \Gamma_b(\mathcal{E}, Q_\phi) \tag{5}$$

where $Q_\phi \in \Delta(X)$ is given by $Q_\phi(x) = \mathbb{P}_F[\phi(\theta) \leq x]$ and is the equilibrium distribution over actions induced by ϕ. An optimal evidence structure maximizes the total value over the space of all regular evidence structures:

$$\sup_{\mathcal{E}} \ \mathcal{R}(\mathcal{E}) - \Gamma_f(\mathcal{E}) \tag{6}$$

3 Main Results

We now study optimal evidence structures and menus under front-end and back-end costs.

3.1 Front-End Costs Imply Coarse Contracts

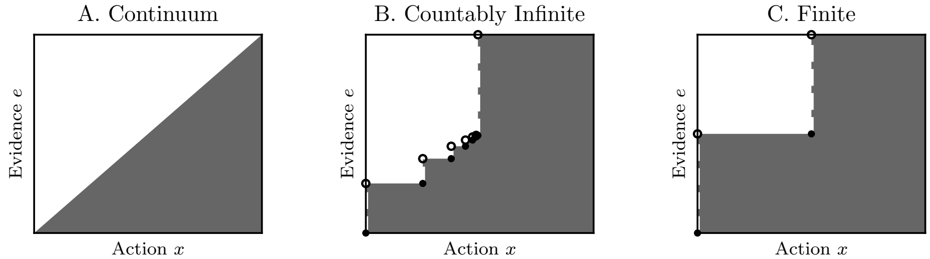
We first define *coarseness*. For any non-empty set A, we denote its cardinality by $|A| \in \mathbb{N} \cup \{\infty\}$. We let $\mathcal{E}(X) = \{\mathcal{E}(x)\}_{x \in X} \subseteq 2^{\Omega}$ denote the image of the evidence correspondence $\mathcal{E}$, and we recall that $\mathcal{M} = \{(y_i, \tau_i)\}_{i \in I} \subseteq X \times \mathbb{R}$ denotes a menu of actions and payments.

Definition 2 (Coarseness). *An evidence structure $\mathcal{E}$ is coarse if $|\mathcal{E}(X)| \in \mathbb{N}$. A menu $\mathcal{M}$ is coarse if $|\mathcal{M}| \in \mathbb{N}$.*

A coarse evidence structure partitions the actions into finitely many classes, within which actions cannot be distinguished. In Figure 3, we illustrate a coarse evidence structure and two non-coarse evidence structures. A coarse menu specifies only a finite number of contingencies: promised actions and corresponding payments if the agent cannot be shown to have

[8] It suffices that this fine is at least $\max_{x \in X, \theta \in \Theta} u(x, \theta)$.

Figure 3: Coarse and Non-Coarse Evidence Structures



$\mathcal{E}(X) = \{[0, x]\}_{x\in[0,1]}$
$|\mathcal{E}(X)| = \infty$
Not Coarse

$\mathcal{E}(X) = \{[0, \frac{1}{2} - \frac{1}{2^k}]\}_{k=1}^{\infty} \cup \{[0, 1]\}$
$|\mathcal{E}(X)| = \infty$
Not Coarse

$\mathcal{E}(X) = \{\{0\}, [0, \frac{1}{2}], [0, 1]\}$
$|\mathcal{E}(X)| = 3$
Coarse

Note: Each panel shows an example evidence correspondence, with an action space $X = [0, 1]$ and an evidence space $\Omega = [0, 1]$. Panel A has the cardinality of the continuum and Panel B has the cardinality of the set of natural numbers, both of which are infinite. Panel C has a finite cardinality, and is therefore coarse by Definition 2.

deviated from the promise. Therefore, the coarseness of the evidence structure has a natural interpretation as representing the *vagueness* of a contract: many possible actions generate the same evidence, and so many actions are legally consistent with the contract.

Toward stating our first main result, we define the maximum concavity of the virtual surplus function $\bar{J}_{xx} = \max_{x\in X,\theta\in\Theta} |J_{xx}(x, \theta)|$, the minimum supermodularity of the virtual surplus function $\underline{J}_{x\theta} = \min_{x\in X,\theta\in\Theta} J_{x\theta}(x, \theta)$, the maximum density of types $\bar{f} = \max_{\theta\in\Theta} f(\theta)$, and the minimal front-end cost of distinguishing any two actions $\underline{g}_f = \min_{x,z\in X} g_f(x, z)$. Under our maintained assumptions, we have that $\bar{J}_{xx}, \underline{J}_{x\theta}, \bar{f}, \underline{g}_f > 0$. We first consider the case with only front-end costs and no back-end costs, *i.e.*, $g_f \in \mathcal{G}$ and $g_b \equiv 0$.

Theorem 1 (Front-end Costs Imply Coarseness)**.** *If the principal only faces front-end costs, then an optimal evidence structure exists and* every *optimal evidence structure $\mathcal{E}^*$ is coarse:*

$$|\mathcal{E}^*(X)| \leq 6 + \left\lfloor 3\frac{\bar{x}\bar{J}_{xx}^2\bar{f}}{\underline{g}_f\underline{J}_{x\theta}} \right\rfloor \tag{7}$$

There exists an optimal menu $\mathcal{M}^*$ that is coarse with the same bound on its cardinality.

Under front-end costs, contracts are coarse. Underlying these coarse contracts are evidence structures that can only differentiate among a finite number of equivalence classes of actions. This optimal imprecision in the contract structure is reminiscent of the vague language (*e.g.*, "best efforts," "reasonable care," and "good faith") that Scott and Triantis (2005) argue is "commonplace in commercial contracts."

Our bound on the extent of richness in contracting is invariant to the slope of the virtual surplus function in the agent's action, or the *stakes* of contracting. In this way, the bound rationalizes why high-stakes settings where billions of dollars are on the line may feature contracts that are comparably vague to settings in which the stakes are trivial. In Section 4, we provide intuition for the exact form of the bound. In our application, we moreover study an explicit class of problems and find that this bound is tight up to an affine transformation.[9]

We also highlight two immediate remarks regarding Theorem 1. First, even arbitrarily small front-end costs can induce the qualitative property of coarseness. That is, if Theorem 1 holds given a particular front-end cost described by a cost of distinguishing g_f, then it also holds under the cost κg_f for any arbitrarily small $\kappa \in \mathbb{R}_{++}$. Second, the coarseness result holds despite the fact that the front-end cost of perfectly discerning evidence structures is finite. Thus, the result does *not* rely on the infeasibility of continuum evidence structures: instead, it shows their suboptimality under even arbitrarily small front-end costs.

3.2 Contractual Completeness Under Pure Back-End Costs

We next study the case with only back-end costs ($g_b \in \mathcal{G}$ and $g_f \equiv 0$), which will reveal a sharp contrast to the previous result. To do so, we define the modified virtual surplus

$$H(x,\theta) = J(x,\theta) - \int_x^{\overline{x}} g_b(x,z)\,\mathrm{d}z \tag{8}$$

We say that H is non-trivial if there exists $\theta \in \Theta$ such that $H(\cdot,\theta)$ has a unique maximizer that is, moreover, in the interior of X. This condition rules out only cases in which the principal prefers to allocate all agents either to 0 or $\bar{x}$.

Theorem 2 (Back-End Costs Do Not Imply Coarseness)**.** *If the principal faces only back-end costs, then an optimal evidence structure $\mathcal{E}^*$ with cardinality $|\mathcal{E}^*(X)| = \infty$ exists. Moreover, if H is non-trivial, then every optimal evidence structure $\mathcal{E}^*$ and every optimal menu $\mathcal{M}^*$ have cardinality $|\mathcal{E}^*(X)| = |\mathcal{M}^*| = \infty$.*

Under back-end costs, contracts specify a continuum of contingencies that are tailored to different types of agents. This is despite the presence of transaction costs that, at first impression, may appear similar to the front-end costs which we considered in Theorem 1. We provide a detailed economic intuition for the origin of this difference in Section 4.

3.3 The General Case: Coarseness and the Optimal Contract

We next show that, in the general case with *both* front-end and back-end costs, the qualitative properties of optimal contracts and evidence are as in Theorem 1.

[9] That is, the numbers 6 and 3 in the bound may not be optimal, but the remaining term is.

Corollary 1. *If H is strictly quasi-concave or strictly quasi-convex, then every optimal evidence structure $\mathcal{E}^*$ is coarse and there exists an optimal menu $\mathcal{M}^*$ that is coarse.*

When H is strictly quasi-concave in x for all $\theta \in \Theta$, the result follows from the same arguments as Theorem 1, replacing J with H. When H is strictly quasi-convex, the principal assigns every type to actions 0 and $\bar{x}$, so all optimal menus have at most two elements.

Remark 1 (A General Richness Bound)**.** When H is strictly quasi-concave, adapting the proof of Theorem 1 immediately yields the same bound with $\bar{H}_{xx} = \max_{x\in X,\theta\in\Theta} |H_{xx}(x,\theta)|$ replacing $\bar{J}_{xx}$ but with all other terms unchanged. Moreover, if g_b is symmetric—or distinguishing x from y is exactly as costly as it is to distinguish y from x—then:

$$|\mathcal{E}^*(X)| \leq 6 + \left\lfloor \frac{3\bar{x}\left(\bar{J}_{xx} + 3\bar{g}_{b,x} + \bar{x}\bar{g}_{b,xx}\right)^2 \bar{f}}{\underline{g}_f \underline{J}_{x\theta}} \right\rfloor \tag{9}$$

where $\bar{g}_{b,x} = \max_{x\in X} |g_{b,x}(x,x)|$ is the maximum slope of g_b *on the diagonal* and $\bar{g}_{b,xx} = \max_{x,z\in X} |g_{b,xx}(x,z)|$ is the maximum curvature of g_b. An immediate consequence is that, differently from front-end costs, when g_b is a constant function, $g_b(x,z) \equiv \kappa$, the presence of back-end costs does not even affect the general bound on richness. △

We finally establish the structure of optimal contracts with front-end and back-end costs.

Proposition 1. *Suppose that H is strictly quasi-concave or strictly quasi-convex, and fix an optimal menu $\mathcal{M}^* = \{(x_k, T(x_k))\}_{k=1}^{K^*}$ where $0 = x_1 < \ldots < x_k < \ldots < x_{K^*} = \bar{x}$. Then, for all $k \in \{2,\ldots,K^*-1\}$:*

$$\begin{aligned}&\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k,\theta)\,\mathrm{d}F(\theta) + \left(F(\hat{\theta}_{k+1}) - F(\hat{\theta}_k)\right)\left(g_b(x_k,x_k) - \int_{x_k}^{\bar{x}} g_{b,x}(x_k,z)\,\mathrm{d}z\right)\\ &= G_f(x_{k+1},x_k) - G_f(x_k,x_k) - \int_{x_{k-1}}^{x_k} g_f(x,x_k)\,\mathrm{d}x\end{aligned} \tag{10}$$

where $G_f(w,x) = \int_0^w g_f(x,z)\,\mathrm{d}z$ and $\hat{\theta}_k$ is given by: (i) $\hat{\theta}_1 = 0$ and $\hat{\theta}_{K+1} = 1$, (ii) $\hat{\theta}_k$ is the unique solution to $H(x_k,\hat{\theta}_k) - H(x_{k-1},\hat{\theta}_k) = 0$ for $k \in \{3,\ldots,K-1\}$, and (iii) $\hat{\theta}_2$ and $\hat{\theta}_K$ are the unique solution to the same equality if one exists and are otherwise 0 and 1, respectively. Furthermore, the corresponding transfer payments, for all $k \in \{1,\ldots,K^*\}$, are given by:

$$T(x_k) = \mathbb{I}[k \geq 2] \sum_{j=2}^{k} \left[u(x_j,\hat{\theta}_j) - u(x_{j-1},\hat{\theta}_j)\right] \tag{11}$$

Proof. See Appendix B.3. □

The first part of the result, in particular Equation 10, defines a second-order difference equation that must be satisfied by the optimal coarse menu. We derive this from a necessary first-order condition for the principal optimizing over the space of finite menus and considering a marginal increase in the menu item x_k. The first term on the left-hand-side measures the marginal increase in the principal's payoff and revenue from transfers. The second captures the marginal reduction in back-end costs, which scales with the measure of agents who are allocated to action x_k. The term on the right-hand-side measures the marginal change in front-end costs. While in principle this difference equation is quite complicated, we show in Section 5 that it takes a tractable form in a canonical application and enables computation of and comparative statics for the optimal coarse contract. The second part of the result completes the description of the optimal menu by describing the optimal tariff.

Remark 2 (The Optimal Number Of Self-Enforcing Recommendations)**.** We have described the solution to the problem up to characterizing the optimal number of self-enforcing recommendations K^* and selecting among solutions to the first-order condition from Proposition 1 (if there are multiple). In this remark, we provide a practical method to find this number that we employ in our application. First, for any $K \in \{2, \ldots, B\}$, where B is the bound from Remark 1, define the set of candidate optima with K self-enforcing recommendations as those that solve the first-order conditions from Proposition 1:

$$\mathcal{O}_K := \left\{ \mathbf{x} \in X^K : 0 = x_1 < \ldots < x_K = \overline{x} \text{ and Equation 10 holds} \right\} \tag{12}$$

Second, for every K, let $\mathcal{V}(K)$ denote the total value of the principal when they are restricted to evidence structures that can enforce only a vector of points $\mathbf{x} \in \mathcal{O}_K$, with the convention that $\mathcal{V}(K) = -\infty$ when $\mathcal{O}_K = \emptyset$. For $K = K^*$, this value coincides with the value of the original problem; moreover, we know that $K^* \leq B$. Third, we have that K^* solves the original problem if and only if $K^* \in \arg\max_{K \in \{2,\ldots,B\}} \mathcal{V}(K)$ which can be solved in linear time in the richness bound, B. $\triangle$

4 Proofs of Main Results

We now sketch the proof for the main results, up to some technical steps that are relegated to the Appendix. Our argument moves backward through the timeline of the model (Figure 2). We first describe how the agent acts given the potential for punishment under an evidence structure (4.1). We then describe the principal's choice of the menu, fixing an evidence structure (4.2). These arguments will suffice to prove Theorem 2. We next turn to characterizing the optimal evidence structure under front-end costs (Theorem 1), first establishing existence (4.3), then presenting the core economic arguments bounding the benefits and cost of "precise" contracting (4.4), and finally combining these arguments to establish the coarseness property (4.5). In a final subsection, we discuss how our results and proof strategy relate to other work (4.6).

4.1 Contractibility in the Shadow of the Law

We first make a simple observation about how the evidence structure and potential for punishment shape the agent's action. After selecting (y, τ) from a menu $\mathcal{M}$, the agent can take all of the following actions without being proved to have breached the contract:

$$C_{\mathcal{E}}(y) = \{x \in X : \mathcal{E}(x) \subseteq \mathcal{E}(y)\} \tag{13}$$

We call $C_{\mathcal{E}} : X \rightrightarrows X$ the *contractibility correspondence* induced by the evidence structure $\mathcal{E}$. The contractibility correspondence represents the set of actions to which the agent optimally restricts themselves in the so-called *shadow of the law*: in anticipation of a large punishment for legally provable deviations, the agent will avoid them. We call a contractibility correspondence C regular if there exists a regular evidence structure $\mathcal{E}$ such that $C = C_{\mathcal{E}}$.

The next lemma characterizes regular contractibility correspondences in terms of sets of *self-enforcing recommendations*, *i.e.*, actions from which an agent would not deviate:

Lemma 1. *A contractibility correspondence C is regular if and only if there exist two closed sets $\underline{D} \subseteq X$ and $\overline{D} \subseteq X$ such that $0 \in \underline{D}$, $0, \overline{x} \in \overline{D}$ and for all $y \in X$:*

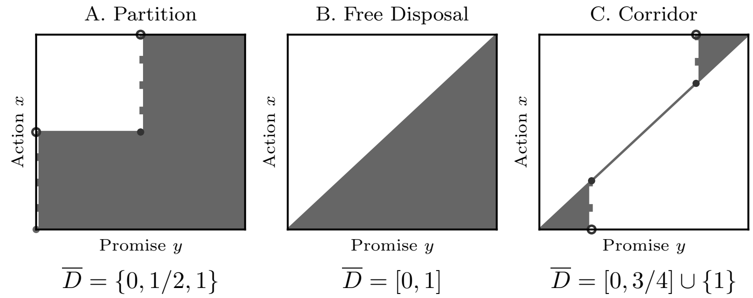
$$C(y) = \left[\max_{z \leq y : z \in \underline{D}} z, \min_{z \geq y : z \in \overline{D}} z\right] \tag{14}$$

Moreover, given C, $\underline{D}$ and $\overline{D}$ are unique and given by $\underline{D} = \{\min C(y)\}_{y \in X}$ and $\overline{D} = \{\max C(y)\}_{y \in X}$.

Proof. See Appendix C.1. □

For any regular evidence structure $\mathcal{E}$, we can uniquely define its sets of upper and lower self-enforcing recommendations $\overline{D}_{\mathcal{E}}$ and $\underline{D}_{\mathcal{E}}$. We call these the sets of self-enforcing recommendations because they are those that an agent with monotone increasing and decreasing preferences over final actions, respectively, would follow. The result furthermore implies that regular contractibility correspondences are intervals bounded by their upper and lower envelopes, $\overline{\delta}(y) = \max C(y)$ and $\underline{\delta}(y) = \min C(y)$. The monotonicity of these functions is necessary by monotonicity of the evidence structure. In turn, these functions are enough

Figure 4: Examples of Contractibility Correspondences



Note: This Figure illustrates the contractibility correspondences generated by the three regular evidence structures illustrated earlier in Figure 1 (Panels i.-iii.). The action space is $X = [0, 1]$. Under each example, we give the corresponding set of self-enforcing recommendations.

to recover the sets of self-enforcing recommendations which are uniquely defined by their images, $\overline{D} = \overline{\delta}(X)$ and $\underline{D} = \underline{\delta}(X)$.

Geometrically, Lemma 1 implies that, at any point, $\overline{\delta}$ and $\underline{\delta}$ either coincide with the identity line, or they are flat. To illustrate this, Figure 4 shows the three contractibility correspondences that are generated by the example regular evidence structures shown earlier in Figure 1. Visually, contractibility correspondences can be composed from three basic building blocks: "disposal" ("lower triangles"), "creation" ("upper triangles"), or complete indistinguishability ("boxes").

Focusing on the sets of *upper* self-enforcing recommendations is enough, as we show next:

Lemma 2. *For any regular evidence structure $\mathcal{E}$ such that $\underline{D}_{\mathcal{E}} \nsubseteq \{0, \overline{x}\}$, there exists a regular evidence structure $\mathcal{E}'$ with $\underline{D}_{\mathcal{E}'} \subseteq \{0, \overline{x}\}$ and such that $\mathcal{R}(\mathcal{E}') - \Gamma_f(\mathcal{E}') \geq \mathcal{R}(\mathcal{E}) - \Gamma_f(\mathcal{E})$, with the previous inequality being strict unless the principal faces only back-end costs.*

Proof. See Appendix B.1. □

The simple reason is that agents with monotone increasing preferences never prefer to deviate to lower actions, and building an evidence structure that distinguishes higher actions from lower actions is costly. Thus, for the remainder of the analysis, it is without loss of optimality to set $\underline{D}_{\mathcal{E}} = \{0\}$ and focus on $\overline{D}_{\mathcal{E}}$. With this, it is convenient to define $\overline{\mathcal{D}} \subseteq 2^X$ as the set of closed subsets of X that contain 0 and $\overline{x}$.

We finally observe that, in light of the arguments above regarding the shadow of the law, our model has an alternative interpretation in terms of partitioning actions. In particular, the function $\overline{\delta} : X \to X$ defines a (monotone) partition of the action space, mapping each recommendable action to a final action with which the agent would comply. Viewed this way, the evidence structure is a technology for enforcing this partition. Our assumptions on the form of transaction costs, motivated by the legal context of defining and compiling evidence for external arbitration, can therefore be understood as inducing a particular cost over partitions.

4.2 Optimal Contracts and the Value of Evidence Structures

We now turn to the principal's "Contract Design" (Step 2 in Figure 2), the choice of the menu $\mathcal{M} = \{(y_i, \tau_i)\}_{i \in I}$. The agent's ability to deviate from the promised action complicates this problem. That is, ours is a problem of both adverse selection and *ex post* moral hazard of a particular form that is governed by the evidence structure. Nonetheless, the following lemma shows that the solution to this problem can be described as a pointwise maximization of the modified virtual surplus function, subject to the constraint that actions are within the set of self-enforcing actions:

Lemma 3. *For every regular evidence structure $\mathcal{E}$ such that $\underline{D}_{\mathcal{E}} \subseteq \{0, \bar{x}\}$, we have $R(\mathcal{E}) = \mathcal{H}(\overline{D}_{\mathcal{E}})$, where:*

$$\mathcal{H}(\overline{D}) = \int_{\Theta} \max_{x \in \overline{D}} H(x, \theta) \,\mathrm{d}F(\theta) \qquad \forall \overline{D} \in \overline{\mathcal{D}} \tag{15}$$

Proof. See Appendix A.1. □

We prove this result in three steps. In the first step, we characterize implementable *final action* functions $\phi : \Theta \to X$ given $\mathcal{E}$. Each ϕ is implementable given $\mathcal{E}$ if and only if it is monotone increasing and maps into the set of self-enforcing recommendations: $\phi(\Theta) \subseteq \overline{D}_{\mathcal{E}}$. These properties are intuitively *necessary* because of standard arguments for local incentive compatibility and the fact that agents will always take the highest action that is feasible under the shadow of the law. A more technical argument is required to show they are *sufficient*, given that agents can contemplate global "double deviations": type θ could, in principle, imitate type θ', receive their recommended action $\phi(\theta')$, and then choose a feasible final action $x \in C_{\mathcal{E}}(\phi(\theta'))$ that differs from $\phi(\theta')$.

In the second step, we show that, for every $\mathcal{E}$ and every ϕ that is implementable given $\mathcal{E}$, we can rewrite the back-end cost as

$$\Gamma_b(\mathcal{E}, Q_\phi) = \int_{\Theta} \int_{\phi(\theta)}^{\bar{x}} g_b(\phi(\theta), z) \,\mathrm{d}z \,\mathrm{d}F(\theta) \tag{16}$$

where $Q_\phi = F \circ \phi^{-1}$ is the distribution over actions induced by ϕ. This equality follows from a change of variable and the fact that $\phi(\Theta) \subseteq \overline{D}_{\mathcal{E}}$. Moreover, it establishes a key fact: the back-end cost does not depend on the structure of $\mathcal{E}$ when evaluated at one of the induced *equilibrium* distributions of actions. A direct implication of this is that the back-end cost induces a cost $\int_x^{\bar{x}} g_b(x, z) \,\mathrm{d}z$ for every action x that is taken by the agent in equilibrium.

This is akin to an additional production cost borne by the principal, but with a different economic interpretation.

In the third step, we express the transfer function $t : \Theta \to \mathbb{R}$ in terms of ϕ using local incentive compatibility and individual rationality. This allows us to define the modified virtual surplus function $H(x, \theta) = J(x, \theta) - \int_x^{\overline{x}} g_b(x, z)\,\mathrm{d}z$, inclusive of the "as-if production cost" induced by the back-end cost of evidence. Finally, we derive the pointwise formulation of Equation 15 by observing that the pointwise maximization has a monotone solution, due to the supermodularity of J, and thus H, in (x, θ). If we had assumed that H were strictly concave (but not necessarily supermodular), then at this point one could apply the techniques of Toikka (2011) to characterize the value.[10]

The Solution Under Back-end Costs. Theorem 2 follows as a corollary of Lemma 3. To see why, it is immediate that $\mathcal{H}(X) \geq \mathcal{H}(\overline{D})$ for all $\overline{D} \in \overline{\mathcal{D}}$. In the absence of front-end costs, any $\mathcal{E}$ such that $\overline{D}_{\mathcal{E}} = X$ is optimal. The remainder of the proof in Appendix A.2 shows that it is *strictly* optimal for the principal to choose an evidence structure and menu with the cardinality of the continuum when H is non-trivial. Intuitively, the principal wishes to allocate a continuum of different actions to a positive measure subset of agents. Restricting them to an evidence structure with a lower cardinality would therefore induce a strict loss.

4.3 The Existence of an Optimal Evidence Structure

We next establish the existence of an optimal evidence structure in the presence of both front-end and back-end costs. Because of Lemmas 1 and 2, it suffices to show the existence of an optimal set of upper self-enforcing recommendations $\overline{D}$.

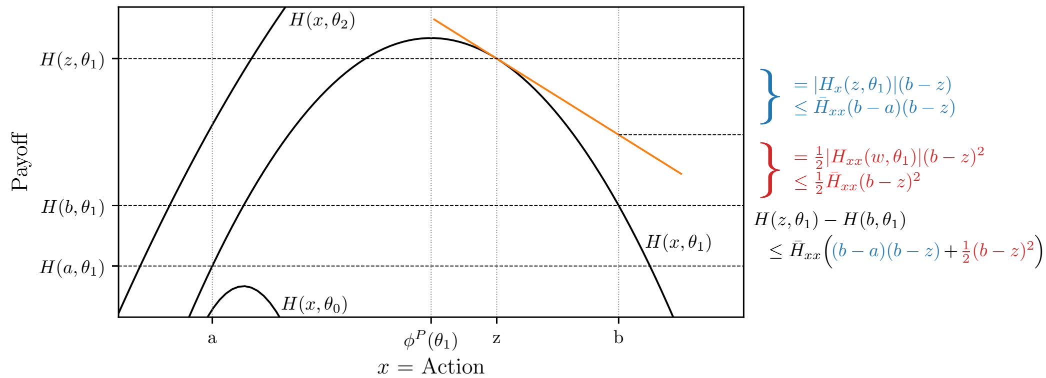
We have shown in Lemma 3 that we can express the value of an evidence structure via the function $\mathcal{H} : \overline{\mathcal{D}} \to \mathbb{R}$. By application of Berge's maximum theorem, this function is continuous in the Hausdorff topology. We can also express the front-end cost of contractibility associated with a regular evidence structure $\mathcal{E}$ with $\underline{D}_{\mathcal{E}} \subseteq \{0, \bar{x}\}$ as the function $\tilde{\Gamma}_f : \overline{\mathcal{D}} \to \mathbb{R}$ defined as $\tilde{\Gamma}_f(\overline{D}_{\mathcal{E}}) = \Gamma_f(\mathcal{E})$. We show that this function is continuous in the Hausdorff topology. The existence of an optimal evidence structure finally follows from Weierstrass' theorem because $\overline{\mathcal{D}}$ is compact in the Hausdorff topology.

Lemma 4. *$\mathcal{H} : \overline{\mathcal{D}} \to \mathbb{R}$ and $\tilde{\Gamma}_f : \overline{\mathcal{D}} \to \mathbb{R}$ are continuous in the Hausdorff topology. Therefore, an optimal evidence structure $\mathcal{E}^*$ exists.*

Proof. See Appendix A.3. □

[10] While it would be relatively simple to adapt our subsequent arguments to the non-supermodular case, we do not do so in the interests of space.

Figure 5: Illustrating the Payoff Gains of Precise Contractibility (Lemma 5)



Note: This figure illustrates one step in deriving the bound in Lemma 5. The black curves illustrate the virtual surplus function for three types $\theta_0 < \theta_1 < \theta_2$, and type θ_1 is such that the principal's preferred allocation is $\phi^P(\theta_1) \in (a,b)$. We annotate the relevant calculations for the first-order and second-order terms of the bound in blue and red, respectively. The orange line is tangent to $H(\cdot,\theta_1)$ at $x = z$, representing the first-order term in a Taylor expansion of $H(\cdot,\theta_1)$ around z. The action $w \in (z,b)$ used in the bound of the second-order term comes from Taylor's remainder theorem.

4.4 The Costs and Benefits of Precision

We next use a variational approach to derive the optimality of coarse evidence structures and menus as well as a bound on their cardinality. The key step is to show that an evidence structure with self-enforcing recommendations $a < b$, but no other $x \in (a,b)$, is strictly better than one that adds a self-enforcing recommendation $a < z < b$, provided that a and b are sufficiently close. We focus on the case of Corollary 1 in which there are both front-end and back-end costs, and H is strictly quasi-concave. Of course, if $g_b \equiv 0$ as in Theorem 1, then $H = J$ is strictly quasi-concave by assumption.

Benefits. We remind that $\bar{H}_{xx} = \max_{x,\theta} |H_{xx}(x,\theta)|$, $\underline{H}_{x\theta} = \min_{x,\theta} H_{x\theta}(x,\theta)$, and $\bar{f} = \max_\theta f(\theta)$. We define the function that maximizes modified virtual surplus as $\phi^P(\theta) = \arg\max_{x\in X} H(x,\theta)$, which is unique (by strict quasi-concavity), increasing, and moreover strictly increasing on the interior of X (by strict supermodularity).

Lemma 5. *Fix any $\overline{D} \in \overline{\mathcal{D}}$ and any $a, b \in \overline{D} \cap \phi^P(\Theta)$ such that $a < b$ and $\overline{D} \cap (a,b) = \emptyset$. For all $z \in (a,b)$, we have that:*

$$\mathcal{H}(\overline{D} \cup \{z\}) - \mathcal{H}(\overline{D}) \leq \frac{3}{2} \frac{\bar{H}_{xx}^2 \bar{f}}{\underline{H}_{x\theta}} (b-a)(b-z)(z-a) \tag{17}$$

Proof. See Appendix A.4. $\square$

We sketch the derivation of this inequality below. When the action z is added between a and b, agents fall into three groups: agents who continue to take their previous action, agents who previously took action a and now take action z, and agents who previously took action b and now take action z. To describe which types fall into the latter two cases, we define the type that the principal is indifferent to assign to $x, y \in \phi^P(\Theta)$ with the function $\hat{\theta} : \phi^P(\Theta)^2 \to \Theta$. Agents $\theta \in [\hat{\theta}(a, z), \hat{\theta}(a, b)]$ are now assigned to z rather than a and agents $\theta \in [\hat{\theta}(a, b), \hat{\theta}(z, b)]$ are now assigned to z rather than b. Hence, the total change in value is:

$$\mathcal{H}(\overline{D} \cup \{z\}) - \mathcal{H}(\overline{D}) = \int_{\hat{\theta}(a,z)}^{\hat{\theta}(a,b)} (H(z, \theta) - H(a, \theta))\, \mathrm{d}F(\theta) + \int_{\hat{\theta}(a,b)}^{\hat{\theta}(z,b)} (H(z, \theta) - H(b, \theta))\, \mathrm{d}F(\theta) \tag{18}$$

To bound this object, we first bound the integrands, focusing for simplicity on $H(z, \theta) - H(b, \theta)$. We illustrate this procedure in Figure 5. Specifically, because $\phi^P(\theta) \in [a, b]$ for all $\theta \in [\hat{\theta}(a, b), \hat{\theta}(z, b)]$ due to strict quasi-concavity, we can apply Taylor's remainder theorem twice to conclude that $|H(z, \theta) - H(b, \theta)| \leq \bar{H}_{xx} \left((b - a)(b - z) + \frac{1}{2}(b - z)^2\right)$. Intuitively, when concavity $\bar{H}_{xx}$ is higher, the principal gains more by tailoring actions to types.

We then bound the limits of integration. By applying the implicit function theorem to $\hat{\theta}$, we can show that $\hat{\theta}(z, b) - \hat{\theta}(a, b) \leq \frac{1}{2} \frac{\bar{H}_{xx}}{\underline{H}_{x\theta}}(z - a)$. If H is either very concave or not very supermodular, then many types of agents are allocated to the new contractible action z.

We can then combine all of the information to place the following upper bound on the second integral: $\frac{1}{2} \frac{\bar{H}_{xx}^2 \bar{f}}{\underline{H}_{x\theta}}(z - a) \left((b - a)(b - z) + \frac{1}{2}(b - z)^2\right)$. Performing the same steps for the first integral and factorizing yields the bound.

Costs. We next bound from below the marginal front-end cost of adding an isolated point. We remind that $\underline{g}_f = \min_{x,z \in X} g_f(x, z) > 0$.

Lemma 6. *Fix any $\overline{D} \in \overline{\mathcal{D}}$ and any $a, b \in \overline{D}$ such that $a < b$ and $\overline{D} \cap (a, b) = \emptyset$. For all $z \in (a, b)$, we have that:*

$$\tilde{\Gamma}_f(\overline{D} \cup \{z\}) - \tilde{\Gamma}_f(\overline{D}) \geq \underline{g}_f(b - z)(z - a) \tag{19}$$

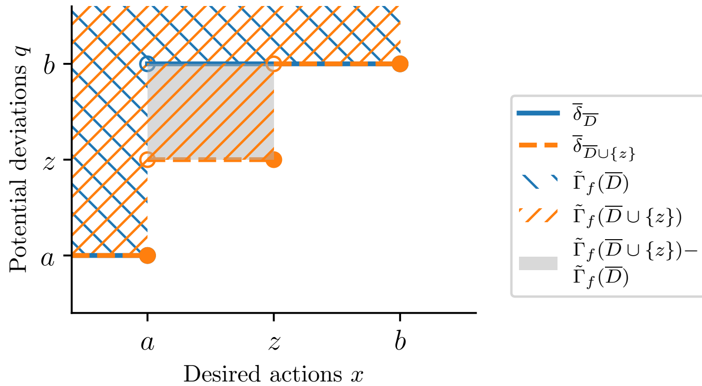
Proof. See Appendix A.5. $\square$

To understand this bound, we first re-express the front-end cost as the sum of costs across recommendations. We show that:

$$\Gamma_f(\mathcal{E}) = \int_X \left(\int_0^{\underline{\delta}_{\mathcal{E}}(x)} g_f(x, q)\, \mathrm{d}q + \int_{\overline{\delta}_{\mathcal{E}}(x)}^{\bar{x}} g_f(x, q)\, \mathrm{d}q \right) \mathrm{d}x = \int_X \int_{\overline{\delta}_{\mathcal{E}}(x)}^{\bar{x}} g_f(x, q)\, \mathrm{d}q\, \mathrm{d}x \tag{20}$$

where $\underline{\delta}_{\mathcal{E}} = \min C_{\mathcal{E}}$, $\overline{\delta}_{\mathcal{E}} = \max C_{\mathcal{E}}$. Intuitively, the principal pays to distinguish each desired action x from potential deviations $q < \underline{\delta}_{\mathcal{E}}(x)$ (first inner integral) and $q > \overline{\delta}_{\mathcal{E}}(x)$ (second inner

Figure 6: Illustrating the Cost of Precise Contractibility (Lemma 6)



Note: This figure illustrates the cost-saving bound of Lemma 6 in the case of constant costs of distinguishing, or $g_f \equiv \kappa > 0$. The blue solid line illustrates $\overline{\delta}$ for a $\overline{D}$ which contains the isolated points $a < b$, and the orange dashed line illustrates $\overline{\delta}$ for $\overline{D} \cup \{z\}$ for some $z \in (a, b)$. The colored hashed areas illustrate the front-end costs under each, and the grey rectangle is proportional to the difference in front-end costs.

integral). But, as shown earlier, the principal sets $\underline{\delta}_{\mathcal{E}}(x) = 0$ to avoid the cost of policing potential deviations that are undesirable to the agent (Lemma 2), so only the second term remains. Graphically, we can therefore understand the front-end cost as an integral of the area above the contractibility correspondence (*i.e.*, above $\overline{\delta}_{\mathcal{E}} = \max C_{\mathcal{E}}$) weighted by g_f.

To build intuition, we derive the bound in the simple case of constant costs of distinguishing ($g_f \equiv \kappa > 0$) and we illustrate this in Figure 6. In this case, the front-end cost is proportional to the area above $\overline{\delta}$. For each desired action x, the principal pays a cost proportional to the measure of potential deviations q from which x must be distinguished. The additional cost from adding self-enforcing recommendation z comes from distinguishing potential deviations $q \in (z, b]$ from desired actions $x \in (a, z]$. Thus, the cost is equal to $\kappa(b - z)(z - a)$, illustrated by the gray shading in Figure 6. To derive the bound under non-constant costs, we simply replace κ with the minimum value of g_f.

4.5 Establishing Coarseness

We finally combine our bounds on the costs and benefits of "precise" contractibility to establish the optimality of coarseness as well as the explicit bound:

Lemma 7. *If $\overline{D} \in \overline{\mathcal{D}}$ contains two points $a, b \in \overline{D} \cap \phi^P(\Theta)$ such that $0 < b - a < \frac{2}{3}\frac{g_f \underline{H}_{x\theta}}{\bar{H}^2_{xx}\bar{f}}$ and $\overline{D} \cap (a, b) \neq \emptyset$, then $\overline{D}$ is not optimal. Moreover, any optimal $\overline{D}^* \in \overline{\mathcal{D}}$ satisfies:*

$$|\overline{D}^*| \leq 6 + \left\lfloor 3\frac{\bar{x}\bar{H}^2_{xx}\bar{f}}{g_f \underline{H}_{x\theta}} \right\rfloor \tag{21}$$

Proof. See Appendix A.6. $\square$

To build intuition for the first part of the result, let us first assume that $\overline{D} \cap (a, b)$ contains a single point. Because a and b are close enough by assumption, Lemmas 5 and 6 yield that $\overline{D}$ cannot be optimal. Economically, this follows from the observation that the revenue benefits of adding an additional self-enforcing recommendation are an order of magnitude smaller than the front-end costs of doing so.

Consider now the case where $\overline{D} \cap (a, b)$ is an arbitrary set, possibly countably infinite, uncountably infinite and dense, or even uncountable infinite and nowhere dense (*e.g.,* a fractal set). Next, remove all the points in $\overline{D}$ between a and b to arrive at $\overline{D} \setminus (a, b)$. We prove that this yields an improvement via the following observation: we can reconstruct the points in $\overline{D} \cap (a, b)$ iteratively by adding one point at a time. Each time we add a point, the value must strictly decrease due to our previous argument. As the value and cost are continuous in the Hausdorff topology, the value of this sequence of sets converges to the value of the original set. This procedure therefore constructed a strict improvement to $\overline{D}$, demonstrating its suboptimality. As an optimum exists, we therefore know it must be finite.

We derive the explicit bound by making two additional observations. First, we show that outside of $\phi^P(\Theta) \subseteq X$, there can at most be four points in any optimal $\overline{D}^*$. Second, we observe that the previous observations ensure that any optimal $\overline{D}$ is a finite set that we may enumerate as $\{x_1, \ldots, x_L\}$ for some $L \in \mathbb{N}$. The necessary condition on the spacing of points translates to the family of inequalities $x_{k+1} - x_{k-1} \geq \frac{2}{3} \frac{g_f \underline{H}_{x\theta}}{\bar{H}^2_{xx} \bar{f}}$ for all $k \in \{2, \ldots, L-1\}$. The number of points that can fit in $[0, \bar{x}]$ and satisfy this restriction is bounded above by $2\left(1 + \left\lfloor \bar{x} / \left(\frac{2}{3} \frac{g_f \underline{H}_{x\theta}}{\bar{H}^2_{xx} \bar{f}}\right) \right\rfloor\right)$. Combining this information yields the bound.

4.6 Relationship To Other Work on Finite Menus

Before proceeding to our application, we briefly comment on the technical relationship between our results and those in other analyses of design problems with finite menus.

Costs of Contracts and Finite Contracts. An influential literature in contract theory has studied how costs of writing contracts can give rise to optimally incomplete contracts (Dye, 1985; Bajari and Tadelis, 2001; Battigalli and Maggi, 2002). In these papers, the cost of a contract depends on the number of elements on the contract, a form of a front-end cost in the language of Scott and Triantis (2005). In our setting, we would write these costs as $\tilde{\Gamma}_f(\overline{D}) = \kappa_f |\overline{D}|$. These costs make infinite contracts infinitely costly and thus led to the Tirole (1999) critique: while writing linear contracts may be costly, it is not plausible to believe that such costs are infinite. Our approach is different, emphasizing that the cardinality of costs may not fully capture the costs of evidence. Indeed, under our front-end costs of distinguishing, infinite contracts can be arbitrarily cheap, and the asymptotic marginal cost of adding another self-enforcing recommendation can be zero. Yet, infinite contracts remain sub-optimal. We briefly illustrate these points using an example:

Remark 3 (Coarseness Obtains Despite Vanishing Marginal Costs)**.** Consider a constant cost of distinguishing actions, $g_f \equiv \kappa_f > 0$. For any finite number of action recommendations $K \in \mathbb{N}$, the cost of having a uniform grid of K self-enforcing recommendations is:

$$\tilde{\Gamma}_f\left(\left\{\frac{k-1}{K-1}\bar{x}\right\}_{k=1}^{K}\right) = \frac{\kappa_f \bar{x}^2}{2}\frac{K-2}{K-1} \tag{22}$$

which is a *strictly concave* function of the number of self-enforcing recommendations K. As $K \to \infty$, we have $\tilde{\Gamma}_f \to \kappa_f \bar{x}^2/2$. Thus, even when the marginal cost saving of removing contractibility at infinity is zero, contracts can nevertheless be optimally coarse. △

Our results show that complete contracts can be suboptimal even when they are feasible, overcoming this Tirole (1999) critique. Moreover, our results demonstrate a subtlety in arguing that costs of contracts will necessarily yield incompleteness; this conclusion hinges on the nature of costs and whether they are incurred at the *ex ante* or the *ex post* stage.

Technical Approaches to Prove Finiteness of Menus. Here, we connect Theorem 1 to other results in the literature of mechanism and information design that derive the finiteness of optimal contracts, menus, or experiments. As explained, our main technical argument can be summarized in three steps: (i) deriving a lower bound on the cost of adding contractibility points; (ii) deriving an upper bound on the benefit of doing so; and (iii) applying a finite approximation result. The first step is the least related to prior work, since it pertains to the variational analysis of the front-end cost of distinguishing introduced in our model.

The second step is related to a literature in information theory that studies the optimal discretization of a continuous signal (see *e.g.,* Lloyd, 1982). Wilson (1989) used a similar approach to bound the loss of a principal screening an agent by using only finite menus with a fixed number of items. These results are refined and extended to the multidimensional case in Bergemann, Yeh, and Zhang (2021). We contribute by providing a bound that depends on the *structure* of the menu and the distances between adjacent triples of points, as opposed to just the menu's cardinality. This is important because any cardinality-based bound would be insufficient to derive our results due to the richness of our costs.

The third step is similar to the approximation arguments used by Bergemann and Pesendorfer (2007), Lou (2022), and Bergemann, Heumann, and Morris (2025) to derive the optimality of finite information structures in simultaneous information and mechanism design problems. Unlike these papers, we also provide an explicit bound on the extent of coarseness, and we begin by proving the existence of an optimal evidence structure without first restricting to the coarse ones. We also remark that our approach is different from alternative approaches based on duality (*e.g.,* Jung, Kim, Matějka, and Sims, 2019; Amador, Bagwell, and Carpizo, 2025) or variational arguments used in the optimal-delegation literature that rely on specific functional forms and solely derive the suboptimality of continuous intervals (*e.g.,* Alonso and Matouschek, 2008; Dovis, Kirpalani, and Sublet, 2024).

Finally, we note that applying the tools developed in this paper to study costly design in alternative economic settings, such as delegation and information disclosure, may be a potentially interesting avenue for further research.

5 Application to Procurement

We now apply our results to study procurement with endogenous evidence. The agent is a contractor with private information about their productivity and the principal is a purchaser that bears both front-end and back-end transaction costs. We describe in closed form the optimal evidence structure and menu, which features a discrete number of payments corresponding to discrete tiers of effort (or, equivalently, quality). We show that front-end and back-end costs play distinct roles: the former determines the coarseness of the menu, while the latter distorts down equilibrium effort and payment. We further find that, due to its coarseness, the optimal payment structure can be rigid in the face of small changes to productivity, while changing discontinuously in the face of large ones. Finally, we briefly describe additional applications to other screening problems.

5.1 Set-up: Procurement with Evidence of Effort

A contractor (the agent) supplies a service to a purchaser (the principal). The contractor's payoff from providing effort level $e \in E = [0, 1]$ is

$$\tilde{u}(e, \vartheta) = -\alpha(1 - \vartheta)e - \beta\frac{e^2}{2} \tag{23}$$

where $\vartheta \sim \tilde{F} = U[0, 1]$ is the contractor's privately observed productivity, $\alpha > 0$ is a parameter that increases productivity, especially so for higher types, and $\beta > 0$ is a parameter that scales curvature in effort costs. The contractor's final payoff is $\tilde{u}(e, \vartheta) + w$, where w is a payment from the purchaser. High effort leads to a higher payoff for the purchaser:

$$\tilde{\pi}(e) = \eta e \tag{24}$$

where $\eta > 0$ represents the return to effort. The purchaser's total payoff (exclusive of evidence costs) is $\tilde{\pi}(e) - w$.

The purchaser faces front-end and back-end costs of the constant form, $g_f(x, y) \equiv \kappa_f$ and $g_b(x, y) \equiv \kappa_b$, where $\kappa_f, \kappa_b > 0$. The former represent the *ex ante* costs of foreseeing contingencies and codifying their difference from one another: employing an evidence structure that enables the distinction of effort level e from e' will cost κ_f. The latter represents *ex post* costs of generating evidence of what action was taken and perhaps certifying this to an external arbitrator—that is, every time the agent puts in effort e, if the evidence generated allows effort level e to be distinguished from e', then the principal incurs a cost κ_b.

We introduce the simplifying assumption $\beta \leq \eta - \kappa_b \leq 2\alpha$ to ensure that the full action space is relevant for the problem: the purchaser will desire high effort from the highest productivity types and wish to exclude the lowest productivity types from undertaking the project. Our general results naturally do not hinge on this assumption, but it simplifies the closed-form description of the optimal contract.

The purchaser designs a menu of contracts, pairs of effort levels e and payments $w(e)$, as well as a regular evidence structure, $\mathcal{E} : E \to [0, 1]$, to maximize its expected payoff net of front-end and back-end costs.

5.2 Optimal Contracts Specify Finitely Many Effort Levels

We now apply our results to solve the purchaser's problem. We use the change of variables $x = 1 - e$ ("shirking") and $\theta = 1 - \vartheta$ ("unproductiveness") to bring the model under the assumptions of Section 2 (see Appendix B.4 for the formal argument). The purchaser's modified virtual surplus function is

$$H(x, \theta) = (2\alpha\theta + \beta - \eta + \kappa_b)x - \beta\frac{x^2}{2} \tag{25}$$

which is strictly concave and strictly supermodular. Hence, Corollary 1 implies that any optimal evidence structure is coarse and there exists an optimal menu that is finite. We can therefore optimize over a number $K \in \mathbb{N}$ of distinct shirking levels $0 = x_1 < \ldots < x_k < \ldots < x_K = 1$ (equivalently, effort levels).

Proposition 1 implies that optimal shirking levels solve a first-order condition that trades off the *ex post* benefits from precise screening against the *ex ante* costs of specifying the contract. Plugging the form of H and of the front-end cost into Equation 10, we have for each interior self-enforcing recommendation indexed by $k \in \{2, K^* - 1\}$:

$$\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} (2\alpha\theta + \beta(1 - x_k) - \eta + \kappa_b)\, \mathrm{d}\theta = \kappa_f(x_{k-1} + x_{k+1} - 2x_k) \tag{26}$$

where $\hat{\theta}_k = \frac{\beta}{4\alpha}(x_k + x_{k-1}) + \frac{\eta - \kappa_b - \beta}{2\alpha}$ are the types between which the purchaser is indifferent to assign x_k or x_{k-1}. This reduces to a second-order difference equation:

$$(x_{k+1} + x_{k-1} - 2x_k)\left[\frac{\beta^2}{16\alpha}(x_{k+1} - x_{k-1}) - \kappa_f\right] = 0 \tag{27}$$

with boundary conditions $x_0 = 0$ and $x_K = 1$. To solve for the optimal menu, we compute the solutions of this difference equation for each candidate K. Using the algorithm described in Remark 2, we solve for the optimal K^*. To do so, we observe that the difference equation has two families of solutions: a uniform grid over the points and an "alternating grid" with even spacings between odd points and even points. Explicit computation of the second-order conditions of the principal's problem demonstrates that the alternating grid is not a maximum. Thus, we can characterize the optimal menu in closed form:

Proposition 2 (Optimal Menu)**.** *The purchaser optimally offers the following menu:*

$$e_k = \frac{k-1}{K^*-1} \qquad w(e_k) = \frac{1}{2}\frac{k-1}{K^*-1}\left(\frac{\beta}{2}\frac{k-1}{K^*-1} + \eta - \kappa_b\right) \qquad k \in \{1, \dots, K^*\} \tag{28}$$

where K^* satisfies $|K^* - \tilde{K}| < 1$ and

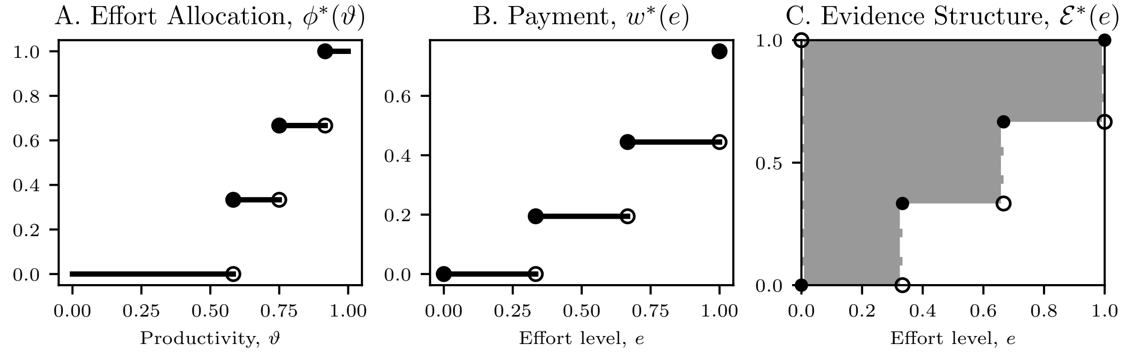
$$\tilde{K} = 1 + \frac{\beta^2}{12\alpha\kappa_f} \tag{29}$$

Moreover, K^* decreases in α, increases in β, decreases in κ_f, and does not depend on η or κ_b. If $\frac{\beta^2}{16\alpha\kappa_f} > 1$, then $K^* \geq 3$.

Proof. See Appendix B.4. □

In the optimal menu, an example of which we illustrate in Figure 7, only a finite number of effort levels and corresponding payments are explicitly specified. After promising any one of these effort levels (except the highest), the contractor can, in principle, put in any level of effort within a range without having provably breached the contract (Panel C). But, given any of these efforts, the contractor would receive the same compensation (Panel B); as a result, they keep their promise by putting in the minimal effort consistent with the terms (Panel A). We interpret this result as a description of real-world contracts that specify

Figure 7: Optimal Procurement Contracts with Costly Evidence



Note: This figure illustrates the optimal contract from Proposition 2 with $\alpha = \beta = 1$, $\eta = \kappa_b = 1/2$, and $\kappa_f = 1/32$. The optimal contract has $K^* = 4$. The first panel shows the assignment ϕ^*, the second panel shows the payment w^*, and the third shows the evidence structure $\mathcal{E}^*$.

obligations under vague contingencies like the agent's "best efforts," "reasonable care," or "good faith."

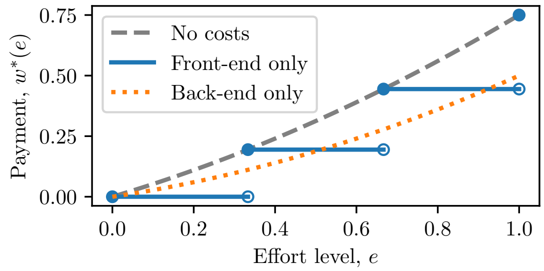
The optimality of uniformly spaced effort levels arises due to symmetries in the benefits and costs of more precise contracting. To understand symmetric benefits, we observe that the second derivative $H_{xx} = -\beta$ is constant as a function of (x, θ). Starting from perfect contractibility, the opportunity cost of adding contractibility in some interval of the action space is the same *regardless* of where that interval is located. This is for two reasons. First, as virtual surplus is quadratic in this model, the purchaser has an equal opportunity cost of forgoing differentiation for high-output versus low-output contractors. Second, because the optimal assignment function is linear, the same measure of types is affected. The corresponding symmetry in costs arises because, for constant costs of distinguishing, producing evidence to distinguish any pair of effort levels has the same cost.

5.3 The Distinct Roles of Front-End and Back-End Costs

We next contrast the economic role of front-end and back-end costs. To bring this into sharpest relief, Figure 8 shows the optimal payment schedule under three different assumptions for costs: no costs ($\kappa_f = \kappa_b = 0$), front-end costs only ($\kappa_f > 0$, $\kappa_b = 0$), and back-end costs only ($\kappa_f = 0$, $\kappa_b > 0$). As observed in our general analysis (Theorems 1 and 2), the front-end cost is essential for the qualitative prediction of a coarse menu. Under either the "no cost" and "back-end only" scenarios, the optimal contract specifies a continuum of effort levels and payments, although the exact schedule differs between the cases.

Economically, the front-end and back-end costs play separate roles in the comparative statics of the optimal contract. The number of contracted effort levels depends on the front-

Figure 8: Comparing Front-end and Back-end Costs



Note: This figure illustrates the different implications of front-end and back-end costs for the optimal menu. It shows the optimal payment schedule $w^*(e)$ in a parametrization with $\alpha = \beta = \eta = 1$ and three different calibrations for the costs of evidence: "No costs" with $\kappa_f = \kappa_b = 0$ (grey dashed line), "Front-end only" with $\kappa_f = 1/32$ and $\kappa_b = 0$ (blue solid line), and "Back-end only" with $\kappa_f = 0$ and $\kappa_b = 1/2$.

end cost parameter κ_f but not the back-end cost parameter κ_b, even in the case with both costs (Equation 28). The latter property arises because the back-end cost enters as if it is a linear production cost (see Equation 25). Such a linear cost scales the total surplus of the contract, but does not affect the cost or benefit of more *precise* contracting, which is instead governed by the curvature of the disutility of effort. In our general analysis, the same logic underlies the comparative statics of our richness bound, which depends on the back-end cost only through derivatives of g_b, which are zero in the example (see Remark 1).

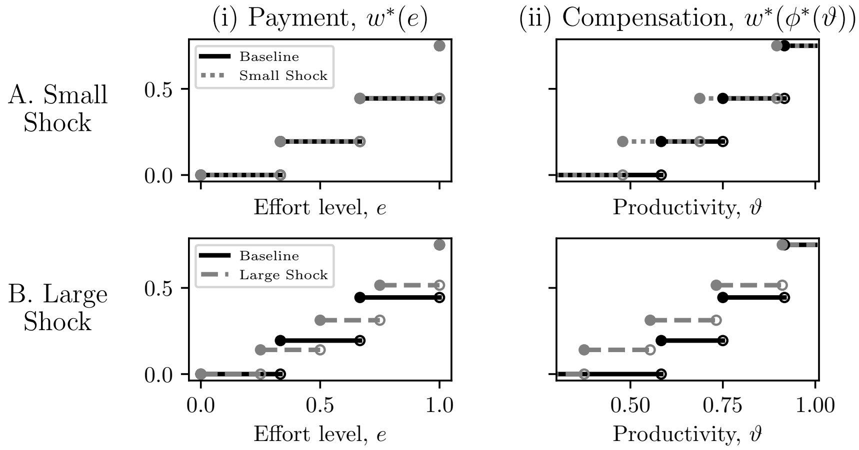
The payments themselves, conditional on the contracted effort levels, depend on the back-end cost parameter κ_b and not the front-end cost parameter κ_f. This is the reason why, in Figure 8, the "front-end only" and "no costs" contracts offer the same payments for the four effort levels that arise in equilibrium under front-end costs, while the "back-end only" contract specifies a lower payment for each effort level. The back-end cost, like a production cost, distorts downward the effort level assigned to each productivity type; because lower quality contractors are delivering each effort level when back-end costs are introduced, the payment schedule shifts downward. The front-end cost, by contrast, affects payments only if it affects information rents. With quadratic utility, the effect on information rents is zero.

5.4 Comparative Statics: Rigidity in Contracts

We now study how the optimal menu $\{e_k, w(e_k)\}_{k=1}^{K^*}$ responds to changes in productivity. The parameter η can be described as a *neutral* shifter of productivity. The parameter α can be described as a *type-augmenting* shifter of productivity, as it increases productivity more for agents with higher θ. For this analysis, we assume that parameters $(\alpha, \beta, \eta, \kappa_b, \kappa_f)$ are such that there is a unique K^*. This is true for almost all such vectors of parameters.[11]

[11] This follows because if there are multiple optimal values for K^*, then they must differ by one. Given this, it is simple to show that, for a situation in which K^* is not unique, an indifference equation of the

Figure 9: How Payment Schedules Respond to Productivity Changes



Note: This figure illustrates the comparative statics of the optimal contract in response to type-augmenting productivity (α), as per Corollary 2. We plot the wage structure ($w^*(e)$) and the compensation for each agent ($w^*(\phi^*(\vartheta))$) for a baseline case ($\alpha = 1, K^* = 4$; black solid line), a "small shock" with slightly higher productivity ($\alpha = 0.8, K^* = 4$; gray dotted line), and a "large shock" with much higher productivity ($\alpha = 0.7, K^* = 5$; gray dashed line). We fix $\beta = 1$, $\eta = \kappa_b = 1/2$, and $\kappa = 1/32$. The graph of transfers is truncated at $\vartheta = 0.3$, as lower types are excluded and receive no wages in all cases.

Corollary 2. *The following statements are true about the optimal menu:*

1. *The optimal effort levels are invariant to η, and payments continuously increase in η.*
2. *For every α, there exists a neighborhood $\mathcal{A}$ of α such that the optimal effort levels and payments are invariant to α' for $\alpha' \in \mathcal{A}$, and there exist different optimal effort levels and payments for $\alpha' \notin \mathcal{A}$.*

Proof. See Appendix B.5 $\square$

Neutral productivity shocks affect total surplus, but they do not affect the purchaser's *marginal* incentives to write a more precise contract. This property was also echoed in our more general analysis through the lack of dependence of the richness bound on *first derivatives* of the surplus function and, therefore, the first derivatives of u or π (Theorem 1 and Corollary 1). In practice, this is consistent with the motivating observation that vague contracts are ubiquitous even in high-value settings, like billion-dollar commercial contracts.

Type-augmenting productivity shocks have a more subtle effect on the payment structure, which depends discontinuously on their size. In response to small shocks, both the number of effort levels and the payments for those effort levels are unaffected. When type-augmenting productivity increases by sufficiently small amounts, agents' wages may increase, but only through discrete jumps across fixed steps. We illustrate such a scenario in the first row (Panel A) of Figure 9.

following form must hold: $\frac{\beta^2}{24\alpha\kappa_f}\Lambda(K^*) = 1$, where Λ is a polynomial.

By contrast, in response to larger changes, the entire payment schedule can change: that is, the purchaser offers a different and potentially non-nested set of effort levels and corresponding payments. That is, in response to large increases in type-augmenting productivity, effort levels are destroyed and created, and some individual contractors may even be reassigned to lower effort levels. We illustrate such a scenario in the second row (Panel B) of Figure 9: in particular, in the right panel, note that the new compensation schedule (gray dashed line) is sometimes above and sometimes below the old schedule (black solid line).

In this way, the model generates size-dependent rigidity in contracts: contracts are perfectly rigid in response to small shocks but feature discrete transitions in response to large shocks. This is true even though our analysis imposes no cost of adjusting the menu.

5.5 Additional Applications

Pay Grades and Performance Bands in Labor Procurement. In the context of firms procuring workers, our findings can help rationalize the ubiquity of pay grades that coarsely group workers to have common salaries. In his survey of wage-setting practices for US manufacturing and services firms, Bewley (1999) observes that piece rates that continuously vary compensation with output are relatively rare. A commonly mentioned problem with piece rates is the "cost of establishing the rates." Instead, an alternative arrangement is a *grade and step system* whereby the full set of labor tasks is segmented into discrete grades (job titles) and, within each grade, discrete steps that correspond to different salaries. Our model rationalizes such a system as an optimal response to even very small front-end costs of "establishing the rates."

Quality Segmentation in Input Procurement. We can re-interpret our model such that the principal is a purchasing firm, the agent is a supplier whose costs are given by Equation 23, and the costly action is to produce an input of a given quality. Our result implies that a supplier contract specifies a coarse menu of quality levels and corresponding payments. Our result is consistent with Asanuma's (1989) description of input sourcing by automobile and machinery manufacturers in Japan. Asanuma (1989) describes how purchasing firms segment suppliers into three categories differentiated by the quality of their inputs and contract in sharply different ways with suppliers in different categories.

Grading and Enforcement in Government Regulation. As in Baron and Myerson (1982), the agent is a firm that is a monopolist in its market and the principal is a regulator. The regulator imposes subsidies or fines that depend on the quality of the monopolist's output, but the monopolist has private information about its demand. The government pays a cost to design an evidence structure that legally formalizes a definition of quality. In this setting, the model predicts that the government specifies coarse quality grades, which are common in public health regulations. For example, U.S. regulations governing salmonella for raw poultry products place firms into one of three discrete categories based on the percentage of *Salmonella* tests of their meat products relative to the legally permissible level (U.S. Department of Agriculture, Food Safety and Inspection Service, 2021): less than 50% (Category 1), greater than 50% but less than 100% (Category 2), and greater than 100% (Category 3). These categories are publicly recorded and form the basis for escalating enforcement measures, such as intensified inspection and suspension of inspection services.

Penalties and Quality Conditions for Rental Goods. The Mussa and Rosen (1978) nonlinear pricing model fits into our abstract setting. A monopolist sells a service (*e.g.,* a car or vacation rental) that can be utilized to different extents. The monopolist chooses a menu of both utilization levels and prices, as in the standard nonlinear pricing problem. The monopolist faces a cost of higher utilization, akin to the production cost in Mussa and Rosen (1978). Moreover, they must write a contract that describes what levels of utilization by the buyers are acceptable. Contractibility is costly because the monopolist has to describe the acceptable levels of utilization of the good—for example, what constitutes a unit in "good" versus "bad" condition. Our results imply that utilization is contracted upon in tiers: for instance, discrete grades of condition for a car or vacation rental. This lines up with common practice. For example, the Europcar terms of service for the United Kingdom specify discrete condition levels for car returns and corresponding fees. Concretely, according to the terms and conditions of the rental contract (Europcar, 2024), if the front bumper of a Mini/Economy rental has a dent of less than 2cm, between 2cm and 5cm, between 5cm and 15cm, or a dent larger than 15cm, then the corresponding fees are £0, £542, £694, £738.

6 Conclusion

In this paper, we introduce a model of contractibility design. Our analysis has two premises. The first is that contracts are only enforceable to the extent that the principal can prove that the agent deviated from the terms of the contract. The second is that the codification and generation of evidence that can be used to provide such proof potentially entail front-end or back-end costs. Front-end costs yield coarse contracts while back-end costs yield complete contracts. In the presence of both costs, we show how the agent's and principal's preferences affect the structure and potential coarseness of contracts. We argue that our model generates insights into the nature of procurement contracts, worker hiring, government regulation, and monopoly pricing. In particular, we rationalize why even very small front-end costs of writing contracts may give rise to incomplete contracts. Moreover, we show that the incompleteness of contracts can be invariant to the economic stakes involved in a decision, rationalizing why high-value and low-value contracts alike are often vague.

A Proofs of the Main Results

The proofs of Lemmas 1 and 2 are mechanical and we defer them to Appendices C.1 and B.1. However, in what follows, we exploit that every regular evidence structure $\mathcal{E}$ affects the contracting problem solely via the induced regular contractibility correspondence $C = C_{\mathcal{E}}$ as explained in Section 4.1. Therefore, we often start with a regular C as the primitive object.

A.1 Proof of Lemma 3

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

Part 1: Implementation. We begin by establishing a general taxation principle with partial contractibility induced by an evidence structure. To do this, we re-cast the inner problem of designing the menu (Equation 5) as the following mechanism design problem. Given a fixed contractibility correspondence C, the revelation principle allows us to restrict to direct and truthful mechanisms.[12] Thus, a mechanism is a triple (ϕ, ξ, T) comprising a recommendation $\xi : \Theta \to X$, a final action or outcome $\phi : \Theta \to X$, and a tariff $T : X \to \overline{\mathbb{R}}$. The tariff and the recommendation jointly determine the transfer between the principal and the agent $T(\xi(\theta))$. The recommendation and the final action must be consistent with the contractibility correspondence, that is, $\phi(\theta) \in C(\xi(\theta))$. This, together with incentive constraints, determines the set of implementable mechanisms.

Definition 3 (Implementable Mechanism)**.** *A mechanism (ϕ, ξ, T) is implementable given a contractibility correspondence C if the following three conditions are satisfied:*

1. *Obedience:*

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

2. *Incentive Compatibility:*

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

[12] Following the standard approach in mechanism design, we select the principal's preferred equilibrium and restrict to deterministic mechanisms. We omit the formal proof of the revelation principle in this case as it is standard and closely follows the steps in Myerson (1982).

3. Individual Rationality:

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

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

Given a regular contractibility correspondence C, we say that $T : X \to \bar{\mathbb{R}}$ is monotone with respect to C if $T(y) \geq T(x)$ for all $x, y \in X$ such that $x \in C(y)$. Next, we establish that monotonicity of the tariff with respect to C is necessary and sufficient for implementability.

Lemma 8 (C-Monotone Taxation Principle)**.** *Fix a regular contractibility correspondence C. A final action function ϕ is implementable given C if and only if there exists a tariff $T : X \to \bar{\mathbb{R}}$ that is monotone with respect to C and such that:*

$$\phi(\theta) \in \arg\max_{x \in X} \{u(x,\theta) - T(x)\} \tag{30}$$

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. See Appendix B.2. □

With this result in hand, we now show:

Lemma 9 (Implementation)**.** *A final action function ϕ is implementable under $C = [\underline{\delta}, \overline{\delta}]$, with self-enforcing recommendation sets $\underline{D} = \underline{\delta}(X)$ and $\overline{D} = \overline{\delta}(X)$, if and only if it is monotone increasing and such that: (i) if agent preferences are monotone increasing, then $\phi(\Theta) \subseteq \overline{D}$, (ii) if preferences are monotone decreasing, then $\phi(\Theta) \subseteq \underline{D}$. Moreover, ϕ is supported by $\xi = \phi$ and tariff:*

$$T(x) = T(0) + u(x, \phi^{-1}(x)) - \int_0^{\phi^{-1}(x)} u_\theta(\phi(s), s)\,\mathrm{d}s \tag{31}$$

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

Proof. **(Only If for First Part)** If ϕ is implementable, then there exists (ξ, T) that support ϕ. By Lemma 8, we may take that $\xi = \phi$. By (IC) and Lemma 8, there exists a transfer function $t : \Theta \to \mathbb{R}$ given by $t(\theta) = T(\phi(\theta))$ such that $u(\phi(\theta),\theta) - t(\theta) \geq u(\phi(\theta'),\theta) - t(\theta')$ for all $\theta, \theta' \in \Theta$. As u is strictly single-crossing, Proposition 1 in Rochet (1987) then implies that ϕ is monotone. Without loss of generality, consider the case with monotone increasing preferences and toward a contradiction suppose that $\phi(\theta) \notin \overline{D}$. Deviating to $\overline{\delta}(\phi(\theta)) > \phi(\theta)$ is a strict improvement for the agent. Thus, if ϕ is implementable, then it is monotone, and $\phi(\Theta) \subseteq \overline{D}$ (or $\phi(\Theta) \subseteq \underline{D}$ with monotone decreasing preferences) holds.

(If For First Part) Without loss of generality, we again 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 $t : \Theta \to \mathbb{R}$ as

$$t(\theta) = Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s \tag{32}$$

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

$$T(x) = \inf_{\theta' \in \Theta} \{t(\theta') : x \in C(\phi(\theta'))\} \tag{33}$$

Fix $x, y \in X$ such that $x \in C(y)$. By transitivity of C, for all $\theta \in \Theta$, if $y \in C(\phi(\theta))$, then $x \in C(\phi(\theta))$. This shows that $\{\theta \in \Theta : y \in C(\phi(\theta))\} \subseteq \{\theta \in \Theta : x \in C(\phi(\theta))\}$. Therefore, applying the construction of T, $T(y) \geq T(x)$. 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 8 that ϕ is implementable.

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

$$u(\phi(\theta), \theta) - T(\phi(\theta)) \geq u(\phi(\theta), \theta) - t(\theta) = \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s - Z \tag{34}$$

Note that the right-hand side of this last equation is monotone increasing in θ since it is continuously differentiable with derivative $u_\theta(\phi(\theta), \theta) = \int_0^{\phi(\theta)} u_{x\theta}(z, \theta)\,\mathrm{d}z \geq 0$ for all $\theta \in \Theta$, owing to the fact that u is supermodular. 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) satisfies Equation 30. We first prove that, for all $\theta, \theta' \in \Theta$:

$$u(\phi(\theta), \theta) - t(\theta) \geq \max_{x \in C(\phi(\theta'))} u(x, \theta) - t(\theta') \tag{35}$$

This is a variation of the standard reporting problem under the final action 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, there is no strict gain to any agent of reporting θ' and consuming $x = \phi(\theta')$. Thus, there are no profitable single deviations under (ϕ, t).

We now must rule out double deviations. Suppose that θ imitates θ' and plans to take final action $x \neq \phi(\theta')$. As $\phi(\theta') \in \overline{D}$ (in the monotone increasing case), $x < \phi(\theta')$. But in that case, simply taking action $\phi(\theta')$ is better. But then this is a single deviation, which we have ruled out. The same logic applies in the monotone decreasing case. To derive the tariff, we can simply set $T(x) = t(\phi^{-1}(x))$. □

Part 2: Control Problem. We now use this characterization of implementation to turn the principal's problem into an optimal control problem:

Lemma 10. *Fix a regular $\mathcal{E}$ with corresponding regular contractibility correspondence C and induced $(\underline{D}, \overline{D})$. When agents have monotone increasing preferences, any optimal final action function solves:*

$$\begin{aligned} \mathcal{H}(\overline{D}) = \max_{\phi} \quad & \int_{\Theta} H(\phi(\theta), \theta) \, \mathrm{d}F(\theta) \\ s.t. \quad & \phi(\theta') \geq \phi(\theta), \phi(\theta) \in \overline{D}, \quad \theta, \theta' \in \Theta : \theta' \geq \theta \end{aligned} \tag{36}$$

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

Proof. We begin by eliminating the proposed allocation and transfers from the objective function of the principal. From the proof of Lemma 9, we have that transfers for any incentive compatible triple (ξ, ϕ, t) are given by:

$$t(\theta) = Z + u(\phi(\theta), \theta) - \int_0^{\theta} u_{\theta}(\phi(s), s) \, \mathrm{d}s \tag{37}$$

for some constant $Z \in \mathbb{R}$. Thus, any ξ that supports ϕ leads to the same principal 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 $\phi(\theta) \in \overline{D}$.

We next re-express the back-end costs. We start from the definition,

$$\Gamma_b(\mathcal{E}, Q) = \int_X \gamma_{g_b}(\mathcal{E}, x) \, \mathrm{d}Q(x) = \int_X \int_{\mathcal{E}(x)^c \cap \mathcal{E}(z) \neq 0} g_b(x, z) \, \mathrm{d}z \, \mathrm{d}Q(x) \tag{38}$$

By definition of C, we can write the domain of the inner integral as $\{z : z \notin C(x)\}$. Moreover, from Lemma 1, $C(x) = [\underline{\delta}(x), \overline{\delta}(x)]$, for some upper and lower envelope functions $\overline{\delta}, \underline{\delta}$. Therefore,

$$\Gamma_b(\mathcal{E}, Q) = \int_X \left(\int_0^{\underline{\delta}(x)} g_b(x, z) \, \mathrm{d}z + \int_{\overline{\delta}(x)}^{\bar{x}} g_b(x, z) \, \mathrm{d}z \right) \mathrm{d}Q(x) \tag{39}$$

Without loss of generality, we focus on the case of increasing preferences in which it is without loss of optimality to set $\underline{\delta} = 0$ (Lemma 2). We moreover observe, in light of the implementation argument above, that $x = \overline{\delta}(x)$ for any action x that is chosen in equilibrium. Using these simplifications, as well as a change of variables from $x, Q_\phi(x)$ to $\theta, F(\theta)$, we write

$$\Gamma_b(\mathcal{E}, Q_\phi) = \int_\Theta \int_{\phi(\theta)}^{\overline{x}} g_b(\phi(\theta), z)\,\mathrm{d}z\,\mathrm{d}F(\theta) \tag{40}$$

Using this expression for total transfer revenue, and the characterization of implementation from Lemma 9, we write the principal's problem as

$$\begin{aligned}
\max_{\phi, Z} \; & \int_\Theta \left(\pi(\phi(\theta), \theta) + Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s - \int_{\phi(\theta)}^{\overline{x}} g_b(\phi(\theta), z)\,\mathrm{d}z \right) \mathrm{d}F(\theta) \\
\text{s.t.} \quad & \phi(\theta') \geq \phi(\theta), \; \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \\
& u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s \right) \geq 0 \quad \forall \theta \in \Theta
\end{aligned} \tag{41}$$

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

$$\int_0^1 \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s\,\mathrm{d}F(\theta) = \int_0^1 \frac{(1 - F(\theta))}{f(\theta)} u_\theta(\phi(\theta), \theta)\,\mathrm{d}F(\theta) \tag{42}$$

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

$$\begin{aligned}
\max_{\phi, Z} \; & \int_\Theta (H(\phi(\theta), \theta) + Z)\,\mathrm{d}F(\theta) \\
\text{s.t.} \quad & \phi(\theta') \geq \phi(\theta), \; \phi(\theta) \in \overline{D} \quad \forall \theta, \theta' \in \Theta : \theta' \geq \theta \\
& u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s \right) \geq 0 \quad \forall \theta \in \Theta
\end{aligned} \tag{43}$$

where we define $H(x, \theta) = \pi(x, \theta) + u(x, \theta) - \frac{1 - F(\theta)}{f(\theta)} u_\theta(x, \theta) - \int_x^{\bar{x}} g_b(x, z)\,\mathrm{d}z$. It follows that it is optimal to set $Z \in \mathbb{R}$ as large as possible such that:

$$V(\theta) = u(\phi(\theta), \theta) - \left(Z + u(\phi(\theta), \theta) - \int_0^\theta u_\theta(\phi(s), s)\,\mathrm{d}s \right) \geq 0 \quad \forall \theta \in \Theta \tag{44}$$

We know that $V'(\theta) = u_\theta(\phi(\theta), \theta) \geq 0$ as we have already shown that $u(x, \cdot)$ is monotone over Θ. Thus, the tightest such constraint occurs when $\theta = 0$. Hence, the maximal Z must satisfy $V(0) = -Z \geq 0$. This implies that Z is optimally 0 and ensures that the (IR) constraint holds for all types. Hence, the principal's program is:

$$\begin{aligned} \max_{\phi} \quad & \int_{\Theta} H(\phi(\theta),\theta)\,\mathrm{d}F(\theta) \\ \text{s.t.} \quad & \phi(\theta') \geq \phi(\theta),\ \phi(\theta) \in \overline{D} \quad \forall \theta,\theta' \in \Theta : \theta' \geq \theta \end{aligned} \tag{45}$$

This completes the proof. $\square$

Part 3: The Optimal Contract. We first solve the pointwise problem in the control problem from Lemma 10 and then verify that this solution is monotone. The pointwise problem is $\max_{x\in\overline{D}} H(\phi(\theta),\theta)$, where the maximum exists as H is continuous and $\overline{D}$ is compact. We define

$$\overline{\phi}(\theta) = \min\{x \in \overline{D} : x \geq \phi^P(\theta)\} \quad \text{and} \quad \underline{\phi}(\theta) = \max\{x \in \overline{D} : x \leq \phi^P(\theta)\} \tag{46}$$

As H is strictly quasi-concave, this maximum is either $\overline{\phi}(\theta)$ or $\underline{\phi}(\theta)$. Define $\Delta H(\theta) = H(\overline{\phi}(\theta),\theta) - H(\underline{\phi}(\theta),\theta)$ for all $\theta \in \Theta$. When $\Delta H(\theta) > 0$, it is $\overline{\phi}(\theta)$. When $\Delta H(\theta) < 0$, it is $\underline{\phi}(\theta)$. When $\Delta H(\theta) = 0$, either is optimal. Thus, if it is monotone, the claimed solution is optimal (as it is supported on $\overline{D}$). Moreover, this pointwise solution is monotone by Theorem $4'$ in Milgrom and Shannon (1994). The claim that $\xi^* = \phi^*$ and the formula for the optimal tariff follow from applying Lemma 9.

A.2 Remainder of the Proof of Theorem 2

Under only back-end costs, $\mathcal{R}(\mathcal{E}) = \mathcal{H}(\overline{D}_{\mathcal{E}})$ and $\Gamma_f(\mathcal{E}) = 0$. Therefore, we can re-write the program that defines an optimal evidence structure (if it exists), Equation 6, as $\sup_{\mathcal{E}} \mathcal{H}(\overline{D}_{\mathcal{E}})$. From Lemma 10, it is immediate that $\mathcal{H}(X) \geq \mathcal{H}(\overline{D})$ for any $\overline{D} \in \overline{\mathcal{D}}$. Moreover, there exists an evidence structure $\mathcal{E}^*$ such that $\overline{D}_{\mathcal{E}^*} = X$: since Ω has the same cardinality as X, there exists a bijection $\omega : X \to \Omega$ and the corresponding evidence structure can be constructed as $\mathcal{E}^*(x) = \{\omega(x)\}$ for each $x \in X$. Thus, $|\mathcal{E}^*| = \infty$.

We next show that, if H is non-trivial, then every optimal evidence structure $\mathcal{E}^*$ and every optimal menu $\mathcal{M}^*$ have cardinality $|\mathcal{E}^*(X)| = |\mathcal{M}^*| = \infty$. We first observe that, if $|\mathcal{E}^*(X)| < \infty$, then necessarily $|\mathcal{M}^*| < \infty$. Moreover, letting $\overline{D}^* = \cup_{i\in I}\{y_i\}$, where we write $\mathcal{M}^* = \{(y_i,\tau_i)\}_{i\in I}$, we observe that $|\overline{D}^*| < \infty$. It is therefore necessary only to contradict the optimality of an $\overline{D}$ such that $|\overline{D}| < \infty$. Suppose such an $\overline{D}$ maximized $\mathcal{H}$. Because H is non-trivial, there exists some $\theta^* \in \Theta$ such that $\arg\max_{x\in X} H(x,\theta^*)$ exists, is a singleton, and is on the interior of X. By Berge's maximum theorem (given the continuity of H), the correspondence $\hat{H}(\theta) = \arg\max_{x\in X} H(x,\theta)$ is upper hemicontinuous; moreover, due to strict quasi-concavity of H, it is single-valued in some neighborhood $I_{\theta^*} \subset \Theta$ near θ^*. By Topkis' theorem (as H is strictly supermodular), $\hat{H}$ is strictly increasing. Thus, $\hat{H}(I_{\theta^*}) \subseteq X$ is an open interval. Moreover, for almost all $\theta \in I_{\theta^*}$, $\hat{H}(\theta) \notin \overline{D}$: otherwise, it would be the case that $|\overline{D}| = \infty$. Using the fact that the density of types is bounded away from zero,

$$\mathcal{H}(\overline{D}') - \mathcal{H}(\overline{D}) \geq \int_{I_{\theta^*}} \left(\max_{x \in \overline{D}'} H(x, \theta) - \max_{x \in \overline{D}} H(x, \theta) \right) \mathrm{d}F(\theta) > 0 \tag{47}$$

which completes the proof.

A.3 Proof of Lemma 4

As a consequence of Lemma 1, we can uniquely represent each regular $\mathcal{E}$ by $(\underline{D}, \overline{D}) \in \underline{\mathcal{D}} \times \overline{\mathcal{D}}$ (and vice-versa), where $\underline{\mathcal{D}}$ is the collection of closed subsets of X that contain 0. Moreover, as a consequence of Lemma 2, it is without loss of optimality to focus on the case in which $\underline{D} = \{0\}$. Finally, in Lemma 10, we defined the value function $\mathcal{H} : \overline{\mathcal{D}} \to \mathbb{R}$ such that, for all regular evidence correspondences $\mathcal{E}$, $\mathcal{R}(\mathcal{E}) = \mathcal{H}(\overline{D}_{\mathcal{E}})$. Thus, it suffices to show that

$$\sup_{\overline{D} \in \overline{\mathcal{D}}} \mathcal{H}(\overline{D}) - \tilde{\Gamma}_f(\overline{D}) \tag{48}$$

admits a solution. To show existence, we now argue that (i) the domain $\overline{\mathcal{D}}$ is compact in the Hausdorff topology, (ii) $\mathcal{H}$ is continuous in the Hausdorff topology, and (iii) $\tilde{\Gamma}_f$ is lower semi-continuous in the Hausdorff topology. The result then follows from Weierstrass' Theorem. Compactness of $\overline{\mathcal{D}}$ follows from the stronger statement in Lemma 13 that $\underline{\mathcal{D}} \times \overline{\mathcal{D}}$ is compact in the Hausdorff topology (see Appendix C.2). We next show continuity of $\mathcal{H} : \overline{\mathcal{D}} \to \mathbb{R}$. By applying Berge's maximum theorem to the program,

$$\tilde{\mathcal{H}}(\overline{D}, \theta) = \max_{x \in \overline{D}} H(x, \theta) \tag{49}$$

we have that the map $(\overline{D}, \theta) \mapsto \tilde{\mathcal{H}}(\overline{D}, \theta)$ is continuous in the appropriate product topology. Moreover, $\tilde{\mathcal{H}}$ is bounded, due to the boundedness of u (and its first derivatives), π, and g_b. Thus, the dominated converge theorem implies that $\overline{D} \mapsto \mathcal{H}(\overline{D})$ is continuous in the Hausdorff topology. Finally, lower semi-continuity of $\tilde{\Gamma}_f$ follows from the stronger claim in Proposition 3 that the front-end cost, when expressed as a function of $(\underline{D}, \overline{D})$, is continuous in the product Hausdorff topology (see Appendix C.2).

A.4 Proof of Lemma 5

Let ϕ_z^* and ϕ^* denote the optimal allocation under $\overline{D} \cup \{z\}$ and $\overline{D}$, respectively. By Lemma 10, the difference in values under these sets of self-enforcing recommendations is

$$\mathcal{H}(\overline{D} \cup \{z\}) - \mathcal{H}(\overline{D}') = \int_{\Theta} (H(\phi_z^*(\theta), \theta) - H(\phi^*(\theta), \theta)) \, \mathrm{d}F(\theta) \tag{50}$$

Define $\hat{\theta} : \phi^P(\Theta)^2 \to \Theta$ as the type for which the principal is indifferent between assigning x or y, *i.e.*, the unique solution to $\frac{H(x,\theta)-H(y,\theta)}{x-y} = 0$ when $x \neq y$ and as the unique solution to $H_x(x, \theta) = 0$ when $x = y$. This function is well defined (because $x, y \in \phi^P(\Theta)$), symmetric, and jointly continuous. Moreover, by the implicit function theorem and the fact that H is twice continuously differentiable and strictly supermodular, it follows that for all $x \in \phi^P(\Theta)$, $\hat{\theta}(x, \cdot)$ is continuously differentiable over $[\phi^P(0), x)$.

As shown in the proof of Proposition 1, the following assignment function is optimal:

$$\phi_z^*(\theta) = \begin{cases} z & , \theta \in (\hat{\theta}(a, z), \hat{\theta}(z, b)), \\ \phi^*(\theta) & , \text{otherwise.} \end{cases} \tag{51}$$

Further noting that $\phi^*(\theta) = a$ for $\theta \in (\hat{\theta}(a, z), \hat{\theta}(a, b))$ and $\phi^*(\theta) = b$ for $\theta \in (\hat{\theta}(a, b), \hat{\theta}(z, b))$,

$$\mathcal{H}(\overline{D} \cup \{z\}) - \mathcal{H}(\overline{D}) = \int_{\hat{\theta}(a,z)}^{\hat{\theta}(a,b)} (H(z, \theta) - H(a, \theta)) \, \mathrm{d}F(\theta) + \int_{\hat{\theta}(a,b)}^{\hat{\theta}(z,b)} (H(z, \theta) - H(b, \theta)) \, \mathrm{d}F(\theta) \tag{52}$$

We now proceed in three steps to bound this difference.

Step 1: Bounding the Integrands. We first derive an upper bound for $|H(x, \theta) - H(y, \theta)|$, for any $x, y \in X$ and $\theta \in \Theta$. Using Taylor's remainder theorem:

$$H(x, \theta) = H(y, \theta) + H_x(y, \theta)(x - y) + \frac{1}{2} H_{xx}(w(\theta), \theta)(x - y)^2 \tag{53}$$

for some $w(\theta) \in [x, y] \cup [y, x]$. We further apply Taylor's remainder theorem to take a first-order expansion of $H_x(y, \theta)$:

$$H_x(y, \theta) = H_x(\phi^P(\theta), \theta) + H_{xx}(v(\theta), \theta)(y - \phi^P(\theta)) = H_{xx}(v(\theta), \theta)(y - \phi^P(\theta)) \tag{54}$$

where the first equality defines the point $v(\theta) \in [y, \phi^P(\theta)] \cup [\phi^P(\theta), y]$ and the second uses the fact that $H_x(\phi^P(\theta), \theta) = 0$ by definition, since ϕ^P maximizes H. Combining these expansions:

$$\begin{aligned} |H(x,\theta) - H(y,\theta)| &\leq |H_x(y,\theta)||x-y| + \frac{1}{2}|H_{xx}(w(\theta),\theta)|(x-y)^2 \\ &\leq |H_{xx}(v(\theta),\theta)||y-\phi^P(\theta)||x-y| + \frac{1}{2}|H_{xx}(w(\theta),\theta)|(x-y)^2 \\ &\leq \bar{H}_{xx}\left(|y-\phi^P(\theta)||x-y| + \frac{1}{2}(x-y)^2\right) \end{aligned} \tag{55}$$

We now apply this result to bound the integrands:

$$\begin{aligned} H(z,\theta) - H(a,\theta) \leq |H(z,\theta) - H(a,\theta)| &\leq \bar{H}_{xx}\left(|z-\phi^P(\theta)||a-z| + \frac{1}{2}(a-z)^2\right) \\ &\leq \bar{H}_{xx}\left((b-a)(z-a) + \frac{1}{2}(z-a)^2\right) \end{aligned} \tag{56}$$

where in the final line we use the inequality $|z - \phi^P(\theta)| \leq b - a$, which in turn follows from the fact that $z, \phi^P(\theta) \in [a, b]$.[13]

By applying the same arguments to the second integrand, we obtain that:

$$H(z,\theta) - H(b,\theta) \leq \bar{H}_{xx}\left((b-a)(b-z) + \frac{1}{2}(b-z)^2\right) \tag{57}$$

Step 2: Bounding the Limits of Integration. By the implicit function theorem, we have that, for $y < x$:

$$\hat{\theta}_y(x,y) = \frac{H_x(y,\hat{\theta}(x,y))}{H_\theta(x,\hat{\theta}(x,y)) - H_\theta(y,\hat{\theta}(x,y))} \tag{58}$$

For the denominator, we observe that:

$$H_\theta(x,\hat{\theta}(x,y)) - H_\theta(y,\hat{\theta}(x,y)) = \int_y^x H_{x\theta}(t,\hat{\theta}(x,y))\,\mathrm{d}t \geq \underline{H}_{x\theta}(x-y) \tag{59}$$

For the numerator, we apply Taylor's remainder theorem:

$$H(x,\hat{\theta}(x,y)) = H(y,\hat{\theta}(x,y)) + H_x(y,\hat{\theta}(x,y))(x-y) + \frac{1}{2}H_{xx}(w,\hat{\theta}(x,y))(x-y)^2 \tag{60}$$

[13] To see the latter, note that $\theta \in (\hat{\theta}(a,z), \hat{\theta}(a,b))$; $\hat{\theta}(a,b) \leq \hat{\theta}(b,b) = {\phi^P}^{-1}(b)$ and $\hat{\theta}(a,b) \geq \hat{\theta}(a,a) = {\phi^P}^{-1}(a)$; and ϕ^P is increasing.

for some $w \in [y,x]$, and recall that $H(x,\hat{\theta}(x,y)) = H(y,\hat{\theta}(x,y))$. From this, we obtain that:

$$H_x(y,\hat{\theta}(x,y)) = -\frac{1}{2}H_{xx}(w,\hat{\theta}(x,y))(x-y) \leq \frac{1}{2}\bar{H}_{xx}(x-y) \tag{61}$$

Combining these facts, we obtain that:

$$\hat{\theta}_y(x,y) \leq \frac{\frac{1}{2}\bar{H}_{xx}(x-y)}{\underline{H}_{x\theta}(x-y)} = \frac{1}{2}\frac{\bar{H}_{xx}}{\underline{H}_{x\theta}} \tag{62}$$

A symmetric argument yields the same bound when $x < y$. Turning to the limits of integration, we apply the mean-value theorem to obtain:

$$\hat{\theta}(a,b) - \hat{\theta}(a,z) \leq \frac{1}{2}\frac{\bar{H}_{xx}}{\underline{H}_{x\theta}}(b-z) \tag{63}$$

By symmetry of $\hat{\theta}$ that $\hat{\theta}(z,b) - \hat{\theta}(a,b) = \hat{\theta}(b,z) - \hat{\theta}(b,a)$, we obtain the analogous bound.

Step 3: Bounding the Value. Combining steps 1 and 2, we conclude that

$$\begin{aligned}
&\mathcal{H}(\overline{D}\cup\{z\}) - \mathcal{H}(\overline{D}) = \int_{\hat{\theta}(a,z)}^{\hat{\theta}(a,b)} (H(z,\theta)-H(a,\theta))\,\mathrm{d}F(\theta) + \int_{\hat{\theta}(a,b)}^{\hat{\theta}(z,b)} (H(z,\theta)-H(b,\theta))\,\mathrm{d}F(\theta) \\
&\leq \bar{H}_{xx}\left(\int_{\hat{\theta}(a,z)}^{\hat{\theta}(a,b)}\left((b-a)(z-a)+\frac{1}{2}(z-a)^2\right)\mathrm{d}F(\theta) + \int_{\hat{\theta}(a,b)}^{\hat{\theta}(z,b)}\left((b-a)(b-z)+\frac{1}{2}(b-z)^2\right)\mathrm{d}F(\theta)\right) \\
&\leq \bar{H}_{xx}\bar{f}\left(\left(\hat{\theta}(a,b)-\hat{\theta}(a,z)\right)\left((b-a)(z-a)+\frac{1}{2}(z-a)^2\right) + \left(\hat{\theta}(z,b)-\hat{\theta}(a,b)\right)\left((b-a)(b-z)+\frac{1}{2}(b-z)^2\right)\right) \\
&\leq \frac{1}{2}\frac{\bar{H}_{xx}^2\bar{f}}{\underline{H}_{x\theta}}\left((b-z)\left((b-a)(z-a)+\frac{1}{2}(z-a)^2\right)+(z-a)\left((b-a)(b-z)+\frac{1}{2}(b-z)^2\right)\right) \\
&= \frac{1}{2}\frac{\bar{H}_{xx}^2\bar{f}}{\underline{H}_{x\theta}}(b-z)(z-a)\left((b-a)+\frac{1}{2}(z-a)+(b-a)+\frac{1}{2}(b-z)\right) = \frac{3}{2}\frac{\bar{H}_{xx}^2\bar{f}}{\underline{H}_{x\theta}}(b-a)(b-z)(z-a)
\end{aligned} \tag{64}$$

A.5 Proof of Lemma 6

From Lemma 1, $C_{\mathcal{E}}(x) = [\underline{\delta}_{\mathcal{E}}(x), \overline{\delta}_{\mathcal{E}}(x)]$. Thus,

$$\Gamma_f(\mathcal{E}) = \int_X \left(\int_0^{\underline{\delta}_{\mathcal{E}}(x)} g_f(x,q)\,\mathrm{d}q + \int_{\overline{\delta}_{\mathcal{E}}(x)}^{\bar{x}} g_f(x,q)\,\mathrm{d}q\right)\mathrm{d}x \tag{65}$$

Moreover, by Lemma 2, we can restrict attention to the case in which $\underline{\delta}_{\mathcal{E}} = 0$. We therefore define the induced cost over self-enforcing recommendations, $\tilde{\Gamma}_f : \overline{\mathcal{D}} \to \mathbb{R}$, as

$$\tilde{\Gamma}_f(\overline{D}) = \int_X \int_{\overline{\delta}_{\overline{D}}(x)}^{\bar{x}} g_f(x,q)\,\mathrm{d}q\,\mathrm{d}x \tag{66}$$

where we define $\overline{\delta}_{\overline{D}}(x) = \min_{q \geq x: q \in \overline{D}} q$. Recall that $\overline{D} \cap (a,b) = \emptyset$ and fix $z \in X$. We next observe that

$$\overline{\delta}_{\overline{D} \cup \{z\}}(x) = z\mathbb{I}[x \in (a,z]] + \overline{\delta}_{\overline{D}}(x)\left(1 - \mathbb{I}[x \in (a,z]]\right) \tag{67}$$

and, moreover, $\overline{\delta}_{\overline{D}}(x) = b$ for all $x \in (a,b]$. Thus,

$$\tilde{\Gamma}_f(\overline{D} \cup \{z\}) - \tilde{\Gamma}_f(\overline{D}) = \int_a^z \left(\int_z^{\bar{x}} g_f(x,q)\,\mathrm{d}q - \int_b^{\bar{x}} g_f(x,q)\,\mathrm{d}q \right) \mathrm{d}x = \int_a^z \int_z^b g_f(x,q)\,\mathrm{d}q\,\mathrm{d}x \tag{68}$$

The inequality follows by observing that $g_f(x,q) \geq \underline{g}_f := \min_{x,q \in X} g_f(x,q) > 0$, and hence

$$\tilde{\Gamma}_f(\overline{D} \cup \{z\}) - \tilde{\Gamma}_f(\overline{D}) \geq \int_a^z \int_z^b \underline{g}_f\,\mathrm{d}q\,\mathrm{d}x = \underline{g}_f(b-z)(z-a) \tag{69}$$

A.6 Proof of Lemma 7

We first establish the suboptimality of any $\overline{D}$ with the properties described in the statement. Define $S = \overline{D} \cap [a,b]$. Because S is closed, by Corollary 3.90 in Aliprantis and Border (2006) there exists a sequence $\{S_n\}_{n \in \mathbb{N}}$ of finite subsets of S converging to S in the Hausdorff topology. Because $\overline{D} \cap (a,b) \neq \emptyset$, without loss of generality, one can take this sequence such that $S_1 = \{a, b, x_1\}$ and $S_{n+1} = S_n \cup \{x_{n+1}\}$ for some sequence of points $\{x_n\}_{n \in \mathbb{N}}$ in $S \setminus \{a,b\}$. For all $n \in \mathbb{N}$, define $\overline{D}_n = \left(\overline{D} \setminus (a,b)\right) \cup S_n$ and observe that each of these sets is closed and that $\overline{D}_n$ converges to $\left(\overline{D} \setminus (a,b)\right) \cup S = \overline{D}$ in the Hausdorff topology. Moreover, define $V_n = \mathcal{H}\left(\overline{D}_n\right) - \tilde{\Gamma}_f\left(\overline{D}_n\right)$ for all n. In the proof of Lemma 4, we showed that both $\mathcal{H}$ and $\tilde{\Gamma}_f$ are continuous in the Hausdorff topology, yielding that $V_n \to \mathcal{H}(\overline{D}) - \tilde{\Gamma}_f(\overline{D})$.

We next prove that, for all $n \in \mathbb{N}$, we have $V_{n-1} > V_n$, where $S_0 = \{a,b\}$, $\overline{D}_0 = \overline{D} \setminus (a,b)$, and $V_0 = \mathcal{H}\left(\overline{D}_0\right) - \tilde{\Gamma}_f\left(\overline{D}_0\right)$. Fix any $n \in \mathbb{N}$ and observe that there exists two points $a_n, b_n \in S_{n-1}$ such that $0 < b_n - a_n < \frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}_{xx}^2 \bar{f}}$, and $S_n \cap (a_n, b_n) = \{x_n\}$. In turn, this implies that $a_n, b_n \in \overline{D}_{n-1} \cap \phi^P(\Theta)$ and that $\overline{D}_{n-1} \cap (a_n, b_n) = \emptyset$. Combining Lemmas 5 and 6, with $a = a_n$, $b = b_n$, and $z = x_n$, we have

$$\begin{aligned} V_n - V_{n-1} &= \left(\mathcal{H}(\overline{D}_{n-1} \cup \{x_n\}) - \tilde{\Gamma}_f(\overline{D}_{n-1} \cup \{x_n\})\right) - \left(\mathcal{H}(\overline{D}_{n-1}) - \tilde{\Gamma}_f(\overline{D}_{n-1})\right) \\ &\leq \frac{3}{2}\frac{\bar{H}_{xx}^2 \bar{f}}{\underline{H}_{x\theta}}(b_n - a_n)(b_n - x_n)(x_n - a_n) - \underline{g}_f(b_n - x_n)(x_n - a_n) \\ &= \underline{g}_f(b_n - x_n)(x_n - a_n)\left(\frac{3}{2}\frac{\bar{H}_{xx}^2 \bar{f}}{\underline{g}_f \underline{H}_{x\theta}}(b_n - a_n) - 1\right) < 0 \end{aligned} \tag{70}$$

where the very last inequality observes that $b_n - a_n < b - a < \frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}_{xx}^2 \bar{f}}$. By passing to the limit, we thus have $\mathcal{H}(\overline{D}_0) - \tilde{\Gamma}_f(\overline{D}_0) = V_0 > V_1 > \ldots > V_n > \ldots \geq \lim_n V_n = \mathcal{H}(\overline{D}) - \tilde{\Gamma}_f(\overline{D})$.

Thus, as desired, a $\overline{D}$ with the maintained property is not optimal.

We finally prove the finiteness of any optimal $\overline{D}$ as well as the bound on its cardinality. Fix an optimal set of self-enforcing recommendations $\overline{D}^* \in \overline{\mathcal{D}}$; at least one such $\overline{D}^*$ exists according to Lemma 4. All $\overline{D} \in \overline{\mathcal{D}}$ contain 0 and $\overline{x}$. We now establish that $\overline{D}^*$ can contain at most one point in $(0, \phi^P(0)]$ and at most one point in $[\phi^P(1), \overline{x})$. Toward a contradiction, suppose that there are two points $a, b \in \overline{D} \cap (0, \phi^P(0)]$ such that $a < b$. As H is strictly quasi-concave and strictly supermodular, we know that $H(x, \theta)$ is increasing in x for all $x \in (0, \phi^P(0)]$ and $\theta \in \Theta$. Hence, $H(b, \theta) > H(a, \theta)$ for all $\theta \in \Theta$. It follows that $\mathcal{H}(\overline{D}^*) - \mathcal{H}(\overline{D}^* \setminus (0, b)) = 0$. At the same time, by Lemma 6, we have that $\tilde{\Gamma}_f(\overline{D}^*) - \tilde{\Gamma}_f(\overline{D}^* \setminus (0, b)) \geq \underline{g}_f(b - a)a$. Thus, it is a strict improvement to modify $\overline{D}^*$ to $\overline{D}^* \setminus (0, b)$, a contradiction. An entirely symmetric argument for points above $\phi^P(1)$ establishes that there can be at most one element $a \in \overline{D}^*$ such that $a \geq \phi^P(1)$. Thus, $|\overline{D}^* \cap \left([0, \phi^P(0)] \cup [\phi^P(1), \overline{x}]\right)| \in \{2, 3, 4\}$.

By Lemma 7, we know that $\overline{D}^* \cap \phi^P(\Theta)$ cannot contain two points a and b that are closer than $\frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}^2_{xx}\bar{f}}$ and such that there exists a $z \in \overline{D}^* \cap (a, b)$. From this, it is immediate that $\overline{D}^* \cap \phi^P(\Theta)$ is a finite set. We let $L = |\overline{D} \cap \phi^P(\Theta)|$ and enumerate the L points as $\{x_1, \ldots, x_L\}$ with $\phi^P(0) \leq x_1 < x_2 < \ldots < x_k < \ldots < x_L \leq \phi^P(1)$. Applying Lemma 7, we have that $x_{k+1} - x_{k-1} \geq \frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}^2_{xx}\bar{f}}$. Under this constraint, it is possible to fit at most $2 + 2\left\lfloor \bar{x} / \left(\frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}^2_{xx}\bar{f}}\right)\right\rfloor$ points between any subset of $[0, \bar{x}]$. To see this constructively, for odd k set $x_1 = 0$ and then set $x_{2n+1} = x_{2n-1} + \frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}^2_{xx}\bar{f}}$ until $x_{2n+1} > \bar{x}$. For even k, start by setting $x_2 = \eta$, for some small $\eta > 0$ and then set $x_{2n} = x_{2n-2} + \frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}^2_{xx}\bar{f}}$ until $x_{2n} > \bar{x}$. This construction ensures that the feasibility constraint binds for all k, both even and odd. The even indexed points hit $\bar{x}$ first (unless $x_{2k+1} = \bar{x}$ for some k, in which you hit $\bar{x}$ one index sooner). This construction allows for at most $2\left(1 + \left\lfloor \bar{x} / \left(\frac{2}{3}\frac{\underline{g}_f \underline{H}_{x\theta}}{\bar{H}^2_{xx}\bar{f}}\right)\right\rfloor\right) \leq 2 + \left\lfloor 3\frac{\bar{x}\bar{H}^2_{xx}\bar{f}}{\underline{g}_f \underline{H}_{x\theta}}\right\rfloor$ points.

References

Al-Najjar, N. I., L. Anderlini, and L. Felli (2006): "Undescribable events," *The Review of Economic Studies*, 73(4), 849–868.

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

Alonso, R., and N. Matouschek (2008): "Optimal delegation," *The Review of Economic Studies*, 75(1), 259–293.

Amador, M., K. Bagwell, and L. Carpizo (2025): "Generalized Delegation," Working paper.

Anderlini, L., and L. Felli (1994): "Incomplete written contracts: Undescribable states of nature," *The Quarterly Journal of Economics*, 109(4), 1085–1124.

——— (1999): "Incomplete contracts and complexity costs," *Theory and decision*, 46(1), 23–50.

Asanuma, B. (1989): "Manufacturer-supplier relationships in Japan and the concept of relation-specific skill," *Journal of the Japanese and International Economies*, 3(1), 1–30.

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

Ball, I., and D. Kattwinkel (2019): "Probabilistic verification in mechanism design," *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 389–390.

Baron, D. P., and R. B. Myerson (1982): "Regulating a monopolist with unknown costs," *Econometrica*, pp. 911–930.

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

——— (2008): "Costly contracting in a long-term relationship," *The RAND Journal of Economics*, 39(2), 352–377.

Bergemann, D., T. Heumann, and S. Morris (2025): "Screening with Persuasion," Working paper.

Bergemann, D., and M. Pesendorfer (2007): "Information structures in optimal auctions," *Journal of Economic Theory*, 137(1), 580–609.

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

Bewley, T. F. (1999): *Why Wages Don't Fall During a Recession*. Harvard University Press.

Bolton, P., and M. Dewatripont (2004): *Contract Theory*. MIT Press, Cambridge, MA.

Che, Y.-K., and D. B. Hausch (1999): "Cooperative investments and the value of contracting," *American Economic Review*, 89(1), 125–147.

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

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

Dovis, A., R. Kirpalani, and G. M. Sublet (2024): "On the Optimal Allocation of Policy-Making," Discussion paper, National Bureau of Economic Research.

Dye, R. A. (1985): "Costly contract contingencies," *International Economic Review*, pp. 233–250.

Europcar (2024): "Terms and Conditions of Hire," Accessed January 25, 2025, from https://assets.ctfassets.net/wmdwnw6l5vg5/26z4Ocvms7JVQfOcOHgCmn/5876a9c86c9b46cc1af13571f53c1670/CGL_EN_GB.pdf.

Green, J. R., and J.-J. Laffont (1986): "Partially verifiable information and mechanism design," *The Review of Economic Studies*, 53(3), 447–456.

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

Hart, O., and J. Moore (1999): "Foundations of incomplete contracts," *The Review of Economic Studies*, 66(1), 115–138.

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

Hart, S., I. Kremer, and M. Perry (2017): "Evidence games: Truth and commitment," *American Economic Review*, 107(3), 690–713.

Holmström, B. (1977): "On incentives and control in organizations," Ph.D. thesis, Stanford University.

——— (1979): "Moral hazard and observability," *The Bell Journal of Economics*, pp. 74–91.

Holmström, B., and P. Milgrom (1987): "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica*, pp. 303–328.

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.

Krähmer, D., and R. Strausz (2025): "Unidirectional incentive compatibility," *Journal of Economic Theory*, p. 106051.

Laffont, J.-J., and D. Martimort (2009): *The Theory of Incentives: the Principal-Agent Model*. Princeton University Press, Princeton, NJ.

Lloyd, S. (1982): "Least squares quantization in PCM," *IEEE transactions on information theory*, 28(2), 129–137.

Lou, Y. (2022): "Optimal Delegation with Information Manipulation," Working Paper 4298900, SSRN.

Maskin, E., and J. Tirole (1999): "Unforeseen contingencies and incomplete contracts," *The Review of Economic Studies*, 66(1), 83–114.

Milgrom, P., and C. Shannon (1994): "Monotone comparative statics," *Econometrica*, pp. 157–180.

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.

New York City Bar (2014): "Contract Litigation: Breach of Contract," NYC Bar Association website, Legal Editor; retrieved from https://www.nycbar.org/get-legal-help/article/business-and-corporate-law/contract-litigation/.

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

Scott, R. E., and G. G. Triantis (2005): "Incomplete contracts and the theory of contract design," *Case Western Reserve Law Review*, 56, 187.

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

Simon, H. A. (1951): "A formal theory of the employment relationship," *Econometrica: Journal of the Econometric Society*, pp. 293–305.

Spier, K. E. (1992): "Incomplete contracts and signalling," *The RAND Journal of Economics*, pp. 432–443.

Squintani, F. (2019): "Optimal Overspecified Contracts," Working paper.

Thomson Reuters (2024): "Legal drafting challenges, risks and opportunities in a year of transformation," July 11. Retrieved from: https://legal.thomsonreuters.com/blog/legal-drafting-challenges-risks-and-opportunities/.

Tirole, J. (1999): "Incomplete contracts: Where do we stand? (Walras-Bewley Lecture)," *Econometrica*, 67(4), 741–781.

Toikka, J. (2011): "Ironing without control," *Journal of Economic Theory*, 146(6), 2510–2526.

U.S. Department of Agriculture, Food Safety and Inspection Service (2021): "FSIS Directive 10250.2: Performance Standards – Salmonella Verification Program for Raw Poultry Products," Technical directive, U.S. Department of Agriculture, Food Safety and Inspection Service, Accessed: October 10, 2025.

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

——— (1993): *Nonlinear Pricing*. Oxford University Press, New York.

Supplemental Appendix to *Contractibility Design* by Corrao, Flynn, and Sastry

B Omitted Proofs

B.1 Proof of Lemma 2

In Lemma 12 in Appendix C.1, we show that a regular contractibility correspondence C can be equivalently represented via (i) two sets of self-enforcing recommendations $\underline{D}, \overline{D}$ (see Lemma 1 in the main text); or via (ii) $C(y) = [\underline{\delta}(y), \overline{\delta}(y)]$ for all $y \in X$, where $\underline{\delta} : X \to X$ is an upper semi-continuous increasing function and $\overline{\delta} : X \to X$ is a lower semi-continuous increasing function such that: (i) $y \in [\underline{\delta}(y), \overline{\delta}(y)]$, (ii) $\underline{\delta}(x) = \underline{\delta}(y)$ for all $x \in [\underline{\delta}(y), y)$, (iii) $\overline{\delta}(x) = \overline{\delta}(y)$ for all $x \in (y, \overline{\delta}(y)]$, and (iv) $\overline{\delta}(0) = 0$.

There are only two $\underline{\delta}$ functions satisfying the properties of Lemma 1 and such that $\underline{\delta}(X) \subseteq \{0, \overline{x}\}$: (i) $\underline{\delta}(x) = 0$ and (ii) $\underline{\delta}(x) = \overline{x}\mathbb{I}[x = \overline{x}]$. Assume by contradiction that $\underline{D} \not\subseteq \{0, \overline{x}\}$, and therefore $\underline{\delta}$ does not correspond to one of these two functions. Thus, there exists an $x_0 \in (0, \overline{x}]$ such that $\underline{\delta}(x_0) \in (0, \overline{x})$. Consider replacing this $\underline{\delta}$ with $\underline{\delta}' = 0$, defining a corresponding new evidence structure $\mathcal{E}'$. In Lemma 9 in Appendix A.1, we show that implementable actions and conditionally optimal tariffs depend only on the upper self-enforcing recommendations, summarized by the set $\overline{D}$ and function $\overline{\delta}$. Therefore, it suffices to consider the effect on front-end and back-end costs. For front-end costs, we observe (see the proof of Lemma 6 in Appendix A.5) that

$$\Gamma_f(\mathcal{E}) = \int_X \left[\int_0^{\underline{\delta}_{\mathcal{E}}(x)} g_f(x, z) \, \mathrm{d}z + \int_{\overline{\delta}_{\mathcal{E}}(x)}^{\bar{x}} g_f(x, z) \, \mathrm{d}z \right] \mathrm{d}x \tag{71}$$

Considered as a functional of $\underline{\delta}, \overline{\delta}$, Γ_f is monotone. Moreover, $\underline{\delta}_{\mathcal{E}} \geq \underline{\delta}_{\mathcal{E}'}$. Thus, $\Gamma_f(\mathcal{E}) \geq \Gamma_f(\mathcal{E}')$. Moreover, this inequality is clearly strict if $g_f > 0$, as $\overline{\delta}_{\mathcal{E}} > \overline{\delta}_{\mathcal{E}'}$ for a positive measure of recommendations.

Similarly, for the back-end costs, we observe (see the proof of Lemma 10 in Appendix A.1) that

$$\Gamma_b(\mathcal{E}, Q) = \int_X \left[\int_0^{\underline{\delta}_{\mathcal{E}}(x)} g_b(x, z) \, \mathrm{d}z + \int_{\overline{\delta}_{\mathcal{E}}(x)}^{\bar{x}} g_b(x, z) \, \mathrm{d}z \right] \mathrm{d}Q(x) \tag{72}$$

and an analogous argument applies to show that $\Gamma_b(\mathcal{E}, Q) \geq \Gamma_b(\mathcal{E}', Q)$, for all Q. This suffices to show that the variant evidence structure $\mathcal{E}'$ implements the same outcomes and payments at lower front-end and back-end costs. Completing the proof.

B.2 Proof of Lemma 8

(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 supports ϕ. In particular, observe that (O) implies that $\phi(\theta) \in C(\xi(\theta))$ for all $\theta \in \Theta$. Next define $\hat{T} : X \to \bar{\mathbb{R}}$ as:

$$\hat{T}(x) = \inf_{y \in X} \{T(y) : x \in C(y)\} \tag{73}$$

We next show that ϕ is also supported by $(\phi, \hat{T})$. By (O) of (ϕ, ξ, T), we have $u(\phi(\theta), \theta) \geq u(x, \theta)$ for all $x \in C(\phi(\theta)) \subseteq C(\xi(\theta))$ (by transitivity) and for all $\theta \in \Theta$, yielding (O) of $(\phi, \phi, \hat{T})$. By (IR) of (ϕ, ξ, T) and the definition of $\hat{T}$, we have

$$u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \geq u(\phi(\theta), \theta) - T(\xi(\theta)) \geq 0 \tag{74}$$

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 exist $\theta \in \Theta$ and $y \in X$ such that

$$\max_{x \in C(y)} u(x, \theta) - \hat{T}(y) > u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \tag{75}$$

By the definition of $\hat{T}$, there exists a sequence $z_n \in X$ such that $y \in C(z_n)$ for all n and $T(z_n) \downarrow \hat{T}(y)$. Thus, there exists n large enough such that

$$\begin{aligned} \max_{x \in C(z_n)} u(x, \theta) - T(z_n) &\geq \max_{x \in C(y)} u(x, \theta) - T(z_n) > u(\phi(\theta), \theta) - \hat{T}(\phi(\theta)) \\ &\geq u(\phi(\theta), \theta) - T(\xi(\theta)) = \max_{x \in C(\xi(\theta))} u(x, \theta) - T(\xi(\theta)) \end{aligned} \tag{76}$$

The first inequality follows from $C(y) \subseteq C(z_n)$ since $y \in C(z_n)$. The second strict inequality follows from Equation 75 and the fact that $T(z_n) \downarrow \hat{T}(y)$. The third 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 30 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 $\{\hat{x} \in X : x \in C(\hat{x})\} \subseteq \{\hat{x} \in X : y \in C(\hat{x})\}$, yielding that $\hat{T}(y) \leq \hat{T}(x)$, as desired.

(If) We now establish sufficiency. Suppose that there exists a tariff $T : X \to \bar{\mathbb{R}}$ that is monotone with respect to C and such that Equation 30 holds and $u(\phi(\theta), \theta) - T(\phi(\theta)) \geq 0$ for all $\theta \in \Theta$. We will show that (ϕ, ϕ, T) is implementable. (IR) is immediately satisfied.

Next, we show that (IC) is satisfied. Suppose, toward a contradiction, that it were not. That is, there exist $\theta \in \Theta$, $y \in X$, and $x \in C(y)$ such that

$$u(x,\theta) - T(y) > \max_{\hat{x} \in C(\phi(\theta))} u(\hat{x},\theta) - T(\phi(\theta)) \geq u(\phi(\theta),\theta) - T(\phi(\theta)) \tag{77}$$

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

$$u(x,\theta) - T(y) > u(\phi(\theta),\theta) - T(\phi(\theta)) \geq u(x,\theta) - T(x) \implies T(x) > T(y) \tag{78}$$

where the second inequality uses the fact that $\phi(\theta)$ solves the program in Equation 30. Finally, we show that (O) is satisfied. Toward a contradiction, assume that there exists $\theta \in \Theta$ and $x \in C(\phi(\theta))$ such that $u(x,\theta) > u(\phi(\theta),\theta)$. However, by monotonicity of T in C, we know that $T(\phi(\theta)) \geq T(x)$. Thus, $u(x,\theta) - T(x) > u(\phi(\theta),\theta) - T(\phi(\theta))$, yielding a contradiction to IC, which we just showed. This proves sufficiency.

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

B.3 Proof of Proposition 1

By Corollary 1, there always exists an optimal menu with finite actions, supported by a set of self-enforcing recommendations $\overline{D} = \{x_1, \ldots, x_K\}$ for some $K \in \mathbb{N}$ and such that $x_1 < x_2 < \ldots x_K$.

We next derive the optimal allocation and payments. As shown in the last step of the proof of Lemma 3, for any given $\overline{D}$,

$$\phi^*(\theta) = \begin{cases} \overline{\phi}(\theta) = \min\{x \in \overline{D} : x \geq \phi^P(\theta)\}, & \text{if } H(\overline{\phi}(\theta),\theta) > H(\underline{\phi}(\theta),\theta), \\ \underline{\phi}(\theta) = \max\{x \in \overline{D} : x \leq \phi^P(\theta)\}, & \text{otherwise.} \end{cases} \tag{79}$$

We now specialize this using the structure of $\overline{D}$. As H is strictly single-crossing, $H(x_k,\theta) - H(x_{k-1},\theta) = 0$ has no solution if and only if (i) $H(x_k,0) - H(x_{k-1},0) > 0$ and (ii) $H(x_k,1) - H(x_{k-1},1) < 0$. As H is strictly quasi-concave, if $H(x_k,0) - H(x_{k-1},0) > 0$, then $H(\cdot,0)$ is strictly increasing at x_{k-1}, and therefore at all x_j for $j \leq k-1$. Thus, if $H(x_k,0) - H(x_{k-1},0) > 0$ holds for k, it holds for all $j \leq k$. Define $\underline{k} = \max\{k \in \{1,\ldots,K\} : H(x_k,0) - H(x_{k-1},0) > 0\}$, with the convention that $\underline{k} = 1$ if this set is empty. Similarly, if $H(x_k,1) - H(x_{k-1},1) < 0$, then $H(\cdot,1)$ is strictly decreasing at x_k. Thus, if $H(x_k,1) - H(x_{k-1},1) < 0$ holds for k, it holds for all $j \geq k$. Define $\overline{k} = \min\{k \in \{1,\ldots,K\} : H(x_k,1) - H(x_{k-1},1) < 0\}$, with the convention that $\overline{k} = K$ if this set is empty. As H is strictly single crossing, $\overline{k} > \underline{k}$. We now have that $H(x_k,\theta) - H(x_{k-1},\theta) = 0$ has a solution if and only if $k \in \{\underline{k}+1,\dots,\overline{k}-1\}$ (if $\overline{k} = \underline{k}+1$, then this set is empty). For all $k \geq \overline{k}$, we have that $\hat{\theta}_k = 1$. For all $k \leq \underline{k}$, we have that $\hat{\theta}_k = 0$. For all $k \in \{\underline{k}+1,\dots,\overline{k}-1\}$, we have that $\hat{\theta}_k$ is the unique solution to $H(x_k,\hat{\theta}_k) = H(x_{k-1},\hat{\theta}_k)$. As H is strictly quasi-concave, we know that $\phi^P(\hat{\theta}_k) \in (x_{k-1}, x_k)$, which implies that $\underline{\phi}(\hat{\theta}_k) = x_{k-1}$ and $\overline{\phi}(\hat{\theta}_k) = x_k$. Thus, 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 31 from Lemma 9, we have that:

$$\begin{aligned}
T(x_k) &= u(x_k,\hat{\theta}_k) - \mathbb{I}[k \geq 2] \sum_{j=1}^{k-1} \int_{\hat{\theta}_j}^{\hat{\theta}_{j+1}} u_\theta(x_j, s)\,\mathrm{d}s \\
&= u(x_k,\hat{\theta}_k) - \mathbb{I}[k \geq 2] \sum_{j=1}^{k-1} \left[u(x_j,\hat{\theta}_{j+1}) - u(x_j,\hat{\theta}_j)\right] \\
&= u(x_1,0) + \mathbb{I}[k \geq 2] \sum_{j=2}^{k} \left[u(x_j,\hat{\theta}_j) - u(x_{j-1},\hat{\theta}_j)\right]
\end{aligned} \tag{80}$$

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

We next derive the necessary first-order condition. We first observe that, given a set of self-enforcing recommendations $\{x_1,\dots,x_K\}$, the principal's payoff inclusive of back-end costs can be written as

$$\mathcal{H}(\{x_1,\dots,x_K\}) = \sum_{k=1}^{K} \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} H(x_k,\theta)\,\mathrm{d}F(\theta) \tag{81}$$

and, similarly, the front-end cost can be written as

$$\begin{aligned}
\tilde{\Gamma}_f(\{x_1,\dots,x_K\}) &= \int_X \int_{\overline{\delta}_{\overline{D}}(x)}^{\bar{x}} g_f(x,z)\,\mathrm{d}z\,\mathrm{d}x = \sum_{k=1}^{K-1} \int_{x_k}^{x_{k+1}} \int_{x_{k+1}}^{\bar{x}} g_f(x,z)\,\mathrm{d}z\,\mathrm{d}x \\
&= \sum_{k=1}^{K-1} \int_{x_k}^{x_{k+1}} \left(G_f(\bar{x},x) - G_f(x_{k+1},x)\right)\mathrm{d}x
\end{aligned} \tag{82}$$

A necessary condition for optimality is that payoffs cannot be locally improved by perturbing any action x_k, for $k \in \{2,\dots,K-1\}$. That is,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathcal{H}(\{x_1,\dots,x_k+\varepsilon,\dots,x_K\})|_{\varepsilon=0} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\tilde{\Gamma}_f(\{x_1,\dots,x_k+\varepsilon,\dots,x_K\})|_{\varepsilon=0} \tag{83}$$

where we observe that both derivatives are well-defined. The left-hand-side of this is

$$\begin{aligned}
&\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} H_x(x_k^*, \theta)\,\mathrm{d}F(\theta) + \frac{\partial}{\partial x_k^*}\hat{\theta}_k \left(H(x_k^*, \hat{\theta}_k) - H(x_{k-1}^*, \hat{\theta}_k)\right) f(\hat{\theta}_k) \\
&\quad + \frac{\partial}{\partial x_k^*}\hat{\theta}_{k+1} \left(H(x_{k+1}^*, \hat{\theta}_{k+1}) - H(x_k^*, \hat{\theta}_{k+1})\right) f(\hat{\theta}_{k+1}) \\
&= \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} H_x(x_k^*, \theta)\,\mathrm{d}F(\theta)
\end{aligned} \tag{84}$$

where, in the second equality, we use the fact that $H(x_k^*, \hat{\theta}_k) = H(x_{k-1}^*, \hat{\theta}_k)$ by definition. Moreover, using the definition $H(x, \theta) = J(x, \theta) - \int_x^{\overline{x}} g_b(x, z)\,\mathrm{d}z$, we observe that:

$$H_x(x, \theta) = J_x(x, \theta) - \int_x^{\bar{x}} g_{b,x}(x, z)\,\mathrm{d}z + g_b(x, x) \tag{85}$$

where the latter two terms do not depend on θ. Thus,

$$\begin{aligned}
&\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} H_x(x_k^*, \theta)\,\mathrm{d}F(\theta) = \\
&\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} J_x(x_k^*, \theta)\,\mathrm{d}F(\theta) + \left(F(\hat{\theta}_{k+1}) - F(\hat{\theta}_k)\right)\left(g_b(x_k, x_k) - \int_{x_k}^{\bar{x}} g_{b,x}(x_k, z)\,\mathrm{d}z\right)
\end{aligned} \tag{86}$$

The right-hand-side of Equation 83 can be written as

$$\begin{aligned}
&(G_f(\bar{x}, x_k) - G_f(x_k, x_k)) - \int_{x_{k-1}}^{x_k} g_f(x, x_k)\,\mathrm{d}x - (G_f(\bar{x}, x_k) - G_f(x_{k+1}, x_k)) \\
&= G_f(x_{k+1}, x_k) - G_f(x_k, x_k) - \int_{x_{k-1}}^{x_k} g_f(x, x_k)\,\mathrm{d}x
\end{aligned} \tag{87}$$

We obtain the desired Equation 10 by plugging Equations 86 and 87 into Equation 83.

B.4 Proof of Proposition 2

We split the argument into three parts. We first map the problem to one that satisfies the general assumptions of Section 2. We next calculate the optimal contract in the transformed problem. We finally map the allocation and tariff back to the original problem.

Step 1: Transformation of the Problem. We define $x = 1 - e \in X = [0, 1]$ as the agent's shirking and $\theta = 1 - \vartheta \sim F = U[0, 1]$ as their unproductivity. We define transformed preferences for the agent, which satisfy all of our maintained assumptions:

$$
\begin{aligned}
u(x,\theta) &= \tilde{u}(1-x, 1-\theta) - \tilde{u}(1, 1-\theta) \\
&= -\alpha\theta(1-x) - \beta\frac{(1-x)^2}{2} - \left(\alpha\theta - \frac{\beta}{2}\right) \\
&= (\alpha\theta + \beta)x - \beta\frac{x^2}{2}
\end{aligned}
\tag{88}
$$

We define the transformed payoff for the principal,

$$
\pi(x) = \tilde{\pi}(1-x) - \tilde{\pi}(1) = \eta(1-x) - \eta = -\eta x \tag{89}
$$

We next re-write the front-end and back-end costs in terms of shirking. We first observe that an evidence structure over effort induces an evidence structure over shirking defined by $\mathcal{E}^x : X \to [0,1]$, where $X = [0,1]$ and, for each $x \in [0,1]$, $\mathcal{E}^x(x) = \mathcal{E}(1-x)$. Moreover, $\overline{\delta}_{\mathcal{E}^x}(x) = 1 - \underline{\delta}_{\mathcal{E}}(1-x)$ and $\underline{\delta}_{\mathcal{E}^x}(x) = 1 - \overline{\delta}_{\mathcal{E}}(1-x)$, for each $x \in [0,1]$. Thus, in some abuse of notation, we can write

$$
\Gamma_f(\mathcal{E}) = \tilde{\Gamma}_f(\mathcal{E}^x) = \kappa_f \int_E \left(\underline{\delta}_{\mathcal{E}^x}(x) + (1 - \overline{\delta}_{\mathcal{E}^x}(x))\right) \mathrm{d}x \tag{90}
$$

An analogous argument applies to the back-end cost.

The (e, ϑ) problem satisfies the assumptions of Lemma 10 in the case of monotone *decreasing* preferences. In this case, the result implies that the principal problem given a set of self-enforcing efforts $\underline{D} \subseteq E$ is

$$
\begin{aligned}
\tilde{\mathcal{H}}(\underline{D}) := \max_{\tilde{\phi}} \quad & \int_0^1 \tilde{H}(e,\vartheta)\,\mathrm{d}\tilde{F}(\vartheta) \\
\text{s.t.} \quad & \tilde{\phi}(\vartheta') \geq \tilde{\phi}(\vartheta), \tilde{\phi}(\vartheta) \in \underline{D}, \quad \vartheta, \vartheta' \in [0,1] : \vartheta' \geq \vartheta
\end{aligned}
\tag{91}
$$

where the modified virtual surplus function is

$$
\begin{aligned}
\tilde{H}(e,\vartheta) &= \tilde{u}(e,\vartheta) + \tilde{\pi}(e) - \frac{1-\tilde{F}(\vartheta)}{\tilde{f}(\vartheta)}\tilde{u}_\vartheta(e,\vartheta) - \kappa_b e \\
&= -\alpha(1-\vartheta)e + \eta e - \beta\frac{e^2}{2} - (1-\vartheta)\alpha e - \kappa_b e \\
&= -\alpha\theta(1-x) + \eta(1-x) - \beta\frac{(1-x)^2}{2} - \theta\alpha(1-x) - \kappa_b(1-x) \\
&= (2\alpha\theta + \beta - \eta + \kappa_b)x - \beta\frac{x^2}{2} + \left(\eta - \kappa_b - \frac{\beta}{2} - 2a\theta\right)
\end{aligned}
\tag{92}
$$

Using this substitution, we observe the following mapping to the (x, θ) problem:

$$\tilde{\mathcal{H}}(\underline{D}) - \left(\eta - a - \kappa_b - \frac{\beta}{2}\right) = \mathcal{H}(\overline{D}) = \max_{\phi} \int_0^1 \left((2\alpha\theta + \beta - \eta + \kappa_b)\phi(\theta) - \beta\frac{\phi(\theta)^2}{2}\right) \mathrm{d}F(\theta)$$
$$\text{s.t.} \quad \phi(\theta') \geq \phi(\theta), \phi(\theta) \in \overline{D}, \quad \theta, \theta' \in [0,1] : \theta' \geq \theta \tag{93}$$

where $\overline{D} = \{(1-e) \in X : e \in \underline{D}\}$. Combining this with the earlier observation about front-end costs (Equation 90), we obtain that the original problem is equivalent to solving the transformed problem with modified surplus $H(x, \theta) = (2\alpha\theta + \beta - \eta + \kappa_b)x - \beta\frac{x^2}{2}$. Moreover, this function is strictly quasi-concave.

Corollary 1 implies that any optimal contractibility correspondence can be represented by a coarse set of self-enforcing shirking recommendations, $\overline{D} = \{x_1, \ldots, x_K\}$. In the transformed problem, these map to a set of self-enforcing effort recommendations $\{1-x_1, \ldots, 1-x_K\}$. To apply Proposition 1, we must also establish that (without loss of optimality) $x_1 = 0$ and $x_K = \overline{x} = 1$. *Excludability* guarantees that $e_1 = 0$ (zero effort, or full shirking) is perfectly contractible. An additional simple argument is required to establish that (without loss of optimality) we can restrict attention to the case in which $e_K = 1$ (full effort or zero shirking) is also contractible. Imagine it were not, and optimal contractibility were represented by an evidence structure $\mathcal{E}$. Then, construct a variant evidence structure $\mathcal{E}'$ that perturbs $\mathcal{E}$ such that $e = 1$ is self-enforcing. Since the evidence space $\Omega = [0,1]$ has the cardinality of the continuum, this is always possible by having $e = 1$ generate a unique (singleton) piece of evidence. We observe that $\Gamma_f(\mathcal{E}') = \Gamma_f(\mathcal{E})$ and $\Gamma_b(\mathcal{E}', Q) = \Gamma_b(\mathcal{E}, Q)$ for all Q, as both costs, when re-expressed with $(\underline{\delta}, \overline{\delta})$ as their arguments, are L_1-continuous. But under $\mathcal{E}'$ there are weakly more more self-enforcing recommendations. Thus, $\mathcal{E}'$ obtains higher value. Hence, it is without loss of optimal to restrict attention to contractibility containing $e = 1$ $(x = 0)$ as a self-enforcing recommendation. Together, this argument shows that we can apply Proposition 1 to the transformed problem.

Step 2: Optimal Contract in the Transformed Problem. We next apply Proposition 1 to the transformed problem. The modified virtual surplus function in this setting is given by Equation 25.

We first solve for the candidate optimal contract that solves the variational first-order condition in Proposition 1 for each K. This first-order condition for $k \in \{2, K-1\}$ is

$$\int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} (2\alpha\theta + \beta(1 - x_k) - \eta + \kappa_b) \, \mathrm{d}\theta - \kappa_f(-2x_k + x_{k-1} + x_{k+1}) = 0 \tag{94}$$

where $\hat{\theta}_k = \frac{\beta}{4\alpha}(x_k + x_{k-1}) + \frac{\eta-\kappa_b-\beta}{2\alpha}$. Calculating the integral and evaluating at $\hat{\theta}_k$, we write

$$\begin{aligned}\kappa_f(-2x_k + x_{k-1} + x_{k+1}) &= \left[\alpha\theta^2 + (\beta - \eta + \kappa_b - \beta x_k)\theta\right]_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \\ &= \frac{\beta^2}{16\alpha}(x_{k+1} - x_{k-1})(x_{k+1} + x_{k-1} - 2x_k)\end{aligned} \tag{95}$$

This condition can be re-arranged to obtain

$$(x_{k+1} + x_{k-1} - 2x_k)\left[\frac{\beta^2}{16\alpha}(x_{k+1} - x_{k-1}) - \kappa_f\right] = 0 \tag{96}$$

This equation has two solutions,

$$x_k = \frac{x_{k+1} + x_{k-1}}{2}, \qquad\qquad x_{k+1} = x_{k-1} + \Delta \tag{97}$$

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

We first consider case 1, the "uniform grid." From the boundary conditions, $x_1 = 0$ and $x_K = 1$. Thus, $x_k = \frac{x_{k+1}+x_{k-1}}{2} \implies x_k = \frac{k-1}{K-1}$. We can verify that this is a local maximum by checking the Hessian is negative definite at this solution. We calculate that:

$$\begin{aligned}\frac{\partial^2\mathcal{H}}{\partial x_k^2} &= \text{Hess}^{\mathcal{H}}_{k-1,k-1} = -\frac{\beta^2}{4\alpha(K-1)} + 2\kappa_f =: \Lambda \\ \frac{\partial^2\mathcal{H}}{\partial x_k \partial x_{k+1}} &= \text{Hess}^{\mathcal{H}}_{k,k-1} = \text{Hess}^{\mathcal{H}}_{k-1,k} = \frac{\beta^2}{8\alpha(K-1)} - \kappa_f = -\frac{1}{2}\Lambda\end{aligned} \tag{98}$$

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

$$\lambda_k = \Lambda\left(1 + \cos\left(\frac{k-1}{K}\pi\right)\right) \tag{99}$$

for $k \in \{2, \dots, K-1\}$. As $\cos\left(\frac{k-1}{K}\pi\right) > -1$ for all such k, we have that $\text{sgn}(\lambda_k) = \text{sgn}(\Lambda)$. Thus, the Hessian is negative definite if and only if $K < \bar{K} = 1 + \frac{\beta^2}{8\alpha\kappa_f}$. We will later verify that this holds whenever K is set optimally, confirming the optimality of the uniform grid.

We next consider case 2, the alternating grid. In this case, even points form a uniform grid with spacing $\Delta = \frac{16\alpha\kappa_f}{\beta^2}$ and the odd points form a uniform grid with spacing $\Delta = \frac{16\alpha\kappa_f}{\beta^2}$. When K is odd, given the boundary conditions that $x_1 = 0$ and $x_K = 1$, we have that this is possible only when $K = 2 + \frac{2}{\Delta}$, which is itself only possible when $\frac{\beta^2}{8\alpha\kappa_f}$ is an odd integer. When K is even, the solution must be $x_k = \frac{k-1}{2}\Delta$ for k odd, and $x_k = 1 - \frac{K-k}{2}\Delta$ for k even.

This is possible for any even $K < 2 + \frac{2}{\Delta}$.

We next show that the alternating grid (for odd or even K) is *not* a local maximum of the objective function. For a local maximum, a necessary condition is that the Hessian is negative semidefinite. We will show the existence of a vector $x \in \mathbb{R}^{K-2}$ such that $v \neq 0$ an $v'\text{Hess}^{\mathcal{H}}v > 0$, which implies that $\text{Hess}^{\mathcal{H}}$ is not negative semidefinite. To do this, we first calculate and simplify the second derivatives:

$$\begin{aligned} \frac{\partial^2 \mathcal{H}}{\partial x_k^2} &= \text{Hess}^{\mathcal{H}}_{k-1,k-1} = -\frac{\beta^2}{8\alpha}\Delta + 2\kappa_f = 0 \\ \frac{\partial^2 \mathcal{H}}{\partial x_k \partial x_{k+1}} &= \text{Hess}^{\mathcal{H}}_{k,k-1} = \text{Hess}^{\mathcal{H}}_{k-1,k} = \frac{\beta^2}{8\alpha}(x_{k+1} - x_k) - \kappa_f \end{aligned} \tag{100}$$

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

$$v_k'\text{Hess}^{\mathcal{H}}v_k = 2\left(\kappa_f - \frac{\beta^2}{8\alpha}(x_{k+1} - x_k)\right) \tag{101}$$

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 $v_k'\text{Hess}^{\mathcal{H}}v_k > 2\left(\kappa_f - \frac{\Delta\beta^2}{16\alpha}\right) > 0$ and, as desired, we have shown that the Hessian is not negative definite. Next, we consider the case in which $K = 4$. In this case, we take two candidate vectors. The first is $u = z_1 + z_2$, and we observe $u'\text{Hess}^{\mathcal{H}}u = 2(\frac{\beta^2}{8\alpha}(x_3 - x_2) - \kappa_f)$. The second is $v_1 = z_1 - z_2$, and we observe $v_1'\text{Hess}^{\mathcal{H}}v_1 = 2(\kappa_f - \frac{\beta^2}{8\alpha}(x_3 - x_2)) = -u'\text{Hess}^{\mathcal{H}}u$. We finally note that, if $u'\text{Hess}^{\mathcal{H}}u = v_1'\text{Hess}^{\mathcal{H}}v_1 = 0$, then $x_3 - x_2 = \frac{8\alpha\kappa_f}{\beta^2} = \frac{\Delta}{2}$ and the alternating grid collapses to the uniform grid.

We next derive the profit and optimal tariff evaluated at the uniform-grid solution:

Lemma 11. *The value to the monopolist of a K-item contract supported on a uniform grid can be written as $V(K) = \hat{\Pi}(K) - \hat{\Gamma}_f(K)$ where*

$$\hat{\Pi}(K) = \frac{\beta - \eta + \kappa_b + 2\alpha}{4\alpha}(2\alpha - \eta + \kappa_b) + \frac{\beta^2}{48\alpha}\frac{(2K-3)(2K-1)}{(K-1)^2} \quad \text{and} \quad \hat{\Gamma}_f(K) = \frac{\kappa_f}{2}\frac{K-2}{K-1} \tag{102}$$

Moreover, the optimal allocation is supported by the tariff

$$T^*(x_k) = \frac{1}{2}\frac{k-1}{K^*-1}\left(\beta + \eta - \kappa_b - \frac{b}{2}\frac{k-1}{K^*-1}\right) \tag{103}$$

Proof. We observe that the principal's payoff, inclusive of back-end costs (but exclusive of front-end costs), can be written as

$$\mathcal{H}\left((x_k)_{k=1}^K\right) = \sum_{k=1}^K \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} H(x_k, \theta)\,\mathrm{d}F(\theta) \tag{104}$$

Using this, we calculate

$$\begin{aligned}\hat{\Pi}(K) &= \sum_{k=1}^K \int_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \left((2\alpha\theta + \beta - \eta + \kappa_b)x_k - \beta\frac{x_k^2}{2}\right)\mathrm{d}\theta = \sum_{k=1}^K \left[\alpha x_k \theta^2 - x_k\left(\eta - \kappa_b - \beta + \frac{\beta}{2}x_k\right)\theta\right]_{\hat{\theta}_k}^{\hat{\theta}_{k+1}} \\ &= \frac{\beta}{2\alpha(K-1)}\sum_{k=2}^{K-1}\left(\alpha x_k(\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k\left(\eta - \kappa_b + \frac{\beta}{2}(x_k - 1)\right)\right) + (1 - \hat{\theta}_K)\left(\alpha(1 + \hat{\theta}_K) + \frac{\beta}{2} - \eta + \kappa_b\right)\end{aligned} \tag{105}$$

where, in the second line, we use that $\hat{\theta}_{k+1} - \hat{\theta}_k = \frac{\alpha}{2\beta(K-1)}$ for $k < K$ and that $\hat{\theta}_{K+1} = 1$ and $x_K = 1$. We simplify the summation term as

$$\sum_{k=2}^{K-1}\left(\alpha x_k(\hat{\theta}_{k+1} + \hat{\theta}_k) - x_k\left(\alpha + \eta - \kappa_b + \frac{\beta}{2}(x_k - 1)\right)\right) = \frac{\beta}{12(K-1)}(K-2)(2K-3) \tag{106}$$

where we use that $\hat{\theta}_k + \hat{\theta}_{k+1} = 1 - \frac{\beta - \eta + \kappa_b}{\alpha} + \frac{\beta}{\alpha}x_k$. To simplify the second term, we observe that $\hat{\theta}_K = \left(\frac{\eta - \kappa_b - \beta}{2\alpha}\right) + \frac{\beta}{4\alpha}\left(1 + \frac{K-2}{K-1}\right)$ and therefore

$$(1 - \hat{\theta}_K)\left(\alpha(\hat{\theta}_K + 1) + \frac{\beta}{2} - \eta + \kappa_b\right) = \frac{\beta - \eta + \kappa_b + 2\alpha}{4\alpha}(2\alpha - \eta + \kappa_b) + \frac{\beta^2}{16\alpha(K-1)^2}(2K-3) \tag{107}$$

Putting this together and simplifying, we write:

$$\hat{\Pi}(K) = \frac{\beta - \eta + \kappa_b + 2\alpha}{4\alpha}(2\alpha - \eta + \kappa_b) + \frac{\beta^2}{48\alpha}\frac{(2K-3)(2K-1)}{(K-1)^2} \tag{108}$$

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

$$\hat{\Gamma}(K) = \kappa_f \sum_{k=1}^{K-1}\frac{1}{K-1}\frac{k-1}{K-1} = \frac{\kappa_f}{(K-1)^2}\sum_{k=1}^{K-1}(k-1) = \frac{\kappa_f}{2}\frac{K-2}{K-1} \tag{109}$$

We finally compute the tariff. Starting from Equation 11,

$$\begin{aligned}T^*(x_k) &= \mathbb{I}[k \geq 2] \sum_{j=2}^{K} \left[u(x_j, \hat{\theta}_j) - u(x_{j-1}, \hat{\theta}_j)\right] = \mathbb{I}[k \geq 2] \sum_{j=2}^{k} \left[(\alpha\hat{\theta}_j + \beta)(x_j - x_{j-1}) + \frac{\beta}{2}(x_{j-1}^2 - x_j^2)\right] \\ &= \frac{1}{2}\frac{k-1}{K^*-1}\left(\beta + \eta - \kappa_b - \frac{\beta}{2}\frac{k-1}{K^*-1}\right)\end{aligned} \tag{110}$$

where we substitute in the expressions for $\hat{\theta}_j$ and x_j, simplify at each step, and evaluate at $K = K^*$. □

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

$$V'(K) = \frac{\beta^2}{24\alpha(K-1)^3} - \frac{\kappa_f}{2(K-1)^2} \tag{111}$$

We observe that $V'(K) > 0$ if and only if $K < \tilde{K} := \frac{\beta^2}{12\alpha\kappa_f} + 1$. We now prove that $|K^* - \tilde{K}| < 1$. If $K^* - \tilde{K} > 1$, then we know that $V(K^*-1) > V(K^*)$ as $V' < 0$ for all $K^* - 1 < K < K^*$; this contradicts optimality. Similarly, if $\tilde{K} - K^* > 1$, we know that $V(K^*+1) > V(K^*)$ as as $V' > 0$ for all $K^* < K < K^* + 1$; this contradicts optimality. Recall that we needed to check if the Hessian was negative definite. This is true so long as $K^* < \bar{K}$. As $\bar{K} = \frac{4}{3}\tilde{K}$, this holds whenever $\tilde{K} \geq 3$. It remains to check when $\tilde{K} \in (2,3)$ and $K^* = 3$. Direct calculation shows that indifference between $K = 2$ and $K = 3$ occurs when $\kappa_f = \frac{\beta^2}{16\alpha}$. At this point, $\tilde{K} = 7/3$. Thus, whenever $K^* > 2$ is strictly optimal (which is when $\kappa_f < \frac{\beta^2}{16\alpha}$), we have that $K^* < \bar{K}$. The comparative statics follow from standard monotone comparative statics arguments, after the observations that $V_{K\alpha} < 0$, $V_{K\beta} > 0$, $V_{K\eta} = 0$, and $V_{K\kappa_b} = 0$. Finally, $V(3) - V(2) = \frac{1}{4}\left(\frac{\beta^2}{16\alpha} - \kappa_f\right)$. Thus, whenever $\kappa_f < \frac{\beta^2}{16\alpha}$ we have that $V(3) > V(2)$, which implies that $K^* \geq 3$.

Step 3: Mapping Back to the Original Problem. Since $e = 1 - x$, we observe that the optimal contract can be supported on $e_k = 1 - x_{K-k} = \frac{k-1}{K^*-1}$ with K^* self-enforcing recommendations. We next observe that, in the original problem, the IR constraint binds for type $\vartheta = 0$, who always (for any K^*) takes action $x = 1$ or $e = 0$ and receives direct payoff $\tilde{u}(0,0) = 0$. Therefore, $\tilde{T}(e_k) = T(x_{K^*-k}) - C$ where C solves $T(1) - C = 0$, and hence $C = \frac{1}{2}\left(\frac{\beta}{2} + \eta - \kappa_b\right)$. Therefore, we calculate

$$\begin{aligned}\tilde{T}(e_k) &= \frac{1}{2}\left(1 - \frac{k-1}{K^*-1}\right)\left(\beta + \eta - \kappa_b - \frac{b}{2}\left(1 - \frac{k-1}{K^*-1}\right)\right) - \frac{1}{2}\left(\frac{\beta}{2} + \eta - \kappa_b\right) \\ &= -\frac{1}{2}\frac{k-1}{K^*-1}\left(\eta - \kappa_b + \frac{\beta}{2}\frac{k-1}{K^*-1}\right)\end{aligned} \tag{112}$$

We finally translate the transfers into wages by reversing the sign of the payments: $w(e_k) = -\tilde{T}(e_k)$ for all k. That is, a positive wage corresponds to a negative transfer from the agent (worker) to the principal (firm). This completes the proof.

B.5 Proof of Corollary 2

The first part of the statement follows immediately from the characterization of the optimal menu in Proposition 2.

We now prove the second part. By assumption on $(\alpha, \beta, \eta, \kappa_b, \kappa_f)$, there exists a maximal neighborhood $\mathcal{B}$ of α such that, for all $\alpha' \in \mathcal{B}$, there exists a unique optimal K^* at $(\alpha, \beta, \eta, \kappa_b, \kappa_f)$. Let $\mathcal{A}$ be the set of $\alpha' \in \mathcal{B}$ such that $K^*(\alpha') = K^*(\alpha)$, where $K^*(\alpha')$ is the unique optimal K^* at parameter vector $(\alpha', \beta, \eta, \kappa_b, \kappa_f)$. We observe from Lemma 11 that $V = \hat{\Pi} - \hat{\Gamma}$ is submodular in α. This implies that K^* is monotone in α and therefore that $\mathcal{A} = \mathcal{B}$. By Proposition 2, we have that $\{w^*(e_k)\}_{k=1}^{K^*}$ is invariant to α conditional on K^*. Moreover, again by Proposition 2, $\{w^*(e_k)\}_{k=1}^{K^*}$ is different from $\{w^*(e_k)\}_{k=1}^{K}$ for any $K \neq K^*$ and, as $\mathcal{B}$ is the maximal neighborhood of α such that K^* is unique, whenever $\alpha' \notin \mathcal{B}$, there exists an optimal $K^{**} \neq K^*$.

C Representations of Evidence and Their Properties

In this appendix we derive several equivalent representations of evidence and establish their topological properties.

C.1 Representing Evidence

In this appendix, we prove Lemma 1. It follows as a consequence of the following, more general equivalence that we exploit in some of our proofs.

Lemma 12. *The following four properties of a contractibility correspondence $C : X \rightrightarrows X$ are equivalent:*

1. *C is a regular contractibility correspondence, that is, $C = C_{\mathcal{E}}$ for some regular evidence structure $\mathcal{E} : X \rightrightarrows \Omega$ (see definition of $C_{\mathcal{E}}$ in Equation 13).*
2. *C satisfies (i) Reflexivity: for every $y \in X$, $y \in C(y)$; (ii) Excludability: $C(0) = \{0\}$; (iii) Transitivity: for every $x, y, z \in X$, if $x \in C(y)$ and $z \in C(x)$, then $z \in C(y)$; (iv) 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; (v) Continuity: C is closed valued and lower hemicontinuous.*

3. *We have $C(y) = [\underline{\delta}(y), \overline{\delta}(y)]$ for all $y \in X$, where $\underline{\delta} : X \to X$ is an upper semi-continuous increasing function and $\overline{\delta} : X \to X$ is a lower semi-continuous increasing function such that: (i) $y \in [\underline{\delta}(y), \overline{\delta}(y)]$, (ii) $\underline{\delta}(x) = \underline{\delta}(y)$ for all $x \in [\underline{\delta}(y), y)$, (iii) $\overline{\delta}(x) = \overline{\delta}(y)$ for all $x \in (y, \overline{\delta}(y)]$, and (iv) $\overline{\delta}(0) = 0$.*
4. *There exist two closed sets $\underline{D} \subseteq X$ and $\overline{D} \subseteq X$ such that $0 \in \underline{D}$, $0, \overline{x} \in \overline{D}$ and for all $y \in X$:*
$$C(y) = \left[\max_{z \leq y : z \in \underline{D}} z, \min_{z \geq y : z \in \overline{D}} z \right] \tag{113}$$

Moreover, given C, $\underline{D}$ and $\overline{D}$ are unique and given by $\underline{D} = \underline{\delta}(X)$ and $\overline{D} = \overline{\delta}(X)$.

Proof. To prove this, we show that: 1 implies 2, 2 implies 3, 3 implies 4, and 4 implies 1.

1 implies 2. By Equation 13, it is immediate that any contractibility correspondence C that is induced by some evidentiary correspondence $\mathcal{E}$ is transitive, reflexive, closed-valued, and lower hemicontinuous. Observe that definitive evidence of exclusion implies that $C(0) = \{0\}$, yielding excludability of C. It remains only to show monotonicity of C. Fix $y' \geq y$, $x \in C(y)$, and $x' \in C(y')$. If $x' \geq x$, then $\max\{x, x'\} = x' \in C(y')$ and $\min\{x, x'\} = x \in C(y)$. Suppose now that $x' < x$. We now show that $\mathcal{E}(x) \subseteq \mathcal{E}(y')$ and $\mathcal{E}(x') \subseteq \mathcal{E}(y)$, which yields monotonicity of C by the fact that these claims imply that $\max\{x, x'\} = x \in C(y')$ and $\min\{x, x'\} = x' \in C(y)$. By definition we have that $\mathcal{E}(x) \subseteq \mathcal{E}(y)$, $\mathcal{E}(x') \subseteq \mathcal{E}(y')$, $\mathcal{E}(y') \geq_{SSO} \mathcal{E}(y)$, and $\mathcal{E}(x) \geq_{SSO} \mathcal{E}(x')$. Fix $\omega \in \mathcal{E}(x)$ and $\omega' \in \mathcal{E}(x')$. If $\omega \leq \omega'$, as $\mathcal{E}(x) \geq_{SSO} \mathcal{E}(x')$, we have that $\omega = \min\{\omega, \omega'\} \in \mathcal{E}(x')$. As $\mathcal{E}(x') \subseteq \mathcal{E}(y')$, this implies that $\omega \in \mathcal{E}(y')$. If $\omega > \omega'$, as $\omega \in \mathcal{E}(x) \subseteq \mathcal{E}(y)$ and $\omega' \in \mathcal{E}(x') \subseteq \mathcal{E}(y')$, we know that $\omega = \max\{\omega, \omega'\} \in \mathcal{E}(y')$ by the fact that $\mathcal{E}(y') \geq_{SSO} \mathcal{E}(y)$. These steps imply that $\mathcal{E}(x) \subseteq \mathcal{E}(y')$. Now fix $\omega \in \mathcal{E}(x')$ and $\omega' \in \mathcal{E}(x)$. If $\omega > \omega'$, we have that $\omega = \max\{\omega, \omega'\} \in \mathcal{E}(x)$ as $\mathcal{E}(x) \geq_{SSO} \mathcal{E}(x')$. As $\mathcal{E}(x) \subseteq \mathcal{E}(y)$, this implies that $\omega \in \mathcal{E}(y)$. If $\omega \leq \omega'$, as $\omega \in \mathcal{E}(x') \subseteq \mathcal{E}(y')$ and $\omega' \in \mathcal{E}(x) \subset \mathcal{E}(y)$, we have that $\omega = \min\{\omega, \omega'\} \in \mathcal{E}(y)$ as $\mathcal{E}(y') \geq_{SSO} \mathcal{E}(y)$. These steps establish that $\mathcal{E}(x') \subseteq \mathcal{E}(y)$.

2 implies 3. Let $C : X \rightrightarrows X$ be a regular contractibility correspondence and define $\underline{\delta}(y) = \min C(y)$ and $\overline{\delta}(y) = \max C(y)$ for all $y \in X$. By the fact that C is closed-valued, $\overline{\delta}$ and $\underline{\delta}$ exist. By monotonicity, we have that $\underline{\delta}$ and $\overline{\delta}$ are increasing functions. By reflexivity, we know that $y \geq \underline{\delta}(y)$ and $y \leq \overline{\delta}(y)$ for all y. Moreover, by Lemma 17.29 in Aliprantis and Border (2006), $\overline{\delta}$ is lower semi-continuous and $\underline{\delta}$ is upper semi-continuous.

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

We next show parts (ii), (iii), and (iv). Fix $x, y \in X$ and assume that $x \in [\underline{\delta}(y), \overline{\delta}(y))$, which implies $x \in C(y)$. We start with part (ii), and mirror the argument for part (iii). 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 transitivity. For part (iii), consider the same scenario but reversed. Suppose $x > y$. As C is monotone, we know that $\overline{\delta}(x) \geq \overline{\delta}(y)$. Imagine this held with a strict inequality. Then there would exist $z \in (\overline{\delta}(y), \overline{\delta}(x))$ such that $z \in C(x)$ and $z \notin C(y)$, while $x \in C(y)$. This violates transitivity. It is immediate that $\overline{\delta}(0) = 0$ by excludability as $C(0) = \{0\}$.

3 implies 4. We first observe the following for $\underline{\delta}, \overline{\delta}$ with the relevant properties: for all $\underline{z} \in \underline{\delta}(X)$ and $\overline{z} \in \overline{\delta}(X)$, it holds $\underline{\delta}(\underline{z}) = \underline{z}$ and $\overline{\delta}(\overline{z}) = \overline{z}$. To show this, let $\underline{z} = \underline{\delta}(x)$ for some $x \in X$. We have that $\underline{z} \in [\underline{\delta}(x), x]$. If $\underline{z} = x$, then we have that $\underline{\delta}(\underline{z}) = \underline{\delta}(x) = \underline{z}$. If $\underline{z} < x$, given property (ii), we must have $\underline{\delta}(\underline{z}) = \underline{\delta}(x) = \underline{z}$. The proof for $\overline{z} \in \overline{\delta}(X)$ is symmetric, using property (iii).

Define $\underline{D} = \underline{\delta}(X)$ and $\overline{D} = \overline{\delta}(X)$. First, observe that

$$\max_{z \leq x : z \in \underline{D}} z = \max_{z \leq x : z \in \underline{\delta}(X)} z \geq \underline{\delta}(x) \tag{114}$$

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 the earlier argument we have that $x = \underline{\delta}(x) < \underline{z}$, yielding a contradiction. If instead $\underline{z} < x$, then by the earlier argument and the property (ii) 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 $\min_{z \geq x : z \in \overline{D}} z = \overline{\delta}(x)$. Next, observe that necessarily we have $\underline{\delta}(0) = 0$, $\overline{\delta}(\overline{x}) = \overline{x}$, and $\overline{\delta}(0) = 0$ proving that $0 \in \underline{D}$ and $0, \overline{x} \in \overline{D}$. Finally, we need to show that $\underline{D}$ and $\overline{D}$ are closed. Take a sequence $z_n \in \underline{D}$ such that $z_n \to z$. Given that X is closed, we have that $z \in X$ and therefore $\underline{\delta}(z) \leq z$. Given that every z_n is in $\underline{D}$, the property defined above implies that $\underline{\delta}(z_n) = z_n$ for all n. Given that $\underline{\delta}$ is upper semi-continuous, it follows that $z = \lim_{n \to \infty} z_n = \lim_{n \to \infty} \underline{\delta}(z_n) \leq \underline{\delta}(z)$. This implies that $z = \underline{\delta}(z)$ (as $z \geq \underline{\delta}(z)$) and therefore that $z \in \underline{D}$. This shows that $\underline{D}$ is closed. A symmetric argument shows that $\overline{D}$ is closed.

4 implies 1. We show this by construction. For the argument, we take $\Omega = X$; as Ω has the cardinality of the continuum, this is without loss of generality, as the construction below can be composed with an isomorphism from X to Ω. We define

$$\mathcal{E}(x) = \left[\max_{z \leq x : z \in \underline{D}} z, \min_{z \geq x : z \in \overline{D}} z\right] =: \left[\underline{\delta}(x), \overline{\delta}(x)\right] \tag{115}$$

where, for convenience, we define $\underline{\delta}, \overline{\delta}$ in the second equality. We observe immediately that these functions have the three properties from part 3 of the statement. It is immediate to see that both these functions are monotone increasing, such that $\underline{\delta}(x) \leq x \leq \overline{\delta}(x)$, and respectively upper semi-continuous and lower semi-continuous by Lemma 17.30 in Aliprantis and Border (2006). To see the remaining properties, observe that the correspondences $x \rightrightarrows \{z \in \underline{D} : z \leq x\}$ and $x \rightrightarrows \left\{z \in \overline{D} : z \geq x\right\}$ are both upper hemicontinuous. Next, assume that $y \in [\underline{\delta}(x), x)$ and let $z = \underline{\delta}(x)$. We have $\underline{\delta}(y) \leq z$ by monotonicity. Moreover, by assumption $z \leq y$ and $z \in \underline{D}$, so that $z \leq \underline{\delta}(y)$ by definition. We then must have $z = \underline{\delta}(y)$. Symmetrically, assume that $y \in (x, \overline{\delta}(x)]$ and let $z = \overline{\delta}(x)$. We have $\overline{\delta}(y) \leq z$ by monotonicity. Moreover, by assumption $z \geq y$ and $z \in \overline{D}$, so that $z \geq \overline{\delta}(y)$ by definition. We then must have $z = \overline{\delta}(y)$. Finally, as $0 \in \overline{D}$, we have that $\overline{\delta}(0) = 0$.

We now show that the induced evidence structure satisfies definitive evidence of exclusion, monotonicity, and the continuity property. Definitive evidence of exclusion follows from $\overline{\delta}(0) = 0$, and $\overline{\delta}(x) \geq x > 0$ for any $x \neq 0$. Thus, $\mathcal{E}(0) = \{0\}$ and $\mathcal{E}(x) \nsubseteq \mathcal{E}(0)$ for any $x \neq 0$.

We next show monotonicity. To remind, this property implies that, for all $x, x' \in X$ and $e, e' \in \Omega$, if $e' \geq e$ and $x' \geq x$, then

$$\mathbb{I}[e' \in \mathcal{E}(x')]\mathbb{I}[e \in \mathcal{E}(x)] \geq \mathbb{I}[e' \in \mathcal{E}(x)]\mathbb{I}[e \in \mathcal{E}(x')] \tag{116}$$

This amounts to showing that the intersection of conditions on the right-hand-side implies the intersection of conditions on the left-hand-side. If both right-hand-side conditions hold, then $\underline{\delta}(x) \leq e' \leq \overline{\delta}(x)$ and $\underline{\delta}(x') \leq e \leq \overline{\delta}(x')$. Using the monotonicity of $\underline{\delta}, \overline{\delta}$, we have that $e \leq e' \leq \overline{\delta}(x) \leq \overline{\delta}(x')$ and $\underline{\delta}(x) \leq \underline{\delta}(x') \leq e \leq e'$. Thus, $\underline{\delta}(x) \leq e \leq \overline{\delta}(x)$ and $\underline{\delta}(x') \leq e' \leq \overline{\delta}(x')$. By implication, as desired, $e' \in \mathcal{E}(x')$ and $e \in \mathcal{E}(x)$.

We next show continuity of $\mathcal{E}$. Fix $x, y \in X$. First, consider a sequence $x_n \to x$ with $\mathcal{E}(x_n) \subseteq \mathcal{E}(y)$ for all n, that is, $\underline{\delta}(y) \leq \underline{\delta}(x_n) \leq \overline{\delta}(x_n) \leq \overline{\delta}(y)$ for all n. The continuity properties of $\underline{\delta}$ and $\overline{\delta}$ imply that $\mathcal{E}(x) \subseteq \mathcal{E}(y)$. Second, assume $\mathcal{E}(x) \subseteq \mathcal{E}(y)$ and fix a sequence $y_n \to y$. By the semicontinuity properties of $\underline{\delta}$ and $\overline{\delta}$, we obtain that $\limsup \underline{\delta}(y_n) \leq \underline{\delta}(y) \leq \underline{\delta}(x) \leq \overline{\delta}(x) \leq \overline{\delta}(y) \leq \liminf \overline{\delta}(y_n)$. Therefore there exists a subsequence y_{n_k} such that $\underline{\delta}(y_{n_k}) \leq \underline{\delta}(x) \leq \overline{\delta}(x) \leq \overline{\delta}(y_{n_k})$ for all k. Finally, define $x_k = x$ for all k to obtain the desired result. $\square$

C.2 The Topology of Sets of Self-Enforcing Recommendations

Let $\underline{\Delta} \times \overline{\Delta}$ denote the space of pairs of functions $(\underline{\delta}, \overline{\delta})$ that have all the properties in Lemma 1. Recall that we endow this set with the relative topology induced by the product L_1 topology over pairs of integrable functions. Also, recall that $\underline{\mathcal{D}}$ denotes the collection of closed subsets of X that contain 0 and $\overline{\mathcal{D}}$ denotes the collection of closed subsets of X that contain 0 and $\overline{x}$. Let $\underline{\mathcal{D}} \times \overline{\mathcal{D}}$ denote the set of pairs of self-enforcing recommendations sets, $(\underline{D}, \overline{D})$. Recall that we have endowed this space with the product topology induced by the Hausdorff topology on each collection of sets.

Lemma 13. *The set $\underline{\mathcal{D}} \times \overline{\mathcal{D}}$ is compact.*

Proof. We show that each of $\underline{\mathcal{D}}$ and $\overline{\mathcal{D}}$ is compact in the Hausdorff topology. Specifically, we explicitly establish the compactness of $\overline{\mathcal{D}}$ and observe that an entirely symmetric argument applies to establish the compactness of $\underline{\mathcal{D}}$. With this, the compactness of the product space $\underline{\mathcal{D}} \times \overline{\mathcal{D}}$ follows from Tychonoff's theorem.

Observe that $\overline{\mathcal{D}} \subset \mathfrak{F}$, where $\mathfrak{F}$ is the collection of nonempty closed sets of X. Theorem 3.85 in Aliprantis and Border (2006) establishes that $\mathfrak{F}$ is compact in the Hausdorff topology. Fix a sequence $\{\overline{D}_n\}_{n\in\mathbb{N}} \subseteq \overline{\mathcal{D}}$. By the compactness of $\mathfrak{F}$, this sequence must have a subsequence $\overline{D}_{n_k}$ converging to some $\overline{D} \in \mathfrak{F}$ in the Hausdorff topology. Furthermore, we have that $0, \overline{x} \in \overline{D}_{n_k}$ for all k. Hence, by Hausdorff convergence of $\{\overline{D}_n\}_{n\in\mathbb{N}}$, we have that $0, \overline{x} \in \overline{D}$. Since $\overline{D}$ is a closed subset of X that contains 0 and $\overline{x}$, we have $\overline{D} \in \overline{\mathcal{D}}$. Finally, because the initial sequence was arbitrarily chosen, this implies that $\overline{\mathcal{D}}$ is compact. □

Lemma 1 implies that $\underline{\Delta} \times \overline{\Delta}$ and $\underline{\mathcal{D}} \times \overline{\mathcal{D}}$ are isomorphic via the maps:

$$\begin{aligned} (\underline{\delta}, \overline{\delta}) &\mapsto (\underline{D}_{\underline{\delta}}, \overline{D}_{\overline{\delta}}) := (\underline{\delta}(X), \overline{\delta}(X)) \\ (\underline{D}, \overline{D}) &\mapsto (\underline{\delta}_{\underline{D}}(x), \overline{\delta}_{\overline{D}}(x)) := \left(\max\{z \in \underline{D} : z \le x\}, \min\left\{z \in \overline{D} : z \ge x\right\}\right) \end{aligned} \tag{117}$$

We next prove that Hausdorff's convergence of sets of self-enforcing recommendations implies L_1-convergence of the corresponding envelope functions to the envelope functions induced by the limit sets.

Lemma 14. *Fix two sequences $\{(\underline{D}_n, \overline{D}_n)\}_{n\in\mathbb{N}} \subseteq \underline{\mathcal{D}} \times \overline{\mathcal{D}}$ and $\{(\underline{\delta}_n, \overline{\delta}_n)\}_{n\in\mathbb{N}} \subseteq \underline{\Delta} \times \overline{\Delta}$ such that $(\underline{\delta}_n, \overline{\delta}_n) = (\underline{\delta}_{\underline{D}_n}, \overline{\delta}_{\overline{D}_n})$ for all $n \in \mathbb{N}$. If $(\underline{D}_n, \overline{D}_n) \to (\underline{D}, \overline{D})$, then $(\underline{\delta}_n, \overline{\delta}_n) \to (\underline{\delta}_{\underline{D}}, \overline{\delta}_{\overline{D}})$.*

Proof. Fix two sequences as in the statement and assume that $(\underline{D}_n, \overline{D}_n) \to (\underline{D}, \overline{D})$ in the product Hausdorff topology. We only show that $\overline{\delta}_n \to \overline{\delta}_{\overline{D}}$ in L_1, as $\underline{\delta}_n \to \underline{\delta}_{\underline{D}}$ follows from an entirely symmetric argument. Together, these two facts imply convergence in the L_1 product topology.

For notational simplicity, denote $\overline{\delta}_{\overline{D}} = \overline{\delta}$ and let $X_{\overline{\delta}} \subseteq X$ denote the collection of points at which $\overline{\delta}$ is continuous. Because δ is non-decreasing, it follows that $X \setminus X_{\overline{\delta}}$ is at most countable. We next show that $\overline{\delta}_n(x) \to \overline{\delta}(x)$ for all $x \in X_{\overline{\delta}}$. Before showing this, we observe that the previous claim concludes the argument because

$$\lim_{n\to\infty} \int_X |\overline{\delta}_n(x) - \overline{\delta}(x)| \, \mathrm{d}x = \lim_{n\to\infty} \int_{X_{\overline{\delta}}} |\overline{\delta}_n(x) - \overline{\delta}(x)| \, \mathrm{d}x = \int_{X_{\overline{\delta}}} \lim_{n\to\infty} |\overline{\delta}_n(x) - \overline{\delta}(x)| \, \mathrm{d}x = 0 \tag{118}$$

where the first equality follows from the fact that $X_{\overline{\delta}}$ has full measure, the second follows from the dominated convergence theorem, and the last follows from the claim.

Next, fix $x \in X_{\overline{\delta}}$. Clearly, if $x = \overline{x}$, then $\overline{\delta}_n(x) \to \overline{\delta}(x)$ since they are all equal to $\overline{x}$. Thus, we next always assume that $x < \overline{x}$. We split the rest of the proof of the claim into two parts depending on whether x is inside or outside $\overline{D}$.

1. Assume that $x \in X_{\overline{\delta}} \setminus \overline{D}$ and define $\overline{D}_n(x) = \overline{D}_n \cap [x, \overline{x}]$ for all n as well as $\overline{D}(x) = \overline{D} \cap [x, \overline{x}]$. We first show that $\overline{D}_n(x) \to \overline{D}(x)$ in the Hausdorff topology. Let $Li(\overline{D}_n(x))$ and $Ls(\overline{D}_n(x))$ respectively denote the topological limit inferior and the topological limit superior of the sequence $\{\overline{D}_n(x)\}_{n\in\mathbb{N}}$.[14] Next, fix $z \in \overline{D}(x)$ and observe that $z > x$ because $x \in X_{\overline{\delta}} \setminus \overline{D}$. Because $z \in \overline{D}$ and $\overline{D}_n \to \overline{D}$, it follows that there exists a sequence $z_n \in \overline{D}_n$ such that $z_n \to z$. In particular, we must have $z_n > x$ for all n large enough, yielding that $z_n \in \overline{D}_n(x)$ for all n large enough, hence that $z \in Li(\overline{D}_n(x))$. Because z was arbitrarily chosen, this implies that $\overline{D}(x) \subseteq Li(\overline{D}_n(x))$. Next, fix $z \in Ls(\overline{D}_n(x))$. By definition of Ls, there exists a sequence $\{z_k\}_{k\in\mathbb{N}}$ such that $z_k \to z$ and $z_k \in \overline{D}_{n_k}(x)$ along a subsequence parametrized by k. Because $\overline{D}_{n_k} \to \overline{D}$ in the Hausdorff topology by assumption and $z_k \in \overline{D}_{n_k}$ for all k, it follows that $z \in \overline{D}$. Similarly, because $z_k \in [x, \overline{x}]$ for all k, it follows that $z \in [x, \overline{x}]$, yielding that $z \in \overline{D}(x)$. Because z was arbitrarily chosen, this implies that $Ls(\overline{D}_n(x)) \subseteq \overline{D}(x)$ and overall that $\overline{D}(x) = Li(\overline{D}_n(x)) = Ls(\overline{D}_n(x))$. Theorem 3.93 in Aliprantis and Border (2006) then yields that $\overline{D}_n(x) \to \overline{D}(x)$ in the Hausdorff topology. Finally, because $\overline{\delta}_n(x) = \min \overline{D}_n(x)$ for all n and $\overline{\delta}(x) = \min \overline{D}(x)$, Theorem 17.31 in Aliprantis and Border (2006) implies that $\overline{\delta}_n(x) \to \overline{\delta}(x)$.

2. Assume that $x \in X_{\overline{\delta}} \cap \overline{D}$ and define $\overline{D}_n(x)$ for all n and $\overline{D}(x)$ as above. Also, let $\operatorname{int} \overline{D}$ and $\partial\overline{D}$ respectively denote the interior and the boundary points of $\overline{D}$. First, observe that $\delta(x) = x$ because $x \in \overline{D}$. We next show that, for every $\varepsilon > 0$, there exists a point $x^\varepsilon \in \overline{D}(x)$ such that $0 < |x^\varepsilon - x| \leq \varepsilon$. We split the proof of this claim into two cases:

[14] See for example Definition 3.80 in Aliprantis and Border (2006).

a. If $x \in \operatorname{int} \overline{D}$, then the claim is immediately true because $\operatorname{int} \overline{D}$ is open.

b. Assume now that $x \in \partial \overline{D}$ and, by contradiction, that there exists $\varepsilon > 0$ such that for all $z \in \overline{D}(x)$ we have either $|z - x| = 0$ or $|z - x| > \varepsilon$. Because $x < \overline{x}$, there must be $z \in \overline{D}(x)$ with $|z - x| > 0$, hence for all such z we must have $|z - x| > \varepsilon$. In turn, this implies that $\overline{D} \cap [x, x + \varepsilon] = \{x\}$, and hence that $\overline{\delta}(y) > x + \varepsilon$ for all $y \in (x, x + \varepsilon)$. With this, we have that δ is discontinuous at x, yielding a contradiction.

Next, fix $\varepsilon > 0$ and $x^\varepsilon \in \overline{D}(x)$ as above. Because $x^\varepsilon \in \overline{D}$, Hausdorff convergence implies that there exists a sequence $x_n^\varepsilon \in \overline{D}_n$ such that $x_n^\varepsilon \to x^\varepsilon$. Moreover, because $x^\varepsilon > x$, for n large enough we must have $x_n^\varepsilon > x$, hence that $x_n^\varepsilon \in \overline{D}_n$. Therefore, for all n large enough, we have

$$x \leq \overline{\delta}_n(x) = \min \overline{D}_n(x) \leq x_n^\varepsilon \tag{119}$$

Passing to the limits, we have

$$x \leq \liminf_n \overline{\delta}_n(x) \leq \limsup_n \overline{\delta}_n(x) \leq x^\varepsilon \tag{120}$$

By taking $\varepsilon \to 0$ we conclude that $\lim_n \overline{\delta}_n(x) = x = \overline{\delta}(x)$, as desired. □

Before stating the main result of this section, observe that each cost function $\Gamma_{\mathcal{E}}$ induces cost functions in the spaces of envelope functions and sets of self-enforcing recommendations (by Lemma 1). In particular, we write these induced costs as $\Gamma_\Delta : \underline{\Delta} \times \overline{\Delta} \to [0, \infty]$ defined over the envelope functions of regular contractibility correspondences and $\Gamma_{\mathcal{D}} : \underline{\mathcal{D}} \times \overline{\mathcal{D}} \to [0, \infty]$ defined over the sets of self-enforcing recommendations of regular contractibility correspondences. Of course these induced costs are related by the identity $\Gamma_{\mathcal{D}}(\underline{D}, \overline{D}) = \Gamma_\Delta(\underline{\delta}_{\underline{D}}, \overline{\delta}_{\overline{D}})$.

Proposition 3. *If a cost function Γ_Δ is lower semi-continuous (resp. continuous) in the product L_1 topology, then $\Gamma_{\mathcal{D}}$ is lower semi-continuous (resp. continuous) in the product Hausdorff topology.*

Proof. The result immediately follows from Lemma 14. □

This proposition is useful in our analysis as it is simple to verify that front-end costs are continuous in the product L_1 topology when written in terms of envelopes of contractibility correspondences (as they are simply integrals of these functions). However, it is not simple to directly verify Hausdorff continuity of the cost in the space of self-enforcing recommendations. This result provides that conclusion, which is an essential step in our existence proof.

D Extension to Stochastic Evidence

The task of the arbitrator is to determine if the agent violated the agreed terms. They can do so if the set of evidence generated by the agent's action $\mathcal{E}(x)$ contains an element that could not have been generated by the agent doing what they said they would do $\mathcal{E}(y)$. In our main analysis, we assumed that this process was deterministic and error-free. However, it is natural to imagine that an arbitrator may make mistakes in judging the evidence, perhaps because the evidence itself is also a noisy signal. These mistakes may further be of two varieties: type I errors, where the agent is adjudged to have violated the terms of the contract when the true evidence does not support that conclusion; and type II errors, where the agent is adjudged to not have acted in contravention to the contract when the evidence in fact proves that they violated the terms of the contract. In this extension, we show that our main analysis and results extend to a setting in which punishments are bounded and an arbitrator is subject to both such type I and type II errors.

Model. The model and timeline are as in our baseline model, with two modifications. First, in the arbitration phase, after the agent has acted, the arbitrator makes type I and type II errors. Formally, when $\mathcal{E}(x) \subseteq \mathcal{E}(y)$, i.e., there is no evidence to support that the agent deviated, there is a probability $\alpha \in [0,1]$ that the arbitrator wrongly concludes that the agent shall be punished. Moreover, when $\mathcal{E}(x) \cap \mathcal{E}(y)^c \neq \emptyset$, i.e., there is evidence that proves the agent deviated, there is a probability $1-\beta \in [0,1]$ that the arbitrator wrongly concludes that the agent shall not be punished. Second, the financial punishment that the arbitrator levies is given by some $\bar{P}$.

Analysis with Stochastic Arbitration. Under this structure, an agent who acts in accordance with the contract can expect to lose $\alpha\bar{P}$ in utility from being wrongfully punished, in addition to forfeiting their promised payment. Similarly, an agent who fails to act in accordance with the contract can expect to lose $\beta\bar{P}$ in utility. So long as $\beta\bar{P} > \bar{u} \equiv \max_{x\in X,\theta\in\Theta} u(x,\theta)$, it is immediate that no agent will ever participate and then deviate.

The more delicate issue is the possibility of type I errors, as some agents who participate and act faithfully will nevertheless be punished *on path*, in which circumstance they also do not receive their promised payment. Thus, it is as if agents who participate in the mechanism now have a utility function given by $u(x,\theta) - (1-\alpha)t - \alpha\bar{P}$ and the principal's payoffs are given by $\pi(x,\theta) + (1-\alpha)t + \alpha\bar{P}$. It may now be optimal to exclude some agents from participation in the mechanism. However, conditional on this optimal choice of exclusion, our remaining analysis is unchanged.

These claims are summarized in the following Proposition:

Proposition 4. *If $\beta \bar{P} > \bar{u}$, then the claims of Theorem 1, Theorem 2,*[15] *and Corollary 1 hold.*

Proof. We first show that when $\beta \bar{P} > \bar{u}$, type II errors have no effect on the principal's problem. Suppose that the agent participates in the mechanism and selects an $(y, \tau) \in \mathcal{M}$. The maximum possible gain from choosing an action $x \in X$ is $\bar{u}$. The expected loss in utility from choosing an action $x \notin C_{\mathcal{E}}(y)$ is $\beta \bar{P}$. Hence, as $\beta \bar{P} > \bar{u}$, an agent will always choose an action $x \in C_{\mathcal{E}}(y)$. Thus, as the agent never violates the terms of the contract on path, the presence of type II errors has no effect on principal or agent payoffs.

We now study how type I errors affect the principal's problem. As discussed before, these transform the principal and agent gross payoffs into the following form $\tilde{\pi}(x, \theta) = \pi(x, \theta) + \alpha \bar{P}$ and $\tilde{u}(x, \theta) = u(x, \theta) - \alpha \bar{P}$ conditional on the agent participating. This violates our maintained assumption that $\tilde{\pi}(0, \theta) = \tilde{u}(0, \theta) = 0$ for all $\theta \in \Theta$ but does not affect any of our other maintained assumptions on the gross payoffs. We must therefore adapt Lemma 3 to handle this change. A mechanism now comprises (ϕ, ξ, r, T), where $r : \Theta \to \{0, 1\}$ is a mapping from types to inclusion, with $r(\theta) = 1$ denoting inclusion and $r(\theta) = 0$ denoting exclusion. The agent's IR constraint now changes to $u(\phi(\theta), \theta) - (1 - \alpha)T(\xi(\theta)) \geq \alpha \bar{P}$ if $r(\theta) = 1$. The IC constraint is the same, except the agent's payoff is replaced by $u(x, \theta) - (1 - \alpha)T(y)$ and it only need hold when $r(\theta) = 1$. The Obedience constraint is unchanged, except that it need only hold when $r(\theta) = 1$. Finally, we require that the agent's decision not to participate is incentive compatible:

$$r(\theta) = 0 \implies \mathcal{L}(\theta) \equiv \max_{y \in X, x \in C(y)} u(x, \theta) - (1 - \alpha)T(y) - \alpha \bar{P} \leq 0 \tag{121}$$

$\mathcal{L}(\theta)$ is a strictly increasing function of θ. Hence, $r(\theta)$ must be monotone increasing, and so there exists a $\bar{\theta} \in [0, 1]$ such that $r(\theta) = \mathbb{I}[\theta \geq \bar{\theta}]$.

Lemmas 8 and 9 must be updated to replace the IR constraints with the modified IR constraint. Lemma 9 holds as stated except that the tariff must instead be given be:

$$(1 - \alpha)T(x) = \mathbb{I}[\exists \theta \in \Theta : r(\theta) = 1, x = \phi(\theta)] \left(Z + u(x, \phi^{-1}(x)) - \int_{\bar{\theta}}^{\phi^{-1}(x)} u_\theta(\phi(s), s) \, \mathrm{d}s \right) \tag{122}$$

[15] Here, the definition of non-triviality must be strengthened to include the hypothesis that $H(\phi^P(\theta), \theta) > 0$ at some θ for which there is a unique and interior maximum.

with $Z \leq -\alpha\bar{P}$. In Lemma 10, the principal's problem then becomes:

$$\begin{aligned}
&\max_{\phi,\bar{\theta},Z\leq-\alpha\bar{P}} \int_{\bar{\theta}}^{1} \left(\pi(\phi(\theta),\theta) + \alpha\bar{P} + Z + u(\phi(\theta),\theta) - \int_{\bar{\theta}}^{\theta} u_\theta(\phi(s),s)\,\mathrm{d}s - \int_{\phi(\theta)}^{\overline{x}} g_b(\phi(\theta),z)\,\mathrm{d}z\right)\mathrm{d}F(\theta)\\
&\text{s.t.}\quad \phi(\theta') \geq \phi(\theta),\ \phi(\theta) \in \overline{D} \quad \forall \theta,\theta' \in \Theta : \theta' \geq \theta\\
&\qquad u(\phi(\theta),\theta) - \left(\alpha\bar{P} + Z + u(\phi(\theta),\theta) - \int_{\bar{\theta}}^{\theta} u_\theta(\phi(s),s)\,\mathrm{d}s\right) \geq 0 \quad \forall \theta \geq \bar{\theta}
\end{aligned} \tag{123}$$

It follows that it is optimal to set $Z = -\alpha\bar{P}$. Thus, this problem reduces to:

$$\begin{aligned}
&\max_{\phi,\bar{\theta}} \int_{\bar{\theta}}^{1} \left(\pi(\phi(\theta),\theta) + u(\phi(\theta),\theta) - \int_{\bar{\theta}}^{\theta} u_\theta(\phi(s),s)\,\mathrm{d}s - \int_{\phi(\theta)}^{\overline{x}} g_b(\phi(\theta),z)\,\mathrm{d}z\right)\mathrm{d}F(\theta)\\
&\text{s.t.}\quad \phi(\theta') \geq \phi(\theta),\ \phi(\theta) \in \overline{D} \quad \forall \theta,\theta' \in \Theta : \theta' \geq \theta
\end{aligned} \tag{124}$$

Adapting the standard integration by parts argument, we then obtain that:

$$\mathcal{R}(\overline{D}) = \max_{\bar{\theta}} \int_{\bar{\theta}}^{1} \left(\max_{x\in\overline{D}} H(x,\theta)\right)\mathrm{d}F(\theta) \tag{125}$$

The fact that H is strictly supermodular is what allows us to drop the monotonicity constraint. We next show that the map $\theta \mapsto \hat{H}_{\overline{D}}(\theta) \equiv \max_{x\in\overline{D}} H(x,\theta)$ is strictly increasing in θ. Specifically, consider any selection $\phi_D(\theta) \in \arg\max_{x\in\overline{D}} H(x,\theta)$, and observe that

$$\hat{H}'_{\overline{D}}(\theta) = J_\theta(\phi_D(\theta),\theta) = J_\theta(0,\theta) + \int_0^{\phi_D(\theta)} J_{x\theta}(z,\theta)\mathrm{d}z = \int_0^{\phi_D(\theta)} J_{x\theta}(z,\theta)\mathrm{d}z > 0 \tag{126}$$

where the first equality follows from the Envelope theorem, the second by the fundamental theorem of calculus, the third equality follows from our normalization assumption on u and π, and the last strict inequality from the strict supermodularity of J. With this, we have that $\bar{\theta}$ is the unique solution to $\max_{x\in\overline{D}} H(x,\theta) = 0$ if one exists, and is otherwise one if $\max_{x\in\overline{D}} H(x,\theta) < 0$ for all $\theta \in \Theta$ and zero if $\max_{x\in\overline{D}} H(x,\theta) > 0$ for all $\theta \in \Theta$. By non-triviality (with the stated strengthening in Footnote 15), the same arguments as those of Theorem 2 imply that $\overline{D} = X$ strictly maximizes this objective function.

We now establish that the conclusion Lemma 7 holds as stated. We define $\bar{\theta}(\overline{D})$ as the solution $\bar{\theta}$ that we just described. We also define $\hat{x}(\overline{D}) \equiv \max\left(\arg\max_{x\in\overline{D}} H(x,\bar{\theta}(\overline{D}))\right)$. It is immediate that if $\overline{D} \cap (0,\hat{x}(\overline{D})) \neq \emptyset$, then $\overline{D} \setminus (0,\hat{x}(\overline{D}))$ yields a strictly lower front-end cost and the same value of $\mathcal{R}$. Hence, any optimal $\overline{D}$ contains no elements between 0 and $\hat{x}(\overline{D})$. Moreover, for any (a,b) such that $\hat{x}(\overline{D}) < a < b$, we have that $\bar{\theta}(\overline{D} \setminus (a,b)) = \bar{\theta}(\overline{D})$.

Thus, the arguments of Lemma 7 hold as stated. Theorem 1 and Corollary 1 follow. $\square$

We also flag that this proof addresses the variant of the model that we flagged in Footnote 6. In particular, if we allow the principal not to pay back-end costs for the agents that it excludes, then it may wish to exclude agents whenever $\max_{x \in X} H(x, \theta) < 0$. The analysis above shows that this does not affect our conclusions for how front and back-end costs affect optimal contracts.

Discussion. This analysis shows that the possibility of noisy evidence and random arbitration do not affect the conclusions of our analysis. In this framework, it is further natural to consider the behavior of a principal with costs of controlling α and β and who can perhaps also write $\bar{P}$ into the contract. It would also be interesting to extend this framework to potentially allow the error rate to depend on the structure of evidence itself. We leave an analysis of these issues to future work.

E Extending the Results to More General Costs

In this appendix, we describe a richer family of front-end and back-end cost functions than in our main analysis and provide general conditions under which the respective conclusions of Theorems 1 and 2 hold.

E.1 Generalizing Front-End Costs

We first define an order over regular evidentiary correspondences, the set of which we denote by $\mathscr{E}$. We say that $\mathcal{E}'$ generates more refined evidence than $\mathcal{E}$, which we denote by $\mathcal{E}' \succsim \mathcal{E}$, if for all $x, y \in X$ such that $\mathcal{E}'(x) \subseteq \mathcal{E}'(y)$ we also have that $\mathcal{E}(x) \subseteq \mathcal{E}(y)$. In words, this means that $\mathcal{E}'$ generates more refined evidence than $\mathcal{E}$ if every time that x cannot be proven by the principal to be inconsistent with y using evidence generated by $\mathcal{E}'$, the same is true if evidence were generated by $\mathcal{E}$. Observe that $\mathcal{E}' \succsim \mathcal{E}$ if and only if $C_{\mathcal{E}'} \subseteq C_{\mathcal{E}}$. Thus, this order over evidence is equivalent to inducing a greater set of self-enforcing recommendations.

We say that a (front-end) cost function $\Gamma : \mathscr{E} \to \mathbb{R}_+$ is monotone if whenever $\mathcal{E}' \succsim \mathcal{E}$, then we have that $\Gamma(\mathcal{E}') \geq \Gamma(\mathcal{E})$. We argue that this is a natural property for a cost to possess: if an evidentiary correspondence generates more refined evidence, then it incurs a higher cost. We now show that monotonicity of Γ justifies writing costs directly over contractibility correspondences. Recall that we let $\mathcal{C}$ denote the set of regular contractibility correspondences. Formally, we define what it means for a cost function to be measurable in the induced contractibility correspondence as:

Definition 4 (C-measurability)**.** *Γ is $C-$measurable if there exists a $\check{\Gamma}: \mathcal{C} \to [0, \infty]$ such that for all $\mathcal{E} \in \mathscr{E}$, we have that $\Gamma(\mathcal{E}) = \check{\Gamma}(C_{\mathcal{E}})$.*

If Γ is $C-$measurable, we call the corresponding $\check{\Gamma}$ the induced cost. We can now state the following result:

Lemma 15. *If Γ is monotone, then it is $C-$measurable.*

Proof. Fix an arbitrary pair of evidentiary correspondences $\mathcal{E}, \mathcal{E}' \in \mathscr{E}$ such that $C_{\mathcal{E}} = C_{\mathcal{E}'}$. We have that $C_{\mathcal{E}} \subseteq C_{\mathcal{E}'}$, which implies that $\mathcal{E} \succsim \mathcal{E}'$. By monotonicity of Γ we have that $\Gamma(\mathcal{E}) \geq \Gamma(\mathcal{E}')$. By the reverse argument, we have that $\Gamma(\mathcal{E}') \geq \Gamma(\mathcal{E})$. Hence, we have that $\Gamma(\mathcal{E}') = \Gamma(\mathcal{E})$ for all $\mathcal{E}, \mathcal{E}' \in \mathscr{E}$ that induce the same contractibility correspondence. Take $\check{\Gamma}(C_{\mathcal{E}}) = \Gamma(\mathcal{E})$. Thus, Γ is $C-$measurable with induced $\check{\Gamma}$. □

This result implies that whenever costs are monotone in the natural sense over underlying evidentiary correspondences, it is without loss of optimality to write the cost in the space of contractibility correspondences directly. It is immediate to see that all the front-end costs of distinguishing that we considered in the main text are monotone.

As in the main analysis, we call a front-end evidentiary cost of distinguishing linear if $g_f(x, y) = \kappa > 0$ and we write this cost function as Γ^κ. We can now define strong monotonicity:

Definition 5 (Strong Monotonicity)**.** *A cost function Γ is strongly monotone if there exists $\varepsilon > 0$ such that, for all $\mathcal{E}, \mathcal{E}' \in \mathscr{E}$ such that $\mathcal{E}' \succsim \mathcal{E}$:*

$$\Gamma(\mathcal{E}') - \Gamma(\mathcal{E}) \geq \varepsilon \left(\Gamma^1(\mathcal{E}') - \Gamma^1(\mathcal{E}) \right) \tag{127}$$

With this, we show that strong monotonicity implies a certain property for the induced cost defined over contractibility correspondences:

Proposition 5. *If Γ is strongly monotone with constant $\varepsilon > 0$, then the induced $\check{\Gamma}$ satisfies:*

$$\check{\Gamma}(C') - \check{\Gamma}(C) \geq \varepsilon \left(\check{\Gamma}^1(C') - \check{\Gamma}^1(C) \right) \tag{128}$$

for all $C' \subseteq C$.

Proof. We first observe that strong monotonicity of Γ implies monotonicity of Γ. Thus, by Lemma 15, we have that Γ is $C-$measurable and therefore has an induced $\check{\Gamma}$. Now fix an arbitrary pair $\mathcal{E}, \mathcal{E}' \in \mathscr{E}$ such that $\mathcal{E}' \succsim \mathcal{E}$, we have that:

$$\check{\Gamma}(C_{\mathcal{E}'}) - \check{\Gamma}(C_{\mathcal{E}}) = \Gamma(\mathcal{E}') - \Gamma(\mathcal{E}) \geq \varepsilon \left(\Gamma^1(\mathcal{E}') - \Gamma^1(\mathcal{E}) \right) = \varepsilon \left(\check{\Gamma}^1(C_{\mathcal{E}'}) - \check{\Gamma}^1(C_{\mathcal{E}}) \right) \tag{129}$$

where the first equality is by $C-$measurability, the first inequality is by strong monotonicity of Γ and the final equality is by definition of the linear front-end cost of distinguishing. Thus, as $\mathcal{E}' \succsim \mathcal{E} \iff C_{\mathcal{E}'} \subseteq C_{\mathcal{E}}$, we have established the required condition on $\check{\Gamma}$. $\square$

By Lemma 12, we have that $C \in \mathcal{C}$ if and only if $C(y) = [\underline{\delta}_C(y), \overline{\delta}_C(y)]$ for two semicontinuous, monotone functions with the escalator-step property. We say that $\check{\Gamma} : \mathcal{C} \to \mathbb{R}_+$ is continuous if it is continuous in the product L_1-topology with respect to $(\underline{\delta}_C, \overline{\delta}_C)$.

Given a monotone front-end cost function Γ, we say that Γ is continuous if its induced $\check{\Gamma}$ is continuous. In this case, Proposition 3 implies that the induced cost function over the set of self-enforcing recommendations is continuous in the Hausdorff topology. We now show that strong monotonicity and continuity of Γ are sufficient for the conclusion that optimal contracts are necessarily coarse.

Proposition 6. *Suppose that Γ is continuous and strongly monotone with constant $\varepsilon > 0$ and the principal faces no back-end costs. Any optimal evidence exists and every optimal evidence structure is coarse. Moreover, the bound on the richness of any optimal evidence structure and menu from Theorem 1 holds, with ε replacing $\underline{g}_f$.*

Proof. By the continuity assumption, Lemma 4 applies and so an optimal evidence structure exists. By Proposition 5, Lemma 6 holds with ε replacing $\underline{g}_f$. The remaining arguments follow the proof of Theorem 1. $\square$

E.2 Generalizing Back-End Costs

We consider back-end cost functions $\Gamma : \mathscr{E} \times \Delta(X) \to \mathbb{R}_+$ given by:

$$\Gamma(\mathcal{E}, Q) = \int_X \gamma(\mathcal{E}, x)\, \mathrm{d}Q\,(x) \tag{130}$$

for some $\gamma : \mathscr{E} \times X \to \mathbb{R}_+$. The interpretation here is that $\gamma(\mathcal{E}, x)$ is the cost of generating the evidence according to the action x given the evidence structure $\mathcal{E}$. Given this, Γ is then the expected cost that will be incurred if the distribution over actions that is taken is given by Q. We say that γ is monotone if whenever $\mathcal{E}' \succsim \mathcal{E}$, then we have that $\gamma(\mathcal{E}', x) \geq \gamma(\mathcal{E}, x)$ for all $x \in X$. By the exact same arguments as those in Lemma 15, if γ is monotone, then it is C-measurable. We denote by $\tilde{\gamma}$ the induced cost over contractibility correspondences, *i.e.,* $\tilde{\gamma}(C_{\mathcal{E}}, x) \equiv \gamma(\mathcal{E}, x)$. By monotonicity of γ in $\mathcal{E}$, we also have that $\tilde{\gamma}$ is monotone with respect to set inclusion in the space of regular contractibility correspondences, that is, $C' \subseteq C$ implies that $\tilde{\gamma}(C', x) \geq \tilde{\gamma}(C, x)$ for all $x \in X$. With this, we have that the induced back-end cost is also C-measurable and we denote the induced cost over contractibility correspondences by $\tilde{\Gamma}$, *i.e.*, $\tilde{\Gamma}(C_{\mathcal{E}}, Q) = \int_X \tilde{\gamma}(C_{\mathcal{E}}, x)\,\mathrm{d}Q\,(x) = \Gamma(\mathcal{E}, Q)$. We say that $\tilde{\gamma}$ is *separable* if $\tilde{\gamma}(C, x) = \hat{\gamma}(C(x), x)$ for some $\hat{\gamma} : \Xi \times X \to \mathbb{R}_+$, where $\Xi \subseteq 2^X$ is the collection of closed and convex subsets of X, and we endow it with the Hausdorff topology. We note that as $\tilde{\gamma}$ is monotone with respect to set inclusion, so too is $\hat{\gamma}$. We say that a separable $\tilde{\gamma}$ is continuous if the induced function $\hat{\gamma}$ is continuous with respect to the product topology. Finally, we define $\zeta(x) = \hat{\gamma}([0, x], x)$ and $H(x, \theta) = J(x, \theta) - \zeta(x)$. With these definitions, we can state the following result.

Proposition 7. *Suppose that the principal faces no front-end costs and γ is monotone with an induced $\tilde{\gamma}$ that is separable and continuous. An optimal evidence structure $\mathcal{E}^*$ with cardinality $|\mathcal{E}^*(X)| = \infty$ exists. Moreover, if H is non-trivial, then every optimal evidence structure $\mathcal{E}^*$ and every optimal menu $\mathcal{M}^*$ have cardinality $|\mathcal{E}^*(X)| = |\mathcal{M}^*| = \infty$.*

Proof. If γ is monotone with a separable $\tilde{\gamma}$, then we may write the back-end cost as:

$$\tilde{\Gamma}(C, Q) = \int_X \hat{\gamma}(C(x), x)\,\mathrm{d}Q\,(x) = \int_X \hat{\gamma}([\underline{\delta}(x), \overline{\delta}(x)], x)\,\mathrm{d}Q\,(x) \tag{131}$$

where $\underline{\delta}$ and $\overline{\delta}$ are the upper and lower envelopes of C from the representation of Lemma 12. The arguments of Lemma 2 then apply here and we have that $\underline{\delta}(x) \equiv 0$ is optimal. Using this, we can then write:

$$\begin{aligned}\tilde{\Gamma}(C, Q_\phi) &= \int_X \hat{\gamma}([0, \overline{\delta}(x)], x)\,\mathrm{d}Q_\phi(x) = \int_\Theta \hat{\gamma}([0, \overline{\delta}(\phi(\theta))], \phi(\theta))\,\mathrm{d}F(\theta) \\ &= \int_\Theta \hat{\gamma}([0, \phi(\theta)], \phi(\theta))\,\mathrm{d}F(\theta) = \int_\Theta \zeta(\phi(\theta))\,\mathrm{d}F(\theta)\end{aligned} \tag{132}$$

where the penultimate line follows (as in Lemma 3) from Obedience and the last uses the definition of $\zeta(x) = \hat{\gamma}([0, x], x)$. We can now apply the remaining arguments of Lemma 3 where the statement now holds with $H(x, \theta) = J(x, \theta) - \zeta(x)$. By the hypothesis of continuity of $\tilde{\gamma}$, we have that ζ (and therefore H) is continuous. From this, the arguments from Appendix A.2 apply and the conclusion of Theorem 2 then follows. □

References

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

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