Informative Certification: Screening vs. Acquisition*

Gorkem Celik† and Roland Strausz‡

January 29, 2025

Abstract

We study monopolistic certification in a buyer-seller relationship, explicitly distinguishing between its role as a device for screening versus acquisition. As a screening device, certification discloses soft information about a seller's private information. As an acquistion device, certification discloses hard information about the good's quality. Despite being costless, we show that, optimally, a monopolistic certifier provides non-maximal information-acquisition, while offering maximal screening. Thus, monopolistic certification exhibits no economic distortions as a screening device, resolving all private information, but provides too little hard information as an acquisition device. While feasible and costless, full information acquisition is suboptimal as it requires excessive information rents. Consequently, market inefficiencies remain due to market uncertainty but not due to private information.

Keywords: certification, disclosure, screening, information acquisition, monopolistic distortions

JEL classification: D82

---

*We are most indebted to Roberto Corrao for pointing out the broader implications of our analysis on the general informativeness of certification. As a result, this paper supersedes our earlier paper "Selling Certification of Private and Market Information". In addition, we thank Gary Biglaiser, Li Hao, Curt Taylor, Nikhil Vellodi, Ran Weksler, Boaz Zik, and seminar participants at Paris Dauphine, UNC Chapel Hill, University of British Columbia, University of Colorado Boulder, Berlin School of Economics, and Africa Business School at UM6P Rabat for helpful discussions. Roland Strausz acknowledges financial support from the European Union through the ERC-grant PRIVDIMA (project number 101096682) and the Deutsche Forschungsgemeinschaft through CRC-TRR 190 (project number 280092119).

†ESSEC Business School & THEMA Research Center. Email: celik@essec.edu.

‡Humboldt Universität zu Berlin & Berlin School of Economics. Email: strauszr@hu-berlin.de.

1 Introduction

Certification plays a crucial role in modern markets by publicly disclosing information to market participants. Its importance is evident across various sectors of the economy. Rating agencies, for instance, publicly certify the default risk of financial securities, providing critical information for investors and regulators. Accounting firms certify the financial reports of corporations, ensuring transparency and accountability in business operations. In the digital realm, online recommendation systems allow consumers to publicly certify the quality of previously purchased goods, shaping the decisions of future buyers.

The significance of these certification processes cannot be overstated. Many analysts argue that the 2007-2008 financial crisis might have been averted had rating agencies performed their certification duties more rigorously. Similarly, the 2001 Enron scandal is often attributed to the failure of Arthur Andersen to properly certify the company's financial statements. In e-commerce, some industry observers attribute eBay's decline relative to Amazon's rise to differences in their rating systems, with eBay's system focusing on certifying the transaction while Amazon's emphasizes the certification of the seller's product quality.

Despite a substantial body of theoretical literature on certification, two fundamental issues remain unclear. First, there is ambiguity about whether certifiers disclose new information or only reveal existing, private information. Second, the exact economic mechanism by which certification provides information to market participants is not thoroughly spelled out. These unresolved questions hinder a comprehensive understanding of certification's role and impact in markets.

Regarding the first fundamental question, two opposing views have emerged in the literature. The more traditional strand presumes that certifiers provide information about the private knowledge of an economic agent, such as a seller who is privately informed about their product's quality (see Viscusi, 1978 and Lizzeri, 1999). In this view, certification *levels* the informational playing field by closing the gap between privately informed and uninformed parties. Conversely, a more recent strand of literature based on ideas in information design conceptualizes certification as providing new information previously unknown to any economic party (see Bizzotto, Rüdiger, and Vigier, 2020; Ali et al., 2022; Evans and Park, 2024; Asseyer and Weksler, 2024). This perspective suggests that certification *raises* the informational playing field for all parties involved.

The second fundamental question is closely related to the first, as it concerns the channel through which certification conveys information. When certification reveals new information, it acts as a device of *acquisition*, providing previously unknown data to market participants through a hard, objective information channel, disclosing new facts or new evidence. In contrast, when certification reveals private information, it often does so by means of *screening*, playing the role of a screening device. This process involves revealing the information of a privately informed party to market participants via a soft, more subjective information channel.[1]

In practice, certification often plays both roles simultaneously. Consider the example of certified pre-owned (CPO) cars, where car owners can certify the quality of their vehicle externally before placing it on the market. The seller, as the previous owner, naturally possesses some private information about the car's quality. The certification process will not only confirm this private information but also provide additional objective data, such as precise measurements of tire wear in millimeters or the exact chemical composition of the exhaust emissions. The provision of this latter data is a clear example of certification as a device for information acquisition. To see, instead, its role as a screening device, it is best to consider the buyer's reaction to an absence of any certification. Even though the absence of certification obviously cannot reveal any new objective information, buyers tend to interpret the fact that the seller did not certify as informative about the seller's private information, naturally downgrading their expectations of the car's quality.

This paper aims to elucidate the economic differences and interactions between these two types and roles of certification. Working out these distinctions and their interplay is crucial for developing a more comprehensive understanding of certification. By examining how certification can simultaneously reveal private information and generate new information, and how it functions as both a screening device and an information acquisition tool, we gain deeper insights into the optimal design and economic impact of certification systems in various market contexts.

The main insight of our analysis is that a monopolistic certifier exploits the screening role of certification to the fullest, whereas it exploits its role of information acquisition at most partially. Notably, we obtain this result despite the absence of any explicit costs of certification. This implies that the certifier's optimal certification structure exhibits maximal screening, fully disclosing all private information, while it does not exhibit full information acquisition, leaving market uncertainties.

This result is both surprising and important, as it corrects a potentially misleading intuition. Conventional wisdom might suggest that screening should be more costly for the certifier than

[1] See Corrao (2024) for a recent study of this soft information channel in more general mediation problems.

information acquisition. This is because, to induce truthful revelation of the seller's private information through screening, the certifier must respect incentive compatibility constraints. Screening thus exhibits an inherent, indirect cost of disclosure—the information rents associated with incentive compatibility. In contrast, information acquisition does not require any incentive compatibility. The indirect costs associated with screening would seem to suggest that, all other things equal, the certifier should prefer the information acquisition role of certification over its screening role. Paradoxically, our analysis reveals that this intuition is incorrect.

The resolution of this paradox lies in how the information rents due to incentive compatibility affect the certifier's incentives. Our analysis uncovers that the information rents associated with incentive-compatible screening are increasing in the amount of information acquisition that the certification induces. It is precisely for this reason that full information acquisition is suboptimal for the certifier. In fact, the certifier can fully screen private information without conceding any information rents if the certification does not provide any information acquisition.

Our main finding has important policy implications. While monopolistic certification does not exhibit any distortions as a screening device, it does exhibit distortions as a device for information acquisition, providing too little new information from a social point of view. In other words, monopolistic certification leads to a complete leveling of the informational playing field, eliminating all private information, but it does not eliminate all market uncertainty, even though doing so would increase aggregate surplus. This suggests that regulatory interventions in certification markets might be more effectively targeted at encouraging the production and disclosure of new information rather than at the revelation of existing private information. Thus, our analysis provides a theoretical foundation for understanding why certifiers might perform well at eliminating information asymmetries while simultaneously underperforming in reducing overall market uncertainty.

Formally, we derive our results in a natural and generic buyer-seller model that can explicitly distinguish between the two types and roles of certification. In particular, we consider a buyer, who faces uncertainty about the good's quality, and a seller, who has only partial information about the good's quality. In this setup of both private information and market uncertainty, we introduce a third party—a certifier—who has the technology to publicly disclose information about the seller's private information as well as the remaining market uncertainty.

We thereby take the seller's private information as also fundamental to the certifier. That is, while the certifier's informative signal may, in the spirit of information design, depend arbitrarily on the true underlying quality (the fundamental state), it cannot directly depend on the seller's private information about this state (the seller's type). However, in the spirit of monopolistic screening, the certifier may screen between different seller-types by offering a menu of different pairs of certification rules and associated prices. Thus, the certifier controls the acquisition of market information by its ability to design any information signal that conditions on the true underlying quality. In addition, the certifier controls the disclosure of the seller's private information by determining the degree to which the certification menu screens seller-types.

While the economic literature on certification is vast (see Dranove and Jin, 2010 for a survey), we are unaware of existing work that explicitly distinguishes between the two natural roles of certification—screening and acquisition.

Indeed, starting with Viscusi (1978), most of the literature on certification exclusively focuses on the role of certification in reducing informational asymmetries. Implicitly, this literature considers certification models of "full private information", in which the seller is fully informed about her quality. As a result, these models cannot capture a certifier's ability to acquire information that is unknown to the seller.[2] Consequently, the certifier's role of acquisition is moot.

By contrast, a more recent literature abstracts from any private information, focusing on the certifier's role of acquiring hard information (e.g., Bizzotto, Rüdiger, and Vigier, 2020; Ali et al., 2022; Evans and Park, 2024; Asseyer and Weksler, 2024). Applying techniques from information design, these models study a certifier, who can reveal information about an unknown state concerning which the parties are uninformed symmetrically. Because of the absence of any private information, the certifier's role of screening is moot.

Studies that, similar to us, consider a seller with "partial private information" are Rosar (2017), Bergemann, Bonatti, and Smolin (2018), Kartik, Lee, and Suen (2021), Ichihashi and Smolin (2023), and Weksler and Zik (2023, 2024). These papers, however, do not explicitly distinguish between the two distinct roles of certification—screening and acquisition—that we focus on. In particular, Rosar (2017), Kartik, Lee, and Suen (2021, Section II), and Weksler and Zik (2023, 2024) study the role of certification with partial private information but limit the certifier's ability to screen. More specifically, they do not allow the certifier to use maximal-screening menus, which we show to be optimal. In Bergemann, Bonatti, and Smolin (2018), there is no role for disclosing private information, because information is sold to the privately

[2] For instance, Hancart (2024) studies a certifier, who can offer a menu of certification contracts to screen for private information, but as the private information is fully informative about the state, an acquisition of any unknown information does not take place.

informed party rather than revealed publicly to the market. Similarly, there is no role for disclosing the seller's private information to buyers in Ichihashi and Smolin (2023) because buyers are already fully informed about their value for the product so that revealing the seller's private information is uninformative to all market participants.

With its insights on the optimality of maximal screening and the suboptimality of maximal information-acquisition, the paper also contributes to an extensive literature that identifies conditions under which verifiable information is fully revealed (e.g., Okuno-Fujiwara et al., 1990; Seidmann and Winter, 1997; Koessler and Renault, 2012; Hagenbach et al., 2014). In particular, we provide the insight that verifiable information that is initially unavailable in the form of private information is typically not fully revealed.

2 The Setup

Allocations and Payoffs. We consider a monopolistic seller (she), who sells a quantity x of a divisible good to a single buyer (he). The seller's quality ω is either high (h) or low (l), i.e., $\omega \in \{h, l\}$. As we explicitly want to consider a seller who does not fully know her quality, we assume that the cost of production is independent of quality so that the seller cannot induce the quality from observing her costs. For simplicity, we take this common cost to be zero.

Consuming a quantity x of quality level ω at price p yields the buyer the utility $U = u_\omega(x) - p$. The buyer's marginal value is strictly decreasing and is higher for high quality than for low quality. Taking $u_\omega(x)$ to be twice differentiable, we thus assume $u''_\omega(x) < 0$ and the single crossing condition $u'_h(x) > u'_l(x)$. We further assume that the marginal value of an initial unit is strictly positive, $u'_\omega(0|\omega) > 0$, whereas for large enough quantities the marginal value of consumption exceeds the seller's (zero) costs, $\lim_{x\to\infty} u'_\omega(x|\omega) < 0$. It follows that the buyer's utility-maximizing quantity x^*_ω for product quality ω satisfies the first order condition $u'_\omega(x^*_\omega) = 0$, equalizing marginal benefits to marginal costs. Our assumptions on $u_\omega(x)$ imply $x^*_h > x^*_l > 0$. Finally, we assume that $u_l(x) < u_h(x)$ for all $x \in [x^*_l, x^*_h]$.

Because an economic allocation (x, p) of quality ω yields the buyer a net utility of $U = u_\omega(x) - p$ and the seller a profit of $\Pi = p$, the value x^*_ω is not only the buyer-optimal level but also the efficient level of consumption with quality level ω. Quality information is therefore efficiency-relevant as it allows to determine the efficient level of consumption.

We denote by $u^*_\omega \equiv u_\omega(x^*_\omega)$ the surplus associated with the efficient quantity in state ω. To simplify the exposition, we normalize utilities and assume without loss that $u^*_l = 0$ and $u^*_h = 1$.

We further assume that, as an outside option, the buyer can obtain the good at the low quality level for sure from a competitive fringe. Hence, the buyer's outside option matches the utility level $u_l^* = 0$.

Information structure. We let $\bar{\theta} \in (0, 1)$ denote the ex ante probability that the seller's quality is high, i.e., $\mathbb{P}\{\omega = h\} = \bar{\theta}$ and $\mathbb{P}\{\omega = l\} = 1 - \bar{\theta}$. The buyer is uninformed about the quality and $\bar{\theta}$ therefore signifies his prior of facing a seller with high quality. By contrast, the seller is better but not fully informed about the quality of her good. In particular, the seller receives an informative signal $\theta \in \Theta \equiv [0, 1]$ about ω such that it leads her to update her prior from $\bar{\theta}$ to θ. The seller observes the signal θ privately so that θ represents the seller's type. We denote the cdf of the seller's type by $F(\theta)$ and assume that its pdf, $f(\theta)$, exists and exhibits full support, i.e., $f(\theta) > 0$ for all $\theta \in \Theta$.[3] Because quality is binary, the seller's private information is one-dimensional, allowing us to avoid the multidimensional-screening complications when studying certification.

Certification. We assume that there is a monopolistic certifier (it) with a technology to generate public certificates. These certificates can be informative about the product's quality ω due to the acquisition of new, *hard* information and due to a screening of the seller, which we view as more subjective, *soft* information.

Following information design, we capture the certifier's information acquisition by allowing the certifier to commit to a *certification structure* $\sigma = (C, \pi^l, \pi^h)$. That is, a certification structure σ consists of a set C that represents a (finite) set of public certificates, together with the two probability vectors $\pi^l = (\pi_c^l)_{c \in C}$ and $\pi^h = (\pi_c^h)_{c \in C}$, where π_c^ω denotes the probability that the certificate $c \in C$ realizes in state $\omega \in \{h, l\}$. Because for a given structure σ, the certificate c is drawn objectively given state ω, it represents hard, objective information. Moreover, a *certification contract* $\gamma = (\sigma, t)$ consists of a certification structure $\sigma = (C, \pi^l, \pi^h)$ together with a price t which the seller has to pay to the certifier.

Following monopolistic screening, we capture the certifier's disclosure of private information by allowing the certifier to screen the different seller-types. That is, the certifier publicly commits to some menu $\Gamma = \{\gamma_i\}_{i \in I}$ of different certification contracts $\gamma_i = (\sigma_i, t_i)$ from some

[3] It is without loss to assume that the signal coincides with the seller's updated belief as we can represent any signal this way. Formally, it means that the signal θ has the cumulative distribution $F^\omega(\theta)$ in state ω and $F(\theta) \equiv \bar{\theta} F^h(\theta) + (1 - \bar{\theta}) F^l(\theta)$ has support Θ and is such that its expectation equals the prior $\bar{\theta}$. Moreover, after receiving the signal θ, the seller's belief updates to θ. That is, the pdfs $f^l(\theta)$ and $f^h(\theta)$ are such that $\theta/(1-\theta) = \bar{\theta} f^h(\theta)/(1-\bar{\theta}) f^l(\theta)$ for all $\theta \in \Theta$. The full support assumption is for technical convenience and not crucial.

set I. From the menu Γ, the privately informed seller can select a menu item $\gamma_i \in \Gamma$ or reject the menu altogether. If the seller selects a menu item, the seller's pick is publicly disclosed. Because the seller has the ability to pick any item from Γ regardless of her true type, the information conveyed by her pick will only be determined in equilibrium. In contrast to the certificate c, the seller's pick therefore represents soft information.

A menu $\Gamma = \{(\sigma_i, t_i)\}_{i \in I}$ induces the following *certification game* between seller and buyer:

$t = 1$: Seller publicly selects some contact $\gamma_i = (\sigma_i, t_i)$ from the menu Γ, or not;

$t = 2$: Certificate $c \in C$ is publicly drawn according to the picked certification structure σ_i (if any);

$t = 3$: Seller offers a quantity x at a price p to buyer;

$t = 4$: Buyer accepts or rejects;

$t = 5$: Payoffs result.

A perfect Bayesian Equilibrium (PBE) of this game consists of strategies for both players and their beliefs. The strategy of the seller describes her selection from the certification menu as well as her offer (x, p) conditional on the chosen menu-item and the realized certificate. The strategy of the buyer is an accept-or-reject decision conditional on the seller's choice from the menu, the realized certificate, and the seller's offer (x, p). The buyer's equilibrium-path beliefs are consistent with the seller's strategy of certification-contract selection and the probability distributions governing the realization of different certificates.[4]

For this setup, we are interested in characterizing the certifier's profit-maximizing menu, i.e., the menu that induces a certification game that yields the certifier the highest expected profit. Moreover, we are interested in the main economic properties of the certification game that the profit-maximizing menu induces, such as the degree of information revelation and efficiency of trade.

Illustrative Example. To illustrate our results, we use a fully parameterized uniform-quadratic example, where high and low quality are equally likely. We use this example not only to illustrate the insights of our more general model but also to point out the extent to which this natural example is special. For instance, while the optimal certification menu for this example will feature maximal revelation of soft information and no revelation of any hard information, the former feature is general, whereas the latter is not.

[4]Concerning the beliefs, we assume that the buyer's beliefs are *passive* vis-à-vis the seller's offer (x, p) *off the equilibrium path*. This is justified with the observation that the seller's payoff function is independent of the quality of the good and her private information about it.

Example: The seller's type θ is distributed uniformly over $[0, 1]$, implying $\bar{\theta} = 1/2$. For a low quality good, the buyer has the quadratic utility function $u_l(x) = -(1-x)^2$ so that the efficient quantity is $x_l^* = 1$, generating the normalized aggregate surplus of $u_l^* = u_l(1) = 0$. For a high quality good, the buyer has the quadratic utility function $u_h(x) = 1 - (2 - x)^2$ so that the efficient quantity is $x_h^* = 2$, generating the normalized aggregate surplus of $u_h^* = u_h(2) = 1$. Hence, from an ex ante perspective, the efficient solution generates an expected surplus of $\mathbb{E}_\omega\{u_\omega^*\} = \bar{\theta}u_h(x_h^*) + (1 - \bar{\theta})u_l(x_l^*) = 1/2$. □

3 Benchmarks and a Theorem

In this section, we present four natural benchmarks that inform about the key differences between soft and hard information. These benchmarks also illustrate the main intuition of a theorem that captures our main result that we prove in the remainder of the paper.

1. No Certification. Suppose there is no certifier. In this case, a seller θ cannot credibly convey her private information to the buyer. Because the seller's outside option and costs are type-independent, there is no adverse selection effect in the sense of Akerlof (1970). Hence, consistent with the buyer's passive beliefs, the buyer, upon observing a seller's offer (x, p), continues to expect high quality with probability $\bar{\theta}$ and low quality with probability $1 - \bar{\theta}$. Consequently, he expects a quantity x to yield a (gross) utility

$$\bar{u}(x) \equiv \mathbb{E}_\omega\{u_\omega(x)\} = \bar{\theta}u_h(x) + (1 - \bar{\theta})u_l(x).$$

Hence, it is sequentially rational for the buyer to accept an offer (x, p) if and only if $p \leq \bar{u}(x)$. Thus, for some quantity x, the monopolistic seller optimally charges $p = \bar{u}(x)$.

It follows that her profit-maximizing quantity, $\bar{x}$, maximizes $\bar{u}(x)$, yielding her the revenue $\bar{\Pi} \equiv \bar{u}(\bar{x})$. Clearly, the expressions $\bar{x}$ and $\bar{\Pi}$ depend on the ex ante probability, $\bar{\theta}$, that quality is high. For future reference, it is helpful to express this dependence explicitly. Thus, we write the buyer's optimal quantity at a belief θ as

$$\hat{x}(\theta) \equiv \arg\max_x \theta u_h(x) + (1 - \theta)u_l(x) \tag{1}$$

and the associated revenue function as

$$\hat{v}(\theta) \equiv \bar{u}(\hat{x}(\theta)) = \theta u_h(\hat{x}(\theta)) + (1-\theta)u_l(\hat{x}(\theta)). \tag{2}$$

Note that $0 < x_l^* = \hat{x}(0) \leq \hat{x}(\bar{\theta}) = \bar{x} \leq \hat{x}(1) = x_h^*$ and $0 = u_l^* = \hat{v}(0) \leq \bar{\Pi} = \hat{v}(\bar{\theta}) \leq \hat{v}(1) = u_h^* = 1$.

The following lemma plays a crucial role in our analysis.

Lemma 1 *The revenue function $\hat{v}(\theta)$ is strictly increasing and strictly convex in θ.*

Intuitively, the positive slope of $\hat{v}$ reflects that a high θ represents "good news" about the good's quality. Its strict convexity expresses that quality information is efficiency-relevant so that revealing additional information about the quality has a strictly positive social value.

Example: Applying this to our running example, the buyer's expected utility from a quantity x given a belief θ equals

$$\bar{u}_h(x) = \theta u_h(x) + (1-\theta)u_l(x) = (x-1)(1+2\theta-x).$$

It follows that the utility maximizing quantity equals $\hat{x}(\theta) = 1+\theta$, yielding the buyer the expected utility θ^2, which the seller can fully extract. Hence, in our example with quadratic utilities, we have

$$\hat{v}(\theta) = \theta^2,$$

confirming that $\hat{v}$ is increasing and convex. It then follows $\bar{x} = 1+\bar{\theta}$ and $\bar{\Pi} = \bar{\theta}^2$. □

2. Certification Without Soft or Hard Information. Next consider the presence of a monopolistic certifier, who however cannot provide any information to market participants, neither in the form of hard nor in the form of soft information. Effectively, the certifier can only use the non-informative certificate structure $\sigma^u \equiv (\{c_\emptyset\}, 1, 1)$, which, irrespective of actual quality ω, always generates the uninformative certificate $c_\emptyset$.

We verify first that, in line with Lizzeri (1999), the certifier can, by using the non-informative certification structure, extract the entire surplus that seller-types $\theta > 0$ generate.[5]. To see the full surplus extraction, suppose the certifier charges a certification price $t = \bar{\Pi}$ for providing

[5]Because in our context quality information is socially valuable, the uninformative certification structure σ^u does however not maximize aggregate surplus. Hence, more precisely, but fully in line with Lizzeri (1999), the uninformative certification structure σ^u allows the certifier to extract all (non-maximal) rents in excess of what the worst seller-type $\theta = 0$ obtains from the buyer. See also Kartik, Lee, and Suen (2021).

a fully uninformative certificate. That is, the certifier offers the uninformative pooling menu $\Gamma^u \equiv \{(\sigma^u, \bar{\Pi})\}$, containing only the unformative certification structure σ^u at the single price $t = \bar{\Pi}$.

The uninformative menu Γ^u sustains the following equilibrium outcome. All seller-types $\theta > 0$ apply for the certification, while seller-type $\theta = 0$ does not. Upon seeing no certification, the buyer correctly anticipates $\theta = 0$ and upon seeing a certificate, the buyer correctly anticipates $\mathbb{P}\{\omega = h\} = \bar{\theta}$.[6] Given these Bayes-consistent beliefs, it is then indeed optimal for a seller-type $\theta = 0$ not to apply for certification and sell quantity x_l^* at price $p = 0$, while it is optimal for seller-types $\theta > 0$ to apply for certification and, subsequently, sell a quantity $\hat{x}(\bar{\theta})$ of the certified good at the price $p = \bar{\Pi}$. In this outcome, the certifier fully extracts the generated surplus $\bar{\Pi}$ and the buyer and all seller-types obtain zero payoffs.

Hence, without providing any information, the certifier still manages to appropriate the entire aggregate surplus, obtaining profits $\Pi^u \equiv \bar{\Pi}$. The resulting aggregate surplus is however not maximized, as the induced quantity $\hat{x}(\bar{\theta})$ is inefficient.

3. Certification with only Soft Information. Next consider a monopolistic certifier that can disclose the seller's private information through screening, but cannot acquire additional hard information about the product's quality ω. That is, the certifier only uses the uninformative certification structure σ^u, as in the first benchmark.

We first argue that screening enables the certifier to *fully* disclose the seller's private information by offering the seller a menu of uninformative tests at different prices t. To show this, we consider a certification menu Γ that only contains the uninformative certification structure σ^u, but offers it at different prices. In particular, this menu allows the seller to buy the uninformative σ^u for any price in between 0 and 1, i.e., $\Gamma^s \equiv \{(\{\sigma^u, \hat{v}(\theta))\}_{\theta \in [0.1]}$.

We make the, perhaps at first sight paradoxical, claim that the menu Γ^s is "incentive compatible". That is, given Γ^s, it is optimal for any seller-type θ to reveal her type honestly by paying exactly $\hat{v}(\theta)$ to the certifier for obtaining the uninformative certificate σ^u. To the buyer, the seller's pick from the menu Γ^s fully reveals the seller-type θ.[7]

To substantiate our claim, we argue that Γ^s induces a certification game in which it is an equilibrium for any seller-type $\theta > 0$ to buy the certificate σ^u at a price $\hat{v}(\theta)$ and, subsequently, sell to the buyer the quantity $\hat{x}(\theta)$ at a price $\hat{v}(\theta)$ so that, in this equilibrium, the seller ends

[6] Hence, treating seller-type $\theta = 0$ as a type who does not certify clarifies that the equilibrium does not depend on some construed out-of-equilibrium belief when the buyer does not see any certification.

[7] Clearly, the menu Γ^s supports other equilibrium outcomes as well, but to illustrate the main point of this benchmark, we focus on the incentive compatible one.

up with a profit of zero. Moreover, seller-type $\theta = 0$ does not buy any certification, and sells the quantity $x_l^* = \hat{x}(0)$ at the price $\hat{v}(0) = 0$. Upon seeing the seller picking the certificate θ, the buyer has the Bayes-consistent belief $\mathbb{P}\{\omega = h\} = \theta$ so that it is optimal to accept the seller's offer $(x, p) = (\hat{x}(\theta), \hat{v}(\theta))$. For a good that remains uncertified, the buyer has the Bayes-consistent belief $\mathbb{P}\{\omega = h\} = 0$, and, hence, accepts the offer $(x, p) = (\hat{x}(0), \hat{v}(0))$. Given the buyer's specified behavior, it is indeed (weakly) optimal for the seller to reveal her type θ by picking her type-specific menu-item $(\sigma^u, \hat{v}(\theta))$ from Γ^s, as any item from Γ^s yields any seller-type a payoff of zero.

Because, in this equilibrium of Γ^s, the screening yields the certifier the transfer $\hat{v}(\theta)$ from a seller-type $\theta > 0$, the certification contract with full disclosure yields the certifier the revenue

$$\Pi^s \equiv \mathbb{E}_\theta\{\hat{v}(\theta)\} > \hat{v}(\mathbb{E}_\theta\{\theta\}) = \hat{v}(\bar{\theta}) = \Pi^u,$$

where the inequality follows from the strict convexity of the revenue function $\hat{v}(\theta)$ as established in Lemma 1.

We therefore conclude that, at least with respect to the uninformative certification structure σ^u, the certifier profits from providing complete soft information in the form of full screening. Moreover, the benchmark shows that the certifier still manages to extract the full surplus that the seller and buyer generate. This latter observation implies that the certifier does not need to concede any information rents to the seller for inducing full screening. Hence, for the specific case that the certification structure does not reveal any hard information, the intuition as suggested in the introduction that screening exhibits an inherent cost because of ensuring incentive compatibility is not compelling.

4. Certification with only Hard Information. Finally, consider a certifier who, using information acquisition, only provides hard information and no soft information, thus offering a pooling menu. In particular, suppose the certifier uses a certification scheme with a maximal degree of hard information, fully disclosing the good's quality ω. Effectively, the certifier can do so with the fully informative certification structure $\sigma^f \equiv (\{c_l, c_h\}, (1, 0), (0, 1))$, where certificate c_l obtains for sure when quality is low ($\omega = l$), while for high quality ($\omega = h$), only certificate c_h obtains. Disregarding in this benchmark any soft information so that the menu consists of only one fully informative contract, we thus consider the hard information menu $\Gamma^h = (\sigma^f, t^f)$, where we still have to specify the price t^f.

From an aggregate surplus perspective, the certification structure σ^f is highly attractive as it induces a certification game which results in the efficient quantity x_l^* when quality is low, and the efficient quantity x_h^* when quality is high. Also note that with such maximal information acquisition, disclosure of the seller's private information is superfluous; perfect hard information renders any soft information uninformative.

To the certifier, providing maximal hard information to all seller-types is however unattractive. This is so because a seller-type θ obtains at most $(1-\theta)\hat{v}(0)+\theta\hat{v}(1)=\theta$ from buying σ^f. Consequently, if the certifier wants to induce all seller-types $\theta>0$ to participate in Γ^h, paying a price t^f for σ^f, then we must have $t^f=0$; the certifier does not obtain any profits. Hence, we conclude that while feasible, it is not optimal for the certifier to maximize hard information and induce all seller-types to accept it.

A Theorem. The latter benchmark implies that a certifier faces a trade-off between profit maximization and information acquisition, while the third benchmark indicates that there is no such trade-off between profit maximization and screening. Taken together, the two suggest the following theorem.

Theorem 1 *The certifier's profit-maximizing menu of certification contracts reveals the seller's private information fully and reveals additional market information at most partially.*

In the remainder of this paper, we prove this theorem. In doing so, we identify further properties of the certifier's optimal menu. In particular, we show the crucial role that the curvature of the convex revenue function, $\hat{v}(\theta)$, plays—i.e., whether its average slope is larger or smaller than the geometric mean of its slopes.

We prove the theorem in two steps. First, we focus, in Section 4, on maximal-screening menus which provide a maximal degree of soft information. We derive the profit maximizing one among these menus and establish its properties. In particular, we show that, in general, optimal maximal-screening menus must also provide some degree of hard information by non-trivial information acquisition, but, in line with Benchmark 4, it is never strictly optimal that they provide hard information to the fullest extent. Second, we show, in Section 5, that certification menus that fully screen the seller's private information are indeed optimal.

Example: We close this section by analyzing the different benchmarks for our running example. We already established for the first benchmark, where buyers do not learn more about quality than its ex ante probability $\bar{\theta}$, our example generates aggregate surplus $\hat{v}(\bar{\theta})=1/4$. In Benchmark 2, the certifier extracts this surplus completely; the certifier's maximum profits are 1/4. For Benchmark 3, where the certifier fully discloses the seller's private information to the buyer but acquires no hard information, the certifier's payoff and ex ante expected aggregate surplus increase to $\mathbb{E}_\theta\{\hat{v}(\theta)\} = \int_0^1 \theta^2 d\theta = 1/3$. Finally, in the last benchmark of acquiring maximal hard information, a surplus of $\hat{v}(0) = 0$ obtains for low quality and the surplus $\hat{v}(1) = 1$ obtains for high quality. From an ex ante perspective, the expected aggregate surplus is $\bar{\theta}\hat{v}(0) + (1-\bar{\theta})\hat{v}(1) = 1/2$, while a seller-type θ expects the (gross) payoff θ and the certifier's payoff is zero. Note that the certifier appropriates the entire surplus in Benchmark 2 and 3, but does not get any rents in Benchmark 4, as seller-types arbitrarily close to 0 have no willingness to pay for the certification. □

4 Optimal Maximal-Screening Menus

We start with studying menus that, in equilibrium, fully disclose the seller's private information. We call such menus *maximal-screening* menus and denote them with Γ^S. In Section 5, we show that these menus are indeed optimal for the monopolistic certifier.

Without loss, we can capture maximal-screening menus by an incentive compatible, direct menu of certification contracts from which the seller publicly picks her type-specific certification contract, thereby revealing her private information to the buyer. We denote a direct menu contract by

$$\Gamma^D \equiv \{(C(\theta^r), \pi^l(\theta^r), \pi^h(\theta^r), t(\theta^r)\}_{\theta^r \in \Theta},$$

and the set of all direct menus as $\mathcal{M}^D$.

Hence, in this section, we restrict attention to direct menus Γ^D in $\mathcal{M}^D$ that induce a game in which it is optimal for a seller-type θ to report her type honestly, so that her type gets fully disclosed.[8]

By definition, a maximal-screening menu resolves the seller's private information completely so that, after the chosen certification contract is publicly revealed, the buyer and seller have the identical beliefs $\theta^B = \theta^S = \theta$. Hence, a maximal-screening menu is special in that, in any PBE, it holds that after the seller reports some θ and the information provider reveals the certificate

[8]To be sure, as we assume that the certifier only discloses publicly the chosen certification contract but not the seller's message θ^r, a maximal-screening menu also has to satisfy the (trivial) requirement that all items in the menu differ in the sense that $(C(\theta), \pi^l(\theta), \pi^h(\theta), t(\theta)) = (C(\theta'), \pi^l(\theta'), \pi^h(\theta'), t(\theta'))$ implies $\theta = \theta'$. This requirement can be satisfied trivially by labeling certificates in $C(\theta)$ and $C(\theta')$ differently even if the number of certificates in $C(\theta)$ and $C(\theta')$, and the probability vectors $(\pi^l(\theta), \pi^h(\theta))$ and $(\pi^l(\theta'), \pi^h(\theta'))$ do not differ.

c, the buyer's belief θ^B and the seller's belief θ^S coincide, and by Bayes' rule satisfy

$$\theta_c(\theta) \equiv \frac{\theta \pi_c^h(\theta)}{\theta \pi_c^h(\theta) + (1-\theta)\pi_c^l(\theta)} = \mathbb{P}\{\omega = h | c, \theta\}. \tag{3}$$

This means that, at the selling stage $t = 3$, there is no asymmetric information. That is, given (θ, c), the buyer and seller both expect a quantity x to yield the buyer the utility

$$\mathbb{E}_\omega\{u_\omega(x)|\theta, c\} = \theta_c(\theta) u_h(x) + (1 - \theta_c(\theta)) u_l(x).$$

Consequently, it is optimal for the buyer to accept an offer (x, p) if and only if $p \leq \mathbb{E}_\omega\{u_\omega(x)|\theta, c\}$.

Thus, given (θ, c), the seller's optimal offer (x, p) at stage 3 is straightforward. Given a quantity x, she optimally charges $p = \mathbb{E}_\omega\{u_\omega(x)|\theta, c\}$. Her profit-maximizing quantity equals $\hat{x}(\theta_c(\theta))$ with profits $\hat{v}(\theta_c(\theta))$ as defined by (1). Consequently, a maximal-screening menu Γ^S which generates a certificate c, yields a seller-type θ the revenue $\hat{v}(\theta_c(\theta))$ as defined in (2).

We next formalize incentive compatibility of a direct menu Γ^D, which is crucial for determining the certifier's optimal menu. For a maximal-screening menu, a seller-type θ expects the payoff $\hat{v}(\theta_c(\theta^r))$ from a certificate c after reporting some θ^r. Hence, when seller-type θ reports θ^r, she expects to obtain an expected profit of

$$\begin{aligned} \tilde{V}(\theta^r|\theta) &\equiv \sum_c \left[\theta \pi_c^h(\theta^r) + (1-\theta)\pi_c^l(\theta^r)\right] \hat{v}(\hat{\theta}_c(\theta^r)) - t(\theta^r) \\ &= \tilde{V}(\theta^r|\theta^r) + (\theta - \theta^r) I(\theta^r), \end{aligned}$$

where we can interpret

$$I(\theta) \equiv \sum_c [\pi_c^h(\theta) - \pi_c^l(\theta)] \hat{v}(\hat{\theta}_c(\theta)) \tag{4}$$

as the marginal information rent of seller-type θ.

Denoting by $V(\theta) \equiv \tilde{V}(\theta|\theta)$ the seller-type θ's rent from revealing her type truthfully, we have that Γ^D is incentive compatible with truth-telling if and only it holds for all $\theta, \theta^r \in [0, 1]$ that

$$V(\theta) \geq V(\theta^r) + (\theta - \theta^r) I(\theta^r). \tag{IC_{θ,θ^r}}$$

Moreover, a direct menu Γ^D is individual rational if for all $\theta \in \Theta$

$$V(\theta) \geq 0. \tag{IR_θ}$$

Hence, we can express a maximal-screening menu Γ as a direct menu Γ^D that satisfies (IC_{θ,θ^r}) and (IR_θ).

Because the seller's offer does not leave any rents to the buyer, the certifier's profit, $\Pi^c(\Gamma)$, from a maximal-screening menu $\Gamma \in \mathcal{M}^D$ is the difference between the surplus

$$S(\theta) \equiv \sum_j [\theta\pi_j^h(\theta) + (1-\theta)\pi_j^l(\theta)]\hat{v}(\hat{\theta}_j(\theta)) \tag{5}$$

and the seller's rent, $V(\theta)$. That is,

$$\Pi^c(\Gamma) = \int_0^1 [S(\theta) - V(\theta)]\ dF(\theta). \tag{6}$$

Hence, the certifier's optimal maximal-screening menu is a direct one that solves the following maximization program:

$$\mathcal{P}^S : \max_{\Gamma \in \mathcal{M}^D} \Pi^c(\Gamma) \text{ s.t. } (IC_{\theta,\theta^r}) \text{ and } (IR_\theta) \text{ for all } \theta, \theta^r \in \Theta. \tag{7}$$

We denote a solution to the program by $\hat{\Gamma}^S$ and its associated value by $\hat{V}^S = \Pi^c(\hat{\Gamma}^S)$.

To simplify this problem, we first characterize incentive compatibility and individual rationality for maximal-screening menus. To do so, we say that a direct menu Γ^D satisfies *monotonicity* if

$$I(\theta) \text{ is increasing;} \tag{MON}$$

and satisfies *payoff-equivalence* if

$$V(\theta) = \int_0^\theta I(\tau)d\tau + V(0). \tag{PE}$$

The following lemma characterizes a menu's incentive compatibility and its individual rationality in terms of monotonicity and payoff equivalence.

Lemma 2 *A direct menu Γ^D is incentive compatible if and only if it satisfies both (MON) and (PE). An incentive compatible direct menu Γ^D exhibits an increasing V. An incentive compatible direct menu Γ^D is individually rational if and only if $V(0) \geq 0$.*

The lemma motivates our interpretation of $I(\theta)$ as the marginal rent of a seller-type θ. Using an integration by parts, it also allows us to rewrite the certifier's objective (6) so that we can simplify $\mathcal{P}^S$ to

$$\hat{\mathcal{P}}^S: \max_{\Gamma \in \mathcal{M}^D} \hat{\Pi}^c(\Gamma) \equiv \int_0^1 \left[S(\theta) - \frac{1-F(\theta)}{f(\theta)} I(\theta) \right] dF(\theta) \text{ s.t. } (MON). \quad (8)$$

This simplified problem however still maximizes over a large set of direct menus. In particular, it is unclear what the optimal number of certificates is. In a non-divisible, unit-good model in which there are only the two allocations—to buy or not to buy—it follows already from feasibility considerations alone that it suffices to consider only two certificates, as this leads to at most two posteriors. Instead, in our setup with continuous quantities the feasibility considerations alone do not imply any limits on the optimal number of certificates.[9]

To address the question of the optimal number of certificates, we first derive the following result.

Lemma 3 *A maximal-screening menu with at most 3 certificates is optimal.*

We prove the lemma by showing that we can always replicate the value associated with an optimal maximal-screening menu that uses more than 3 certificates with a maximal-screening menu of at most 3 certificates. The main idea behind the proof is that if there is a solution that uses more than 3 posteriors, then we can reduce it to at most 3 posteriors. More specifically, the reduction to 3 posteriors follows because we can express the reduction as the outcome of a linear optimization problem with 3 constraints. The lemma then follows from the general insight that there is a solution to a linear problem that has at most as many non-zero entries as there are constraints.

Because one of the three constraints expresses the requirement that the linear optimization does not violate the monotonicity condition, the proof of the lemma suggests that if the monotonicity is not binding then only 2 posteriors suffice. In this case, one can use the following two-step procedure for finding optimal certification menus. As a first step, we solve program $\hat{\mathcal{P}}^S$ disregarding the monotonicity condition and with respect to certificate menus that involve 2 certificates only. As we disregard the monotonicity condition in this step, it requires only pointwise maximization. If this solution satisfies the monotonicity condition (MON), then, as we confirm in Lemma 4, it also represents an optimal maximal-screening certification menu in

[9]More specifically, the sufficiency of two posteriors in a unit good model reflects the revelation principle's insight that an optimal direct mechanism can be interpreted as providing incentive compatible recommendations over possible allocations—whether the good is bought or not. With continuous quantities, there is a continuum of possible allocations rather than only two and hence the revelation principle provides no limits on the number of possible recommendations.

general. In case the found solution violates the neglected monotonicity constraint (MON), we take the second step and solve the program $\hat{\mathcal{P}}^S$ with respect to menus that contain 3 certificates, including the monotonicity condition explicitly.

To demonstrate formally that this two step procedure is correct and also to operationalize it, we next study 2-certificate menus in more detail. Because 2-certificate menus exhibit $C(\theta) = \{c_0, c_1\}$ for all θ, we can characterize a 2-certificate menu by a menu of triples $\{(t(\theta), \pi^l(\theta), \pi^h(\theta))\}_\theta$, expressing the transfer t together with the probabilities π^ω that, for quality ω, the certificate c_1 is revealed.

Based on a relabeling argument, it is without loss to assume that $\pi^l(\theta) \leq \pi^h(\theta)$ so that certificate c_0 is more indicative of low quality and certificate c_1 is more indicative of high quality. Indeed, if the seller reports θ, thus selecting the pair $(\pi^l(\theta), \pi^h(\theta))$ from a 2-certificate menu, then this induces the belief

$$\hat{\theta}_0(\theta) = \frac{\theta(1-\pi^h(\theta))}{\theta(1-\pi^h(\theta)) + (1-\theta)(1-\pi^l(\theta))} \text{ and } \hat{\theta}_1(\theta) = \frac{\theta\pi^h(\theta)}{\theta\pi^h(\theta) + (1-\theta)\pi^l(\theta)} \tag{9}$$

after observing, respectively, a certificate c_0 and c_1, whenever this certificate obtains with a strictly positive probability. This confirms that $\pi^l(\theta) \leq \pi^h(\theta)$ implies $\hat{\theta}_0(\theta) \leq \theta \leq \hat{\theta}_1(\theta)$ for all θ.

Inverting the expressions in (9), we obtain the probabilities $\pi^h(\theta)$ and $\pi^l(\theta)$ that, for a given prior θ, generate the posteriors $q = \hat{\theta}_0$ after seeing certificate $c = c_0$ and $p = \hat{\theta}_1$ after seeing certificate $c = c_1$:

$$\pi^h(\theta) = \frac{(\theta-q)p}{(p-q)\theta} \text{ and } \pi^l(\theta) = \frac{(\theta-q)(1-p)}{(p-q)(1-\theta)}. \tag{10}$$

For 2-certificate contracts, it is easier to work directly with the posteriors (p, q) rather than their inducing probability (π^l, π^h). Hence, we express an incentive compatible 2-certificate menu Γ_2^D as a triple consisting of the transfer plus a pair of posterior beliefs (p, q) with $0 \leq q \leq \theta \leq p \leq 1$:

$$\Gamma_2^D = \{(t(\theta), q(\theta), p(\theta))\}_{\theta\in\Theta}.$$

We denote the set of 2-certificate menus with maximal-screening by Γ_2^D.

Defining

$$I_2(p, q|\theta) \equiv \frac{[p-\theta][\theta-q]}{(1-\theta)\theta} \frac{[\hat{v}(p) - \hat{v}(q)]}{(p-q)}, \tag{11}$$

we can further exploit the simple structure of 2-certificate menus to simplify the monotonicity requirement; the summation expression in (MON) reduces to $I_2(p(\theta), q(\theta)|\theta)$. It follows from

Lemma 2 that a 2-certificate menu Γ_2^D is incentive compatible if and only if

$$I_2(p(\theta), q(\theta)|\theta) \text{ is increasing in } \theta. \tag{MON_2}$$

Rewriting the objective of program $\hat{\mathcal{P}}^S$ for the specific case of 2-certificate menus in terms of the two posteriors, we obtain

$$V_2 \equiv \int_0^1 \left[S_2(p(\theta), q(\theta)|\theta) - \frac{1 - F(\theta)}{f(\theta)} I_2(p(\theta), q(\theta)|\theta) \right] dF(\theta) \tag{12}$$

where

$$S_2(p, q|\theta) \equiv \frac{\theta - q}{p - q} \hat{v}(p) + \frac{p - \theta}{p - q} \hat{v}(q). \tag{13}$$

That is, the function $S_2(p, q|\theta)$ expresses the surplus generated by a type θ who induces the posteriors q and p from a 2-certificate menu and the function $I_2(p, q|\theta)$ captures the associated marginal information rent when scaled by the inverse hazard rate.

Accordingly, we can express the first step of the procedure as follows. Find a pair of posterior functions $(p(\theta), q(\theta))$ with $p(\theta) \geq \theta \geq q(\theta)$ that maximizes the virtual surplus expression V_2 point-wise. That is, we solve for each $\theta \in \Theta$ the problem

$$\mathcal{R}_\theta^S : \max_{(p,q): q \leq \theta \leq p} S_2(p, q|\theta) - \frac{1 - F(\theta)}{f(\theta)} I_2(p, q|\theta).$$

The next lemma confirms that if the resulting pair $(\hat{p}(\theta), \hat{q}^*(\theta))$ satisfies the monotonicity constraint that $I_2(\hat{p}(\theta), \hat{q}(\theta), \theta)$ is increasing in θ, it solves $\hat{\mathcal{P}}^S$.

Lemma 4 *Suppose $(\hat{p}(\theta), \hat{q}(\theta))$ with value $\hat{V}_2$ solves the unconstrained optimization problem $\mathcal{R}_\theta^S$ for each $\theta \in \Theta$. If $(\hat{p}(\theta), \hat{q}(\theta))$ satisfies MON_2, then $\hat{V}_2$ is the value of program $\hat{\mathcal{P}}^S$, i.e. $\hat{V}^S = \hat{V}_2$.*

The proof of Lemma 4 mirrors the proof of Lemma 3, showing that, with the help of a linear maximization problem that now only consists of two constraints, one can replicate any solution with more than 2 posteriors by a maximal-screening menu with at most 2 posteriors. Hence, the lemma validates our 2-step procedure.

Illustrating its strengths, we now apply our 2-step procedure to our running example with the result that the first step alone already yields the optimal maximal-screening menu.

Example: Applying our first step, we first need to find, for each $\theta \in \Theta$, a pair (p^*, q^*) that is a solution to

$$\max_{(p,q):q\leq\theta\leq p} S_2(p,q|\theta) - \frac{1-F(\theta)}{f(\theta)} I_2(p,q|\theta) = 2\theta(p+q) - 3qp - p^2 - q^2 + \frac{pq}{\theta}(p+q).$$

First order conditions of this quadratic optimization problem yields the optimum $p^*(\theta)=\theta$ and $q^*(\theta)=0$ so that by (11), we have

$$I_2(p^*(\theta), q^*(\theta)|\theta) = I_2(\theta, 0|\theta) = 0,$$

which satisfies the monotonicity condition (MON_2) trivially. Hence, the optimal maximal-screening menu consists of a certification contract that does not involve any acquisition of hard information. Our general treatment will however show that the latter result is a special feature of the uniform-quadratic structure. □

Lemma 4 implies that problems, for which the monotonicity condition is non-binding, possess a high degree of tractability in two respects. First, they only require the consideration of 2-certificate menus. Second, they only require a point-wise maximization rather than a full-fledged dynamic optimization. Consequently, it would be helpful to characterize sufficient conditions on the model's primitives guaranteeing such that (MON) is non-binding. While in classical monopolistic screening problems, such sufficient conditions often exist in the form of assumptions only on the distribution of types, establishing a sufficient condition in our certification context requires also assumptions involving the revenue function $\hat{v}(\theta)$.[10] These latter conditions turn out to be rather strong; for instance, our uniform-quadratic example, for which we just showed that (MON) does not bind, does not satisfy these sufficient conditions.

We next show that, irrespective of the optimality of 2-certificate menus, the certifier's optimal certification structure depends crucially on the curvature of $\hat{v}$. In particular, on whether the geometric mean of its slopes is larger or smaller than the average of its slopes. To make this precise, consider a pair $p, q \in [0,1]$ with $p > q$ so that $\sqrt{\hat{v}'(p)\hat{v}'(q)}$ express the geometric

[10] To be precise, a sufficient condition is a decreasing inverse hazard rate $[1-F(\theta)]/f(\theta)$ together with a non-negative cross-partial derivative, $B_{I\theta}(I,\theta)$, of the function

$$B(I,\theta) \equiv \max_{\{p,q:q\leq\theta\leq p\wedge I_2(p,q)=I\}} S_2(p,q).$$

mean of the slopes at p and q. Moreover, let

$$w(p,q) \equiv \frac{\hat{v}(p) - \hat{v}(q)}{p - q} \tag{14}$$

express the average slope of $\hat{v}$ between the two points p and q. Note that the convexity of $\hat{v}$ implies that both the inequality $\hat{v}'(q) < w(p,q)$ and the inequality $\hat{v}'(p) > w(p,q)$ hold. By comparing the geometric mean to the average slope, we obtain a measure about which of these two inequalities dominates. That is, we evaluate the sign of the expression

$$A(p,q) \equiv \sqrt{\hat{v}'(p)\hat{v}'(q)} - w(p,q). \tag{15}$$

Using this expression, we get the following insight.

Lemma 5 *There exists an optimal maximal-screening menu, $\hat{\Gamma}^S = \{\sigma(\theta), t(\theta)\}_{\theta \in \Theta}$, such that for any $\theta \in \Theta$, and for any two distinct posteriors p and q with $p > q$ induced by $\sigma(\theta)$, it holds*

$$\begin{aligned} &A(p,q) \leq 0 \quad \textit{if } q = 0 \textit{ and } p < 1; \\ &A(p,q) = 0 \quad \textit{if } q > 0 \textit{ and } p < 1; \\ &A(p,q) \geq 0 \quad \textit{if } q > 0 \textit{ and } p = 1. \end{aligned} \tag{16}$$

To understand the intuition behind the lemma and its implications, consider a seller type θ who chooses a certification contract that could induce either p or q as a posterior. The conditions stated in the lemma correspond to the first order conditions that are necessary for maximizing the total surplus $S(\theta)$ subject to keeping the marginal information rent $I(\theta)$ fixed. If p and q do not satisfy these first order conditions, then a new certification contract can be constructed by modifying these two posteriors. This new contract leaves the same information rent $I(\theta)$ for the seller type θ and therefore continues to satisfy the monotonicity conditions of the program. Because the new contract yields a higher total surplus $S(\theta)$, while leaving the same information rents to the seller, it must yield the certifier a higher profit.

Note that these optimality conditions are independent of the seller type θ who chooses the contract generating the posteriors. Instead, they only refer to the geometric mean and the average of the slopes of function $\hat{v}$, evaluated between these posteriors.

Lemma 5 leads to an alternative, more practical sufficient condition for the optimality of 2-certificate menus than the one based on non-binding monotonicity constraints, as discussed above. More specifically, when the sign of $A(p,q)$ is independent of (p,q), then there exists an optimal menu relying on only two certificates. Stated more intuitively, the condition requires that the order of the geometric mean and its average slope does not depend on (p, q). Because the condition then also pins down one of the posteriors at the optimum, we report this finding as an independent result even though it is a direct implication of Lemma 5.

Corollary 1 *Suppose for all $p > q$, the geometric mean of the slopes at q and p lies below its average slope, i.e., $A(p, q) < 0$. Then an optimal maximal-screening menu is a 2-certificate menu with $q^*(\theta) = 0$. Suppose for all $p > q$, the geometric mean of the slopes at q and p exceeds its average slope, i.e., $A(p, q) > 0$. Then an optimal maximal-screening menu is a 2-certificate menu with $p^*(\theta) = 1$.*

Hence, any contract involving two certificates that would update the belief on the quality to two distinct but non-degenerate posteriors $p, q \in (0, 1)$ is suboptimal if $A(p, q)$ is of constant sign. The result follows directly form the explanation above that if a certification menu induces a seller-type θ to choose such a contract, the certifier can raise its profits by moving the posteriors down- or upward, depending on whether the geometric mean is smaller or larger than its average. Such an improvement only fails if one of the posteriors is already at the boundary of the feasible set $[0, 1]$.

To be clear, the two cases expressed in the corollary do not necessarily imply that a certification by hard information is optimal. Indeed, this would only be so if the fully revealing certificate is sent with a strictly positive probability, which the corollary does not claim. To the contrary, the corollary allows this probability to be zero, implying that the second certificate is actually sent with probability 1 so that the seller's and the buyer's beliefs about the state remain constant—the certification does not reveal any hard information. We demonstrate this subtle but important point with the help of our example.

Example: For our running example, we established that $\hat{v}(\theta) = \theta^2$, and, hence it follows that $A(p, q) = -(p - q)^2 < 0$ for all $p > q$. By Corollary 1, one of the two certificates in the optimal menu exhibits a degenerate posterior $q^*(\theta) = 0$, confirming the first step of the analysis as presented previously. This result suggests that, for this example, fully revealing the good's low quality ($\omega = l$) is optimal. However, when the seller's types are uniformly distributed ($F(\theta) = \theta$) on interval $[0, 1]$, an optimal certification menu sends this certificate of revealing the low quality with probability 0, confirming the first step of the analysis as presented above that $p^*(\theta) = \theta$. Effectively, the optimal contract boils down to a single-certificate contract that involves no acquisition of hard information but only discloses the seller's type θ to the buyer through screening. It is however important to point out that this is a special feature of the uniform distribution. For instance, when the type distribution is $F(\theta) = \sqrt{\theta}$, both certificates are sent with strictly positive probabilities for each seller-type in the interior of support $[0, 1]$ and, hence, display a positive degree of information gathering and fully reveal the state $\omega = l$ with a strictly positive probability.[11] Moreover, it is also important to note that, in addition to the uniform distribution, the quadratic utility in the example is also crucial for optimality of $q^*(\theta) = 0$. Indeed, the lemma indicates that this is not a general property of optimal maximal-screening contracts, as it depends on the specific functional form of $\hat{v}$.[12] □

Because the sign of $A(p, q)$ depends on two variables and can therefore be difficult to check, we provide a more practical sufficient condition that depends only on one variable. In particular, consider the following scaled curvature measure which measures how the slope, $\hat{v}'(\theta)$, changes relative to its curvature, $\hat{v}''(\theta)$:

$$\hat{A}(\theta) \equiv \frac{\hat{v}'(\theta)\sqrt{\hat{v}'(\theta)}}{\hat{v}''(\theta)}.$$

Lemma 6 *Suppose the scaled curvature measure $\hat{A}(\theta)$ is strictly increasing for $\theta > 0$. Then $A(p, q) < 0$ for all $p > q$. Suppose the scaled curvature measure $\hat{A}(\theta)$ is strictly decreasing for $\theta > 0$. Then $A(p, q) > 0$ for all $p > q$.*

With the help of this lemma, it is, for instance, straightforward to check that whenever the revenue function $\hat{v}(\theta)$ corresponds to a convex power function of the form $\hat{v}(\theta) = \theta^\alpha$ with $\alpha > 1$, then it holds $A(p, q) < 0$ so that an optimal maximal-screening menu is a 2-certificate menu with $q^*(\theta) = 0$.

Example: We conclude this section by showing that Lemma 6 allows us to obtain closed form solutions of optimal maximal-screening menus for settings which extend our example to cases where the revenue function is a convex power function, i.e., $v(\theta) = \theta^\alpha$ with $\alpha \geq 2$. Doing so enables us to identify the quadratic utility structure of our specific example as responsible for the suboptimality of any information gathering. Indeed, for a general power function, it follows

$$\hat{A}(\theta) = \frac{\alpha^{3/2}\theta^{3(\alpha-1)/2}}{\alpha(\alpha-1)\theta^{\alpha-2}} \text{ and } \hat{A}'(\theta) = \frac{(\alpha+1)\sqrt{\alpha\theta^{\alpha-1}}}{2(\alpha-1)} > 0.$$

[11] One may show that, optimally, $q_0^*(\theta) = 0$ and $p^*(\theta) = \frac{1}{4}\sqrt{\theta} + \frac{3}{4}\theta$.

[12] For instance, taking the (convex) function $v(\theta) = \theta/(2-\theta)$ yields $A(p, q) = 0$ for all p, q, implying again that optimal certification involve only two certificates. For this case, starting with any (p, q) pair, as we move these posteriors either up and down while keeping the information rent constant, the surplus function does not change. Accordingly, for each type θ we have a continuum of optimal certification tests, including the two extremes where one of the certificates is fully informative.

Hence, Lemma 6 implies that we can restrict attention to 2-certificate menus inducing the two-point belief $q(\theta) = 0$ and $p(\theta) \geq \theta$ for a seller-type θ. Using (11) and (13), it follows

$$I(p, 0|\theta) = \frac{p - \theta}{1 - \theta} p^{\alpha-1} \text{ and } S(p, 0) = \theta p^{\alpha-1}.$$

With a uniform distribution, a 2-certificate maximal-screening menu that induces the two posteriors $q(\theta) = 0$ and $p(\theta) \geq \theta$ for a seller-type θ, therefore yields the certifier

$$\int_0^1 \left[S(p(\theta), 0) - \frac{1 - F(\theta)}{f(\theta)} I(p(\theta), 0|\theta) \right] dF(\theta) = \int_0^1 [2\theta - p(\theta)] p(\theta)^{\alpha-1} d\theta.$$

Solving this pointwise for each $p(\theta)$, we obtain from first order conditions the optimum

$$p^*(\theta) = \begin{cases} \frac{2(\alpha-1)}{\alpha}\theta & \text{if } \theta \leq \frac{\alpha}{2(\alpha-1)} \\ 1 & \text{otherwise.} \end{cases}$$

For our specific example $\alpha = 2$, we therefore confirm $p^*(\theta) = \theta$, implying that the optimal maximal-screening menu does not reduce any market uncertainty. However, this is specific to the quadratic case $\alpha = 2$; for any $\alpha > 2$, the optimal maximal-screening menu reduces market uncertainty to at least some degree because $p^*(\theta) > \theta$ for all $\theta \in (0, 1)$. □

5 Optimality of Maximal-Screening Menus

The previous section only considers maximal-screening menus that fully reveal the seller's private information to the buyer. In this section, we show that such menus are indeed profit maximizing for the certifier. We demonstrate this by showing that, for any menu that does not fully reveal the seller's private information, the certifier can increase the degree of screening *while keeping the seller's information rents constant.* Because the increased amount of certification information raises aggregate surplus and because the certifier's profit is the difference between aggregate surplus and the seller's information rents, this raises the certifier's profits.

To show this formally, we first introduce notation so that we can capture non-maximal-screening menus in general. Intuitively, such a menu induces different types of sellers to pick the same certification contract $\{C, \pi^l, \pi^h, t\}$ from the menu so that the buyer does not fully learn the seller's type from observing the picked menu item. That is, the certification menu induces some seller-types to pool at the same certification contract.

To formally describe such menus, fix a (possibly non-direct) menu of contracts

$$\Gamma = \left\{C(m), \pi^l(m), \pi^h(m), t(m)\right\}_{m\in M},$$

where M is some message set. That is, from the menu Γ, the seller picks some menu option $m \in M$, committing the certifier to the certification structure $(C(m), \pi^l(m), \pi^h(m))$ for the transfer $t(m)$.

The menu Γ presents the seller with a game in which she first has to select a certification contract by sending some message m. Let $\hat{m} : \Theta \to M$ represent the seller's *messaging strategy*, describing that seller-type θ picks message $\hat{m}(\theta)$. Given a messaging strategy $\hat{m}$, let $\hat{\Theta}(m)$ represent the subset of types that pick message m.[13] That is,

$$\hat{\Theta}(m) \equiv \{\theta|\hat{m}(\theta) = m\}.$$

The non-empty sets $\hat{\Theta}(m)$ form a partition of Θ to which we refer by $\hat{M}$, i.e., $\hat{M} \equiv \{\hat{\Theta}(m)\}_{m\in M}$.

Hence, a menu Γ is *maximal* (with respect to a messaging strategy $\hat{m}$) if $\hat{\Theta}(m)$ is a singleton (or empty) for each $m \in M$. By contrast, a menu is non-maximal if it induces some types to pool and pick the same contract from the menu Γ. We therefore refer to non-maximal menus also as pooling or bunching ones.

A non-maximal menu reveals information about the seller's average type $\bar{\theta}$. To capture such information revelation, let $\bar{\theta} : M \to \Theta$ denote the expected value type associated with a message m as induced by the partition $\hat{M}$. That is,

$$\bar{\theta}^M(m) \equiv \frac{\int_{\theta\in\hat{\Theta}(m)} \theta dF(\theta)}{\int_{\theta\in\hat{\Theta}(m)} dF(\theta)}.$$

The mapping $\bar{\theta}^M(m)$ captures the disclosure information that underlies the certification menu. In particular, a maximal-screening menu exhibits $\bar{\theta}^M(\hat{m}(\theta)) = \theta$ for all $\theta \in \Theta$.

To capture the menu's acquisition of hard information, let $\hat{\theta}_c^M : M \to \Theta$ denote the mapping such that $\hat{\theta}_c^M(m)$ represents the Bayes' consistent expectation of θ given the seller's message $m \in \cup_\theta \hat{M}(\theta)$ and a certificate $c \in C(m)$. I.e, for any $m \in M$ and $c \in C(m)$, we have

$$\hat{\theta}_c^M(m) \equiv \frac{\bar{\theta}^M(m)\pi_c^h(m)}{\bar{\theta}^M(m)\pi_c^h(m) + (1 - \bar{\theta}^M(m))\pi_c^l(m)}.$$

[13] We focus on sellers sending a deterministic message (we think this without loss).

Thus for a non-maximal-screening menu, the change in beliefs from $\bar{\theta}$ to $\bar{\theta}^M(m)$ represents the disclosure via soft information, whereas the change in beliefs from $\bar{\theta}^M(m)$ to $\hat{\theta}_c^M(m)$ expresses the extent to which the certification menu provides hard information.

Note that if $\hat{\Theta}(m)$ is not a singleton, then, at stage $t = 3$, the buyer is facing a seller who is privately informed about his type $\theta \in \hat{\Theta}(m)$. Hence, when the seller offers a contract (p, x) to the buyer, the buyer may interpret this offer to be informative about the seller's type. That is, we formally have an informed principal problem sustaining multiple equilibrium outcomes. It is straightforward to see that the seller's optimal equilibrium outcome is supported by the buyer's belief being independent of the seller's offer, i.e., they equal $\bar{\theta}^M(m)$ for any offer (x, p).

In this case, the seller optimally offers the quantity $x^m = \hat{x}(\bar{\theta}^M(m))$ at the price $p^m = \mathbb{E}_\omega\{u_\omega(x^m)|\bar{\theta}^M(m)\}$, independent of his type θ. While there are equilibrium outcomes in which the seller obtains less, there are no equilibrium outcomes in which some seller-type θ obtains more.[14]

Equipped with this notation, we can extend the notion of incentive compatibility to non-maximal menus. In particular, we say that a menu Γ together with the seller's messaging strategy $\hat{m}$ is incentive compatible if for any type θ, the message $\hat{m}(\theta)$ yields type θ the highest payoff among all messages $m \in M$. That is, for each type $\theta \in \Theta$, the following incentive compatibility constraint is satisfied

$$\begin{aligned} \sum_c \left[\theta\pi_c^h(\hat{m}(\theta)) + (1-\theta)\pi_c^l(\hat{m}(\theta))\right] \hat{v}(\hat{\theta}_c^M(\hat{m}(\theta))) - t(\hat{m}(\theta)) \geq \\ \sum_c \left[\theta\pi_c^h(m) + (1-\theta)\pi_c^l(m)\right] \hat{v}(\hat{\theta}_c^M(m)) - t(m), \quad \forall m \in M, \end{aligned} \tag{17}$$

where we set $\hat{\theta}_c^S(m) = 0$ for any off-equilibrium message m with $\hat{\Theta}(m) = \emptyset$.

While, for generality, this representation of incentive compatibility explicitly allows for menus comprising of contracts that are not picked in equilibrium, these off-equilibrium choices are inconsequential, both in terms of payoffs and for supporting $\hat{m}$ as an equilibrium. It is therefore without loss to restrict attention to menus that do not contain contracts that are, in equilibrium, not picked, i.e., $\hat{\Theta}(m) \neq \emptyset$ for all $m \in M$. As a result, we may take the partition $\hat{M}$ as a primitive of a general (i.e., possibly non-maximal) certification menu and take the range

[14] E.g. the buyer having beliefs $\bar{\theta}^M(m)$ only if the seller offers quantity $\hat{x}$ with at a price $p = p^m - \delta$ (with $\delta > 0$ small) and believes to face the lowest possible type otherwise, sustains an equilibrium in which all seller-types in $\hat{\Theta}(m)$ obtain the lesser profit $p^m - \delta$. By contrast, there is no equilibrium with a profit higher for some types, since this higher payoff then has to be obtained for all types, while also requiring that the buyer has a belief exceeding $\bar{\theta}^M(m)$ for any offer that these types make, contradicting Bayes' rule that on average this must equal $\bar{\theta}^M(m)$.

of $\hat{m}$ to coincide with M.

Consequently, we say that a (general) certification menu is incentive compatible with respect to messaging strategy $\hat{m} : \Theta \to M$ if for all $\theta, \theta^r \in \Theta$ we have

$$\hat{V}(\theta) \geq \hat{V}(\theta^r) + (\theta - \theta^r) I^{\hat{m}}(\theta), \qquad (\widehat{IC}_{\theta,\theta^r})$$

where

$$\hat{V}(\theta) \equiv \sum_c \left[\theta \pi_c^h(\hat{m}(\theta)) + (1-\theta)\pi_c^l(\hat{m}(\theta))\right] \hat{v}(\hat{\theta}_c^M(\hat{m}(\theta))) - t(\hat{m}(\theta)) \qquad (18)$$

expresses seller-type θ's utility of sending her message $\hat{m}(\theta)$, and

$$I^{\hat{m}}(\theta) \equiv \sum_c \left[\pi_c^h(\hat{m}(\theta)) - \pi_c^l(\hat{m}(\theta))\right] \hat{v}(\hat{\theta}_c^M(\hat{m}(\theta))) \qquad (19)$$

expresses the marginal information rent of type θ. Note that $(\widehat{IC}_{\theta,\theta^r})$ differs from (IC_{θ,θ^r}) in that for a non-maximal menu, we have to keep track of the message $\hat{m}(\theta)$ that seller-type θ sends.

Lemma 7 *A (general) certification menu* Γ *is incentive compatible with respect to messaging strategy* $\hat{m} : \Theta \to M$ *if and only if it satisfies both*

i) monotonicity

$$I^{\hat{m}}(\theta) \text{ is increasing in } \theta. \qquad (\widehat{MON})$$

ii) and payoff-equivalence

$$\hat{V}(\theta) = \int_0^\theta I^{\hat{m}}(\tau) d\tau + \hat{V}(0). \qquad (\widehat{PE})$$

The lemma extends Lemma 2 in that for a direct maximal-screening menu we implicitly have the identity $\hat{m}(\theta) = \theta$ as the messaging strategy, in which case $(\widehat{MON})$ and $(\widehat{PE})$ reduce to (MON) and (PE), respectively.

Representing both maximal and non-maximal screening contracts as direct mechanisms, we can view the certifier as solving a two-step maximization procedure. In the first step, the certifier decides how to partition the type set into subsets that would choose different items from menu Γ. I.e., the certifier determines which mapping $\bar{\theta}(.)$ to implement. In the second step, the certifier will choose what $\left(C, \pi^l, \pi^1, t\right)$ combination to offer to each subset, subject to the monotonicity constraint above to maximize the expected revenue:

$$\tilde{V}^C(\Gamma) \equiv \int_0^1 \sum_c \left\{ \begin{array}{l} \theta\pi_c^h(\hat{m}(\theta)) + (1-\theta)\pi_c^l(\hat{m}(\theta)) \\ \quad -\frac{1-F(\theta)}{f(\theta)} \left[\pi_c^h(\hat{m}(\theta)) - \pi_c^l(\hat{m}(\theta))\right] \end{array} \right\} \hat{v}(\hat{\theta}_c^S(\hat{m}(\theta)))dF(\theta).$$

Maximal-screening amounts to choosing the finest partition in the first step of this procedure.

As the proofs of Lemma 3, Lemma 4, and Lemma 5 readily extend to non-maximal screening partitions, we have:

1. A menu with at most 3-certificate contracts is optimal.
2. If $\widehat{MON}$ does not bind for the optimal 2-certificate menu, then 2-certificate menus are optimal.
3. If, for all $p > q$, it holds $A(p,q) < 0$, then a 2-certificate menu with $q^*(\theta) = 0$ is optimal. If, for all $p > q$, it holds $A(p,q) > 0$, then a 2-certificate menu with $p^*(\theta) = 1$ is optimal.

A general (possibly non-maximal) 2-certificate menu $\hat{\Gamma}_2$ exhibits $C(\theta) = \{c_p, c_q\}$ for any $\theta \in \Theta$. We can fully characterize it by the partition $\hat{M}$ of Θ and a pair of posterior mappings $(p(\theta), q(\theta))$ with $q(\theta) \leq p(\theta)$ and the requirement that $(p(\theta_1), q(\theta_1)) = (p(\theta_2), q(\theta_2))$ if θ_1 and θ_2 come from the same partition cell in $\hat{M}$. The posterior $p(\theta)$ represents the probability that the buyer assigns to the good state $\omega = h$ after certificate c_p, while the posterior $q(\theta)$ represents this probability after certificate c_q. The assumption $q(\theta) \leq p(\theta)$ expresses the labeling convention that certificate p is more indicative of the good state than certificate q.

In terms of posterior pairs $(p(\theta), q(\theta))$ and using (14), we can express the marginal surplus that a seller-type θ generates as

$$\hat{S}(\theta) \equiv \hat{v}(q(\theta)) + [\theta - q(\theta)]w(p(\theta), q(\theta)) = \hat{v}(p(\theta)) - [p(\theta) - \theta]w(p(\theta), q(\theta)). \tag{20}$$

Based on (19) and using (14), we can rewrite type θ's marginal information rent as

$$\hat{I}(\theta) = \frac{p(\theta) - \bar{\theta}(\theta)}{1 - \bar{\theta}(\theta)} \frac{\bar{\theta}(\theta) - q(\theta)}{\bar{\theta}(\theta)} w(p(\theta), q(\theta)). \tag{21}$$

By Lemma 7, the pair $(p(\theta), q(\theta))$ is implementable if and only if the marginal surplus $\hat{I}(\theta)$ is weakly increasing in θ.

Integrating over seller-types, we obtain the surplus from certification and information rents

$$\hat{S} = \int_0^1 \hat{S}(\theta)dF(\theta) \text{ and } \hat{I} = \int_0^1 \frac{1-F(\theta)}{f(\theta)}\hat{I}(\theta)dF(\theta), \tag{22}$$

respectively. Thus, an implementable 2-certificate menu $\hat{\Gamma}_2$ yields the certifier the expected profit $\hat{V}^c = \hat{S} - \hat{I}$.

Lemma 8 *In the class of 2-certificate menus, maximal-screening menus are optimal.*

While the result sounds straightforward, the proof is involved. In particular, it is based on the observation that, whenever a 2-certificate menu is non-maximal then this implies that over some subset $\hat{\Theta} \subseteq \Theta$ all types are bunched in the sense that there exist $\bar{q} \leq \bar{p}$ such that $q(\theta) = \bar{q}$ and $p(\theta) = \bar{p}$ for all $\theta \in \hat{\Theta}$. We show in the proof that such bunching is not optimal because we can raise the surplus by fully separating the types in the subset $\hat{\Theta}$, while keeping the marginal information rents $I(\theta)$ to each seller-type constant. The key of the proof is to construct type-specific posteriors $q(\theta)$ and $p(\theta)$ for all θ so that we obtain a surplus function $\check{S}(\theta)$ that is convex.

It is easiest to see the construction of the alternative separating certification menu when the types in subset $\hat{\Theta}$ are bunched by a certification structure that does not provide any hard information, i.e, when either $\bar{q}$ or $\bar{p}$ equals the average type $\bar{\theta}(\hat{\Theta})$ in $\hat{\Theta}$. Such a certification contract extracts the full surplus $\hat{S} = \hat{v}(\bar{\theta}(\hat{\Theta}))$ without leaving any information rent ($I(\theta) = 0$) for $\theta \in \hat{\Theta}$. The corresponding separating menu fully discloses the types in $\hat{\Theta}$ by still leaving zero information rent to each of them. The surplus function resulting from this separating menu $\check{S}(\theta) = \hat{v}(\theta)$ is indeed convex, implying a higher revenue for the certifier than does the original bunching contract, because $\mathbb{E}\{\hat{v}(\theta)\} > \hat{v}(\mathbb{E}\{\theta\})$.

When a subset of types are bunched by a contract revealing some market information to the buyer ($\bar{q} < \bar{\theta}(\hat{\Theta}) < \bar{p}$), then there exist different separating menus ($q(\theta)$ and $p(\theta)$) resulting in the same marginal information rent $I(\theta)$ as in the original bunching contract. The proof of the lemma identifies one such menu that yields a convex surplus function $\check{S}(\theta)$.

We next extend the insight of the lemma to 3-certificate contracts. We do so by showing that similar to the case of 2-certificate contracts, we can also construct a convex surplus function $\check{S}(\theta)$ for 3-certificate contracts that pool a subset of seller types. The construction is however more involved as the convex function $\check{S}(\theta)$ will typically consist of two parts: an interval of types for which 3-certificates are used and an interval of types for which the surplus function

$\check{S}(\theta)$ uses only two certificates. Because menus with at most 3 certificates are optimal, we finally establish the paper's theorem with the following result.

Proposition 1 *Maximal-screening menus are optimal in general.*

6 Conclusion

The paper develops a model that allows to distinguish the two natural roles of certification — screening and information acquisition, where screening conveys soft information and information acquisition conveys hard information. It demonstrates that a monopolistic certifier optimally engages in full screening, disclosing all available soft information, while engaging in at most a limited degree of acquisition of hard information. Consequently, monopolistic certification eradicates all market inefficiencies due to private information, but market inefficiencies due to market uncertainty remain.

The intuition behind this result is that even though the certifier could eliminate all market uncertainty by acquiring full hard information costlessly, this would be suboptimal as it requires leaving excessive information rents to the partially privately informed seller. While this intuition emphasizes the role of information rents, it runs counter to the alternative but incorrect intuition that because of the need of incentive compatibility and hence information rents, screening should be an inherently more costly channel to disclose information than the acquisition of hard information.

Appendix

This appendix collects the proofs of the propositions and lemmas in the main text.

Proof of Lemma 1: We first prove that $\hat{x}(\theta)$ is strictly increasing. To see this note that $\hat{x}$ is defined by the first order condition

$$\theta u_h'(\hat{x}) + (1-\theta)u_l'(\hat{x}) = 0.$$

Applying the implicit function theorem, it follows

$$u_h'(\hat{x}) + \theta u_h''(\hat{x})\frac{\partial \hat{x}}{\partial \theta} - u_l'(\hat{x}) + (1-\theta)u_l''(\hat{x})\frac{\partial \hat{x}}{\partial \theta} = 0 \Rightarrow \frac{\partial \hat{x}}{\partial \theta} = -\frac{u_h'(\hat{x}) - u_l'(\hat{x})}{\theta u_h''(\hat{x}) + (1-\theta)u_l''(\hat{x})} > 0,$$

where the strict inequality follows from the single crossing condition $u_h'(\hat{x}) > u_l'(\hat{x})$, and the strict concavity of u.

To see that $\hat{v}$ is strictly increasing, note that by the envelope theorem:

$$\hat{v}'(\theta) = \frac{\partial \hat{v}}{\partial \theta} + \frac{\partial \hat{v}}{\partial x}\frac{\partial \hat{x}}{\partial \theta} = \frac{\partial \hat{v}}{\partial \theta} = u_h(\hat{x}(\theta)) - u_l(\hat{x}(\theta)) > 0,$$

where the strict inequality follows from single crossing.

To see that $\hat{v}$ is strictly convex, consider a pair of beliefs $\theta_1 < \theta_2$, and convex combination $\tilde{\theta} = \pi\theta_1 + (1-\pi)\theta_2$ with $\pi \in (0,1)$ so that $\theta_1 < \tilde{\theta} < \theta_2$. Because, as established above, $\hat{x}(\theta)$ is strictly increasing, we have $\hat{x}(\theta_1) < \hat{x}(\tilde{\theta}) < \hat{x}(\theta_2)$. Hence, it follows

$$\begin{aligned}
\hat{v}\,(\theta_1) = \theta_1 u_h(\hat{x}(\theta_1)) + (1-\theta_1)u_l(\hat{x}(\theta_1)) &> \theta_1 u_h(\hat{x}(\tilde{\theta})) + (1-\theta_1)u_l(\hat{x}(\tilde{\theta}));\\
\hat{v}\,(\theta_2) = \theta_2 u_h(\hat{x}(\theta_2)) + (1-\theta_2)u_l(\hat{x}(\theta_2)) &> \theta_2 u_h(\hat{x}(\tilde{\theta})) + (1-\theta_2)u_l(\hat{x}(\tilde{\theta})).
\end{aligned}$$

Taking the average of these two strict inequalities with the weights π and $(1-\pi)$ yields

$$\pi\hat{v}(\theta_1) + (1-\pi)\hat{v}(\theta_2) > \tilde{\theta}u_h(\tilde{x}) + (1-\tilde{\theta})u_l(\tilde{x}) = \hat{v}(\tilde{\theta}).$$

This shows the strict convexity of $\hat{v}$. $\square$

Proof of Lemma 2: To see that incentive compatibility implies (MON), consider (IC_{θ,θ^r}) and $(IC_{\theta^r,\theta})$:

$$V(\theta) \geq V(\theta^r) + (\theta - \theta^r)I(\theta^r);$$

$$V(\theta^r) \geq V(\theta) + (\theta^r - \theta)I(\theta).$$

Together they imply

$$(\theta - \theta^r)[I(\theta) - I(\theta^r)] \geq 0.$$

Hence, incentive compatibility implies (MON).

To see that incentive compatibility implies (PE), rewrite, for $\delta > 0$, the incentive constraint $(IC_{\theta+\delta,\theta})$ as

$$\frac{V(\theta+\delta) - V(\theta)}{\delta} \geq I(\theta)$$

Likewise, for $\delta > 0$, rewrite $(IC_{\theta-\delta,\theta})$ as

$$\frac{V(\theta) - V(\theta-\delta)}{\delta} \leq I(\theta)$$

At a point of differentiability of $V(\theta)$, we have

$$I(\theta) \leq \lim_{\delta \downarrow 0} \frac{V(\theta+\delta) - V(\theta)}{\delta} = V'(\theta) = \lim_{\delta \downarrow 0} \frac{V(\theta) - V(\theta-\delta)}{\delta} \leq I(\theta),$$

implying $V'(\theta) = I(\theta)$ so that (PE) then follows from the fundamental theorem of calculus.

We now show sufficiency of (MON) and (PE) for incentive compatibility. When $\theta > \theta^r$, applying (PE), condition (IC_{θ,θ^r}) can be rewritten as

$$V(\theta) - V(\theta^r) = \int_{\theta^r}^{\theta} I(\tau)d\tau \geq (\theta - \theta^r)I(\theta^r),$$

which follows from (MON), since $I(\tau) \geq I(\theta^r)$ for all $\tau \geq \theta^r$. Similarly, when $\theta < \theta^r$, condition (IC_{θ,θ^r}) reduces to

$$V(\theta) - V(\theta^r) = -\int_{\theta}^{\theta^r} I(\tau)d\tau \geq -(\theta^r - \theta)I(\theta^r), \tag{23}$$

which also holds, since $I(\tau) \leq I(\theta^r)$ for all $\tau \leq \theta^r$.

To see the second statement, note that by (IC_{θ,θ^r}) the result follows directly if for any θ we have

$$I(\theta) = \sum_c \left[\pi_c^h(\theta) - \pi_c^l(\theta)\right] \hat{v}(\hat{\theta}_c(\theta)) \geq 0 \tag{24}$$

To see that this inequality holds, fix θ and, without loss, order the certificates in $C(\theta)$ such that the likelihood ratio $\pi_c^h(\theta)/\pi_c^l(\theta)$ is increasing in c. Hence, $\pi^h(\theta)$ dominates $\pi^l(\theta)$ in the sense of FOSD. Note that because the likelihood ratio is increasing, also $\hat{v}(\theta_c(\theta))$ is increasing in c. As FOSD implies that the expectation of any increasing functions is larger under $\pi^h(\theta)$ than under the FOSD-ed $\pi^l(\theta)$, it holds in particular for the increasing sequence $\hat{v}(\theta_c(\theta))$. Hence, $\sum_c \pi_c^h(\theta)\hat{v}(\hat{\theta}_c(\theta)) \geq \sum_c \pi_c^l(\theta)\hat{v}(\theta_c(\theta))$, which implies (24). Finally, because $V(\theta)$ is increasing, condition (IR_0) is sufficient for (IR_θ) for all θ. □

Proof of Lemma 3: Fix some solution $\hat{\Gamma} = \{(\hat{C}(\theta), \hat{\pi}^l(\theta), \hat{\pi}^h(\theta), \hat{t}(\theta)\}_{\theta\in[0,1]}$ to the problem $\hat{\mathcal{P}}^S$ with associated value $\hat{\Pi}^c(\hat{\Gamma})$. Let $\Theta^+ \subseteq [0,1]$ denote the subset of types θ for which the solution $\hat{\Gamma}$ uses more than 3 certificates. We next argue that if $\Theta^+ \neq \emptyset$, then we can construct an alternative solution $\check{\Gamma}$ that has at most three certificates for each θ which is also a solution to $\hat{\mathcal{P}}^S$ with associated value $\hat{\Pi}^c(\hat{\Gamma})$. To see this, let

$$\hat{S}(\theta) \equiv \sum_c \left[\theta\hat{\pi}_c^h(\theta) + (1-\theta)\hat{\pi}_c^l(\theta)\right]\hat{v}(\hat{\theta}_c(\theta))$$

represent the surplus that a type θ generates under solution $\hat{\Gamma}$. Likewise, let

$$\hat{I}(\theta) = \sum_c [\hat{\pi}_c^h(\theta) - \hat{\pi}_c^l(\theta)]\hat{v}(\hat{\theta}_c(\theta))$$

express the marginal information rent of seller-type θ under solution $\hat{\Gamma}$. We have

$$\hat{\Pi}^c(\hat{\Gamma}) = \int_0^1 \hat{\Pi}(\theta)dF(\theta) \text{ where } \hat{\Pi}(\theta) \equiv \hat{S}(\theta) - \frac{1-F(\theta)}{f(\theta)}\hat{I}(\theta).$$

Next for any $\theta \in \Theta^+$, consider the linear problem indexed by θ of finding a solution $\alpha(\theta) = (\alpha_1(\theta), \ldots, \alpha_N(\theta))$ to:

$$\hat{\mathcal{P}}^\alpha(\theta): \max_{(\alpha_1,\ldots,\alpha_N)\geq 0} \sum_c \alpha_c \left\{\theta\hat{\pi}_c^h(\theta) + (1-\theta)\hat{\pi}_c^l(\theta) - \frac{1-F(\theta)}{f(\theta)}\left[\hat{\pi}_c^h(\theta) - \hat{\pi}_c^l(\theta)\right]\right\}\hat{v}(\hat{\theta}_c(\theta))$$
$$\text{s.t.} \sum_c \alpha_c\hat{\pi}_c^l(\theta) = 1; \sum_c \alpha_c\hat{\pi}_c^h(\theta) = 1; \sum_c \alpha_c[\hat{\pi}_c^h(\theta) - \hat{\pi}_c^l(\theta)]\hat{v}(\hat{\theta}_c(\theta)) = \hat{I}(\theta).$$

As $\alpha_1(\theta) = \ldots = \alpha_N(\theta) = 1$ is feasible in program $\hat{\mathcal{P}}^\alpha(\theta)$, the program's value, $\hat{\Pi}^\alpha(\theta)$, is at least $\hat{\Pi}(\theta)$. Moreover, note that $\hat{\mathcal{P}}^\alpha(\theta)$ is a linear program with only 3 constraints. Consequently, it has a solution $\hat{\alpha}(\theta)$ with at most 3 entries being non-zero. For each $\theta \in \Theta^+$, let these three entries be indexed by the triple $(k(\theta), l(\theta), m(\theta))$. Given the solution $\hat{\alpha}(\theta)$, we construct a 3-certificate menu $\check{\Gamma}$ as follows. For each $\theta \in \Theta^+$ and $\omega \in \{0,1\}$, let $\check{C}(\theta) \equiv \{c_{k(\theta)}, c_{l(\theta)}, c_{m(\theta)}\}$,

$\check{\pi}^{\omega}_{k(\theta)}(\theta) \equiv \hat{\alpha}_{k(\theta)}(\theta)\pi^{\omega}_{k(\theta)}(\theta)$, $\check{\pi}^{\omega}_{l(\theta)}(\theta) \equiv \hat{\alpha}_{l(\theta)}(\theta)\pi^{\omega}_{l(\theta)}(\theta)$, $\check{\pi}^{\omega}_{m(\theta)}(\theta) \equiv \hat{\alpha}_{m(\theta)}(\theta)\pi^{\omega}_{m(\theta)}(\theta)$, $\check{t}(\theta) \equiv t(\theta)$. Now construct $\check{\Gamma} = \{\check{\Gamma}(\theta)\}_{\theta\in[0,1]}$ by setting $\check{\Gamma}(\theta) = (\check{C}(\theta), \check{\pi}^l(\theta), \check{\pi}^h(\theta), \check{t}(\theta))$ for $\theta \in \Theta^+$ and $\check{\Gamma}(\theta) = \hat{\Gamma}(\theta)$ for $\theta \notin \Theta^+$ By construction, the 3-certificate menu $\check{\Gamma}$ satisfies monotonicity, because the marginal information rents are unchanged. It is therefore incentive compatible and yields $\hat{\Pi}^c(\check{\Gamma}) = \int_{\theta\in\Theta^+} \hat{\Pi}^{\alpha}(\theta)dF + \int_{\theta\notin\Theta^+} \hat{\Pi}(\theta)dF \geq \int_0^1 \hat{\Pi}(\theta)dF = \hat{\Pi}^c(\hat{\Gamma})$. But since $\hat{\Gamma}$ itself is optimal, we must also have $\hat{\Pi}^c(\hat{\Gamma}) \geq \hat{\Pi}^c(\check{\Gamma})$, implying that also the 3-certificate menu $\check{\Gamma}$ must be optimal. This proves the claim that 3-certificate menus are optimal. □

Proof of Lemma 4: Let $(\hat{p}(\theta), \hat{q}(\theta))$ with value $\hat{V}_2$ (as defined by (12)) be such that it satisfies (MON_2) and solves $\mathcal{R}^S_\theta$ for each $\theta \in \Theta$. Let $\hat{\Gamma}_2 = \{(\{c_0, c_1\}, \hat{\pi}^l(\theta), \hat{\pi}^h(\theta), \hat{t}(\theta)\}_{\theta\in[0,1]}$ represent the corresponding 2-certificate menu in terms of the probabilities $(\hat{\pi}^l, \hat{\pi}^h)$, satisfying (10) with $p = \hat{p}(\theta)$ and $q = \hat{q}(\theta)$, and the transfer $\hat{t}(\theta)$, yielding seller-type θ the information rent $\hat{I}(\theta)$ induced by $(\hat{p}(\theta), \hat{q}(\theta))$. Recalling (6), let $\Pi^c(\hat{\Gamma}_2)$ represent the certifier's profit of the 2-certificate menu $\hat{\Gamma}$. As it is a solution to $\mathcal{R}^S_\theta$, it corresponds to the value, $V_2^R(\theta)$, of program $\mathcal{R}^S_\theta$, i.e., $\hat{V}_2 = \Pi^c(\hat{\Gamma}_2) = \int_0^1 V_2^R(\theta)dF$.

Next note that $\hat{\Gamma}_2$ is feasible in program $\hat{\mathcal{P}}^S$. Hence, the value $\hat{V}^S$ of program $\hat{\mathcal{P}}^S$ is at least $\hat{V}_2$, i.e. $\hat{V}^S \geq \hat{V}_2$.

Next consider a solution of a relaxed version of $\hat{\mathcal{P}}^S$ that disregards the constraint (MON). Let $\hat{V}^R$ denote the value of this program and note that $\hat{V}^R \geq \hat{V}^S$.

Let $\hat{\Gamma} = \{(\hat{C}(\theta), \hat{\pi}^l(\theta), \hat{\pi}^h(\theta), \hat{t}(\theta)\}_{\theta\in[0,1]}$ denote a solution to the relaxed program. Note that for each θ, the collection $\{(\hat{\pi}^l_c(\theta), \hat{\pi}^h_c(\theta))\}_{c\in C(\theta)}$ solves

$$\hat{V}(\theta) = \max_{\{\pi^l_c, \pi^h_c\}_{c\in C(\theta)}} \sum_c \left[\left\{\theta\pi^h_c + (1-\theta)\pi^l_c - \frac{1-F(\theta)}{f(\theta)}\left[\pi^h_c - \pi^l_c\right]\right\} \hat{v}(\theta_c(\theta))\right]. \tag{25}$$

Hence, $V^R = \int_0^1 \hat{V}(\theta)dF$

Next, consider the linear program of finding a solution $\alpha(\theta) = (\alpha_1, \ldots, \alpha_N)$ to:

$$\mathcal{P}^{\alpha}(\theta) : \max_{(\alpha_1,\ldots,\alpha_N)\geq 0} \sum_c \left[\alpha_c \left\{\theta\hat{\pi}^h_c(\theta) + (1-\theta)\hat{\pi}^l_c(\theta) - \frac{1-F(\theta)}{f(\theta)}\left[\hat{\pi}^h_c(\theta) - \hat{\pi}^l_c(\theta)\right]\right\} \hat{v}(\theta_c(\theta))\right]$$
$$\text{s.t. } \sum_c \alpha_c \pi^l_c(\theta) = 1; \sum_c \alpha_c \pi^h_c(\theta) = 1.$$

As $\alpha_1 = \ldots = \alpha_N = 1$ is feasible in program $\mathcal{P}^{\alpha}(\theta)$, its value, $\hat{V}^{\alpha}(\theta)$, is at least $\hat{V}(\theta)$, i.e. $\hat{V}^{\alpha}(\theta) \geq \hat{V}(\theta)$. Moreover, note that this is a linear optimization problem under two linear constraints. Hence, a solution $\hat{\alpha}(\theta) = (\hat{\alpha}_1(\theta), \ldots, \hat{\alpha}_N(\theta))$ exists with at most two entries in $\hat{\alpha}(\theta)$

as non-zero, say $k(\theta)$ and $l(\theta)$.

Given the solutions $\hat{\alpha}(\theta)$, we construct a 2-certificate menu $\check{\Gamma}$ as follows. Let $\check{C}(\theta) \equiv \{c_{k(\theta)}, c_{l(\theta)}\}$, $\check{\pi}^{\omega}_{k(\theta)}(\theta) \equiv \hat{\alpha}_{k(\theta)}(\theta)\pi^{\omega}_{c}(\theta)$, $\check{\pi}^{\omega}_{l(\theta)}(\theta) \equiv \hat{\alpha}_{l(\theta)}(\theta)\pi^{\omega}_{c}(\theta)$, $\check{t}(\theta) \equiv t(\theta)$, and let $\check{\Gamma} = \{(\check{C}(\theta), \check{\pi}^{l}(\theta), \check{\pi}^{h}(\theta), \check{t}(\theta)\}_{\theta\in[0,1]}$. By construction, this 2-certificate contract has value $V^{\hat{\alpha}} \equiv \int_0^1 \hat{V}^{\alpha}(\theta)dF \geq \int_0^1 \hat{V}(\theta)dF = \hat{V}^R$

But given that the menu $\check{\Gamma}$ is a 2-certificate menu, it cannot generate a higher value than the optimal 2-certificate menu $\hat{\Gamma}_2$, i.e. $\hat{V}_2 \geq V^{\hat{\alpha}}$. We therefore obtain the string of weak inequalities $\hat{V}_2 \geq V^{\hat{\alpha}} \geq \hat{V}^R \geq \hat{V}^S \geq \hat{V}_2$, so that, in fact, they must hold with equality so that $\hat{V}^S = \hat{V}_2$. □

Proof of Lemma 5: Following Lemma 3, we prove the lemma by considering an optimal maximal-screening menu $\hat{\Gamma}^S = \{\hat{\sigma}(\theta), \hat{t}(\theta)\}_{\theta\in\Theta}$ that has at most 3 certificates. That is, for any $\theta \in \Theta$, the certification structure $\hat{\sigma}(\theta) = (\hat{C}(\theta), \hat{\pi}^l(\theta), \hat{\pi}^h(\theta))$ induces at most three distinct posteriors. If $\hat{\sigma}(\theta)$ exhibits only one certificate, then $\hat{\sigma}(\theta)$ does not induce two distinct posteriors and the lemma holds trivially. Thus, to prove the lemma we distinguish between the case that $\hat{\sigma}(\theta)$ induces, with a strict positive probability, exactly 2 distinct posteriors, and the case in which $\hat{\sigma}(\theta)$ induces, with a strict positive probability, exactly 3 distinct posteriors. We prove that for either case, there is an optimal Γ^S satisfying (16).

Before doing so, we define

$$\tilde{A}(p,q) \equiv v'(p)v'(q) - w(p,q)^2$$

so that the sign of $A(p,q)$ coincides with the sign of $\tilde{A}(p,q)$.

Case 1: $\hat{\sigma}(\theta) = (\{\hat{c}_p, \hat{c}_q\}, (1-\hat{\pi}^l_p(\theta), \hat{\pi}^l_p(\theta)), (1-\hat{\pi}^h_p(\theta), \hat{\pi}^h_p(\theta)))$ *with* $\hat{\pi}^h_p(\theta) > 0$, $\hat{\pi}^l_p(\theta) < 1$, *and* $\hat{\pi}^h_p(\theta) > \hat{\pi}^l_p(\theta)$.[15] By Bayes' consistency and (3), it holds $q < \theta < p$ and

$$q = \frac{\theta(1-\hat{\pi}^h_p)}{\theta(1-\hat{\pi}^h_p) + (1-\theta)(1-\hat{\pi}^l_p)}; \quad p = \frac{\theta\hat{\pi}^h_p}{\theta\hat{\pi}^h_p + (1-\theta)\hat{\pi}^l_p}, \tag{26}$$

where we dropped the arguments θ in $\hat{\pi}^h_p(\theta)$ and $\hat{\pi}^l_p(\theta)$.

Solving for $\hat{\pi}^h_p$ and $\hat{\pi}^l_p$ in (26) yields

$$\hat{\pi}^l_p(p,q) = \frac{(1-p)(\theta-q)}{(p-q)(1-\theta)}; \; \hat{\pi}^h_p(p,q) = \frac{p(\theta-q)}{(p-q)\theta}. \tag{27}$$

[15]The conditions $\hat{\pi}^h_p(\theta) > 0$ and $\hat{\pi}^l_p(\theta) < 1$ ensure that the two certificates $\hat{c}_p$ and $\hat{c}_q$ obtain with a strict positive probability, while $\hat{\pi}^h_p(\theta) > \hat{\pi}^l_p(\theta)$ ensures $p > q$, as stated in the lemma.

Using (4), it then follows that with 2 certificates, type θ's marginal information rent $I(p,q|\theta)$ equals

$$I(p,q|\theta) = [\pi_p^h(p,q) - \pi_p^l(p,q)][\hat{v}(p) - \hat{v}(q)] = \frac{(p-\theta)(\theta-q)}{(1-\theta)\theta} w(p,q). \tag{28}$$

Hence, by the implicit function theorem, the information rent does not change if we change p and q such that

$$\frac{\partial p}{\partial q} = -\frac{\partial I/\partial q}{\partial I/\partial p} = \frac{(p-\theta)[(p-\theta)w(p,q) + (\theta-q)v'(q)]}{(\theta-q)[(p-\theta)v'(p) + (\theta-q)w(p,q)]} \geq 0. \tag{29}$$

Hence, for $q > 0$, we can decrease q and p marginally without affecting type θ's information rent. Similarly, we can increase q and p marginally without affecting type θ's information rent if $p < 1$.

To assess the effect of such changes of q and p on the certifier's profits, it suffices to determine how it affects the surplus, because the certifier's profits is the difference between surplus and information rents, and the increase in q and p is such that the latter is kept constant.

Using (27), we can express the surplus (5) with two certificates in two equivalent formulations

$$\begin{aligned} S(p,q|\theta) &= \hat{v}(q) + [\theta(1-q) - (1-\theta)q]w(p,q) \\ &= \hat{v}(p) + [\theta(1-p) - (1-\theta)p]w(p,q). \end{aligned}$$

Consequently,

$$\frac{\partial S}{\partial p} = [\theta(1-q) - (1-\theta)q]\frac{\partial w}{\partial p} \text{ and } \frac{\partial S}{\partial q} = [\theta(1-p) - (1-\theta)p]\frac{\partial w}{\partial q}.$$

It therefore follows that an increase in p and q that leaves type θ's information rent constant changes the surplus by

$$\frac{dS}{dq} = \frac{\partial S}{\partial p}\frac{\partial p}{\partial q} + \frac{\partial S}{\partial q} = \frac{p-\theta}{(p-\breve{\theta})\hat{v}'(p) + (\theta-q)w(p,q)} \tilde{A}(p,q),$$

where the fraction is strictly positive. Hence, the sign of $\tilde{A}(p,q)$ determines the effect on the certifier's profits. As the sign of $\tilde{A}(p,q)$ coincides with the sign of $A(p,q)$, we obtain (16). Because, if $\tilde{A}(p,q) > 0$ and $p < 1$, we can increase the certifier's profits by increasing q and p. And, if, by contrast, $\tilde{A}(p,q) < 0$ and $q > 0$, we can increase the certifier's profits by decreasing q and p.

Case 2: $\hat{\sigma}(\theta) = (\{\hat{c}_1, \hat{c}_2, \hat{c}_3\}, (\hat{\pi}_1^l(\theta), \hat{\pi}_2^l(\theta), \hat{\pi}_3^l(\theta)), (\hat{\pi}_1^h(\theta), \hat{\pi}_2^h(\theta), \hat{\pi}_3^h(\theta)))$ *with* $\hat{\pi}_1^h(\theta) + \hat{\pi}_1^l(\theta) > 0$, $\hat{\pi}_2^h(\theta) + \hat{\pi}_2^l(\theta) > 0$, $\hat{\pi}_3^h(\theta) + \hat{\pi}_3^l(\theta) > 0$.[16] Dropping the argument θ, the certification structure $\hat{\sigma}$ induces posteriors

$$\hat{p}_1 = \frac{\theta\hat{\pi}_1^h}{\theta\hat{\pi}_1^h + (1-\theta)\hat{\pi}_1^l}; \hat{p}_2 = \frac{\theta\hat{\pi}_2^h}{\theta\hat{\pi}_2^h + (1-\theta)\hat{\pi}_2^l}; \hat{p}_3 = \frac{\theta\hat{\pi}_3^h}{\theta\hat{\pi}_3^h + (1-\theta)\hat{\pi}_3^l},$$

where we label the posteriors such that $\hat{p}_1 > \hat{p}_2 > \hat{p}_3$ and, hence, Bayes' consistency implies $\hat{p}_1 > \theta > \hat{p}_3$. These posteriors yield seller-type θ the marginal information rent

$$\hat{I} \equiv [\hat{\pi}_1^h - \hat{\pi}_1^l]v(\hat{p}_1) + [\hat{\pi}_2^h - \hat{\pi}_2^l]v(\hat{p}_2) + [\hat{\pi}_3^h - \hat{\pi}_3^l]v(\hat{p}_3).$$

We show that, if $\hat{\sigma}(\theta)$ is part of an optimal menu, then there is no loss in assuming that the posteriors $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$ satisfy the conditions stated in the lemma (where p and q in the lemma correspond to a pair of posteriors, i.e. $(p, q) \in \{(\hat{p}_1, \hat{p}_2), (\hat{p}_1, \hat{p}_3), (\hat{p}_2, \hat{p}_3)\}$).

To see this, first note that if $\hat{\sigma}$ is optimal, then it necessarily maximizes the certifier's surplus

$$S = [\theta\pi_1^h + (1-\theta)\pi_1^l]v(p_1) + [\theta\pi_2^h + (1-\theta)\pi_2^l]v(p_2) + [\theta\pi_3^h + (1-\theta)\pi_3^l]v(p_3) \tag{30}$$

subject to the following six constraints:

1. The information rent being constant at $\hat{I}$,

$$[\pi_1^h - \pi_1^l]v(p_1) + [\pi_2^h - \pi_2^l]v(p_2) + [\pi_3^h - \pi_3^l]v(p_3) = \hat{I}; \tag{31}$$

2. the feasibility conditions

$$\pi_1^l + \pi_2^l + \pi_3^l = 1; \tag{32}$$

$$\pi_1^h + \pi_2^h + \pi_3^h = 1; \tag{33}$$

[16]The conditions $\hat{\pi}_1^h(\theta) + \hat{\pi}_1^l(\theta) > 0$, $\hat{\pi}_2^h(\theta) + \hat{\pi}_2^l(\theta) > 0$, and $\hat{\pi}_3^h(\theta) + \hat{\pi}_3^l(\theta) > 0$ ensure that each of the three certificates $\hat{c}_1, \hat{c}_2, \hat{c}_3$ obtains with a strict positive probability.

3. the three posterior conditions

$$[\theta\pi_1^h + (1-\theta)\pi_1^l]p_1 = \theta\pi_1^h; \tag{34}$$

$$[\theta\pi_2^h + (1-\theta)\pi_2^l]p_2 = \theta\pi_2^h; \tag{35}$$

$$[\theta\pi_3^h + (1-\theta)\pi_3^l]p_3 = \theta\pi_3^h. \tag{36}$$

The six constraints (30)-(36) form a linear system of the six variables $(\pi_1^h, \pi_1^l, \pi_2^h, \pi_2^l, \pi_3^h, \pi_3^l)$. For a combination (θ, p_1, p_2, p_3) with $1 \geq p_1 > p_2 > p_3 \geq 0$ and $p_1 > \theta > p_3$, define

$$D(p_1, p_2, p_3, \theta) \equiv (p_2 - p_3)(p_1 - \theta)(v(p_1) - v(p_3)) + (p_1 - p_3)(\theta - p_2)(v(p_2) - v(p_3))$$

so that the solution of this linear system with respect to (p_1, p_2, p_3, θ) exhibits

$$\pi_1^h = \frac{p_1}{\theta}\pi_1;\ \pi_1^l = \frac{1-p_1}{1-\theta}\pi_1 \text{ with } \pi_1 \equiv \frac{(\theta-p_2)(\theta-p_3)[v(p_2)-v(p_3)] + (1-\theta)\theta(p_2-p_3)I}{D(p_1,p_2,p_3)};$$

$$\pi_2^h = \frac{p_2}{\theta}\pi_2;\ \pi_2^l = \frac{1-p_2}{1-\theta}\pi_2 \text{ with } \pi_2 \equiv \frac{(p_1-\theta)(\theta-p_3)[v(p_1)-v(p_3)] - (1-\theta)\theta(p_1-p_3)I}{D(p_1,p_2,p_3)};$$

$$\pi_3^h = \frac{p_3}{\theta}\pi_3;\ \pi_3^l = \frac{1-p_3}{1-\theta}\pi_3 \text{ with } \pi_3 \equiv \frac{(p_1-\theta)(p_2-\theta)[v(p_1)-v(p_2)] + (1-\theta)\theta(p_1-p_2)I}{D(p_1,p_2,p_3)}. \tag{37}$$

Substituting these solutions for $(\pi_1^h, \pi_1^l, \pi_2^h, \pi_2^l, \pi_3^h, \pi_3^l)$ into (30), we obtain after a rearrangement of terms that $(\hat{p}_1, \hat{p}_2, \hat{p}_3)$ maximizes the expression

$$\begin{aligned} S(p_1,p_2,p_3,\theta) = \frac{1}{D(p_1,p_2,p_3)} \times \Big[& [\theta-p_2][\theta-p_3][v(p_2)-v(p_3)]v(p_1) \\ & + [p_1-\theta][\theta-p_3][v(p_1)-v(p_3)]v(p_2) \\ & + [p_1-\theta][p_2-\theta][v(p_1)-v(p_2)]v(p_3) \\ & + \{[p_2-p_3]v(p_1) + [p_3-p_1]v(p_2) + [p_1-p_2]v(p_3)\}\theta(1-\theta)I \Big]. \end{aligned} \tag{38}$$

Taking derivatives and using $w(.,.)$ as defined in (14), we can rewrite them as follows

$$\begin{aligned}
\frac{\partial S(p_1,p_2,p_3,\theta)}{\partial p_1} &= \frac{(p_1-p_3)\pi_1}{(p_1-\theta)w(p_1,p_2)+(\theta-p_3)w(p_2,p_3)}[w(p_2,p_3)v'(p_1)-w(p_1,p_3)w(p_1,p_2)];\\
\frac{\partial S(p_1,p_2,p_3,\theta)}{\partial p_2} &= \frac{(p_1-p_3)\pi_2}{(p_1-\theta)w(p_1,p_2)+(\theta-p_3)w(p_2,p_3)}[w(p_1,p_3)v'(p_2)-w(p_2,p_3)w(p_1,p_2)];\\
\frac{\partial S(p_1,p_2,p_3,\theta)}{\partial p_3} &= \frac{(p_1-p_3)\pi_3}{(p_1-\theta)w(p_1,p_2)+(\theta-p_3)w(p_2,p_3)}[w(p_1,p_2)v'(p_3)-w(p_1,p_3)w(p_2,p_3)],
\end{aligned}$$

where all the fractions are strictly positive.

Now suppose posteriors $(\hat{p}_1,\hat{p}_2,\hat{p}_3)$ are optimal.

If any pair $(p,q)\in\{(\hat{p}_1,\hat{p}_2),(\hat{p}_1,\hat{p}_3),(\hat{p}_2,\hat{p}_3)\}$ of these posteriors are interior, then their optimality implies that the two derivatives corresponding to the posteriors evaluated at $(\hat{p}_1,\hat{p}_2,\hat{p}_3)$ equals zero. For the case $(p,q)=(\hat{p}_1,\hat{p}_2)$, it follows

$$w(\hat{p}_2,\hat{p}_3)v'(\hat{p}_1)=w(\hat{p}_1,\hat{p}_3)w(\hat{p}_1,\hat{p}_2) \text{ and } w(\hat{p}_1,\hat{p}_3)v'(\hat{p}_2)=w(\hat{p}_2,\hat{p}_3)w(\hat{p}_1,\hat{p}_2).$$

From which it follows $\tilde{A}(p,q)=0$. Likewise, the cases $(p,q)=(\hat{p}_1,\hat{p}_3)$ and $(p,q)=(\hat{p}_2,\hat{p}_3)$ also imply $\tilde{A}(p,q)=0$.

Similarly, if $q=0$ and $p<1$, then their optimality implies that, evaluated at $(\hat{p}_1,\hat{p}_2,\hat{p}_3)$, we have $\partial S/\partial q\leq 0$ and $\partial S/\partial p=0$. For the case $(p,q)=(\hat{p}_1,\hat{p}_2)$, it follows

$$w(\hat{p}_2,\hat{p}_3)v'(\hat{p}_1)=w(\hat{p}_1,\hat{p}_3)w(\hat{p}_1,\hat{p}_2) \text{ and } w(\hat{p}_1,\hat{p}_3)v'(\hat{p}_2)\leq w(\hat{p}_2,\hat{p}_3)w(\hat{p}_1,\hat{p}_2)$$

From which it follows $\tilde{A}(p,q)\leq 0$. Likewise, the cases $(p,q)=(\hat{p}_1,\hat{p}_3)$ and $(p,q)=(\hat{p}_2,\hat{p}_3)$ also imply $\tilde{A}(p,q)\leq 0$.

Similarly, if $q>0$ and $p=1$, then their optimality implies that, evaluated at $(\hat{p}_1,\hat{p}_2,\hat{p}_3)$, we have $\partial S/\partial q=0$ and $\partial S/\partial p\geq 0$. That is, it holds

$$w(\hat{p}_2,\hat{p}_3)v'(\hat{p}_1)\geq w(\hat{p}_1,\hat{p}_3)w(\hat{p}_1,\hat{p}_2) \text{ and } w(\hat{p}_1,\hat{p}_3)v'(\hat{p}_2)=w(\hat{p}_2,\hat{p}_3)w(\hat{p}_1,\hat{p}_2)$$

From which it follows $\tilde{A}(p,q)\geq 0$. Likewise, the cases $(p,q)=(\hat{p}_1,\hat{p}_3)$ and $(p,q)=(\hat{p}_2,\hat{p}_3)$ also imply $\tilde{A}(p,q)\geq 0$. □

Proof of Lemma 6: Define for $q,p\in[0.1]$ with $p>q$,

$$g(p,q)\equiv v'(p)v'(q) \text{ and } h(p,q)\equiv w(p,q)^2.$$

We extend the domain of $g(p,q)$ and $h(p,q)$ to include $p=q$ by setting $g(p,p)=\hat{g}(p)$ and $h(p,p)=\hat{h}(p)$, where $\hat{g}(p)\equiv v'(p)^2$ and $\hat{h}(p)\equiv \lim_{q\uparrow p} h(q,p)=v'(p)^2$.

Note that $g(p,q)\in[v'(0)^2,v'(1)^2]$ and $h(p,q)\in[v'(0)^2,v'(1)^2]$. Moreover,

$$A(p,q)=\sqrt{g(p,q)}-\sqrt{h(p,q)}$$

so that $A(p,q)=0$ if and only if $g(p,q)=h(p,q)$. As $g(p,p)=h(p,p)=v'(p)^2$, we have, in particular, $A(p,p)=\lim_{q\uparrow p} A(p,q)=0$.

Next, for any $k\in[v'(0)^2,v'(1)^2]$, define functions $q_g(p|k)$ and $q_h(p|k)$ implicitly by

$$g(p,q_g(p|k))=k \text{ and } h(p,q_h(p|k))=k. \tag{39}$$

By the implicit function theorem, we obtain for $p>q$

$$q_g'(p|k)=-\frac{v''(p)v'(q)}{v''(q)v'(p)}\leq 0 \text{ and } q_h'(p|k)=-\frac{v'(p)-w(p,q)}{w(p,q)-v'(q)}\leq 0. \tag{40}$$

In addition, note that we have $\hat{g}(p)=\hat{h}(p)=v'(p)^2$ so that the convexity of $v(\theta)$ implies that $\hat{h}(p)$ and $\hat{g}(p)$ are continuous and increasing with $\hat{h}(0)=\hat{g}(0)=v'(0)^2$ and $\hat{h}(1)=\hat{g}(1)=v'(1)^2$. Hence, for any $k\in[v'(0)^2,v'(1)^2]$, we can find a $p(k)\in[0,1]$ such that

$$q_g(p(k)|k)=q_h(p(k)|k)=p(k).$$

Thus, we have established that for any $k\in[v'(0)^2,v'(1)^2]$, we can find a $p(k)\in[0,1]$ such that $A(p(k),p(k))=0$. Hence, the curves $q_g(p|k)$ and $q_h(p|k)$ intersect at the point $(p(k),p(k))$.

We next argue that, while the curves have the same slope of $q_g'(p(k)|k)=q_h'(p(k)|k)=-1$ at the intersection point $(p,q)=(p(k),p(k))$, $\hat{A}'(\theta)>0$ implies $q_g''(p(k)|k)>q_h''(p(k)|k)$ so that $q_g(p(k)+\varepsilon|k)>q_h(p(k)+\varepsilon|k)$ for all $\varepsilon>0$ small. Indeed, evaluating $q_g'(p|k)$ at $p=q=p(k)$ by taking limits, we obtain

$$q_g'(p(k)|k)=\lim_{\varepsilon\to 0}-\frac{v''(p)v'(p-\varepsilon)}{v''(p-\varepsilon)v'(p)}=-1$$

Likewise, evaluating $q_h'(p|k)$ at $p=q=p(k)$, we obtain after applying Hopital's rule twice

$$q_h'(p(k)|k)=\lim_{\varepsilon\to 0}-\frac{v(p)-v(p-\varepsilon)-\varepsilon v'(p)}{v(p)+v(p-\varepsilon)+\varepsilon v'(p-\varepsilon)}=\lim_{\varepsilon\to 0}-\frac{v''(p-\varepsilon)}{v''(p-\varepsilon)+\varepsilon v'''(p-\varepsilon)}=-1$$

Implicitly differentiating (39) twice, we evaluate the 2nd order derivative of $q_g(p|k)$ at $(p,q) = (p(k), p(k))$ to obtain

$$q_g''(p(k)|k) = \frac{2v''(p(k))}{v'(p(k))} - \frac{2v'''(p(k))}{v''(p(k))}.$$

Evaluating moreover the 2nd order derivative of $q_h(p|k)$ at $(p,q) = (p(k), p(k))$ by taking limits, we obtain after using l'Hopital's rule multiple times that

$$q_h''(p(k)|k) = -\frac{2v'''(p(k))}{3v''(p(k))}.$$

Thus, we obtain

$$q_g''(p(k)|k) > q_h''(p(k)|k) \Leftrightarrow \frac{2[3v''(p)^2 - 2v'(p)v'''(p)]}{3v'(p)v''(p)} \geq 0 \Leftrightarrow 3v''(p)^2 - 2v'(p)v'''(p)) \geq 0 \Leftrightarrow \hat{A}'(\theta) > 0. \tag{41}$$

We next argue that if $\hat{A}(\theta)$ is strictly increasing then the previous result implies that the two curves do not intersect for any $p > p(k)$. We do so by showing that a strictly increasing $\hat{A}$ implies $q_g'(p|k) > q_h'(p|k)$ at any point (p,q) with $p > q$ where the curves intersect (i.e., where $q_g(p|k) = q_h(p|k)$) and, hence, at any intersection point the curve $q_g(p|k)$ cuts $q_h(p|k)$ from above, implying that the two curves have an intersection point only at $(p,q) = (p(k), p(k))$. To see $q_g'(p|k) > q_h'(p|k)$ for any intersection point, note first that at an intersection point (p,q), it holds

$$v'(p)v'(q) = w(p,q)^2. \tag{42}$$

Using (42), it holds for $q_h'(p|k)$ at an intersection point (p,q) that

$$q_h'(p|k) = -\frac{v'(p) - w(p,q)}{w(p,q) - v'(q)} = -\frac{v'(p)w(p,q) - w(p,q)^2}{w(p,q)^2 - w(p,q)v'(q)} = -\frac{v'(p)w(p,q) - v'(p)v'(q)}{v'(p)v'(q) - w(p,q)v'(q)} = \frac{v'(p)}{v'(q)}\frac{1}{q_h'(p|k)}.$$

Hence, it holds

$$q_h'(p|k)^2 = \frac{v'(p)}{v'(q)},$$

which with $q_h'(p|k) \leq 0$ implies

$$q_h'(p|k) = -\frac{\sqrt{v'(p)}}{\sqrt{v'(q)}}.$$

We therefore have

$$q_g'(p|k) > q_h'(p|k) \Leftrightarrow \frac{v''(p)v'(q)}{v''(q)v'(p)} < \frac{\sqrt{v'(p)}}{\sqrt{v'(q)}} \Leftrightarrow \frac{v'(q)\sqrt{v'(q)}}{v''(q)} < \frac{v'(p)\sqrt{v'(p)}}{v''(p)} \Leftrightarrow \hat{A}(q) < \hat{A}(p), \tag{43}$$

which, given that $\hat{A}(\theta)$ is strictly increasing for $\theta > 0$, holds true for any $p > q \geq 0$. This yields the first statement of the lemma.

By contrast, if $\hat{A}(\theta)$ is strictly decreasing all inequalities in (41) and (43) hold in reverse, and we obtain the second statement of the lemma. $\square$

Proof of Lemma 8: Consider a 2-certificate menu $\hat{\Gamma}_2$ that is non-maximal. That is, the partition $\hat{M}$ contains a non-singleton subset $\hat{\Theta} \subseteq \Theta$ so that it holds $p(\theta) = \bar{p}$ and $q(\theta) = \bar{q}$ for all $\theta \in \hat{\Theta}$. Let $\bar{\theta}$ denote the average over $\hat{\Theta}$. Then all types $\theta \in \hat{\Theta}$ receive the same marginal information rent

$$\bar{I} \equiv \frac{\bar{p} - \bar{\theta}}{1 - \bar{\theta}} \frac{\bar{\theta} - \bar{q}}{\bar{\theta}} w(\bar{p}, \bar{q}).$$

Now consider a pair of mappings $(\check{p}(\theta), \check{q}(\theta))$ with $(\check{p}(\bar{\theta}), \check{q}(\bar{\theta})) = (\bar{p}, \bar{q})$ and the function

$$\check{S}(\theta) \equiv \hat{v}(\check{q}(\theta)) + (\theta - \check{q}(\theta))w(\check{p}(\theta) \text{ with } \check{q}(\theta)) = \hat{v}(\check{p}(\theta)) - [\check{p}(\theta) - \theta]w(\check{p}(\theta), \check{q}(\theta)). \tag{44}$$

If $\check{S}$ is convex, it follows

$$\begin{aligned}
\hat{S} &= \int_{\theta \in \hat{\Theta}} [\hat{v}(\bar{q}) + (\theta - \bar{q})w(\bar{p}, \bar{q})]\, dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta) \\
&= \int_{\theta \in \hat{\Theta}} \left[\hat{v}(\bar{q}) + (\bar{\theta} - \bar{q})w(\bar{p}, \bar{q})\right] dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta) \\
&= \int_{\theta \in \hat{\Theta}} \check{S}(\bar{\theta}) dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta) \\
&\leq \int_{\theta \in \hat{\Theta}} \check{S}(\theta) dF(\theta) + \int_{\theta \notin \hat{\Theta}} \hat{S}(\theta) dF(\theta).
\end{aligned}$$

To obtain the result, it therefore suffices to construct a pair of mappings $(\check{p}(\theta), \check{q}(\theta))$ with $(\check{p}(\bar{\theta}), \check{q}(\bar{\theta})) = (\bar{p}, \bar{q})$ so that $\check{S}$ is convex in θ and for all $\theta \in \hat{\Theta}$, it holds

$$\frac{\check{p}(\theta) - \theta}{1 - \theta} \frac{\theta - \check{q}(\theta)}{\theta} w(\check{p}(\theta), \check{q}(\theta)) = \bar{I}. \tag{45}$$

Given $(\check{p}(\bar{\theta}), \check{q}(\bar{\theta})) = (\bar{p}, \bar{q})$, Equality (45) holds for all $\theta \in \hat{\Theta}$ if $(\check{p}(\theta), \check{q}(\theta))$ satisfies the differential equation

$$(\theta - \check{q})[w(\check{p}, \check{q}) + (\check{p} - \theta)w_p(\check{p}, \check{q})]\check{p}' - (\check{p} - \theta)[w(\check{p}, \check{q}) - (\theta - \check{q})w_q(\check{p}, \check{q})]\check{q}' = \psi(\theta, \check{p}, \check{q})w(\check{p}, \check{q}), \tag{46}$$

where we dropped the argument θ and have

$$\psi(\theta,p,q) \equiv (\theta-q)\frac{1-p}{1-\theta}-(p-\theta)\frac{q}{\theta}.$$

We construct $(\check{p}(\theta),\check{q}(\theta))$ as follows. We fix $\check{p}(\bar{\theta})=\bar{p}$ and $\check{q}(\bar{\theta})=\bar{q}$. Then, depending on whether $\psi(\theta,\check{p},\check{q})$ is positive or negative, marginally change only $\check{p}$ or $\check{q}$ so that type θ's marginal information rent remains at $\bar{I}$. In particular,

$$\check{p}'(\theta)=\begin{cases}\frac{\psi(\theta,\check{p},\check{q})}{\theta-\check{q}}\frac{w}{w+(\check{p}-\theta)w_p} & \text{if } \psi(\theta,\check{p},\check{q})\geq 0\\ 0 & \text{otherwise;}\end{cases} \quad \text{and } \check{q}'(\theta)=\begin{cases}\frac{-\psi(\theta,\check{p},\check{q})}{\check{p}-\theta}\frac{w}{w-(\theta-\check{q})w_q} & \text{if } \psi(\theta,\check{p},\check{q})\leq 0\\ 0 & \text{otherwise.}\end{cases} \tag{47}$$

This construction yields two differential equations that define the mappings $\check{p}(\theta)$ and $\check{q}(\theta)$. They are both continuous and (weakly) increasing in θ. They are also continuously differentiable, as $\check{p}'(\theta^0)=\check{q}'(\theta^0)=0$ with θ^0 such that $\psi(\theta^0,\check{p}(\theta^0),\check{q}(\theta^0))=0$. For θ such that $\psi\geq 0$, it also implies that ψ is strictly increasing in θ, as

$$\begin{aligned}\left.\frac{d\psi}{d\theta}\right|_{\psi\geq 0} &= \frac{\partial\psi}{\partial\theta}+\frac{\partial\psi}{\partial\check{p}}\check{p}'(\theta)=\frac{(1-\check{p})(1-\check{q})}{(1-\theta)^2}+\frac{\check{p}\check{q}}{\theta^2}-\left(\frac{\theta-\check{q}}{1-\theta}+\frac{\check{q}}{\theta}\right)\check{p}'(\theta)\\ &> \frac{(1-\check{p})(1-\check{q})}{(1-\theta)^2}+\frac{\check{p}\check{q}}{\theta^2}-\left(\frac{\theta-\check{q}}{1-\theta}+\frac{\check{q}}{\theta}\right)\frac{(1-\check{p})}{(1-\theta)}=\frac{1-\check{p}}{1-\theta}+\frac{\check{q}}{\theta}\frac{\check{p}-\theta}{\theta(1-\theta)}>0.\end{aligned}$$

This implies that ψ changes sign at most once, and only from negative to positive. Hence, there is at most one value θ^0 such that $\psi(\theta^0,\check{p}(\theta^0),\check{q}(\theta^0))=0$. Moreover, $\psi(\theta,\check{p}(\theta),\check{q}(\theta))>0$ for $\theta>\theta^0$ and $\psi(\theta,\check{p}(\theta),\check{q}(\theta))<0$ for $\theta<\theta^0$.

Note that $\check{S}'(\theta)$ is continuous at θ^0.[17] As a result, it suffices to show convexity of $\check{S}$ by showing convexity of $\check{S}$ separately for $\psi<0$ and $\psi>0$. Hence, we distinguish two cases:

Case 1: Consider an interval over which $\psi\geq 0$. In this case $q'(\theta)=0$ and

$$\check{p}'(\theta)=\frac{\psi(\theta,\check{p},\check{q})}{\theta-\check{q}}\frac{w}{w+(\check{p}-\theta)w_p}=\underbrace{\left(\frac{1-\check{p}}{1-\theta}-\frac{(\check{p}-\theta)\check{q}}{(\theta-\check{q})\theta}\right)}_{\in[0,1]}\underbrace{\frac{w}{w+(\check{p}-\theta)\,w_p}}_{\in[0,1]}\in[0,1].$$

Using the middle expression in (44) and $\check{q}'(\theta)=0$, we obtain

$$\check{S}'(\theta)=w+(\theta-\check{q})w_p\check{p}'(\theta); \tag{48}$$

[17]To see that $\check{S}'(\theta)$ is continuous at θ^0 (given its existence), note $\lim_{\theta\uparrow\theta^0}\check{S}'(\theta)=w(\check{p}(\theta^0),\check{q}(\theta^0))-[\check{p}(\theta^0)-\theta^0]w_q(\check{p}(\theta^0),\check{q}(\theta^0))\check{q}'(\theta^0)=w(\check{p}(\theta^0),\check{q}(\theta^0))$ which equals $\lim_{\theta\downarrow\theta^0}\check{S}'(\theta)=w(\check{p}(\theta^0),\check{q}(\theta^0))-[\theta^0-\check{q}(\theta^0)]w_p(\check{p}(\theta^0),\check{q}(\theta^0))\check{p}'(\theta^0)=w(\check{p}(\theta^0),\check{q}(\theta^0))$.

and differentiating once more yields

$$\check{S}''(\theta) = 2w_p\check{p}'(\theta) + (\theta - \check{q})[w_{pp}\check{p}'(\theta)^2 + w_p\check{p}''(\theta)] \tag{49}$$

$$= 2w_p\check{p}'(\theta) - 2\frac{\theta - \check{q}}{\check{p} - \check{q}}w_p\check{p}'(\theta)^2 + \frac{\theta - \check{q}}{\check{p} - \check{q}}v''(\check{p})\check{p}'(\theta)^2 + (\theta - \check{q})w_p\check{p}''(\theta) \tag{50}$$

$$= 2w_p\check{p}'(\theta)\left(1 - \frac{\theta - \check{q}}{\check{p} - \check{q}}\check{p}'(\theta)\right) + \frac{\theta - \check{q}}{\check{p} - \check{q}}v''(\check{p})\check{p}'(\theta)^2 + (\theta - \check{q})w_p\check{p}''(\theta) \tag{51}$$

$$\geq \frac{\theta - \check{q}}{\check{p} - \check{q}}\left(v''(\check{p})\check{p}'(\theta)^2 + (\check{p} - \check{q})w_p\check{p}''(\theta)\right), \tag{52}$$

where (50) follows from substituting out w_{pp}, where the definition of $w(p, q)$ implies

$$w_p = \frac{v'(p) - w}{p - q} \text{ and } w_{pp} = \frac{v''(p) - 2w_p}{p - q}, \tag{53}$$

and (52) follows because the first term in (51) is positive as $w_p \geq 0$, $\check{p}'(\theta) \in [0, 1]$, and $0 < \theta - \check{q} < \check{p} - \check{q}$ so that the term in parenthesis is also positive.

We next show that also the remaining term (52) is positive, implying that $\check{S}''(\theta) \geq 0$ and, hence, $\check{S}$ is convex over the range $\psi \geq 0$. To sign (52), we first derive $\check{p}''(\theta)$ for $\psi \geq 0$. By the product rule, it follows from (47) that

$$\check{p}''(\theta) = \left(\frac{1-\check{p}}{1-\theta} - \frac{\check{p}-\theta}{\theta-\check{q}}\frac{\check{q}}{\theta}\right)\left(\frac{w}{w + (\check{p}-\theta)w_p}\right)' + \left(\frac{1-\check{p}}{1-\theta} - \frac{\check{p}-\theta}{\theta-\check{q}}\frac{\check{q}}{\theta}\right)'\left(\frac{w}{w + (\check{p}-\theta)w_p}\right), \tag{54}$$

$$\geq \frac{2w_p p'(\theta) - 2\frac{\theta-\check{q}}{\check{p}-\check{q}}w_p p'(\theta)^2 - \frac{\check{p}-\theta}{\check{p}-\check{q}}v''(\check{p})p'(\theta)^2}{w + (\check{p}-\theta)w_p} \geq -\frac{(\check{p}-\theta)v''(\check{p})p'(\theta)^2}{(\check{p}-\check{q})(w + (\check{p}-\theta)w_p)}, \tag{55}$$

where the last inequality in (55) obtains because, taken together, the first two terms in the numerator of the first fraction are positive as $\theta - \check{q} < \check{p} - \check{q}$ and $p'(\theta) \in [0, 1]$. To see the first inequality in (55) note that collecting terms after substituting out w_{pp} using (53), we obtain

$$\left(\frac{w}{w + (p-\theta)\,w_p}\right)' = \frac{[w + (p-\theta)w_p]w_p\check{p}'(\theta) - [w_p\check{p}'(\theta) + (\check{p}'(\theta) - 1)w_p + (p-\theta)w_{pp}\check{p}'(\theta)]w}{(w + (p-\theta)\,w_p)^2}$$

$$= \frac{ww_p - 2\frac{\theta-q}{p-q}ww_p p'(\theta) + w_p p'(\theta)(w + (\check{p}-\theta)w_p) - \frac{p-\theta}{p-q}wv''(\check{p})p'(\theta)}{(w + (\check{p}-\theta)w_p)^2}.$$

Moreover, it follows

$$\begin{aligned}
\left(\frac{1-p}{1-\theta}-\frac{p-\theta}{\theta-q}\frac{q}{\theta}\right)' &= \frac{1-p}{(1-\theta)^2}+\frac{(p-q)}{(\theta-q)^2}\frac{q}{\theta}+\frac{p-\theta}{\theta-q}\frac{q}{\theta^2}-\frac{1}{1-\theta}p'(\theta)-\frac{1}{\theta-q}\frac{q}{\theta}p'(\theta) \\
&\geq \frac{1-p}{(1-\theta)^2}+\frac{1}{\theta-q}\frac{1-p}{1-\theta}\frac{q}{\theta}-\frac{1}{1-\theta}p'(\theta)-\frac{1}{\theta-q}\frac{q}{\theta}p'(\theta) \\
&= \frac{1-p}{1-\theta}\left(\frac{1}{1-\theta}+\frac{1}{\theta-q}\frac{q}{\theta}\right)-\left(\frac{1}{1-\theta}+\frac{1}{\theta-q}\frac{q}{\theta}\right)p'(\theta) \\
&\geq \left(\frac{1-p}{1-\theta}-\frac{p-\theta}{\theta-q}\frac{q}{\theta}-p'(\theta)\right)\left(\frac{1}{1-\theta}+\frac{1}{\theta-q}\frac{q}{\theta}\right) \\
&= \left(\frac{w+(p-\theta)w_p}{w}p'(\theta)-p'(\theta)\right)\left(\frac{1}{1-\theta}+\frac{1}{\theta-q}\frac{q}{\theta}\right) \\
&= \frac{(p-\theta)w_p}{w}\left(\frac{1}{1-\theta}+\frac{1}{\theta-q}\frac{q}{\theta}\right)p'(\theta),
\end{aligned}$$

where the first inequality holds since $\frac{p-q}{\theta-q}\geq 1\geq\frac{1-p}{1-\theta}$ and $\frac{p-\theta}{\theta-q}\frac{q}{\theta^2}$ is non-negative, the second one holds since $\frac{p-\theta}{\theta-q}\frac{q}{\theta}$ is non-negative. Substitution of these terms in (54) yields the first inequality in (55) and, after collecting terms, the first fraction.

Using (55), we continue from (52) to obtain

$$\check{S}''(\theta)\geq\frac{\theta-\check{q}}{\check{p}-\check{q}}\left(1-\frac{w_p(\check{p}-\theta)}{w+(\check{p}-\theta)w_p}\right)v''(\check{p})\check{p}'(\theta)^2=\frac{\theta-\check{q}}{\check{p}-\check{q}}\frac{w}{w+(\check{p}-\theta)w_p}v''(\check{p})\check{p}'(\theta)^2\geq 0.$$

This establishes convexity of $\check{S}$ over an interval for which $\psi\geq 0$.

Case 2: Consider an interval over which $\psi<0$. In this case $p'(\theta)=0$ and

$$\check{q}'(\theta)=-\frac{\psi(\theta,\check{p},\check{q})}{\check{p}-\theta}\frac{w}{w+(\theta-\check{q})w_q}=\underbrace{\left(\frac{\check{q}}{\theta}-\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)}_{\in[0,1]}\underbrace{\frac{w}{w+(\theta-\check{q})w_q}}_{\in[0,1]}\in[0,1].$$

We first establish that the definition of $w(p,q)$ implies

$$w_q=\frac{w-v'(q)}{p-q}\text{ and }w_{qq}=\frac{2w_q-v''(q)}{p-q}. \tag{56}$$

By the product rule, we have

$$\begin{aligned}
\check{q}''(\theta) &= \left(\frac{\check{q}}{\theta}-\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\left(\frac{w}{w-(\theta-\check{q})w_q}\right)'+\left(\frac{\check{q}}{\theta}-\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)'\frac{w}{w-(\theta-\check{q})w_q}, &(57)\\
&\leq \frac{2(\check{p}-\check{q})w_q\check{q}'(\theta)-2(\check{p}-\theta)w_q\check{q}'(\theta)^2-(\theta-\check{q})v''(\check{q})\check{q}'(\theta)^2}{[\check{p}-\check{q}][w-(\theta-\check{q})w_q]} &(58)
\end{aligned}$$

where the inequality follows from using (56) to obtain

$$\left(\frac{w}{w-(\theta-\check{q})w_q}\right)' = \frac{(w-(\theta-\check{q})w_q)w_q\check{q}'(\theta)-[w_q\check{q}'(\theta)-(1-\check{q}'(\theta))w_q-(\theta-\check{q})w_{qq}\check{q}'(\theta)]w}{(w-(\theta-\check{q})w_q)^2}$$
$$= \frac{ww_q-2\frac{\check{p}-\theta}{\check{p}-\check{q}}ww_q\check{q}'(\theta)+(w-(\theta-\check{q})w_q)w_q\check{q}'(\theta)-\frac{\theta-\check{q}}{\check{p}-\check{q}}wv''(\check{q})\check{q}'(\theta)}{(w-(\theta-\check{q})w_q)^2},$$

and

$$\begin{aligned}\left(\frac{\check{q}}{\theta}-\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)' &= \left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)q'(\theta)-\frac{\check{q}}{\theta^2}-\frac{\check{p}-\check{q}}{\check{p}-\theta}\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}-\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{(1-\theta)^2}\\
&\le \left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\check{q}'(\theta)-\frac{\check{q}}{\theta^2}-\frac{\check{q}}{\theta}\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\\
&= \left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\check{q}'(\theta)-\left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\frac{\check{q}}{\theta}\\
&\le \left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\check{q}'(\theta)-\left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\left(\frac{\check{q}}{\theta}-\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\\
&= \left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\left(1-\frac{w-(\theta-\check{q})\,w_q}{w}\right)\check{q}'(\theta)\\
&= \frac{(\theta-\check{q})\,w_q}{w}\left(\frac{1}{\theta}+\frac{1}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\right)\check{q}'(\theta),\end{aligned}$$

with using $\frac{\check{p}-\check{q}}{\check{p}-\theta}\ge 1\ge\frac{\check{q}}{\theta}$ and $\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{(1-\theta)^2}$ is non-negative to obtain the first inequality, and using $\frac{\theta-\check{q}}{\check{p}-\theta}\frac{1-\check{p}}{1-\theta}\ge 0$ to obtain the second inequality.

Using the right-hand-side expression in (44) and $\check{p}'(\theta)=0$, we obtain

$$\check{S}'(\theta)=w-(\check{p}-\theta)w_q\check{q}'(\theta); \tag{59}$$

and differentiating once more yields

$$\check{S}''(\theta) = 2w_q\check{q}'(\theta)-(\check{p}-\theta)[w_{qq}\check{q}'(\theta)^2+w_q\check{q}''(\theta)] \tag{60}$$
$$= 2w_q\check{q}'(\theta)\left(1-\frac{\check{p}-\theta}{\check{p}-\check{q}}\check{q}'(\theta)\right)+\frac{\check{p}-\theta}{\check{p}-\check{q}}v''(\check{q})\check{q}'(\theta)^2-(\check{p}-\theta)w_q\check{q}''(\theta) \tag{61}$$
$$\ge 2w_q\check{q}'(\theta)\left(1-\frac{\check{p}-\theta}{\check{p}-\check{q}}\check{q}'(\theta)\right)\left[1-\frac{(\check{p}-\theta)w_q}{w-(\theta-q)w_q}\right]\ge 0. \tag{62}$$

where the second equality follows from substituting out w_{qq} using (53) and the inequality from (58). This establishes convexity of $\check{S}$ over an interval for which $\psi<0$. □

Proof of Proposition 1: We first note that the proof of Lemma 3 extends to non-maximal-screening menus. Hence, optimal certification menus can be composed of contracts exhibiting at most three certificates. If a menu pools some subset of types $\hat{\Theta} \in \Theta$ at a two-certificate contract, then it follows from the proof of Lemma 8 that the certifier's revenue can be improved with an alternative menu fully separating these types. It remains to show that certification menus pooling types at a three-certificate contract that induces three distinct posteriors are suboptimal as well.

To show this, we consider an incentive compatible menu that pools a subset of types $\hat{\Theta} \subset \Theta$ at a certification structure $\hat{\sigma} = (\{\hat{c}_1, \hat{c}_2, \hat{c}_3\}, (\bar{\pi}_1^l, \bar{\pi}_2^l, \bar{\pi}_3^l), (\bar{\pi}_1^h, \bar{\pi}_2^h, \bar{\pi}_3^h))$ such that each of the three certificates obtains with a strict probability: $\bar{\pi}_1^h + \bar{\pi}_1^l > 0$, $\bar{\pi}_2^h + \bar{\pi}_2^l > 0$, $\bar{\pi}_3^h + \bar{\pi}_3^l > 0$. We denote by $\bar{\theta} = \mathbb{E}\{\hat{\Theta}\}$ the expected type over the subset $\hat{\Theta}$ (with possibly $\bar{\theta} \notin \hat{\Theta}$). This certification structure induces posteriors

$$\bar{p}_1 = \frac{\bar{\theta}\bar{\pi}_1^h}{\bar{\theta}\bar{\pi}_1^h + (1-\bar{\theta})\bar{\pi}_1^l};\ \bar{p}_2 = \frac{\bar{\theta}\bar{\pi}_2^h}{\bar{\theta}\bar{\pi}_2^h + (1-\bar{\theta})\bar{\pi}_2^l};\ \bar{p}_3 = \frac{\bar{\theta}\bar{\pi}_3^h}{\bar{\theta}\bar{\pi}_3^h + (1-\bar{\theta})\bar{\pi}_3^l}, \tag{63}$$

such that $\bar{p}_1 > \bar{p}_2 > \bar{p}_3$. Incentive compatibility implies that each type $\hat{\theta} \in \hat{\Theta}$ obtains the marginal information rent

$$\bar{I} = [\bar{\pi}_1^h - \bar{\pi}_1^l]v(\bar{p}_1) + [\bar{\pi}_2^h - \bar{\pi}_2^l]v(\bar{p}_2) + [\bar{\pi}_3^h - \bar{\pi}_3^l]v(\bar{p}_3). \tag{64}$$

A type $\hat{\theta} \in \hat{\Theta}$ generates the surplus

$$\hat{S}(\hat{\theta}) = [\hat{\theta}\bar{\pi}_1^h + (1-\hat{\theta})\bar{\pi}_1^l]v(\bar{p}_1) + [\hat{\theta}\bar{\pi}_2^h + (1-\hat{\theta})\bar{\pi}_2^l]v(\bar{p}_2) + [\hat{\theta}\bar{\pi}_3^h + (1-\hat{\theta})\bar{\pi}_3^l]v(\bar{p}_3)$$

with average $\hat{S}(\bar{\theta})$.

We next construct an alternative incentive compatible menu that coincides with the old one for all $\theta \notin \hat{\Theta}$, but separates all types in $\hat{\Theta}$ such that type $\hat{\theta} \in \hat{\Theta}$ obtains certificate $\hat{c}_i$ in state $\omega \in \{l, h\}$ with probability $\pi_i^\omega(\hat{\theta})$ for $i = 1, 2, 3$, while keeping the marginal information rent constant at $\bar{I}$. The constant information rent ensures that the alternative menu is also incentive compatible. We construct the type-dependent probability system $\{\pi_i^\omega(\hat{\theta})\}$ in such a way that the respective posteriors also remain constant at $(\bar{p}_1, \bar{p}_2, \bar{p}_3)$. It follows from the proof of Lemma 5 (Case 2) that $(\bar{p}_1, \bar{p}_2, \bar{p}_3, \hat{\theta})$ identify $\{\pi_i^\omega(\hat{\theta})\}$ as a solution to a system of 6 linear equation (see (37)). As the solution $\{\pi_i^\omega(\theta)\}$ is continuous in θ and since $\pi_i^\omega(\bar{\theta}) = \bar{\pi}_i^\omega$, we have $\pi_c^\omega(\theta) \in [0, 1]$ for θ close to $\bar{\theta}$. Indeed, as the numerator in $\pi_i^\omega(\theta)$ is quadratic in θ, we find for each certificate $\hat{c}_i$ two thresholds $\theta_i^l < \bar{\theta} < \theta_i^h$ such that $\pi_i^\omega(\theta) \in [0, 1]$ if and only if $\theta \in [\theta_i^l, \theta_i^h]$.

Note that these thresholds are independent of the state ω.

Denoting by $\theta^l \equiv \max_i\{\theta_i^l\}$ the maximal lower threshold and by $\theta^h \equiv \min_i\{\theta_i^h\}$ the minimal upper threshold, it follows that the solution $\{\pi_i^\omega(\theta)\}$ satisfies the individual boundary conditions for any $\theta \in [\theta^l, \theta^h]$. Let $\check{S}(\theta)$ for $\theta \in [\theta^l, \theta^h]$ be the surplus generated by type θ given the solution $\{\pi_i^\omega(\theta)\}$.

$$\check{S}(\theta) = S(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta),$$

where function S is defined in (38). Notice that function $\check{S}$ gives the same value as function $\hat{S}$ when evaluated at the average type in subset $\hat{\Theta}$, that is $\check{S}(\bar{\theta}) = \hat{S}(\bar{\theta})$. Hence, if $\hat{\Theta} \subset [\theta^l, \theta^h]$, we obtain our result by showing that $\check{S}(\theta)$ is convex in θ. Convexity of $\check{S}$ holds if $\check{S}''(\theta)$ is positive for any $\theta \in [\theta^l, \theta^h]$.

By differentiating (38) with respect to θ twice, we find this second derivative as

$$\check{S}''(\theta) = \frac{K(\bar{p}_1, \bar{p}_2, \bar{p}_3)}{D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta)^3}$$

with

$$\begin{aligned} K(\bar{p}_1, \bar{p}_2, \bar{p}_3) =& 2\{(\bar{p}_2 - \bar{p}_3)[v(\bar{p}_1) - v(\bar{p}_3)] - (\bar{p}_1 - \bar{p}_3)[v(\bar{p}_2) - v(\bar{p}_3)]\}\times \\ & \{(\bar{p}_1 - \bar{p}_3)(\bar{p}_2 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)[v(\bar{p}_1) - v(\bar{p}_3)][v(\bar{p}_2) - v(\bar{p}_3)][v(\bar{p}_1) - v(\bar{p}_2)] \\ & - [\bar{p}_1(\bar{p}_2 - \bar{p}_3)(v(\bar{p}_1) - v(\bar{p}_2)) - \bar{p}_3(\bar{p}_1 - \bar{p}_2)(v(\bar{p}_2) - v(\bar{p}_3))]\times \\ & [(1 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)(v(\bar{p}_2) - v(\bar{p}_3)) - (1 - \bar{p}_1)(\bar{p}_2 - \bar{p}_3)(v(\bar{p}_1) - v(\bar{p}_2))]\bar{I}\} \end{aligned} \quad (65)$$

and

$$D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta) = (\bar{p}_2 - \bar{p}_3)(\bar{p}_1 - \theta)(v(\bar{p}_1) - v(\bar{p}_3)) + (\bar{p}_1 - \bar{p}_3)(\theta - \bar{p}_2)(v(\bar{p}_2) - v(\bar{p}_3))$$

To see that $D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta)$ is positive, notice that convexity of v implies $(\bar{p}_2 - \bar{p}_3)[v(\bar{p}_1) - v(\bar{p}_3)] > (\bar{p}_1 - \bar{p}_3)[v(\bar{p}_2) - v(\bar{p}_3)]$. It follows from this inequality that

$$\begin{aligned} D(\bar{p}_1, \bar{p}_2, \bar{p}_3, \theta) &> (\bar{p}_1 - \bar{p}_3)(\bar{p}_1 - \theta)[v(\bar{p}_2) - v(\bar{p}_3)] + (\bar{p}_1 - \bar{p}_3)(\theta - \bar{p}_2)(v(\bar{p}_2) - v(\bar{p}_3)) \\ &= (\bar{p}_1 - \bar{p}_3)(\bar{p}_1 - \theta + \theta - \bar{p}_2)(v(\bar{p}_2) - v(\bar{p}_3)) \\ &= (\bar{p}_1 - \bar{p}_3)(\bar{p}_1 - \bar{p}_2)(v(\bar{p}_2) - v(\bar{p}_3)) \\ &> 0 \end{aligned}$$

Hence, the sign of $\check{S}''$ is positive if (65) is positive. As the sign of (65) coincides with the sign of the term in the second curly brackets of (65), $\check{S}''$ is positive if the following inequality holds

$$
\begin{aligned}
&(\bar{p}_1-\bar{p}_3)(\bar{p}_2-\bar{p}_3)(\bar{p}_1-\bar{p}_2)[v(\bar{p}_1)-v(\bar{p}_3)][v(\bar{p}_2)-v(\bar{p}_3)][v(\bar{p}_1)-v(\bar{p}_2)] \geq \\
&\qquad [\bar{p}_1(\bar{p}_2-\bar{p}_3)(v(\bar{p}_1)-v(\bar{p}_2))-\bar{p}_3(\bar{p}_1-\bar{p}_2)(v(\bar{p}_2)-v(\bar{p}_3))]\times \\
&\qquad\qquad [(1-\bar{p}_3)(\bar{p}_1-\bar{p}_2)(v(\bar{p}_2)-v(\bar{p}_3))-(1-\bar{p}_1)(\bar{p}_2-\bar{p}_3)(v(\bar{p}_1)-v(\bar{p}_2))]\bar{I} \quad (66)
\end{aligned}
$$

For any $\bar{I} \geq 0$, the right hand side of (66) is smaller than

$$
\begin{aligned}
&[\bar{p}_1(\bar{p}_2-\bar{p}_3)(v(\bar{p}_1)-v(\bar{p}_2))][(1-\bar{p}_3)(\bar{p}_1-\bar{p}_2)(v(\bar{p}_2)-v(\bar{p}_3))] \\
&\quad + [\bar{p}_3(\bar{p}_1-\bar{p}_2)(v(\bar{p}_2)-v(\bar{p}_3))][(1-\bar{p}_1)(\bar{p}_2-\bar{p}_3)(v(\bar{p}_1)-v(\bar{p}_2))]\bar{I},
\end{aligned}
$$

which simplifies to

$$
[(\bar{p}_2-\bar{p}_3)(\bar{p}_1-\bar{p}_2)(\bar{p}_1+\bar{p}_3-2\bar{p}_1\bar{p}_3)(v(\bar{p}_1)-v(\bar{p}_2))(v(\bar{p}_2)-v(\bar{p}_3))]\bar{I}. \quad (67)
$$

Hence, it suffices to show that the left hand side of (66) is larger than (67), or, equivalently that

$$
(\bar{p}_1-\bar{p}_3)[v(\bar{p}_1)-v(\bar{p}_3)] \geq (\bar{p}_1+\bar{p}_3-2\bar{p}_1\bar{p}_3)\bar{I}. \quad (68)
$$

Because $\theta \in [\theta^l, \theta^h]$ implies that $\pi_2 = [\theta\bar{\pi}_2^h + (1-\theta)\bar{\pi}_2^l] \geq 0$ for $\theta \in [\theta^l, \theta^h]$, the numerator in the fraction for π_2 in (37) must be positive. That is, it holds

$$
\bar{I} \leq \frac{(\bar{p}_1-\theta)(\theta-\bar{p}_3)}{(1-\theta)\theta}\frac{v(\bar{p}_1)-v(\bar{p}_3)}{\bar{p}_1-\bar{p}_3}.
$$

Hence, inequality (68) holds if

$$
(\bar{p}_1-\bar{p}_3)^2 \geq (\bar{p}_1+\bar{p}_3-2\bar{p}_1\bar{p}_3)\frac{(\bar{p}_1-\theta)(\theta-\bar{p}_3)}{(1-\theta)\theta}. \quad (69)
$$

Because

$$
\max_\theta \frac{(\bar{p}_1-\theta)(\theta-\bar{p}_3)}{(1-\theta)\theta} = \bar{p}_1+\bar{p}_3-2\bar{p}_1\bar{p}_3-2\sqrt{\bar{p}_1\bar{p}_3(1-\bar{p}_1)(1-\bar{p}_3)},
$$

it follows that inequality (69) holds if

$$
(\bar{p}_1-\bar{p}_3)^2 \geq (\bar{p}_1+\bar{p}_3-2\bar{p}_1\bar{p}_3)(\bar{p}_1+\bar{p}_3-2\bar{p}_1\bar{p}_3-2\sqrt{\bar{p}_1\bar{p}_3(1-\bar{p}_1)(1-\bar{p}_3)}). \quad (70)
$$

Using $\alpha = \bar{p}_1(1-\bar{p}_3)$ and $\beta = \bar{p}_3(1-\bar{p}_1)$, (70) rewrites as

$$(\alpha-\beta)^2 \geq (\alpha+\beta)(\alpha+\beta-2\sqrt{\alpha\beta}), \tag{71}$$

where it follows

$$(\alpha-\beta)^2 \geq (\alpha+\beta)(\alpha+\beta-2\sqrt{\alpha\beta}) \Leftrightarrow (\alpha+\beta)^2 \geq 4\alpha\beta \Leftrightarrow (\alpha-\beta)^2 \geq 0. \tag{72}$$

As the latter inequality is true, we conclude that also inequality (66) holds so that $\check{S}$ is convex.

Hence, if $\hat{\Theta} \subset [\theta^l, \theta^h]$, the 3-certificate menu is not optimal for the certifier, because the convexity of $\check{S}$ implies that we can raise the certifier's profit by separating all types in $\hat{\Theta}$, while keeping the three posteriors for each of these types at $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$.

We next argue that, if $\hat{\Theta} \not\subset [\theta^l, \theta^h]$, the 3-certificate menu can also not be optimal. To show this, we derive a convex function $\check{\check{S}}(\theta)$ that is defined for the entire interval Θ by combining the construction of $\check{S}(\theta)$ in this proof with the construction of $\check{S}(\theta)$ as in the proof of Lemma 8.

To develop this argument, first note that $\check{S}$ is defined only on the domain $[\theta^l, \theta^h]$, because for $\theta \notin [\theta^l, \theta^h]$ our construction implies that $\pi_i^\omega(\theta) \notin [0,1]$ for some state $\omega \in \{l, h\}$ and certificate $c \in \{\hat{c}_1, \hat{c}_2, \hat{c}_3\}$.

In particular, we have at θ^l that $\pi_i^\omega(\theta^l) = 0$ for some state $\omega \in \{l, h\}$. This implies that at θ^l, we effectively have a two-certificate menu that induces two of the three posteriors $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$. Hence, following the proof of Lemma 8, we can construct a pair of mappings $(\check{p}_l(\theta), \check{q}_l(\theta))$ satisfying the differential equation (47) with $(\check{p}_l(\theta^l), \check{q}_l(\theta^l))$ corresponding to the only two posteriors among $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ that occur with positive probability at θ^l. This gives then rise to a convex function $\check{S}_l(\theta)$ as defined in (44).

Likewise, we can construct $(\check{p}_h(\theta), \check{q}_h(\theta))$ with $(\check{p}_h(\theta^h), \check{q}_h(\theta^h))$ corresponding to the two posteriors among $\bar{p}_1$, $\bar{p}_2$, $\bar{p}_3$ that still occur with positive probability at θ^h. This yields a function $\check{S}_h(\theta)$ that is defined for all θ in Θ.

Thus, we obtain three convex functions $\check{S}(\theta)$, $\check{S}_l(\theta)$, and $\check{S}_h(\theta)$. The function $\check{S}(\theta)$ is supported by the three (fixed) posteriors $\bar{p}_1, \bar{p}_2, \bar{p}_3$ and has the limited domain $[\theta^l, \theta^h]$. The convex function $\check{S}_l(\theta)$ is defined for all Θ, satisfies $\check{S}(\theta^l) = \check{S}_l(\theta^l)$ and is supported by the two posteriors $\check{p}_l(\theta)$, $\check{q}_l(\theta)$. Likewise, the convex function $\check{S}_h(\theta)$ is also defined for all Θ, satisfies $\check{S}(\theta^h) = \check{S}_h(\theta^h)$ and is supported by the two posteriors $\check{p}_l(\theta)$, $\check{q}_h(\theta)$.

Define

$$\check{\check{S}}(\theta) \equiv \begin{cases} \max\{\check{S}(\theta), \check{S}_l(\theta), \check{S}_h(\theta)\} & \text{if } \theta \in [\theta^l, \theta^h] \\ \max\{\check{S}_l(\theta), \check{S}_h(\theta)\} & \text{otherwise.} \end{cases}$$

Being the maximum over convex functions, $\check{\check{S}}$ is convex. Hence, if $\hat{\Theta} \not\subset [\theta^l, \theta^h]$, the 3-certificate menu is also not optimal, because the convexity of $\check{\check{S}}$ implies that we can raise the certifier's profit by separating all types in $\hat{\Theta}$ with using either 2 or 3 certificates. □

References

Ali, N., N. Haghpanah, X. Lin and R. Siegel. 2022. "How to Sell Hard Information." *The Quarterly Journal of Economics* 137-1: 619-678.

Asseyer, A. and R. Weksler. 2024. "Certification Design with Common Values." *Econometrica* 92-3, 651-686

Bergemann, D., A. Bonatti, and A. Smolin. 2018. "The Design and Price of Information." *American Economic Review* 108, 1-48.

Bizzotto, J., J. Rüdiger, A. Vigier. 2020. "Testing, disclosure and approval." *Journal of Economic Theory* 187, 105002.

Corrao, R. 2024. "Mediation Markets: The Case of Soft Information." mimeo, MIT.

Dranove, D. and G. Z. Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48, 935–963.

Evans, R., Park, I-U. 2024. "Selling Information for Bilateral Trade." mimeo, University of Bristol.

Hagenbach, J., F. Koessler and E. Perez-Richet. 2014. "Certifiable Pre-Play Communication: Full Disclosure." *Econometrica* 82(3), 1093-1131.

Hancart, N. 2024. "The Optimal Menu of Tests." mimeo, University of Oslo.

Ichihashi, S. and A. Smolin. 2023. "Data Provision to an Informed Seller.", mimeo, Queen's University.

Kartik, N., F. Lee, and W. Suen. 2021 "Information validates the prior: A theorem on Bayesian updating and applications," *American Economic Review: Insights* 3(2), 165-182.

Koessler, F. and R. Renault. 2012. "When does a firm disclose product information?." *RAND Journal of Economics* 43(4), 630–649.

Lizzeri, A. 1999. "Information Revelation and Certification Intermediaries." *The RAND Journal of Economics* 302, 214-–231.

Okuno-Fujiwara, A., M. Postlewaite, and K. Suzumura. 1990. "Strategic Information Revelation." *Review of Economic Studies* 57, 25–47.

Rosar, Frank. 2017. "Test design under voluntary participation." *Games and Economic Behavior* 104, 632-655.

Seidmann, D. J. and E. Winter. 1997. "Strategic Information Transmission with Verifiable Messages." *Econometrica* 65, 163–169.

Viscusi, W. 1978. "A Note on 'Lemons' Markets with Quality Certification." *The Bell Journal of Economics* 9: 277–279.

Weksler, R. and B. Zik. 2023. "Disclosure in Markets for Ratings." *American Economic Journal: Microeconomics* 15(3): 501-26.

Weksler, R. and B. Zik. 2024. "Selling Certification to Informed Agents." mimeo, Hebrew University of Jerusalem.