SELLING CERTIFICATION, CONTENT MODERATION, AND ATTENTION

HESKI BAR-ISAAC[†], RAHUL DEB[≬], AND MATTHEW MITCHELL[‡]

ABSTRACT. We introduce a model of content moderation for sale, where a platform can channel attention in two ways: direct steering that makes content visible to consumers and certification that controls what consumers know about the content. The platform optimally price discriminates using both instruments. Content from higher willingness-to-pay providers enjoys higher quality certification and more views. The platform cross-subsidizes content: the same certificate is assigned to content from low willingness-to-pay providers that appeals to consumers and content from higher willingness-to-pay providers that does not. Cross-subsidization can benefit consumers by making content more diverse; regulation enforcing accurate certification may be harmful.

1. INTRODUCTION

The digital environment is overwhelming. There are over a billion websites on the internet, Tiktok hosts billions of videos, more than two trillion posts have been created on Facebook.[1] In such an environment, it is little surprise that attention is a key resource and the platforms that control access to such attention can be increasingly sophisticated and profitable in doing so. This has not escaped economists and by now there are surveys on both digital economics (Goldfarb and Tucker, 2019) and social media (Aridor, Jiménez-Durán, Levy, and Song, forthcoming). Meanwhile regulators have become concerned over how platforms, possibly with monopoly power, might manage attention in ways that favor profits over consumer welfare.

In this paper, we examine how a monopoly platform maximizes profits by governing and charging content providers for two key aspects of attention: the frequency with which posts gets viewed (steering) and how posts are presented to viewers (certification). We broadly refer to these two channels as content moderation. We show that imperfect certification, where the quality differences in content that matter to consumers are obscured, can increase content diversity when views

Date: June 14, 2025.

[†] heski.bar-isaac@rotman.utoronto.ca; Rotman School of Management, University of Toronto
[≬] rahul.deb@bc.edu; Department of Economics, Boston College
[‡] matthew.mitchell@rotman.utoronto.ca; Rotman School of Management, University of Toronto

We would like to thank Mark Armstrong, Itay Fainmesser, Kinshuk Jerath, Byung-Cheol Kim, Marco Ottaviani, Yossi Spiegel, Greg Taylor, and participants at numerous conferences and seminars for their insightful comments.

[1] These estimates are drawn from https://siteefy.com/how-many-websites-are-there/, https://www.usesignhouse.com/blog/tiktok-stats, and https://www.wordstream.com/blog/ws/2017/11/07/facebook-statistics all accessed on June 7, 2024.

are for sale. As a result, consumers can, but need not, benefit from certification for sale relative to enforced perfect certification. This informs recent policy discussion around permitting the sale of certification on platforms.

We analyze a model where a platform sells views and certification to content providers interested in consumers' attention and engagement. The platform can observe whether content providers are good or bad; content from the former is valued by consumers and content from the latter is not. Therefore, from the consumers' perspective, the platform has the perfect ability (but possibly not the incentive) to screen content. Consumers' attention is governed by their intrinsic interest (which makes it ever harder to generate engagement) and by their expectations that any piece of content will be something they value seeing. These beliefs are governed by an understanding of the likelihood that content presented in a given way (or, equivalently, with a given certificate) arises from a good provider. In turn, these beliefs govern consumers' attention which is what content providers—both good and bad—value. We focus on the application of an online platform selling steering and certification because it corresponds to an important recent development in these markets, and in their regulation. However, in the conclusion section, we point to a wider set of applications that are captured by our framework.

Content providers want attention: to be seen and, hopefully, trusted so that their content is engaged with. Being seen is necessary to command attention and get engagement but it is not sufficient. In addition to steering content directly, the platform can attach an arbitrary message (the certificate) in order to provide consumers with information about quality (the likelihood that the content is from a good provider).[2] This corresponds to everything the consumer sees before deciding whether to further inspect the content. The platform cannot tell the willingness to pay for engagement that good providers have, and it price discriminates by offering different bundles of views and messages. Conversely, we assume that bad providers have a known willingness to pay for attention.

One might suppose that the ability to control views alone is sufficient to exercise full monopoly power. We show that this is not the case. The ability to vary (and charge for) certification can generate additional platform profit notably through imperfect certification (whereby content from both good and bad providers are assigned the same certificate). This allows for revenue from content that consumers would prefer not to see. In turn, one might suppose that consumers suffer from a platform's ability to sell certification as well as views. We show that there is an economic force acting against this intuition that can overwhelm it. Specifically, consumers might gain from

[2] In practice, this may, most obviously, include "checkmarks" for verified status as Twitter began to charge for in November 2022. But it could also include position on a page, aspects of display, ancillary information such as "your great aunt Naima likes this post," etc.

seeing good content whose provider might have relatively little value from receiving this attention. Imperfect certification can, in effect, subsidize views of such content in order to to sell views to bad content providers and so can improve content diversity.

To see why content diversity is impacted by imperfect certification, start from the case of perfect certification. This implies that the platform is mandated to use its certification technology to ensure that bad providers receive no views (one can imagine bad content is clearly labeled as such) and consequently, no attention. Under perfect certification, our platform's profit maximization problem of choosing how many views to provide to each content provider and at what price reduces to a classic model of second-degree price discrimination equivalent to the seminal analysis of Mussa and Rosen (1978). Since good content providers vary in their private valuation for engagement, those that value it more, pay more and receive more views. Lower willingness-to-pay providers get fewer views and some providers with a positive willingness-to-pay are excluded entirely.

Now consider the case where there is only one imperfect certificate available, and its quality is fixed. This is equivalent to all content receiving the same message but the aggregate share of good content being known. As a result, for each view that good providers receive, bad providers receive views in some fixed proportion. In this case, views from good providers allow the platform to generate revenue from bad providers. This allows the platform to profit from selling views to lower value good content providers that would be excluded under perfect certification.

In the fully profit maximizing choice of views and certification, higher willingness-to-pay content providers receive both more views and a message that makes their content more trusted. High enough willingness-to-pay good content providers get perfect certification. Lower willingness-to-pay good providers are not excluded but instead assigned lower quality certification. As we argue, this expansion of content diversity (relative to perfect certification) might benefit consumers by making the platform more egalitarian.

To understand why platforms might have changed their approaches to selling attention — notably Twitter's move to charge for "verified status" in November 2022 — we examine comparative statics in the model. In particular, we highlight how a reduction in what platforms can charge for ads can lead to a move away from perfect certification and that cheaper targeting does not affect the quality of certification. In addition, we highlight how the nature of attention affects whether or not platforms engage in imperfect certification. Specifically, we vary the convexity of attention as a function of content quality with the aim of capturing the extent to which consumers are put off by bad content.

Both steering and certification by platforms have come under regulatory scrutiny more generally. The importance of views is central to algorithmic design around ranking which is under increasing regulatory scrutiny, as in Competition and Markets Authority (2022). The recognition that the presentation of some information might affect the extent to which it is deemed worthy of attention is, of course, at the heart of disclosure regulation (in the context of the kind of social media application that inspires our study, see Mitchell (2021), for example) and central to understanding the discussion around disclosure and certification (Dranove and Jin (2010) provides an excellent overview). Content certification for sale has become an issue as well, with the European Commission announcing that they will seek remedies against X for its practice of selling certification through checkmarks.[3]

Our results directly impact the policy debate around certification for sale. While the optimal mechanism always results in weakly greater content diversity relative to perfect certification, consumers are faced with more bad content. Using total engagement as a measure for consumer welfare, we provide conditions under which enforcing perfect certification does, and does not, benefit consumers. We thus show that there is no single correct answer to the question of whether the sale of certification should be prohibited.

The paper is organized as follows. In the next section, we review the literature. Then, we present the model and study several benchmarks: an engagement maximizing planner, followed by profit maximization when restricted to simple certification with only one or two certificates. The one certificate case includes perfect certification as a special case. We then derive the optimal mechanism and develop comparative statics. Since it is an important policy concern, we study the comparison of the optimal mechanism to enforced perfect certification in detail. Throughout, we illustrate the central forces through an example where attention is a linear function, which allows for an explicit and graphical illustration. Finally, via a simple extension, we show that our framework can accommodate bad content that causes damage and social media addiction, and highlight the intuitive effects of both.

1.1. *Related Literature*

In considering a platform that sells attention both through certification and more prominent views, we bring together literatures that have considered each of these aspects separately.

1.1.1. *Platform Steering without Certification.* Our approach is based on solving a second-degree price discrimination problem (in the style of Mussa and Rosen (1978)). Others have considered that the two-sided nature of platforms changes the standard analysis when the platform collects

[3] See https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761

revenue from both sides. Papers include Choi, Jeon, and Kim (2015), Böhme (2016), and Jeon, Kim, and Menicucci (2022).

Choi, Jeon, and Kim (2015) and Böhme (2016) explore how platforms can maximize profits by differentiating prices for different sellers or users, which indirectly steers users. Jeon, Kim, and Menicucci (2022) further expand on this by investigating second-degree price discrimination in monopoly platforms. Focusing on the impact of platform steering on sellers' incentives, Johnson, Rhodes, and Wildenbeest (2023) and Ichihashi and Smolin (2023) argue that steering alters sellers' competitive behavior. By influencing the exposure of sellers' products to consumers, platforms shape the strategic decisions that sellers make. Our model focuses on the social media context, where content is not priced directly by providers to consumers and can be steered to many consumers at low cost. Moreover, the platform sells not only "quantity" in the form of views but also quality (via certification), which affects consumer attention. It is precisely the interaction between these two that is the focus of our analysis.

In the social media context, various papers have studied steering without certification or direct monetary payment for steering. These papers focus on how platforms recommend different types of content to viewers, and how they incentivize content creation. Examples, building on Ghosh and McAfee (2011), include Bhargava (2022), Srinivasan (2023), Qian and Jain (2024), and Chen, Li, and Preuss (2025).

1.1.2. *Certification without Steering.* Dranove and Jin (2010) provide a wide-ranging survey of the literature on certification. In this literature, Lizzeri (1999) is an early contribution that shares our conclusion that imperfect certification is profit-maximizing for the certifier as it allows good but not great sellers to charge a higher price.[4] Among more recent papers, perhaps Bouvard and Levy (2018) is the most related in very clearly highlighting that profit-maximizing certification trades off pooling different types of sellers to earn more from low-quality sellers but diluting quality too much alienates consumers (which in our environment corresponds to receiving less attention). Our certification is mixed with direct steering, so that the certification need not do the steering on its own.

1.1.3. *Indirect Steering through Search.* Direct steering means that the platform in our model must consider the scarcity in the total possible attention (which we capture by a convex cost for the platform in finding relevant viewers for good content). Consequently, our analysis shares features

[4] This literature has developed in several ways. For example, Ali, Haghpanah, Lin, and Siegel (2022) consider uninformed sellers who can conceal that they have been tested. Following the subprime crisis in 2007, a broad literature has considered certification in the credit-rating industry for which useful surveys can be found in White (2010) and Jeon and Lovo (2013).

with the literature that has focused on how platforms sell off scarce slots (including seminal contributions by Edelman, Ostrovsky, and Schwarz (2007), Chen and He (2011), Armstrong and Zhou (2011) and Athey and Ellison (2011), or, more recently, Bar-Isaac and Shelegia (2025) who contrast different sales mechanisms). We share with much of this literature the observation that given the mechanisms through which these positions are sold, consumers draw equilibrium inferences about the quality of offerings associated with their rankings (different certificates, in our work).

Our focus on how profit incentives might lead a platform away from perfect certification and efficiently allocating views is somewhat reminiscent of a literature that examines biased intermediaries and search diversion (De Corniere and Taylor (2014), Hagiu and Jullien (2014), Burguet, Caminal, and Ellman (2015), and De Corniere and Taylor (2019)), though much of this literature is more focused on the consumer search process.

1.1.4. *Content Moderation (Not for Sale).* Content moderation without direct monetary incentives has become an increasingly important area of study as platforms attempt to balance openness with the need to manage harmful or misleading content. Madio and Quinn (2024) study a platform that manages the value of a third party (advertiser) interest in content moderation. Zou, Wu, and Sarvary (2025) like this paper highlights a tradeoff between quality and variety but in an environment where entry and content quality are endogenous and affected by a recommendation system that aims to maximize consumer surplus and does not earn revenue from content providers.

Kominers and Shapiro (2024) explore a sender-receiver game where a platform can moderate content, in the sense of manipulating what is seen by the receiver for any message sent. This conforms to our idea of the general description of content moderation, but in a different modeling setting. Acemoglu, Ozdaglar, and Siderius (2024) study a model where content moderation is about sharing, and how content sharing between consumers might be regulated. Here content moderation is more easily thought of as relating to the content shared between users and not between providers and consumers.

Another form of content moderation that can be considered part of the certification process is disclosure regulation. Here content that is paid for must be combined with a message that indicates that this is the case. Inderst and Ottaviani (2012) examine a general model of disclosure regulation. Mitchell (2021) and Fainmesser and Galeotti (2021) model how disclosure regulations might impact relationships between content providers ("influencers") and consumers ("followers"). Ershov and Mitchell (forthcoming) provide evidence on the impact of this form of content moderation.[5]

2. Model and preliminaries

We study a price discrimination problem of a platform through which content providers reach consumers. We now describe the model in detail starting with the content providers.

Content Providers. Content *providers* can either be *g*ood or *b*ad; good content can be of value to interested consumers (described below), but bad content cannot. There is a continuum of a unit mass of good providers whose private *values* $\theta \in [0, \overline{\theta}] =: \Theta$ are distributed according to $F \in \Delta(\Theta)$ that has a continuous, positive density $f(\cdot) > 0$. θ captures the extent to which a good content provider values engagement. There is an unlimited mass of bad providers who all have the same value for their content being read.[6]

The amount of *engagement* av_g with good content is the product of the number $v_g \in \mathbb{R}_+$ of *interested* views that the platform provides a good content provider and the *attention* $a \in [0,1]$ that these viewers pay to the content. The utility of a good content provider with value θ from a given level of engagement av_g is θav_g.

Consumers never engage with content from bad providers and thus they only value their content being read. A bad content provider receives utility av_b from $v_b \in \mathbb{R}_+$ views that pay a level $a \in [0,1]$ of attention to their content.

There are three differences between good and bad providers. First, good providers care about engagement whereas bad providers only care about being read. This distinction is not yet apparent from the payoffs (since they both depend on the product of attention and views) but will become clear below when we define the platforms costs (in essence, it is more costly for the platform to provide interested views to the good providers). Second, there is no heterogeneity in the marginal valuations of the bad providers.[7] Third, there is a limited mass of good providers but we assume (for realism) that there is an unlimited amount of bad content since, in particular, its generation can be automated. Alternatively, the assumption is consistent with the idea that, since

[5] Papers on rules surrounding deceptive sales practices such as Corts (2013), Corts (2014), Glaeser and Ujhelyi (2010), and Rhodes and Wilson (2018) also highlight the role that some form of rules on messages might play. This also fits our description of content moderation.

[6] Given our assumptions on the costs of the platform that follow, we could equivalently assume that there is a fixed mass of bad providers but that their content can be spread widely.

[7] This assumption is purely for technical convenience since it avoids the complications that arise with multidimensional private information. It is also consistent with the assumption that there are infinitely many bad providers since the platform can sell to those with highest value.

bad content is never of interest to the consumer, a single piece could be reused and shown to many consumers.

Targeting of Content and Platform Costs. The platform can distinguish between good and bad providers.[8] However, the platform does not know good providers' valuations for engagement. The platform chooses the number of views to direct to each provider's content. Directing v_b untargeted views at content from a bad provider costs the platform γv_b where $0 < \gamma < \min\{\overline{\theta}, 1\}$.[9] This opportunity cost reflects, for example, that the platform could direct advertisements at consumers instead of content.

The same opportunity cost is also present when directing views at content from good providers. However, these providers only value engaged users and this is the source of an additional cost: the platform needs to search for such users of whom there is a smaller pool.[10] The platform faces a cost $\gamma v_g + c(v_g)$ of providing v_g interested views to content from a good provider, where $c : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing, strictly convex and differentiable function that satisfies $c(0) = c'(0) = 0$ and $\lim_{v_g \to \infty} c'(v_g) = \infty$. The convexity of c corresponds to the increasing difficulty of targeting the content of a given good provider with interested viewers as the content is shown more times. Therefore, the convexity is at the level of a given content provider.

Consumers. There are a unit mass of consumers who may view content. Each piece of content is accompanied by a message m.[11] Consumers observe the message m and decide whether to read the content to learn more it. Reading requires paying a cost $q \geq 0$, distributed according to strictly increasing, differentiable cumulative distribution function $A(q)$. The consumer receives a payoff of 1 if they decide to read the content and it turns out to be good content that matches their interest (such content generates engagement). They receive a payoff of 0 otherwise. In Section 5 we generalize to negative payoffs for consumers from reading bad content.[12]

[8] For example, platforms can distinguish between good and bad providers exactly as users can but need only do so once on behalf of all potential viewers. Further, in addition to observing content directly, platforms monitor consumer reactions and other measures of engagement.

[9] This assumption ensures that it is not automatically unprofitable for the platform to serve either type of providers. If $\gamma > \overline{\theta}$, good providers cannot be served profitably. If $\gamma > 1$, bad providers cannot be served profitably.

[10] Of course, the platform could send out untargeted views with the hope of randomly finding interested users. Suppose that the unconditional probability of a user being interested in an untargeted post is ρ. The optimal mechanism we derive using the cost function c will remain optimal for a sufficiently low ρ. This is because it is cheaper to target than to pay the opportunity cost of the number of untargeted posts required to generate the same amount of interest.

[11] The message m can be broadly understood as reflecting whatever the consumer uses to form beliefs about the quality of content before reading it and may include location on a page, information such as how many others "liked" a post, their identity, explicit checkmarks etc.

[12] Such a change might reflect that bad content includes scams and other costly outcomes. As will become apparent, allowing for heterogeneity in the extent of harm from bad content, rather than reading costs, would lead to similar effects.

Upon observing a message m, consumers assign probability $\lambda \in [0,1]$ to the content being good and of interest. The expected payoff from reading the content is therefore $\lambda - q$. Consequently, a consumer chooses to read the content if, and only if, $q \leq \lambda$. In other words, $A(\lambda)$ is the likelihood that content labeled with message m is read by consumers. We henceforth refer to A as the *attention function*. Since engagement is the consumer's payoff, the engagement maximizing benchmark is a natural consumer welfare standard that we analyze below. Aside from allowing a welfare calculation, the details of the underlying interpretation of the attention function as a cost calculation embodied in q is inessential. The same predictions arise, under any functional relationship between beliefs and attention that satisfies our assumptions (that is, a strictly increasing, differentiable, cumulative distribution function). These could represent other forms of heterogeneity across and within consumers that drives attention.

Platform Pricing. The platform price discriminates by offering a (direct) *mechanism* to providers. The mechanism consists of four functions. These correspond to the message or *certificate* assigned to a good provider claiming to be of value θ; the number of targeted views that this provider receives; the number of untargeted bad provider views that are assigned to value θ's certificate for each report of θ; and the price that the good provider pays to receive the certificate. As discussed below, the surplus from bad providers is fully extracted and so we do not need incorporate their payments into the mechanism.

These functions are written as follows:

$$M : \Theta \to \mathbb{R},$$

$$V_g : \Theta \to \mathbb{R}_+,$$

$$V_b : \Theta \to \mathbb{R}_+,$$

$$P : \Theta \to \mathbb{R}.$$

Note that the only private information is the value of the good providers so the mechanism is a function of Θ. A good provider whose value is θ pays $P(\theta)$ to receive a certificate $M(\theta)$ and $V_g(\theta)$ targeted views. Additionally, there are $V_b(\theta)$ untargeted views by bad providers that are also assigned certificated $M(\theta)$ for each report of θ.

For every m in the image of M, we use

$$\Lambda(m) = \frac{\mathbb{E}\left[V_g(\theta) \mid M(\theta) = m\right]}{\mathbb{E}\left[V_g(\theta) + V_b(\theta) \mid M(\theta) = m\right]},$$

to denote the fraction of good views assigned to certificate m or the *quality* of the certificate for short. When both the numerator and denominator are zero in the above fraction, $\Lambda(m)$ can be chosen arbitrarily.

The platform's mechanism design problem is

$$\max_{V_g,V_b,M,P} \int_{\Theta} \left[P(\theta) + A(\Lambda(M(\theta)))V_b(\theta) - \gamma(V_b(\theta) + V_g(\theta)) - c(V_g(\theta))\right] f(\theta)d\theta,$$

(1) subject to

$$\theta A(\Lambda(M(\theta)))V_g(\theta) - P(\theta) \geq \max\{\theta A(\Lambda(M(\theta')))V_g(\theta') - P(\theta'), 0\} \qquad \text{for all } \theta, \theta' \in \Theta.$$

The objective function sums the total payments received by the platform net of the costs of providing the views to the content providers. Each good provider of value θ pays the platform $P(\theta)$. Bad providers do not have any private information so the platform simply charges them the utility that they receive from being assigned certificate $M(\theta)$ and receiving $V_b(\theta)$ views, which is $A(\Lambda(M(\theta)))V_b(\theta)$. $A(\Lambda(M(\theta)))$ is the fraction of consumers who pay attention to content marked with a certificate $M(\theta)$. The cost of $V_b(\theta)$ untargeted views is $\gamma V_b(\theta)$. The cost of $V_g(\theta)$ targeted views is $\gamma V_g(\theta) + c(V_g(\theta)))$ due to the additional component associated with targeting. The constraint captures both incentive compatibility and individual rationality constraints for good providers.

The case that $A(1) \leq \gamma$ trivially implies that no views will be directed to bad providers since the costs to the platform would then be higher than the value of the views. Therefore, we henceforth focus on the more interesting case $A(1) > \gamma$ and we normalize $A(1) = 1$ (recall that $A(0) = 0$).

Before we begin analyzing the above problem, a few comments are in order. First, observe that because providers are infinitesimal, a misreport by value θ as a value $\theta' \neq \theta$ does not affect the quality $\Lambda(M(\theta'))$ of the certificate $M(\theta')$. Second, the above problem (1) bears a similarity to the classic work of Mussa and Rosen (1978). The key difference is that the platform is choosing *both* the quality (Λ) and quantity (V_g, V_b) of the product, and that these two are related.

2.1. *Preliminary Analysis and Simplification of the Platform's Problem*

We simplify the platform's problem with the following observation.

Lemma 1. *For any incentive compatible and individually rational mechanism (V_g, V_b, M, P) there exists another incentive compatible and individually rational mechanism $(V_g, \tilde{V}_b, \tilde{M}, P)$ such that, for all $\theta \in \Theta$, $\tilde{M}(\theta) = \theta$ and, both the platform and good providers of every value receive the same payoff as from (V_g, V_b, M, P).*

In words, this lemma states that it is without loss to assign a separate certificate to each value θ. This is because different certificates can have the same quality. So we can take any mechanism in which different values are assigned to the same certificate and construct a new mechanism in which all values have distinct certificates but we reassign the bad providers' views in a way that the quality is constant across certificates. The intuition stems from the linearity of the platform's payoff in V_b; the platform can reallocate the bad provider views V_b freely across different messages with identical qualities such that the ratio of good to bad content is equal to the average level, value-by-value.

This allows us to drop the certification decision M and the function V_b determining the bad provider views and rewrite the platform's problem with it choosing the quality $\Lambda : \Theta \to [0, 1]$ of the certificate for each good provider value θ. This is a consequence of Lemma 1 and the fact that the bad provider views V_b are pinned down by the equation $\Lambda(\theta) = V_g(\theta)/[V_g(\theta) + V_b(\theta)]$. The platform thus solves[13]

$$\max_{V_g,\Lambda,P} \int_\Theta \left[P(\theta) + A(\Lambda(\theta))V_g(\theta)\frac{1-\Lambda(\theta)}{\Lambda(\theta)} - \gamma\frac{V_g(\theta)}{\Lambda(\theta)} - c(V_g(\theta)) \right] f(\theta)d\theta,$$

subject to

$$\theta A(\Lambda(\theta))V_g(\theta) - P(\theta) \geq \max\{\theta A(\Lambda(\theta'))V_g(\theta') - P(\theta'), 0\} \qquad \text{for all } \theta, \theta' \in \Theta.$$

If we interpret $A(\Lambda(\theta))V_g(\theta)$ as the "allocation" when value θ is reported, the incentive compatibility constraint is essentially identical to the standard incentive compatibility constraint of Mussa and Rosen (1978). Thus, we can use the standard characterization of incentive compatibility to eliminate the price function P from the platform's problem and restate it as

$$\max_{V_g,\Lambda} \int_\Theta \left[\left(\phi(\theta) + \frac{1-\Lambda(\theta)}{\Lambda(\theta)}\right) A(\Lambda(\theta))V_g(\theta) - \gamma\frac{V_g(\theta)}{\Lambda(\theta)} - c(V_g(\theta)) \right] f(\theta)d\theta,$$

(2) subject to

$$A(\Lambda(\theta))V_g(\theta) \geq A(\Lambda(\theta'))V_g(\theta') \quad \text{for } \theta \geq \theta',\ \theta, \theta' \in \Theta.$$

In the above objective function

$$\phi(\theta) := \theta - \frac{1-F(\theta)}{f(\theta)}$$

is the standard *virtual value*. The constraint captures the fact that incentive compatibility requires the allocation to be nondecreasing. Note that we also eliminated the individual rationality constraint in the standard way by observing that it is optimal for the platform to provide a utility of

[13] In what follows, note that, in the integrand, a fraction whose numerator and denominator are both zero takes the value zero. We also note that, as written, the problem permits us to choose $\Lambda(\theta) > 0$ and $V_g(\theta) = 0$ (which technically violates the definition of Λ). We do not explicitly prevent this by imposing an additional constraint as, when $V_g(\theta) = 0$, the objective function and the constraints take the same values for any $\Lambda(\theta) \in [0, 1]$. This mild abuse allows us to state results more concisely without changing their economic content.

zero to a good provider of value $\theta = 0$. We refer to a solution of the platform's problem (2) as an *optimal mechanism*.

The following lemma summarizes the simplification of the platform's problem. It states that a pointwise solution that satisfies the monotonicity required for incentive compatibility is an optimal mechanism.

Lemma 2. *Suppose there is a mechanism (V_g^p, Λ^p) such that*

$$(V_g^p(\theta), \Lambda^p(\theta)) \in \underset{v_g, \lambda}{\operatorname{argmax}} \left[\left(\phi(\theta) + \frac{1-\lambda}{\lambda} \right) A(\lambda) v_g - \gamma \frac{v_g}{\lambda} - c(v_g) \right]$$

pointwise maximizes the platform's objective for all $\theta \in \Theta$ and that $A(\Lambda^p(\theta)) V_g^p(\theta)$ is nondecreasing in θ. Then (V_g^p, Λ^p) is an optimal mechanism; that is, it is a solution to the platform's problem (2).

In what follows, we assume the distribution F is such that the virtual value ϕ is strictly increasing. This is a standard technical assumption in mechanism design. This assumption simplifies the derivation of the optimal mechanism; as we will show, it implies that there is a pointwise optimum (V_g^p, Λ^p) as described in Lemma 2 such that $A(\Lambda^p(\theta)) V_g^p(\theta)$ is indeed nondecreasing. This assumption can be dispensed with by employing standard ironing techniques but we choose not to do so because we view the additional technicality (required for this generality) complicates the presentation of our main economic insights.

A useful way to view the pointwise problem in Lemma 2 is to separate the decision of certificate quality from the decision of the quantity of views to allocate to good providers. Let

$$R(\hat{\phi}, \lambda) := \left(\hat{\phi} + \frac{1-\lambda}{\lambda} \right) A(\lambda) - \frac{\gamma}{\lambda}.$$

Then, $\Lambda^p(\theta) = \operatorname{argmax}_\lambda \{R(\phi(\theta), \lambda)\}$ and $V_g^p(\theta) = \operatorname{argmax}_{v_g} \{R(\phi(\theta), \Lambda^p(\theta)) v_g - c(v_g)\}$. The latter problem, of choosing v_g, mimics Mussa and Rosen (1978), but where the maximized R stands in for the effective virtual value (or return) generated by v_g. Notice that when $\lambda = 1$, $R(\phi(\theta), 1) = \phi(\theta) - \gamma$, that is, the function R captures the true virtual value net of the cost γ of an untargeted view. The platform will deviate from perfect certification because this can lead to a higher value of $R(\phi(\theta), \lambda)$. The quantity of views directed at content the of good providers is simply derived by equating the marginal cost of targeting, c', with the return $R(\phi(\theta), \Lambda^p(\theta))$ if it is positive, and is set to zero if $R(\phi(\theta), \Lambda^p(\theta))$ is negative.

3. Benchmarks

In this section, we derive three benchmarks that provide context for the properties of the optimal mechanism. We first solve a planning problem in which the goal is to maximize consumer engagement net of platform costs. We then derive the properties of the optimal mechanism when the platform is restricted to using a constant quality for all Θ. This includes the case of perfect quality—not mixing content from good and bad providers—as desired by the European Commission. This contrasts the maximization of consumer engagement and platform profits. We then derive the properties of the optimal mechanism when the platform is restricted to offering only two certificates (so the quality function can take at most two values). This allows us to transparently demonstrate the value to the platform of selling multiple certificates, which is typically a feature of the solution to platform's problem described in expression (1).

3.1. *Engagement Maximizing Planner*

We consider a planner who wants to maximize the level of engagement on the platform net of platform costs.[14] That is, the planner solves

$$\max_{V_g,\Lambda}\left\{\int_\Theta \left[A(\Lambda(\theta))V_g(\theta)-\gamma\frac{V_g(\theta)}{\Lambda(\theta)}-c(V_g(\theta))\right]f(\theta)d\theta.\right\}$$

The first term in the objective function above is the engagement $A(\Lambda(\theta))V_g(\theta)$ of users with good providers (not the content provider's utility which incorporates the value of that engagement and writes as $\theta A(\Lambda(\theta))V_g(\theta)$). The remaining terms are the costs: both targeting costs for views to good providers, and untargeted views to bad providers (recalling that $V_b(\theta)=V_g(\theta)\frac{1-\Lambda(\theta)}{\Lambda(\theta)}$).

Note that the planner's objective is not the same as the sum of the utilities of the consumers of content and the profit of the platform. To determine the former, we would need to take a stand on the consumer utility from engagement generating content (which we normalized to 1); this combined with the attention function A pins down the distribution of the time costs of content reading q. Since consumers do not pay for content, there is no natural way to relate consumer preferences and firm profits (as there is in standard mechanism design problems). We therefore focus on features that are general to the relative weighting of the costs and benefits for the planner. Indeed, the social surplus should also incorporate the payoff of the content providers but we do not include this as content provider utility is never brought up in the policy debate around certification for sale.

The solution (V_g^e,Λ^e) to the problem of maximizing engagement above is immediate from pointwise optimization (since there are no constraints) and is given by

$$\Lambda^e(\theta)=1,$$
$$V_g^e(\theta)=c'^{-1}(1-\gamma)$$

[14] Under the particular class of attention functions $A(\lambda)=\lambda^\alpha$, user welfare is described by engagement (up to a constant of proportionality). See Appendix B for details.

for all $\theta \in \Theta$.[15]

We flag two intuitive properties of this solution. The first is that no views are directed to bad providers since directing content to them is costly and does not generate engagement. In other words, certification for all $\theta \in \Theta$ is perfect with all certificates having quality one. Second, since the planner only wants to maximize engagement, the same number of views are directed to good providers regardless of their value θ. Such egalitarian traffic will not arise from a profit maximizing platform, since the platform will direct more traffic to providers with higher willingness to pay for those views. Note that these two qualitative properties would be unchanged if we alter the planner's objective to assign different weights to the total engagement and platform costs.

3.2. *One certificate*

In this section, we study a benchmark in which the platform assigns the same message to all types $\theta \in \Theta$ or, in terms of the simplified platform problem in expression (2), the function Λ is restricted to be a constant function that takes some value $\lambda \in [0, 1]$.

We do this for several reasons. First, this corresponds to a realistic form of content moderation whereby platforms do not distinguish between different kinds of content (that is, all content is presented in the same way) but still choose how many views to allocate to different providers. The platform may choose to ban bad provider traffic (perfect certification) or not (imperfect certification). Second, the analysis crisply illustrates how imperfect certification can raise platform profits by expanding the subset of types Θ that the platform profitably serves. In this sense, imperfect certification allows for greater content diversity—–a central theme of our analysis and one that features in the optimal mechanism. Finally, it allows us to characterize the solution if the platform is forced to certify perfectly; that is, no content from bad providers is assigned the same certificate as content from good providers.

We first derive the optimal number of views V^g for an arbitrary quality $\lambda \in [0, 1]$. We then vary the quality λ to demonstrate the cross-subsidization effect of bad content on good content, and to consider a hypothetical policy that limits selling of certification without restricting steering. We view this as the natural benchmark for how the European Commission claims platforms should operate: certification should not be for sale, and certificates should be a clear statement of quality (as, for instance, they claim should be the case on X).[16] They have not taken or suggested any action against platforms that sell traffic in various ways, however.

[15] Recall that we have normalized $A(1) = 1$.

[16] `https://ec.europa.eu/commission/presscorner/detail/en/ip_24_3761`

Fixing a λ, the platform's problem (2) is

$$\begin{aligned}
\Pi^s(\lambda) := \max_{V_g} & \left\{ \int \left(\left(\phi(\theta) + \frac{1-\lambda}{\lambda} \right) A(\lambda) V_g(\theta) - \gamma \frac{V_g(\theta)}{\lambda} - c(V_g(\theta)) \right) f(\theta) d\theta \right\} \\
& \text{subject to} \\
& V_g(\theta) \geq V_g(\theta') \quad \text{for } \theta \geq \theta',\ \theta, \theta' \in \Theta.
\end{aligned} \tag{3}$$

It is immediate here that the pointwise optimum

$$V_g^{s,\lambda}(\theta) = c'^{-1} \left(\max \left\{ \left(\phi(\theta) + \frac{1-\lambda}{\lambda} \right) A(\lambda) - \frac{\gamma}{\lambda}, 0 \right\} \right) = c'^{-1}(\max\{R(\phi(\theta), \lambda), 0\})$$

satisfies the required monotonicity constraint because we assumed ϕ is increasing. For $\theta \in \Theta$ such that $V_g^{s,\lambda}(\theta)$ is not zero, the value is derived from the first-order condition

$$c'(V_g^{s,\lambda}(\theta)) = \left(\phi(\theta) + \frac{1-\lambda}{\lambda} \right) A(\lambda) - \frac{\gamma}{\lambda} = R(\phi(\theta), \lambda). \tag{4}$$

A useful special case is perfect certification, $\lambda = 1$, in which only good providers are assigned views. This sort of perfect certification would avoid a policy maker's complaint that messages are "deceiving." So, suppose that perfect certification were enforced, but steering was still for sale.[17] Then, the optimal number of views under perfect certification is $V_g^{s,1}(\theta) = c'^{-1}\left(\max\{\phi(\theta) - \gamma, 0\}\right).$

This solution also provides a natural connection to classic price discrimination: it exactly mirrors Mussa and Rosen (1978). Since there are only good content providers, the benefit is the virtual value, and the cost is the sum of the opportunity cost γ and the targeting cost c. In this solution, the marginal cost of assigning an additional view to a content provider of value θ is equal to the benefit which is precisely the virtual valuation ϕ (that ensures the appropriate information rents accrue to the good providers). Compared to engagement maximization, where all content gets the same number of views, the monopoly platform creates an asymmetry in the views allocated across θ via the virtual valuation ϕ. Those good providers with higher θ who value being seen more reveal their higher valuation by paying more to enjoy a higher number of engaged views.

Following the first-order condition (4), imperfect certification has two effects. First, it lowers attention and therefore reduces the amount that can be charged to good types; this reduction is proportional to the virtual value. On the other hand, it generates, for every good view, some bad traffic that can be monetized. This effect is constant and raises the payoff from views symmetrically for all virtual values. Therefore imperfect certification increases the value of views relatively more for low virtual values. As a result, imperfect certification makes traffic more egalitarian.

[17] This would seem to be consistent with the European Commission's concern and possible action regarding X, for example.

The following result provides a sufficient condition for the profit maximizing simple certification to be imperfect.

Proposition 1. *Suppose $\overline{\theta}A'(1) < 1 - \gamma$. Then, every optimal single certificate*

$$\lambda^{s^*} \in \underset{\lambda}{\operatorname{argmax}} \{\Pi^s(\lambda)\}$$

satisfies $\lambda^{s^*} < 1$. That is, optimal single certification is imperfect.

If $A'(1)$ is not too big, then the gains from imperfect certification outweigh the costs of lost attention and engagement. The net return to bad provider traffic near perfect certification, $1 - \gamma$, provides the sufficient bound on $A'(1)$.

Imperfect certification, when it is desirable for the platform, raises content diversity. Formally, we define the *content diversity* of a mechanism (V_g, Λ) as the set of good provider values $\{\theta \in \Theta | V_g(\theta) > 0\}$ that are allocated views. Thus, a mechanism (V_g', Λ') has greater content diversity than another mechanism (V_g, Λ) if a larger set of values are served by the former or $\{\theta \in \Theta | V_g'(\theta) > 0\} \supset \{\theta \in \Theta | V_g(\theta) > 0\}$. The following result shows that either the platform prefers perfect certification over a fixed level of imperfect certification, or the imperfect certification increases content diversity.

Proposition 2. *For any $0 < \lambda < 1$, either $\Pi^s(\lambda) < \Pi^s(1)$, or content diversity is higher under λ then under perfect certification.*

This previews our result which will extend to the full mechanism below: enforcing perfect certification, when it is not optimal for the platform, lowers content diversity.

Linear Attention. Throughout the paper, we will use the example $A(\lambda) = \lambda$ of a linear attention function (which corresponds to a uniform distribution of the cost of reading content) to illustrate the underlying forces driving our results. In particular, here, we use it to demonstrate how greater content diversity can arise with imperfect certification.

For a fixed level of certification λ, the corresponding optimal $V_g^{s,\lambda}$ is given by

$$V_g^{s,\lambda}(\theta) = c^{-1}\left(\max\left\{\phi(\theta)\lambda + 1 - \lambda - \frac{\gamma}{\lambda}, 0\right\}\right).$$

Perfect certification implies $\lambda = 1$ and therefore, $V_g^{s,1}(\theta) = c^{-1}(\max\{\phi(\theta) - \gamma, 0\})$.

Figure 1 shows the number of views a good content provider enjoys (on the y-axis) as a function of their virtual value (and therefore, indirectly of their type) and compares perfect certification (depicted by the dashed red line) and imperfect certification with a single certificate of quality $\lambda = \frac{1}{2}$ (depicted by the blue line).

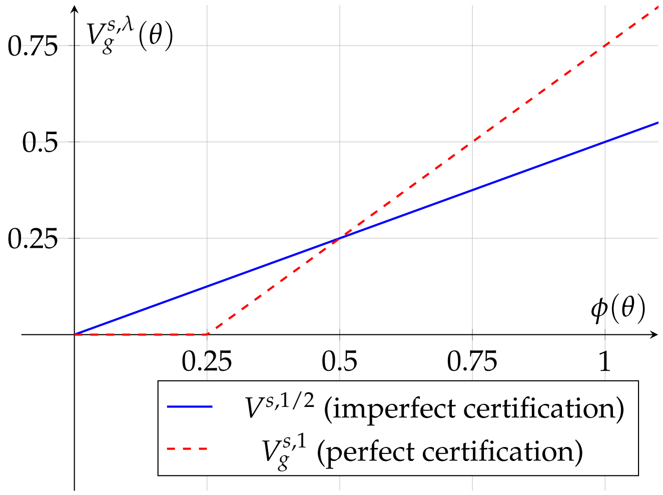


FIGURE 1. Good provider views with simple certification. Parameters: $A(\lambda) = \lambda$, $\gamma = 1/4$ and $c(v_g) = v_g^2/2$.

The figure highlights the two senses discussed above in which imperfect certification results in an allocation of views that is closer to the engagement maximizing benchmark (recall that the engagement maximizing benchmark has identical traffic for all θ). First, note that there is greater content diversity under imperfect certification; specifically, more lower value good providers are served with imperfect certification. This comes from good providers being served at a loss. Good providers with values that satisfy $\phi(\theta)A(\lambda) < \gamma$ are unwilling to make payments high enough to cover their average cost; they are served to keep attention high enough to make it worth serving bad types. Second, traffic is more egalitarian in that the number of views are less sensitive to the virtual value (and therefore the value) with imperfect certification (observe that the red line corresponding to perfect certification is steeper). Of course, these notions of improved content diversity do not translate directly into higher total engagement or welfare. The aggregate comparison is not immediately obvious from the figure since, in shifting from perfect to imperfect certification, some (higher) values receive fewer and other (lower) values receive more views. Finally, all good providers receive less attention conditional on being viewed. Moreover, in this figure, we have not taken a stance on the distribution F of types. We take up welfare in more detail below. However it is immediate that if the mass of F is concentrated on lower values then engagement must be higher with imperfect certification.

3.3. *Two certificates*

Before moving on to our analysis of the general problem, we build some intuition by considering the case of two certificates. This case may also be of independent interest in that it mirrors recent developments in many social media platforms. Specifically, until relatively recently, Instagram and Twitter (now X) had users who were either verified (and their accounts were marked with a check sign) or not (their accounts were unmarked). While they initially did not charge for those providers who obtained a verified status, two certificates is a natural benchmark to study due to this historical precedent. Indeed, when Twitter first started charging for the provision of verified status, they only (in our language) offered two certificates. This section shows how this practice can improve profits for the platform.

The previous subsection showed that fixed imperfect (relative to perfect) certification can lead to higher profits and greater content diversity. The case with two certificates further illustrates how multiple levels of certification can raise the profits for the platform. By using two certificates, the platform can profitably serve low-value good providers, by cross subsidizing them with bad providers, while simultaneously not sacrificing engagement for the views of high value good providers.[18] Furthermore, the two certificate benchmark allows us to demonstrate how the two instruments—certification and steering—interact. We provide conditions under which the optimal policy has two distinct levels of certification, and many levels of steering; simple imperfect certification may be profitable, but is not optimal.

As for the case of a single certificate above, we begin by supposing that the quality associated with the two certificates is exogenously given by $\underline{\lambda}, \overline{\lambda} \in [0,1]$ with $\underline{\lambda} \leq \overline{\lambda}$. We write the two certificate benchmark problem for the platform as

$$\begin{aligned}\Pi^{bin}(\underline{\lambda}, \overline{\lambda}, \hat{\theta}) := \max_{V_g} \Bigg\{ & \int_0^{\hat{\theta}} \left[\left(\phi(\theta) + \frac{1-\underline{\lambda}}{\underline{\lambda}} \right) A(\underline{\lambda}) V_g(\theta) - \gamma \frac{V_g(\theta)}{\underline{\lambda}} - c(V_g(\theta)) \right] f(\theta) d\theta \\ & + \int_{\hat{\theta}}^{\overline{\theta}} \left[\left(\phi(\theta) + \frac{1-\overline{\lambda}}{\overline{\lambda}} \right) A(\overline{\lambda}) V_g(\theta) - \gamma \frac{V_g(\theta)}{\overline{\lambda}} - c(V_g(\theta)) \right] f(\theta) d\theta \Bigg\}, \end{aligned} \tag{5}$$

where $\hat{\theta} \in \Theta$. In the above problem, all values above and below $\hat{\theta}$ are assigned the higher $\overline{\lambda}$ and lower $\underline{\lambda}$ quality certificates respectively.

The optimal two certificate mechanism can be derived by choosing the appropriate $\underline{\lambda}$, $\overline{\lambda}$ and $\hat{\theta}$ to maximize $\Pi^{bin}(\underline{\lambda}, \overline{\lambda}, \hat{\theta})$.[19] In addition to higher value good providers receiving the higher quality certificate, they also receive more targeted views. Thus, as in the single certificate optimum, the platform uses the quantity of views to price discriminate but now also uses the certificate quality to separate higher from lower value good providers. The platform assigns higher-quality certificates

[18] In using certification to soften the incentive constraints for higher types, there is some similarity to Deneckere and McAfee (1996). Of course, in our environment the platform earns revenue (from bad providers) in "damaging" the good rather than incurring costs. More substantively, in Deneckere and McAfee (1996) consumers are constrained to unit demand, whereas we vary both the number of views and the quality of the certificate. This results in different economic effects.

[19] We characterize the two certificate optimal mechanism in Appendix C; we choose not present the result here for brevity. We show that the optimal two certificate mechanism takes the cutoff form assumed in the above problem (5).

to high-value good providers so as not to dilute earnings from these types, while assigning lower-quality certificates to use low-value good providers to use them as a means of earning revenue from the bad providers.

Instead of presenting the two certificate optimum, we derive a sufficient condition under which the platform gets strictly higher profits from using two certificates (as opposed to a single certificate $\underline{\lambda} = \overline{\lambda}$).

Proposition 3. *Suppose $\overline{\theta}A'(1) < 1 - \gamma$. Then, every two certificate optimum*

$$(\underline{\lambda}^*, \overline{\lambda}^*, \hat{\theta}^{bin^*}) \in \underset{\underline{\lambda}, \overline{\lambda}, \hat{\theta}}{\operatorname{argmax}} \left\{ \Pi^{bin}(\underline{\lambda}, \overline{\lambda}, \hat{\theta}) \right\}$$

satisfies $\underline{\lambda}^* < \overline{\lambda}^*$.

This is the same sufficient condition under which the platform chooses imperfect certification when restricted to a single certificate (Proposition 1), so some form of imperfect certification must be optimal. The reason for using multiple certificates is that, in the platform's problem (5), the profit depends on the interaction of the good provider's value with the quality.

We next turn to the fully optimal mechanism in which any number of messages can be used. The forces outlined in both the single and two certificate cases come to the fore: the platform uses different quantities of views to price discriminate coupled with the assignment of distinct quality certificates to effectively extract revenue from bad providers.

4. Certification and steering for sale

4.1. *The optimal mechanism*

We now characterize and analyze the optimal mechanism. The characterization builds on the intuition in Section 3. Relative to the planner's problem, the profit-seeking platform may use imperfect certificates as a means of raising revenue from bad providers and can more profitably price discriminate across good content providers by offering different quantities of targeted views and distinct quality certificates. The following proposition shows that the pointwise optimum satisfies the required conditions to ensure incentive compatibility and both the views V_g and the quality Λ are nondecreasing in θ.[20]

[20] When $V_g^*(\theta) = 0$, the platform makes no profits from value θ so any λ is optimal. For expositional simplicity, we pick the λ that maximizes effective virtual value $R(\phi(\theta), \lambda)$. Further, whenever there are multiple values of λ that maximize $R(\phi(\theta), \lambda)$, we pick the largest solution. This choice is convenient to present results and does not sacrifice any qualitative insight.

Proposition 4. *There is an optimal mechanism* (V_g^*, Λ^*) *solving the platform's problem* (2) *where both* V_g^*, Λ^* *are nondecreasing and satisfy*

$$\Lambda^*(\theta) = \max\left\{\tilde{\lambda} \;\middle|\; \tilde{\lambda} \in \underset{\lambda\in[0,1]}{\operatorname{argmax}}\left\{\left(\phi(\theta)+\frac{1-\lambda}{\lambda}\right)A(\lambda)-\frac{\gamma}{\lambda}\right\}\right\}$$

$$V_g^*(\theta) = c'^{-1}\left(\max\left\{\left[\phi(\theta)+\frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right]A(\Lambda^*(\theta))-\frac{\gamma}{\Lambda^*(\theta)},0\right\}\right)$$

for all $\theta \in \Theta$.

Recall that incentive compatibility is satisfied if the product $A(\Lambda^*(\cdot))V_g^*(\cdot)$ is nondecreasing; thus, one qualitative contribution of the above result is to show that both $\Lambda^*(\cdot)$ and $V_g^*(\cdot)$ are each individually nondecreasing. This implies that good providers with higher values both receive more views and their content is pooled with less bad content. Certification can be perfect (that is, $\Lambda^*(\theta) = 1$) for sufficiently high values θ. Recall that one can interpret this as first choosing λ to maximize the effective virtual value

$$\Lambda^*(\theta) = \max\left\{\tilde{\lambda} \;\middle|\; \tilde{\lambda} \in \underset{\lambda\in[0,1]}{\operatorname{argmax}}\{R(\phi(\theta),\lambda)\}\right\}$$

and then choosing the views in the "standard" way as

$$V_g^*(\theta) = c'^{-1}(\max\{R(\phi(\theta),\Lambda^*(\theta)),0\}).$$

The forces in Section 3.2 that pushed the optimal single certificate mechanism with imperfect (relative to perfect) certification qualitatively closer to planner's optimal solution also apply in the optimal mechanism of Proposition 4. Serving low-value good providers allows the platform to earn more revenue from bad providers. This, in turn, flattens out the relationship between targeted views and good provider valuations leading to more egalitarian content provision relative to perfect certification.

The optimal mechanism of Proposition 4 features greater content diversity than either the single or two certificate benchmarks. To see this, it is instructive to compare the optimal (unrestricted) mechanism to the mechanism that maximizes profits when the platform is restricted to offer only two certificates (derived in Proposition C.1 in the appendix). For any pair of binary qualities $0 < \underline{\lambda} < \overline{\lambda} \leq 1$, it must be the case that

$$\max_{\lambda\in[0,1]}\left\{\left(\phi(\theta)+\frac{1-\lambda}{\lambda}\right)A(\lambda)-\frac{\gamma}{\lambda}\right\} \geq \max_{\lambda\in\{\underline{\lambda},\overline{\lambda}\}}\left\{\left(\phi(\theta)+\frac{1-\lambda}{\lambda}\right)A(\lambda)-\frac{\gamma}{\lambda}\right\}$$

and consequently, the set of values $\{\theta \in \Theta \mid V_g^*(\theta) > 0\} \supseteq \{\theta \in \Theta \mid V_g^{bin}(\theta) > 0\}$ which receive any views at all is a larger set in the optimal mechanism relative to the binary benchmark. This greater content diversity is a result of the optimal mechanism directing traffic towards low-values θ that are unserved under binary certificates. These are values $0 \leq \theta \leq \underline{\theta}$ for which

$$\max_{\lambda \in [0,1]} \left\{ \left(\phi(\theta) + \frac{1-\lambda}{\lambda} \right) A(\lambda) - \frac{\gamma}{\lambda} \right\} > 0$$

where $\underline{\theta}$ satisfies

$$\left(\phi(\underline{\theta}) + \frac{1-\underline{\lambda}}{\underline{\lambda}} \right) A(\underline{\lambda}) - \frac{\gamma}{\underline{\lambda}} = 0.$$

Linear Attention. Returning to the case where the attention function $A(\lambda) = \lambda$ is linear, Proposition 4 implies that the mechanism

$$\Lambda^*(\theta) = \begin{cases} \sqrt{\frac{\gamma}{1-\phi(\theta)}} & \text{if } \phi(\theta) \leq 1-\gamma, \\ 1 & \text{if } \phi(\theta) > 1-\gamma, \end{cases} \tag{6}$$
$$V_g^*(\theta) = \begin{cases} c'^{-1}\left(\phi(\theta)\Lambda^*(\theta) + 1 - \Lambda^*(\theta) - \frac{\gamma}{\Lambda^*(\theta)}\right) & \text{if } \phi(\theta) \geq 1 - \frac{1}{4\gamma} \text{ and} \\ 0 & \text{if } \phi(\theta) < 1 - \frac{1}{4\gamma} \end{cases}$$

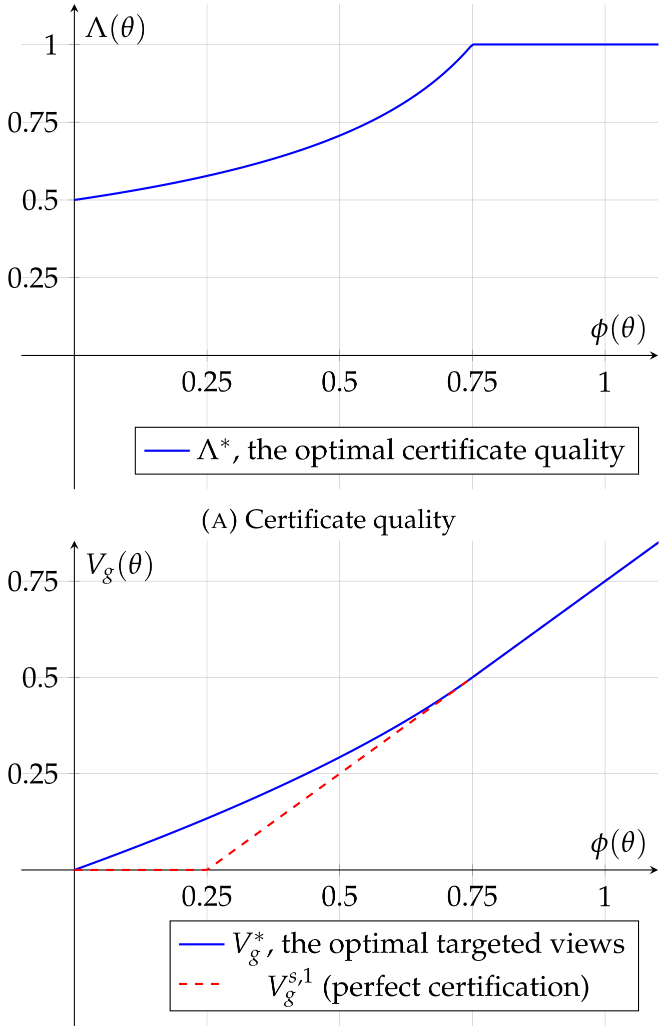
is optimal.

Now observe that $1 - \frac{1}{4\gamma} \geq 1 - \gamma$ when $\gamma \geq 1/2$. Consequently, when $\gamma \geq 1/2$, the optimal mechanism has the property that $\Lambda^*(\theta) = 1$ for all $\theta \in \Theta$ such that $V_g^*(\theta) > 0$. That is, perfect certification is optimal when it is sufficiently costly to supply untargeted views. We therefore focus on the case of $\gamma < 1/2$ in which the optimal mechanism features a variety of levels of certification.

Figure 2 illustrates such a case. Each panel has the good provider's virtual value ϕ on the x-axis. Panel 2a plots the certificate quality Λ^* and Panel 2b plots the number of good provider views V_g^* in the optimal mechanism. Sufficiently high values are all assigned perfect certificates but the platform still price discriminates through the number of views it provides. The platform employs both instruments—imperfect certification and the quantity of views—for the lower values. In contrast to the case with only two certificates, the sale of lower quality certificates to lower values entails a gradual, continuous degradation in certificate quality (rather than a discrete fall).

As a result of using different certificates to price discriminate, content diversity increases relative to perfect certification. In addition, values θ that are assigned imperfect certificates ($\Lambda^*(\theta) < 1$) receive more views than they would under perfect certification. As in the case with a single imperfect certificate, good providers with values θ such that $\phi(\theta)A(\Lambda(\theta)) < \gamma$ are being served at a loss, in terms of direct revenue from the prices charged, in order to generate revenue from bad providers.[21]

[21] Notice that the fact that all good providers served in this example have positive virtual values is an artifact of the specific chosen parameters. For a lower γ, some values θ with negative virtual values $\phi(\theta) < 0$ are also served ($V_g^*(\theta) > 0$) in the optimal mechanism.



(A) Certificate quality

(B) Content Diversity

FIGURE 2. The optimal mechanism. Parameters: $A(\lambda) = \lambda, \gamma = 1/4, c(v_g) = v_g^2/2$.

4.2. *Comparative Statics*

The solution described in Proposition 4 allows us to conduct several comparative statics exercises. They help to explain the circumstances under which imperfect certification is optimal. Lower costs γ for untargeted views or a more concave attention function A make optimal certification more imperfect but a proportionate lower cost of targeting c has no impact.

Costs of untargeted views. We first examine how the quality of the certificates in the optimal mechanism (V_g^*, Λ^*) are affected by the cost γ. Recall, that a natural interpretation for γ, which is the opportunity cost of providing an untargeted view, is lost ad revenue. Thus, the next result shows that when ad revenue falls, good content providers enjoy worse certificates.

Proposition 5. *The quality $\Lambda^*(\theta)$ in the optimal mechanism (V_g^*, Λ^*) is nondecreasing in γ for all $\theta \in \Theta$.*

For intuition, consider a good provider with value $\theta \in \Theta$ that is assigned to an imperfect certificate $\Lambda^*(\theta) \in (0,1)$. Since, $\Lambda^*(\theta)$ maximizes $\left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda}$ (see Proposition 4), it satisfies the first-order condition

$$A(\Lambda^*(\theta)) - \Lambda^*(\theta)^2 \left(\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right) A'(\Lambda^*(\theta)) = \gamma. \tag{7}$$

This equation captures the tradeoffs of including additional bad content with a given certificate. The right hand side is the cost of the additional bad view. If γ is lowered, more bad provider views can be absorbed before the left side balances the right. The benefits of the bad traffic on the left contains two terms. The first is the amount $A(\Lambda^*(\theta))$ that a bad provider pays for this view. But including more bad content reduces the quality of the certificate and therefore the attention; this, in turn, reduces the amount that both good and bad providers are willing to pay, which is captured by the second term.

On the margin, additional bad views reduce the quality by $\frac{\Lambda^*(\theta)^2}{V_g^*(\theta)}$ and therefore the attention by $\frac{\Lambda^*(\theta)^2 A'(\Lambda^*(\theta))}{V_g^*(\theta)}$. The total amount paid for this certificate by both good and bad providers is $\left(\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right) V_g^*(\theta)$. On the margin, the lost revenue from a reduction of quality is given by the product of these two terms which is independent of the revenue from good providers. As a result, the comparative static does not depend on the cost of targeting views c. Intuitively, it is worthwhile to assign more untargeted bad provider views if they are relatively cheaper, since targeted views to good providers involve an additional cost for the platform.

In particular, Proposition 5 and its proof highlight that there are model parameters such that $\Lambda^*(\bar{\theta}) = 1$ for a given γ but $\Lambda^*(\bar{\theta}) < 1$ for some $\gamma' < \gamma$. So falling ad revenue can result in platforms abandoning perfect certification, as was perhaps the case for Twitter.

Costs of targeted views. We conduct a similar comparative static for the cost c of directing interested views at good providers. Again, there is a natural interpretation: more information on viewers, improved algorithms and analytics reduce costs of targeting interested viewers. To consider the effect of such changes, we introduce a parameter $\kappa > 0$ (that only appears in the following discussion) such that the cost of v_g targeted views to good providers is $\kappa c(v_g)$.

Proposition 6. *Let the cost $\kappa c(v_g)$ of interested views v_g be parametrized by $\kappa > 0$. Then, the optimal mechanism (V_g^*, Λ^*) satisfies the following properties.*

(i) *For every $\theta \in \Theta$, the quality $\Lambda^*(\theta)$ is the same for all $\kappa > 0$.*
(ii) *For every $\theta \in \Theta$, the quantity of views $V_g^*(\theta)$ is nonincreasing in κ.*
(iii) *The set of good provider values $\{\theta \in \Theta \mid V_g^*(\theta) > 0\}$ that are served is the same for all $\kappa > 0$.*

Proposition 6 argues that when targeting improves (that is, κ falls) so that it becomes cheaper to find engaged consumers for good content providers, the quality of certificates do not change. Instead, good content providers enjoy more views and the platform earns more revenue from bad providers in proportion. This comparative static is immediate from the expressions for the optimal mechanism in Proposition 4.

Shape of consumer attention function. Lastly, we also examine the effect of making the attention function A more concave or convex. As is clear from the first-order condition (7), the shape of the attention function determines how the platform trades off lower certificate quality against higher revenue from bad providers. Specifically, if the platform optimally chooses perfect certification for a value θ and a given attention function, it will also be optimal for the platform to choose perfect certification at this value for any other attention function that has higher slope $A'(1)$ at quality 1. Simply put, a higher slope implies that it is more costly—in terms of lost attention—for the platform to generate revenue from bad providers.

The concavity or convexity of attention reflect consumer preferences. The attention function can be interpreted as the costs of reading content but, in a more flexible formulation, one might imagine this also reflecting anticipated benefits from good content relative to the costs of bad content. Since the attention function satisfies $A(0) = 0$ and $A(1) = 1$, it lies above the 45 degree line when it is concave and and below when it is convex. Concavity is consistent with consumers who are particularly keen to find good content and suffer relatively little from the inconvenience associated with looking at some bad content. Instead, those with convex attention can be understood as being harmed by even a little bad content. Different kinds of media content might be thought of as differing on this scale: scrolling past bad entertainment content is perhaps only an inconvenience whereas consuming fake news is more harmful. To the extent that concavity or convexity of the attention function captures features of these different kinds of social media, the following result suggests that one would observe a greater amount of bad content on entertainment-oriented social media rather than news-related media.[22]

Proposition 7. *Suppose the attention function $A(\hat{\lambda};\hat{\alpha}) = A(\hat{\lambda})^{\hat{\alpha}}$ is parametrized by a constant $\hat{\alpha} > 0$. Denote the optimal mechanism for a given $\hat{\alpha}$ as $(V^*_{g,\hat{\alpha}}, \Lambda^*_{\hat{\alpha}})$. Then for every $\theta \in \Theta$ and every $0 < \alpha' < \alpha$ with $V^*_{g,\alpha}(\theta) > 0$, we have $V^*_{g,\alpha'}(\theta) > 0$ and $\Lambda^*_{\alpha}(\theta) \geq \Lambda^*_{\alpha'}(\theta)$.*

Note that the attention function in the statement of the proposition is arbitrary so there need not be a unique λ that maximizes $R(\phi(\theta), \lambda)$ for all $\theta \in \Theta$. Nonetheless, we obtain the requisite

[22] Of course, this is a *ceteris paribus* statement and one might expect, for example, that bad providers value views differently across these different kind of media.

monotonicity for $\Lambda^*(\theta)$ (which recall, is defined to be the highest λ that maximizes $R(\phi(\theta), \lambda)$.[23] More generally, the result extends to attention functions and concave transformations for which there is a unique λ that maximizes $R(\phi(\theta), \lambda)$ for all $\theta \in \Theta$ under both the original attention function and the concave transformation.

For intuition, consider a θ such that $\Lambda^*(\theta) \in (0, 1)$ is interior. Recall that $\Lambda^*(\theta)$ must satisfy the first-order condition

$$\left(\phi(\theta) + \frac{1 - \Lambda^*(\theta)}{\Lambda^*(\theta)}\right) A'(\Lambda^*(\theta)) - \frac{A(\Lambda^*(\theta))}{\Lambda^*(\theta)^2} + \frac{\gamma}{\Lambda^*(\theta)^2} = 0.$$

Take an increasing, differentiable, strictly concave function μ such that $\mu(0) = 0$ and $\mu(1) = 1$. Replacing A by the concave transformation $\mu \circ A$ on the left side of the first-order condition, we get

$$\left(\phi(\theta) + \frac{1 - \Lambda^*(\theta)}{\Lambda^*(\theta)}\right) A'(\Lambda^*(\theta))\mu'(A(\Lambda^*(\theta)) - \frac{A(\Lambda^*(\theta))}{\Lambda^*(\theta)^2}\frac{\mu(A(\Lambda^*(\theta))}{A(\Lambda^*(\theta))} + \frac{\gamma}{\Lambda^*(\theta)^2} < 0.$$

This implies that, for the attention function $\mu \circ A$, there are values of $\lambda < \Lambda^*(\theta)$ such that $R(\phi(\theta), \lambda) > R(\phi(\theta), \Lambda^*(\theta))$. To understand the inequality, observe that, relative to the original first-order condition, the first term is scaled by the marginal $\mu'(A(\Lambda^*(\theta))$ and the second term by the average $\frac{\mu(A(\Lambda^*(\theta))}{A(\Lambda^*(\theta))}$. But for strictly concave functions from the unit interval to the unit interval, the average $\frac{\mu(a)}{a}$ is bigger than 1 and $\frac{\mu(a)}{a} > \mu'(a)$; that is, the average is greater than the marginal for all $a \in (0, 1)$. The intuition is similar to that of the comparative static for the cost γ: the concave transformation scales the benefits of bad provider views more than proportionally to its cost, and so the platform generates more revenue from bad providers.

4.3. *Comparison to perfect certification*

Since regulators have suggested that consumers would benefit from enforced perfect certification, it is useful to compare the optimal mechanism with the one studied in the perfect certification benchmark of Section 3.2. Before we state the comparison, recall that, when perfect certification $\lambda = 1$ is enforced, the optimal number of targeted views is given by $V_g^{s,1}(\theta) = c^{-1}(\max\{\phi(\theta) - \gamma, 0\})$. In other words, the platform only serves values $\theta \in \Theta$ such that their virtual values satisfy $\phi(\theta) > \gamma$.

Proposition 8. *The optimal mechanism (V_g^*, Λ^*) must satisfy one of the following two properties.*

(1) *Perfect certification is optimal: $\Lambda^*(\theta) = 1$ for all $\theta \in \Theta$ and consequently, $V_g^*(\theta) > 0$ if, and only if, $\phi(\theta) > \gamma$.*

[23] Proposition A.1 in the appendix provides a more general treatment of comparative statics under parameterized transformations of A.

(2) *Enforced perfect certification reduces content diversity:* $\Lambda^*(\theta) < 1$ *for some* $\theta \in \Theta$ *with* $\phi(\theta) > \gamma$ *and* $V_g^*(\theta) > 0$ *for some* $\theta \in \Theta$ *with* $\phi(\theta) < \gamma$*.*

We already argued (after the statement of Proposition 4) that the optimal mechanism features more content diversity than the single or two certificate benchmarks. What Proposition 8 additionally shows is that, when perfect certification is not optimal, the platform serves *strictly* more values and there is a quality, content-diversity tradeoff. In other words, the optimal mechanism never takes the form that $\Lambda^*(\theta) = 1$ for all $\theta \in \Theta$ with $\phi(\theta) > \gamma$ when $V_g^*(\theta) > 0$ for some $\theta \in \Theta$ with $\phi(\theta) < \gamma$.

This result suggests that the total engagement $\int_\theta A(\Lambda(\theta))V_g(\theta)f(\theta)d\theta$ (our measure of consumer welfare) may be higher or lower with enforced perfect certification depending on whether the value distribution F assigns a greater mass to values above or below $\phi^{-1}(\gamma)$. But note that additionally, for all values $\theta \in \Theta$ with $\phi(\theta) > \gamma$ to which the optimal mechanism assigns a lower quality $\Lambda^*(\theta) < 1$, the number of good provider views $V_g^*(\theta) > V^{s,1}(\theta)$ are strictly higher. This is because, by definition (from Proposition 4) if $\Lambda^*(\theta) < 1$, it implies that

$$\left[\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right] A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)} > \phi(\theta) - \gamma$$

and consequently

$$V_g^*(\theta) = c'^{-1}\left(\left[\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right] A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)}\right) > c'^{-1}\left(\phi(\theta) - \gamma\right) = V_g^{s,1}(\theta).$$

Indeed for θ such that $\phi(\theta) = \gamma$, the number of views under enforced perfect certification $V_g^{s,1}(\theta) = 0$ whereas $V_g^*(\theta) > 0$ and so engagement is higher in a neighborhood of this θ when the platform is allowed to sell imperfect certification. Put differently, even for values in the region $\theta \in \Theta$ with $\phi(\theta) > \gamma$, the effect of enforced perfect certification on engagement is ambiguous.

Linear Attention. We now return to the linear case to study this tradeoff more explicitly and to examine the effect of enforced perfect certification on total engagement. Recall that, with a linear attention function, the quality of certification (see expression (6)) in the optimal mechanism is perfect ($\Lambda^*(\theta) = 1$) for $\theta \in \Theta$ such that $\phi(\theta) > 1-\gamma$. For these θ, there will no difference between the levels of engagement ($V_g^*(\theta)$) from the optimal mechanism and that ($V_g^{s,1}(\theta)$) from enforced perfect certification. For $\theta \in \Theta$ such that $1 - \frac{1}{4\gamma} < \phi(\theta) < \gamma$, the platform assigns zero views ($V_g^{s,1}(\theta) = 0$) under enforced perfect certification but positive views ($V_g^*(\theta) > 0$) in the optimal mechanism. Clearly for these values, engagement is higher in the optimal mechanism. For intermediate values $\gamma < \phi(\theta) < 1-\gamma$ there are offsetting effects.[24] For example, when θ is

[24] Recall that we are focusing on the nontrivial case of $\gamma < 1/2$.

just lower than $1 - \gamma$, the total number of views is higher in the optimal mechanism but this is more than offset by the drop in quality. For these values, engagement is higher under enforced perfect certification. Figure 3 shows the difference in levels of engagement as a function of the virtual value for the parameters of our running example.

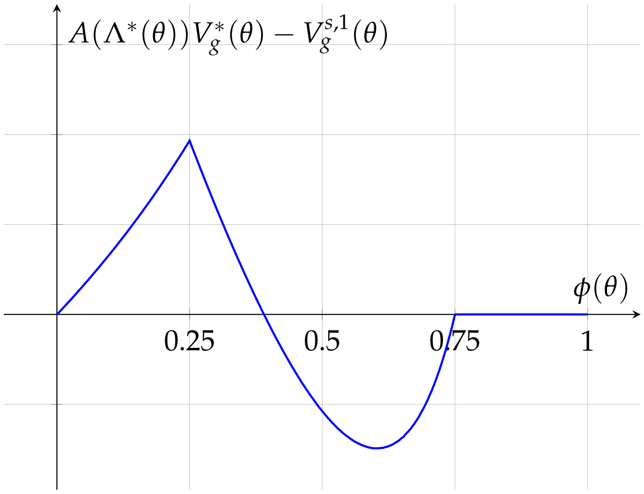


FIGURE 3. Difference in engagement between the optimal mechanism and enforced perfect certification. Parameters: $A(\lambda) = \lambda, \gamma = 1/4, c(v_g) = v_g^2/2$.

Which scenario results in higher total engagement depends on the value distribution F. Loosely speaking, if F is concentrated on values of $\theta \in \Theta$ for which $A(\Lambda^*(\theta))V_g^*(\theta) - V_g^{s,1}(\theta) > 0$, then enforced perfect certification leads to lower total engagement and vice versa. It is easy to construct examples of distributions F such that either case arises.

Small cost γ of untargeted views. It is possible to characterize when enforced perfect certification is and is not beneficial in the special case where, relative to the cost of targeted views $c(v_g)$, the cost γ of untargeted views becomes small.[25] This special case is particularly relevant as there are many platforms whose opportunity cost, in terms of lost advertising revenue per view (our preferred interpretation of γ), is small.

A natural starting point is to evaluate the linear attention case ($A(\lambda) = \lambda$) for which the optimal mechanism is characterized in expression (6). For $\theta \in \Theta$ such that $\phi(\theta) \geq 1$, optimal certification $\Lambda^*(\theta) = 1$ is perfect regardless of the value of γ. Since $\Lambda^*(\theta) = \sqrt{\frac{\gamma}{1-\phi(\theta)}}$ for $\theta \in \Theta$ such that $\phi(\theta) < 1 - \gamma$, this implies that the optimal certification $\Lambda^*(\theta) \to 0$ as $\gamma \to 0$ for all θ such that $\phi(\theta) < 1$. Moreover, the optimal number of views $V_g^*(\theta) = c'^{-1}\left(\phi(\theta)\Lambda^*(\theta) + 1 - \Lambda^*(\theta) - \frac{\gamma}{\Lambda^*(\theta)}\right) \to$

[25] Stated in terms of the probability ρ of finding an interested user with an untargeted view described in footnote 10, we are effectively assuming that the ratio $\frac{\gamma}{\rho}$ is not falling, so that targeting is still necessary for good providers.

$c'^{-1}(1)$ as $\gamma \to 0$ for all θ such that $\phi(\theta) < 1$. Therefore, under the optimal mechanism, the total engagement for values θ such that $\phi(\theta) < 1$ vanishes as γ becomes small. By contrast, the platform provides views $V_g^{s,1}(\theta) = c'^{-1}\left(\max\{\phi(\theta) - \gamma, 0\}\right)$ for all $\theta \in \Theta$ under enforced perfect certification. That is, there is positive total engagement for values θ such that $0 < \phi(\theta) < 1$ even when γ is small. This implies that, for small γ and linear attention, enforced perfect certification unambiguously leads to higher total engagement.

The next result shows that, if we assume that the attention function and the cost of targeted views are power functions, then enforced perfect certification unambiguously improves total engagement only if the attention function is not very concave.

Proposition 9. *Suppose $A(\lambda) = \lambda^\alpha$ where $\alpha \in (0,1)$ and $c(v_g) = \frac{v_g^\sigma}{\sigma}$ where $\sigma > 1$. Then, for all $\theta \in \Theta$, the optimal mechanism (V_g^*, Λ^*) satisfies*

$$\lim_{\gamma \downarrow 0} \Lambda^*(\theta) = 0 \quad \textit{and}$$

$$\lim_{\gamma \downarrow 0} A(\Lambda^*(\theta)) V_g^*(\theta) = \begin{cases} 0 & \textit{if } \alpha \in \left(\frac{1}{\sigma}, 1\right), \\ \alpha^{\frac{1}{\sigma - 1}} & \textit{if } \alpha = \frac{1}{\sigma}, \\ \infty & \textit{if } \alpha \in \left(0, \frac{1}{\sigma}\right). \end{cases}$$

Consequently, there exists a $\underline{\gamma} > 0$, such that for all $\gamma < \underline{\gamma}$ enforced perfect certification leads to higher (lower) total engagement than the optimal mechanism when $\alpha < (>) \frac{1}{\sigma}$.

Proposition 9 characterizes two properties of the optimal mechanism when γ is small. The first is that, when the cost γ of untargeted views becomes small, the platform provides more such views to bad providers and this results in the quality degrading for all certificates. This property might seem intuitive but it relies on the fact that the attention function is concave; in particular, it is not true for the linear attention case we discussed above or the convex attention case we discuss below. But this does not imply that engagement also becomes small since the platform may find it optimal to increase the number of good provider views at a faster rate. Engagement grows unboundedly only when the attention function is sufficiently concave. Intuitively, for small α, the attention function $A(\lambda) = \lambda^\alpha$ is very steep for λ close to zero and consequently, if the platform does not increase the number of engaged views for good providers in combination with the untargeted views for bad providers, this results in a large loss of attention and therefore revenue.

Proposition 9 has two implications. For less concave attention functions, enforced perfect certification is essential since the total engagement goes to zero if the platform is allowed to degrade certificate quality. Conversely, for more concave attention functions, enforced perfect certification can result is large engagement losses because, even though the platform degrades certificate quality, they are willing to spend large amounts on targeted views of good provider content to generate engagement.

The above result restricts attention to concave attention functions so it is natural to ask whether its conclusions extend to convex attention functions. They do not. To see this, recall the first-order condition

$$\Lambda^*(\theta)^\alpha \left[1 - \alpha - (\phi(\theta) - 1)\alpha\Lambda^*(\theta)\right] = \gamma$$

that $\Lambda^*(\theta)$ satisfies if it is interior (to see this, substitute $A(\lambda) = \lambda^\alpha$ into Equation (7)). Consider $\theta \in \Theta$ such that $\phi(\theta) = 0$. When γ is small, the optimal $\Lambda^*(\theta)$ is interior and so the first-order condition holds. This in turn implies that the term in the square brackets in the above expression must be positive so

$$1 - \alpha + \alpha\Lambda^*(\theta) \geq 0 \iff \Lambda^*(\theta) \geq 1 - \frac{1}{\alpha}.$$

Thus, as we take $\gamma \downarrow 0$, the quality $\Lambda^*(\theta)$ is bounded above zero so does not converge to zero as it does with concave attention shown in Proposition 9. Moreover, for this θ, the number of views $V_g^*(\theta)$ is also bounded above zero. This will also be true for θ such that $\phi(\theta)$ is negative but close to zero. Thus, for these types, the optimal mechanism yields strictly positive engagement for small γ but such types are unserved under enforced perfect certification. In other words, case 2 of Proposition 8 applies: for some values of θ, enforced perfect certification may be beneficial for engagement, but it also reduces content diversity in a way that lowers engagement for other θ.[26] Thus, when the attention function is convex, whether or not enforced perfect certification leads to greater total engagement depends on the weights the distribution F places on different values θ.

5. Extension: Addiction and Losses from Bad Content

In this section, we adapt our model in two ways in order to assess two policy-relevant concerns. First that viewing bad content may be costly for consumers, and second that consumers may suffer from digital addiction. We separately implement both into the model in rather straightforward ways and find intuitive results. As bad content becomes more costly, platforms implement cleaner certification; and as consumers are more addicted, certification becomes worse and consumers are exposed to more bad content. We then briefly revisit some of the results in the body of the paper in light of these modifications.

Recall the foundation for the attention function that we described in Section 2. The consumer reads a given piece of content if $\lambda - q > 0$ where λ is the belief that the content is good (the consumer values good content at one and bad content at zero) and q is the cost of reading. We had assumed that the cost q is distributed according to A with $A(0) = 0$ and $A(1) = 1$.

[26] When α grows large, the optimal mechanism converges to the single certificate optimum with perfect certification.

First, suppose reading bad content generates losses $b > 0$ for consumers rather than zero. In this case, a consumer reads a given piece of content when

$$\lambda - (1-\lambda)b - q > 0 \iff q < \lambda - (1-\lambda)b.$$

In effect, this transforms the attention function to become $A_b(\lambda) := A(\max\{\lambda(1+b) - b, 0\})$; where in this expression, the maximum simply ensures that the argument of the attention function lies in the domain $[0,1]$. Note that the function $A_b(\lambda)$ takes value zero for $\lambda \leq \frac{b}{1+b}$ but it is strictly for $\lambda > \frac{b}{1+b}$.

We additionally consider the second possibility that consumers are addicted and experience a loss $z \in (0, A^{-1}(\gamma))$ if they do not read a piece of content. Then, they read if

$$\lambda - q > -z \iff q < \lambda + z.$$

In this case, addiction is equivalent to transforming the attention function to $A_z(\lambda) := A(\min\{\lambda + z, 1\})$; the minimum here serves the same purpose as the maximum does for A_b—ensuring that the argument lies in $[0,1]$. $A_z(\lambda)$ is strictly increasing for $\lambda < 1-z$ and takes value one when $\lambda \geq 1-z$. We assume that $z < A^{-1}(\gamma)$, equivalently $A(z) < \gamma$, as otherwise the addition is high enough to allow the platform to make infinite profits since consumers will be willing to read content that they know is bad for sure.

Note that, even though the functions A_b and A_z are nondecreasing rather than strictly increasing (an assumption we made for the attention function A), the mechanism defined in Proposition 4 is optimal in both cases. We thus use similar notation $(V^*_{g,b}, \Lambda^*_b)$ and $(V^*_{g,z}, \Lambda^*_z)$ respectively to denote the optimal mechanisms in these cases.

The following result shows that platform content is of higher quality when costs b of bad content are higher or consumers have a lower level z of addiction. Note that the condition that $\frac{1}{A(\lambda)}\frac{\partial A(\lambda)}{\partial \lambda}$ is stictly decreasing is satisfied if $A(\lambda) = \lambda^\alpha$ for any $\alpha > 0$. [27]

Proposition 10. *Let $A_{\hat{b}}(\lambda) = A(\max\{\lambda(1+\hat{b}) - \hat{b}, 0\})$ with $\hat{b} > 0$ and $A_{\hat{z}}(\lambda) = A(\min\{\lambda + z, 1\})$ with $0 < \hat{z} < A^{-1}(\gamma)$ where $A(\lambda)$ is an attention function such that $\frac{1}{A(\lambda)}\frac{\partial A(\lambda)}{\partial \lambda}$ is strictly decreasing.*

Then, for every $\theta \in \Theta$ and every $0 < b' < b$ with $V^*_{g,b}(\theta) > 0$, we have $V^*_{g,b'}(\theta) > 0$ and $\Lambda^*_b(\theta) \geq \Lambda^*_{b'}(\theta)$. For every $\theta \in \Theta$ and every $0 < z < z' < A^{-1}(\gamma)$ with $V^*_{g,z}(\theta) > 0$, we have $V^*_{g,z'}(\theta) > 0$ and $\Lambda^*_z(\theta) \geq \Lambda^*_{z'}(\theta)$.

[27] As with Proposition 7, although this condition does not imply that there is a unique optimal mechanism, the condition can be weakened if a unique optimum exists.

Replacing the attention function A with either A_b or A_z does not change any results in the paper with the exception of Proposition 9.[28] In particular, Proposition 8 holds: even if $b > 0$, either enforced perfect certification does not matter, or it decreases content diversity. Greater content diversity remains beneficial for consumers since content is only read when $q < \lambda - (1-\lambda)b$ or equivalently, when the expected payoff is positive. Increasing b might make it more likely that perfect certification is optimal for the platform but it does not change the conclusion that enforced perfect certification is either irrelevant, or comes with the cost of lower content diversity.[29]

6. Conclusion

We develop a simple and tractable model (generalizing Mussa and Rosen (1978)) of steering and certification for sale that yields both positive and normative results.

First, consistent with behavior in practice, a platform can benefit from imperfect certification since doing so enables the platform to earn revenues from bad content. Of course, higher profits can only be obtained if the platform carefully manages the quality of content and thereby consumer attention.

Second, we demonstrate that optimal price discrimination involves offering a menu of distinct levels of certification with different quantities of views. The platform optimally offers lower-quality certificates and fewer views to good providers whose value, and therefore willingness to pay for engagement, is low. Conversely, the platform offers higher views and perfect certification to sufficiently high value good providers. The platform does not sacrifice rents from these high value providers by combining their content with that of bad providers.

Third, we characterize conditions under which consumers are better off facing an unconstrained profit maximizing platform relative to one that is forced to use perfect certification. This occurs through two channels. First, some good providers with low values who do not appear under perfect certification do, in fact, garner views in the optimal mechanism. In essence, the platform subsidizes their views in order to generate revenue from bad providers. Consumers benefit from this increased content diversity. Second, in the optimal mechanism, the quantity of good provider views is less sensitive to their values relative to enforced perfect certification. In this way, the optimal mechanism is closer to the engagement maximizing benchmark in which the values of good content providers are irrelevant. Thus, in contrast to the tenor of recent policy discussions, our results highlight the welfare costs of enforced perfect certification.

[28] This result relies on $b = 0$ allowing λ to optimally be small.

[29] Welfare interpretations here are particularly subtle since when consumers are addicted, revealed preference arguments are more tenuous.

We examine several comparative statics of the optimal mechanism. We show that a lower ad revenue (a natural interpretation of the opportunity cost of allocating viewership) leads to a decrease of certificate quality, or, equivalently, more views being assigned to bad providers. Improved targeting raises platform profits via a higher number of views that can be provided without affecting the quality of certificates. Increasing the convexity of consumers' attention leads to higher quality certificates. Convexity captures the extent to which consumer attention drops off with a higher quantity of bad content at different quality levels.

Our results speak to the ongoing discussion on content moderation in online platforms. Most obviously, our findings suggest that, in principle at least, consumers can benefit from allowing some bad content since it can be used to subsidize more good content and lead to more egalitarian content provision. It is also noteworthy, that the platform to some extent internalizes the harm associated with bad content since this content makes consumers pay less attention, and so from the platform's perspective limits its ability to raise revenue. Consequently, the platform optimally issues higher quality certificates if harms are higher. At an extreme, if the harm is sufficiently high then trivially no consumers will engage with content that may be bad and perfect certification maximizes platform profits with no need for regulatory intervention. There may be alternative reasons for regulatory intervention: most obviously, consumer protection for naïve consumers, and the possibility of externalities as discussed, for example in Bursztyn, Handel, Jimenez, and Roth (2023). Our framework provides a starting point to address such questions in future work.

Although we focus on the application to an online platform moderating content of good and bad type providers, there are other applications that fit our structure. One interpretation is that the bad content is any hidden advertisement the platform can introduce. One can imagine a search platform that can put hidden ads among the organic search results, and separate them from the explicit advertisements.[30] Our model constructs the optimal way to mix these hidden ads into content, and highlights the potential costs of prohibiting this practice.

A final interpretation is that the hidden advertisements are chosen by good content providers, but regulated by an outside force like the platform. An influencer can decide how much content to show that matches their own tastes, and therefore what their followers seek, and how much is not in their followers interest, but is subsidized by an outside advertiser. In that case, the certificate can stand in for a form of disclosure regulation: perfect disclosure regulation corresponds to announcing the type of content post by post, and may not be optimal when steering is for sale. Imperfect disclosure, as often seems to arise, can be better than perfectly enforced disclosure regulations.

[30] Although Google does not explicitly charge for organic traffic, some have argued that having ad business with Google influences organic placement. Similarly Amazon is alleged to favor suppliers that also purchase ancillary services.

Appendix A. Omitted proofs from the text

This section contains the proofs to all the results from the body of the paper. For ease of reference, we restate each result prior to presenting the proof.

Lemma 1. *For any incentive compatible and individually rational mechanism (V_g, V_b, M, P) there exists another incentive compatible and individually rational mechanism $(V_g, \tilde{V}_b, \tilde{M}, P)$ such that, for all $\theta \in \Theta$, $\tilde{M}(\theta) = \theta$ and, both the platform and good providers of every value receive the same payoff as from (V_g, V_b, M, P).*

Proof. Given a mechanism (V_g, V_b, M, P), we will construct another mechanism $(\tilde{V}_b, \tilde{V}_b, \tilde{M}, P)$ with the desired properties stated above.

First, we define $\tilde{M}(\theta) = \theta$. If $\Lambda(M(\theta)) = 0$, we define

$$\tilde{V}_b(\theta) = V_b(\theta)$$

Conversely, if $\Lambda(M(\theta)) > 0$, we define

$$\tilde{V}_b(\theta) = \frac{1 - \Lambda(M(\theta))}{\Lambda(M(\theta))} V_g(\theta) \tag{8}$$

Let $\tilde{\Lambda}$ be the quality associated with mechanism $(V_g, \tilde{V}_b, \tilde{M}, P)$. Observe that, by construction,

$$\tilde{\Lambda}(\tilde{M}(\theta)) = \Lambda(M(\theta)) \qquad \text{for all } \theta \in \Theta. \tag{9}$$

so

$$\theta A(\tilde{\Lambda}(\tilde{M}(\theta'))) V_g(\theta') - P(\theta') = \theta A(\Lambda(M(\theta'))) V_g(\theta') - P(\theta') \qquad \text{for all } \theta, \theta' \in \Theta.$$

In words, good providers of all values have the same payoff as the original mechanism (whether they report truthfully or misreport) and consequently $(V_g, \tilde{V}_b, \tilde{M}, P)$ is incentive compatible and individually rational because the original mechanism (V_g, V_b, M, P) is both.

Take an m in the image of M. If $\Lambda(m) = 0$, then

$$\mathbb{E}[V_b(\theta) | M(\theta) = m] = \mathbb{E}[\tilde{V}_b(\theta) | M(\theta) = m]$$

since V_b and $\tilde{V}_b$ are defined to be equal for such θ. When $\Lambda(m) > 0$, (8) and (9) together imply that

$$\begin{aligned}\mathbb{E}[V_b(\theta) | M(\theta) = m] = \frac{1 - \Lambda(m)}{\Lambda(m)} \mathbb{E}[V_g(\theta) | M(\theta) = m] &= \mathbb{E}\left[\frac{1 - \tilde{\Lambda}(\tilde{M}(\theta))}{\tilde{\Lambda}(\tilde{M}(\theta))} V_g(\theta) \middle| M(\theta) = m\right] \\ &= \mathbb{E}[\tilde{V}_b(\theta) | M(\theta) = m].\end{aligned}$$

Consequently,

$$\int_{\Theta}\left[A(\Lambda(M(\theta)))V_b(\theta)-\gamma V_b(\theta)\right]f(\theta)d\theta=\int_{\Theta}\left[A(\tilde{\Lambda}(\tilde{M}(\theta)))\tilde{V}_b(\theta)-\gamma\tilde{V}_b(\theta)\right]f(\theta)d\theta$$

and so the platform makes the identical profit from (V_g, V_b, M, P) and $(V_g, \tilde{V}_b, \tilde{M}, P)$ as required. □

Proposition 1. *Suppose $\overline{\theta}A'(1) < 1-\gamma$. Then, every optimal single certificate*

$$\lambda^{s^*}\in\underset{\lambda}{\operatorname{argmax}}\left\{\Pi^s(\lambda)\right\}$$

satisfies $\lambda^{s^*}<1$. That is, optimal single certification is imperfect.

Proof. Since $\phi(\overline{\theta})=\overline{\theta}>\gamma$, observe that the optimal number of views $V_g^{s,1}(\theta)=c'^{-1}\left(\max\{\phi(\theta)-\gamma,0\}\right)$ for $\lambda=1$ is such that

$$\int_0^{\overline{\theta}}\left[\left(\phi(\theta)-\gamma\right)V_g^{s,1}(\theta)-c(V_g^{s,1}(\theta))\right]f(\theta)d\theta>0.$$

In words, the assumption $\overline{\theta}>\gamma$ ensures that the platform can make positive profits with perfect certification.

The platform's profit when choosing a single certificate λ and assigning views $V_g^{s,1}(\theta)$ is given by

$$\int_0^{\overline{\theta}}\left[\left(\phi(\theta)+\frac{1-\lambda}{\lambda}\right)A(\lambda)V_g^{s,1}(\theta)-\gamma\frac{V_g^{s,1}(\theta)}{\lambda}-c\left(V_g^{s,1}(\theta)\right)\right]f(\theta)d\theta.$$

Differentiating this expression with respect to λ, we get

$$\int_0^{\overline{\theta}}\left[\left(\phi(\theta)+\frac{1-\lambda}{\lambda}\right)A'(\lambda)-\frac{A(\lambda)}{\lambda^2}+\frac{\gamma}{\lambda^2}\right]V_g^{s,1}(\theta)f(\theta)d\theta.$$

which evaluated at $\lambda=1$ is

$$\int_0^{\overline{\theta}}\left[\phi(\theta)A'(1)-1+\gamma\right]V_g(\theta)f(\theta)d\theta<0$$

since $A'(1)>0$, $\phi(\overline{\theta})A'(1)=\overline{\theta}A'(1)<1-\gamma$ and therefore $\phi(\theta)A'(1)<1-\gamma$ for all $\theta\in[0,\overline{\theta}]$ as $\phi(\theta)$ is strictly increasing. This implies that the platform can make strictly higher profits by picking a certificate $\lambda<1$ and assigning views $V_g^{s,1}(\theta)$ (a mechanism that is incentive compatible) which implies that choosing $\lambda=1$ cannot be optimal. □

Proposition 2. *For any $0<\lambda<1$, either $\Pi^s(\lambda)<\Pi^s(1)$, or content diversity is higher under λ then under perfect certification.*

Proof. Take a $\lambda \in (0,1)$ and let $\tilde{\phi}$ be the solution to

$$R(\tilde{\phi},\lambda) = R(\tilde{\phi},1) \iff \tilde{\phi} - \gamma = \left(\tilde{\phi} + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \gamma/\lambda.$$

Observe that $R(\hat{\phi},1) > R(\hat{\phi},\lambda)$ for $\hat{\phi} > \tilde{\phi}$ and $R(\hat{\phi},1) < R(\hat{\phi},\lambda)$ for $\hat{\phi} < \tilde{\phi}$ because $0 < A(\lambda) < 1$.

If $R(\tilde{\phi},\lambda) = R(\tilde{\phi},1) \leq 0$, then $R(\phi(\theta),1) > R(\phi(\theta),\lambda)$ for all $\theta \in \Theta$ such that $R(\phi(\theta),\lambda) > 0$ since this implies $\phi(\theta) > \tilde{\phi}$. Consequently,

$$\Pi^s(1) = \int_0^{\overline{\theta}} \left(R(\phi(\theta),1)V_g^{s,1}(\theta) - c(V_g^{s,1}(\theta))\right) f(\theta)d\theta > \int_0^{\overline{\theta}} \left(R(\phi(\theta),\lambda)V_g^{s,\lambda}(\theta) - c(V_g^{s,\lambda}(\theta))\right) f(\theta)d\theta = \Pi^s(\lambda)$$

since $V_g^{s,\hat{\lambda}}(\theta) = c'^{-1}\left(\max\{R(\phi(\theta),\hat{\lambda}),0\}\right)$ for all $\hat{\lambda} \in (0,1]$.

If $R(\tilde{\phi},\lambda) = R(\tilde{\phi},1) > 0$, then $R(\phi(\theta),1) \geq 0$ for $\theta \in \Theta$ implies that $R(\phi(\theta),\lambda) > 0$. Moreover, that there are $\theta \in \Theta$ such that $R(\phi(\theta),\lambda) > 0 > R(\phi(\theta),1)$. Consequently

$$\{\theta \in \Theta | V_g^{s,\lambda}(\theta) > 0\} \supset \{\theta \in \Theta | V_g^{s,1}(\theta) > 0\}$$

and so, in this case, content diversity is higher under λ. □

Proposition 3. *Suppose $\overline{\theta}A'(1) < 1 - \gamma$. Then, every two certificate optimum*

$$(\underline{\lambda}^*, \overline{\lambda}^*, \hat{\theta}^{bin^*}) \in \operatorname*{argmax}_{\underline{\lambda},\overline{\lambda},\hat{\theta}} \left\{\Pi^{bin}(\underline{\lambda},\overline{\lambda},\hat{\theta})\right\}$$

satisfies $\underline{\lambda}^* < \overline{\lambda}^*$.

Proof. Suppose, for contradiction, that the two certificate optimum has the property $\underline{\lambda}^* = \overline{\lambda}^*$. Then, by Proposition 1, it must be the case that the quality $\underline{\lambda}^* = \overline{\lambda}^* \in (0,1)$ is interior. This implies that the derivative of the profit

$$\int_0^{\overline{\theta}} \left[\left(\phi(\theta) + \frac{1-\underline{\lambda}^*}{\underline{\lambda}^*}\right) A(\lambda) V_g^{s,\underline{\lambda}^*}(\theta) - \gamma \frac{V_g^{s,\underline{\lambda}^*}(\theta)}{\underline{\lambda}^*} - c\left(V_g^{s,\underline{\lambda}^*}(\theta)\right)\right] f(\theta)d\theta$$

with respect to λ must satisfy

$$\int_0^{\overline{\theta}} \left[\left(\phi(\theta) + \frac{1-\underline{\lambda}^*}{\underline{\lambda}^*}\right) A'(\underline{\lambda}^*) - \frac{A(\underline{\lambda}^*)}{\underline{\lambda}^{*2}} + \frac{\gamma}{\underline{\lambda}^{*2}}\right] V_g^{s,\underline{\lambda}^*}(\theta) f(\theta)d\theta = 0.$$

Now note that $V_g^{s,\underline{\lambda}^*}(\theta)$ in nondecreasing in θ and that there is a $\theta' < \overline{\theta}$ such that $V_g^{s,\underline{\lambda}^*}(\theta) > 0$ for $\theta \in [\theta', \overline{\theta}]$. This combined with the observation that the term in the square brackets above is increasing in θ, implies that there exists a $\tilde{\theta} < \overline{\theta}$ such that

$$\int_{\tilde{\theta}}^{\overline{\theta}} \left[\left(\phi(\theta) + \frac{1-\underline{\lambda}^*}{\underline{\lambda}^*}\right) A'(\underline{\lambda}^*) - \frac{A(\underline{\lambda}^*)}{(\underline{\lambda}^*)^2} + \frac{\gamma}{(\underline{\lambda}^*)^2}\right] V_g^{s,\underline{\lambda}^*}(\theta) f(\theta)d\theta > 0.$$

This contradicts the optimality of $\underline{\lambda}^* = \overline{\lambda}^*$ since the platform can achieve strictly higher profits by keeping the number of views as $V_g^{s,\underline{\lambda}^*}(\theta)$, keeping the certificate below $\tilde{\theta}$ as $\underline{\lambda}^*$ but increasing the quality of the certificate above $\tilde{\theta}$ (a mechanism that is incentive compatible since this change will maintain the requisite monotonicity). This completes the proof of the result. □

Proposition 4. *There is an optimal mechanism* (V_g^*, Λ^*) *solving the platform's problem* (2) *where both* V_g^*, Λ^* *are nondecreasing and satisfy*

$$\Lambda^*(\theta) = \max\left\{ \tilde{\lambda} \;\middle|\; \tilde{\lambda} \in \operatorname*{argmax}_{\lambda\in[0,1]} \left\{ \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda} \right\} \right\}$$

$$V_g^*(\theta) = c'^{-1}\left(\max\left\{ \left[\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right] A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)}, 0 \right\} \right)$$

for all $\theta \in \Theta$.

Proof. We maximize the objective function pointwise and show that the mechanism we obtain satisfies the necessary monotonicity properties to satisfy the incentive compatibility constraints.

First, observe that, if

$$\left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda} \geq \left(\phi(\theta) + \frac{1-\lambda'}{\lambda'}\right) A(\lambda') - \frac{\gamma}{\lambda'} \tag{10}$$

then, for any $V_g(\theta) \in \mathbb{R}_+$, the value of the objective function satisfies

$$\left[\left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda}\right] V_g(\theta) - c(V_g(\theta)) \geq \left[\left(\phi(\theta) + \frac{1-\lambda'}{\lambda'}\right) A(\lambda') - \frac{\gamma}{\lambda'}\right] V_g(\theta) - c(V_g(\theta))$$

and vice versa when inequality (10) is reversed.

Consequently, there is a pointwise optimum that satisfies

$$\Lambda^*(\theta) \in \operatorname*{argmax}_{\lambda\in[0,1]} \left\{ \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda} \right\}.$$

We pick $\Lambda^*(\theta)$ to be the largest above maximizer. Note that

$$\left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda} = (\phi(\theta) - 1)A(\lambda) + \frac{A(\lambda) - \gamma}{\lambda} \quad \longrightarrow \quad -\infty$$

as $\lambda \to 0$ (because $A(\lambda) \to 0$) and therefore $\Lambda^*(\theta) > 0$ for all $\theta \in \Theta$.

We now argue that Λ^* is nondecreasing. By definition,

$$\left(\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right) A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)} \geq \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda}$$

for all $0 < \lambda < \Lambda^*(\theta)$. Then for all $\theta' > \theta$, it must be the case that

$$\left(\phi(\theta') + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right) A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)} \geq \left(\phi(\theta') + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda}$$

for all $0 < \lambda < \Lambda^*(\theta)$ since $A(\Lambda^*(\theta)) > A(\lambda)$ and ϕ is nondecreasing. Consequently, we must have $\Lambda^*(\theta') \geq \Lambda^*(\theta)$.

The function V_g^* as defined in the statement of the proposition is the solution to

$$V_g^*(\theta) = \underset{v_g \in \mathbb{R}_+}{\operatorname{argmax}} \left\{ \left[\left(\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right) A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)} \right] v_g - c(v_g) \right\} \tag{11}$$

and the exact expression is obtained from the first-order condition.

Now observe that, because ϕ is nondecreasing, the function

$$\left(\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right) A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)} = \max_{\lambda \in [0,1]} \left\{ \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda} \right\}$$

is nondecreasing in θ because it is the maximum of nondecreasing functions. This, in turn, implies from (11), that V^* is nondecreasing. Therefore $A(\Lambda^*(\cdot))V_g^*(\cdot)$ is nondecreasing and consequently the pointwise solution is incentive compatible as required. This completes the proof. □

Proposition 5. *The quality $\Lambda^*(\theta)$ in the optimal mechanism (V_g^*, Λ^*) is nondecreasing in γ for all $\theta \in \Theta$.*

Proof. For clarity, in this proof, we use the notation $\Lambda_\gamma^*(\theta)$ to make the dependence on γ explicit.

Recall that $\Lambda_\gamma^*(\theta) > 0$ for all $\theta \in \Theta$. By the definition of Λ_γ^* from Proposition 4, we have

$$\left(\phi(\theta) + \frac{1-\Lambda_\gamma^*(\theta)}{\Lambda_\gamma^*(\theta)}\right) A(\Lambda_\gamma^*(\theta)) - \frac{\gamma}{\Lambda_\gamma^*(\theta)} \geq \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma}{\lambda}$$

$$\iff \gamma \left(\frac{1}{\lambda} - \frac{1}{\Lambda_\gamma^*(\theta)}\right) \geq \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \left(\phi(\theta) + \frac{1-\Lambda_\gamma^*(\theta)}{\Lambda_\gamma^*(\theta)}\right) A(\Lambda_\gamma^*(\theta))$$

for all $\lambda \in (0, \Lambda_\gamma^*(\theta))$.

Therefore, for any $\gamma' > \gamma$, we must have

$$\gamma' \left(\frac{1}{\lambda} - \frac{1}{\Lambda_\gamma^*(\theta)}\right) > \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \left(\phi(\theta) + \frac{1-\Lambda_\gamma^*(\theta)}{\Lambda_\gamma^*(\theta)}\right) A(\Lambda_\gamma^*(\theta))$$

$$\iff \left(\phi(\theta) + \frac{1-\Lambda_\gamma^*(\theta)}{\Lambda_\gamma^*(\theta)}\right) A(\Lambda_\gamma^*(\theta)) - \frac{\gamma'}{\Lambda_\gamma^*(\theta)} > \left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A(\lambda) - \frac{\gamma'}{\lambda}$$

for all $\lambda \in (0, \Lambda_\gamma^*(\theta))$. Therefore, we must have $\Lambda_{\gamma'}^*(\theta) \geq \Lambda_\gamma^*(\theta)$ as required. □

Proposition 6. *Let the cost $\kappa c(v_g)$ of interested views v_g be parametrized by $\kappa > 0$. Then, the optimal mechanism (V_g^*, Λ^*) satisfies the following properties.*

(i) *For every $\theta \in \Theta$, the quality $\Lambda^*(\theta)$ is the same for all $\kappa > 0$.*

(ii) *For every $\theta \in \Theta$, the quantity of views $V_g^*(\theta)$ is nonincreasing in κ.*

(iii) *The set of good provider values $\{\theta \in \Theta \mid V_g^*(\theta) > 0\}$ that are served is the same for all $\kappa > 0$.*

Proof. The first statement follows immediately from Proposition 4 since the cost $\kappa c(\cdot)$ does not enter the expression for Λ^*.

From Proposition 4, $V_g^*(\theta)$ solves

$$V_g^*(\theta) = c^{-1}\left(\frac{1}{\kappa}\max\left\{\left[\phi(\theta) + \frac{1-\Lambda^*(\theta)}{\Lambda^*(\theta)}\right] A(\Lambda^*(\theta)) - \frac{\gamma}{\Lambda^*(\theta)}, 0\right\}\right).$$

The second and third statements follow immediately from this equation and properties of c. Note that $V_g^*(\theta) > 0$ whenever the right hand side of the above equation is greater than 0 and the sign of right hand side does not depend on κ. Clearly the term in the round brackets is nonincreasing in κ so $V_g^*(\theta)$ must be nonincreasing in κ. □

The remaining comparative statics use the following result.

Proposition A.1. *Consider a family of attention functions $A(\hat{\lambda};\hat{\alpha})$ parametrized by some $\hat{\alpha} \in (\underline{\alpha}, \overline{\alpha}) \subset \mathbb{R}_+$ and denote the optimal mechanism for a given $\hat{\alpha}$ as $(V_{g,\hat{\alpha}}^*, \Lambda_{\hat{\alpha}}^*)$. Assume that $A(\hat{\lambda};\hat{\alpha})$ is (i) twice continuously differentiable, (ii) strictly decreasing in $\hat{\alpha}$ and that (iii) $\frac{1}{A(\hat{\lambda};\hat{\alpha})}\frac{\partial A(\hat{\lambda};\hat{\alpha})}{\partial \hat{\alpha}}$ is strictly increasing in $\hat{\lambda}$ whenever $\hat{\lambda} < 1$.*

Then for every $\theta \in \Theta$ and every $\alpha, \alpha' \in (\underline{\alpha}, \overline{\alpha})$ with $\alpha' < \alpha$ and $V_{g,\alpha}^*(\theta) > 0$, we have $V_{g,\alpha'}^*(\theta) > 0$ and $\Lambda_\alpha^*(\theta) \geq \Lambda_{\alpha'}^*(\theta)$.

Proof. Recall that $\Lambda_\alpha^*(\theta)$ is defined to be the greatest λ that maximizes of $R_\alpha(\theta, \lambda)$ (like the mechanism, this notation makes the dependence on α explicit). Since $R_\alpha(\theta, \lambda)$ is continuous in λ and in α (by assumption), the set of λs that maximize $R_\alpha(\theta, \lambda)$ is an upper semi-continuous correspondence in α for each $\theta \in \Theta$. Thus, $\Lambda_\alpha^*(\theta)$ is an upper semi-continuous function in α because it is the maximum of an upper semi-continuous correspondence.

In the proof, we employ the shorthand notation

$$g(\hat{\alpha}, \hat{\lambda}) := \left(\phi(\theta) + \frac{1-\hat{\lambda}}{\hat{\lambda}}\right) A(\hat{\lambda};\hat{\alpha}). \tag{12}$$

Observe that

$$\frac{\partial g(\hat{\alpha}, \hat{\lambda})}{\partial \hat{\alpha}} = \frac{g(\hat{\alpha}, \hat{\lambda})}{A(\hat{\lambda};\hat{\alpha})}\frac{\partial A(\hat{\lambda};\hat{\alpha})}{\partial \hat{\alpha}}.$$

Consider a θ and an α such that $V^*_{g,\alpha}(\theta) > 0$. For all $\alpha' < \alpha$, we have

$$0 < g(\alpha, \Lambda^*_\alpha(\theta)) - \frac{\gamma}{\Lambda^*_\alpha(\theta)} \leq g(\alpha', \Lambda^*_\alpha(\theta)) - \frac{\gamma}{\Lambda^*_\alpha(\theta)} \leq \max_{\lambda \in [0,1]} \left\{ g(\alpha', \lambda) - \frac{\gamma}{\lambda} \right\} = g(\alpha', \Lambda^*_{\alpha'}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)}$$

where the first inequality follows from the fact that $V^*_{g,\alpha}(\theta) > 0$ and the second inequality from the fact that $g(\hat{\alpha}, \hat{\lambda})$ is decreasing in $\hat{\alpha}$. Therefore, $V^*_{g,\alpha'}(\theta) > 0$.

Once again, consider a θ and an α such that $V^*_{g,\alpha}(\theta) > 0$. We will argue that, there exists an $\varepsilon > 0$ such that $\Lambda^*_\alpha(\theta) \geq \Lambda^*_{\alpha'}(\theta)$ for $\alpha' \in (\alpha - \varepsilon, \alpha)$ and $\Lambda^*_\alpha(\theta) \leq \Lambda^*_{\alpha'}(\theta)$ for $\alpha' \in (\alpha, \alpha + \varepsilon)$. This suffices to establish that $\Lambda^*_\alpha(\theta) \geq \Lambda^*_{\alpha'}(\theta)$ for all $\alpha' < \alpha$ (because $V^*_{g,\hat{\alpha}}(\theta) > 0$ for all $\hat{\alpha} \in [\alpha', \alpha]$) and therefore the result.

The proof proceeds in two steps each of which has two sub-cases.

Step 1: We first argue that there exists an $\varepsilon > 0$ such that $\Lambda^*_\alpha(\theta) \geq \Lambda^*_{\alpha'}(\theta)$ for $\alpha' \in (\alpha - \varepsilon, \alpha)$. Obviously, this is true if $\Lambda^*_\alpha(\theta) = 1$ so we only need to consider the case $\Lambda^*_\alpha(\theta) < 1$.

We define the sequence

$$\lambda_n = \sup \left\{ \Lambda^*_{\alpha'}(\theta) \,\middle|\, \alpha \left(1 - \frac{1}{n}\right) \leq \alpha' \leq \alpha \right\}$$

for $n = \underline{n}, \underline{n} + 1, \ldots$ where $\underline{n} > \frac{\alpha}{\alpha - \underline{\alpha}}$ is chosen to be sufficiently large so that $\alpha \left(1 - \frac{1}{n}\right) > \underline{\alpha}$. Observe that this sequence is nonincreasing and bounded so it has a limit. Observe also that, by definition, $\lambda_n \geq \Lambda^*_\alpha(\theta)$ for all n since the supremum is taken to include α.

We split the argument into the two possible cases.

Case 1.1: $\lim_{n\to\infty} \lambda_n > \Lambda^*_\alpha(\theta)$.

Take a quality value $\tilde{\lambda}$ satisfying $\Lambda^*_\alpha(\theta) < \tilde{\lambda} < \lim_{n\to\infty} \lambda_n$. For every $\varepsilon \in (0, 1)$, there exists an $n > \frac{\alpha}{\varepsilon}$ such that $\lambda_n > \hat{\lambda}$. From the definition of λ_n, this implies that there exists an $\alpha' \in \left[\alpha \left(1 - \frac{1}{n}\right), \alpha\right] \subseteq (\alpha - \varepsilon, \alpha]$ such that $\Lambda_{\alpha'} > \tilde{\lambda}$. This violates the upper semi-continuity of the optimal quality $\Lambda^*_\alpha(\theta)$ at α. Thus, this case cannot arise.

Case 1.2: $\lim_{n\to\infty} \lambda_n = \Lambda^*_\alpha(\theta)$.

If there exists an n such that $\lambda_n = \Lambda^*_\alpha(\theta)$, then, for any $0 < \varepsilon < \frac{\alpha}{n}$ we have $\Lambda^*_\alpha(\theta) \geq \Lambda^*_{\alpha'}(\theta)$ for $\alpha' \in (\alpha - \varepsilon, \alpha)$ as required. Thus, for the remainder of the argument of this case, we assume that $\lambda_n > \Lambda^*_\alpha$ for all n.

Take an $n \geq \underline{n}$ for which $\lambda_n < 1$ and note that such an n exists because $\lim_{n\to\infty} \lambda_n = \Lambda^*_\alpha(\theta) < 1$. The cross partial at $(\alpha, \Lambda^*_\alpha(\theta))$ satisfies

$$\frac{\partial^2 g(\alpha, \Lambda^*_\alpha(\theta))}{\partial \hat{\lambda} \partial \hat{\alpha}} = g(\alpha, \Lambda^*_\alpha(\theta)) \frac{\partial}{\partial \hat{\lambda}} \left(\frac{1}{A(\Lambda^*_\alpha(\theta); \alpha)} \frac{\partial A(\Lambda^*_\alpha(\theta); \alpha)}{\partial \hat{\alpha}} \right)$$

$$+\frac{\partial g(\alpha,\Lambda_\alpha^*(\theta))}{\partial\hat{\lambda}}\left(\frac{1}{A(\Lambda_\alpha^*(\theta);\alpha)}\frac{\partial A(\Lambda_\alpha^*(\theta);\alpha)}{\partial\hat{\alpha}}\right)>0.$$

The inequality follows from the facts that (i) $\frac{1}{A(\hat{\lambda};\hat{\alpha})}\frac{\partial A(\hat{\lambda};\hat{\alpha})}{\partial\hat{\alpha}}$ is strictly negative and increasing in $\hat{\lambda}$ (by assumption), (ii) $g(\alpha,\Lambda_\alpha^*(\theta))>0$ because $V_{g,\alpha}^*(\theta)>0$ and (iii) $\frac{\partial g(\alpha,\Lambda_\alpha^*(\theta))}{\partial\hat{\lambda}}<0$ because $\Lambda_\alpha^*(\theta)<1$ (as $\lambda_n<1$) and therefore satisfies the first-order condition

$$\frac{\partial g(\alpha,\Lambda_\alpha^*(\theta))}{\partial\hat{\lambda}}+\frac{\gamma}{\Lambda_\alpha^*(\theta)^2}=0.$$

Consequently, because the cross partial is continuous (by assumption), there exists an $\varepsilon>0$ such that for all (α'',λ'') in an ε-ball $||(\alpha'',\lambda'')-(\alpha,\Lambda_\alpha^*(\theta))||<\varepsilon$ around $(\alpha,\Lambda_\alpha^*(\theta))$, the cross-partial also satisfies $\frac{\partial^2 g(\alpha'',\lambda'')}{\partial\hat{\lambda}\partial\hat{\alpha}}>0$.

Now take an $\alpha'<\alpha$ such that $||(\alpha',\Lambda_{\alpha'}^*(\theta))-(\alpha,\Lambda_\alpha^*(\theta))||<\varepsilon$ and $\Lambda_{\alpha'}^*(\theta)>\Lambda_\alpha^*(\theta)$. Such an α' exists because $\lim_{n\to\infty}\lambda_n=\Lambda_\alpha^*$ and $\lambda_n>\Lambda_\alpha^*$ for all n. For this α', the optimality of $\Lambda_{\alpha'}^*(\theta)$ implies that

$$\begin{aligned}
&g(\alpha',\Lambda_{\alpha'}^*(\theta))-\frac{\gamma}{\Lambda_{\alpha'}^*(\theta)}\geq g(\alpha',\Lambda_\alpha^*(\theta))-\frac{\gamma}{\Lambda_\alpha^*(\theta)}\\
\implies &g(\alpha',\Lambda_{\alpha'}^*(\theta))+\int_{\alpha'}^{\alpha}\frac{\partial g(\hat{\alpha},\Lambda_{\alpha'}^*(\theta))}{\partial\hat{\alpha}}d\hat{\alpha}-\frac{\gamma}{\Lambda_{\alpha'}^*(\theta)}\geq g(\alpha',\Lambda_\alpha^*(\theta))+\int_{\alpha'}^{\alpha}\frac{\partial g(\hat{\alpha},\Lambda_{\alpha'}^*(\theta))}{\partial\hat{\alpha}}d\hat{\alpha}-\frac{\gamma}{\Lambda_\alpha^*(\theta)}\\
\implies &g(\alpha',\Lambda_{\alpha'}^*(\theta))+\int_{\alpha'}^{\alpha}\frac{\partial g(\hat{\alpha},\Lambda_{\alpha'}^*(\theta))}{\partial\hat{\alpha}}d\hat{\alpha}-\frac{\gamma}{\Lambda_{\alpha'}^*(\theta)}> g(\alpha',\Lambda_\alpha^*(\theta))+\int_{\alpha'}^{\alpha}\frac{\partial g(\hat{\alpha},\Lambda_\alpha^*(\theta))}{\partial\hat{\alpha}}d\hat{\alpha}-\frac{\gamma}{\Lambda_\alpha^*(\theta)}\\
\implies &g(\alpha,\Lambda_{\alpha'}^*(\theta))-\frac{\gamma}{\Lambda_{\alpha'}^*(\theta)}>g(\alpha,\Lambda_\alpha^*(\theta))-\frac{\gamma}{\Lambda_\alpha^*(\theta)}.
\end{aligned}$$

The third inequality is a consequence of the fact that $\frac{\partial g(\alpha'',\Lambda_{\alpha'}^*(\theta))}{\partial\hat{\alpha}}>\frac{\partial g(\alpha'',\Lambda_\alpha^*(\theta))}{\partial\hat{\alpha}}$ for all $\alpha''\in[\alpha',\alpha]$ because $\frac{\partial^2 g(\alpha'',\lambda'')}{\partial\hat{\lambda}\partial\hat{\alpha}}>0$ for all $\alpha''\in[\alpha',\alpha]$ and $\lambda''\in[\Lambda_\alpha^*(\theta),\Lambda_{\alpha'}^*(\theta)]$ since these values of (α'',λ'') lie in the ε-ball around $(\alpha,\Lambda_\alpha^*(\theta))$.

This contradicts $\Lambda_\alpha^*(\theta)$ being optimal at α and therefore it is not possible to have $\lambda_n>\Lambda_\alpha^*(\theta)$ for all n. Taken together, we have thus shown that there is an $\varepsilon>0$ such that $\Lambda_\alpha^*(\theta)\geq\Lambda_{\alpha'}^*(\theta)$ for $\alpha'\in(\alpha-\varepsilon,\alpha)$.

Step 2: We now argue that, there exists an $\varepsilon>0$ such that $\Lambda_\alpha^*(\theta)\leq\Lambda_{\alpha'}^*(\theta)$ for $\alpha'\in(\alpha,\alpha+\varepsilon)$. To do so, we employ a similar argument to the first step for the case where the optimum varies continuously, but the case where the function jumps at α requires an additional argument.

First suppose $\Lambda_\alpha^*(\theta)=1$. By the optimality of $\Lambda_\alpha^*(\theta)=1$ for α, we have

$$g(\alpha,\Lambda_\alpha^*(\theta))-\frac{\gamma}{\Lambda_\alpha^*(\theta)}=\phi(\theta)-\gamma\geq g(\alpha,\Lambda_{\alpha'}^*(\theta))-\frac{\gamma}{\Lambda_{\alpha'}^*(\theta)}$$

for all $\alpha'>\alpha$. Suppose, for contradiction, that $\Lambda_{\alpha'}^*(\theta)<1$ for some $\alpha'>\alpha$.

Now $V^*_{g,\alpha}(\theta) > 0$ implies that $\phi(\theta) - \gamma > 0$ which, in turn, implies that $\phi(\theta) > 0$. By assumption, $A(\Lambda^*_{\alpha'}(\theta); \alpha) > A(\Lambda^*_{\alpha'}(\theta); \alpha')$ and this combined with the fact that $\phi(\theta) > 0$ implies $g(\alpha, \Lambda^*_{\alpha'}(\theta)) > g(\alpha', \Lambda^*_{\alpha'}(\theta))$. Consequently,

$$\phi(\theta) - \gamma \geq g(\alpha, \Lambda^*_{\alpha'}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)} \implies \phi(\theta) - \gamma > g(\alpha', \Lambda^*_{\alpha'}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)}$$

violating the optimality of $\Lambda^*_{\alpha'}(\theta)$. To summarize, if $\Lambda^*_{\alpha}(\theta) = 1$, then we must have $\Lambda^*_{\alpha'}(\theta) = 1$ for all $\alpha' > \alpha$ as required.

In what follows, we consider the case of $\Lambda^*_{\alpha}(\theta) < 1$. We define the sequence

$$\lambda_n = \inf\left\{ \Lambda^*_{\alpha'}(\theta) \;\middle|\; \alpha \leq \alpha' \leq \alpha\left(1 + \frac{1}{n}\right) \right\}$$

for $n = \underline{n}, \underline{n}+1, \ldots$ where $\underline{n} > \frac{\alpha}{\overline{\alpha}-\alpha}$ is chosen to be sufficiently large so that $\alpha\left(1 + \frac{1}{n}\right) < \overline{\alpha}$. Observe that this sequence is nondecreasing and bounded so it has a limit. Observe also that, by definition, $\lambda_n \leq \Lambda^*_{\alpha}(\theta)$ for all n.

As with the previous step, we break the argument into the two possible cases.

Case 2.1: $\lim_{n\to\infty} \lambda_n = \Lambda^*_{\alpha}(\theta)$.

If there exists an n such that $\lambda_n = \Lambda^*_{\alpha}(\theta)$, then, for any $0 < \varepsilon < \frac{\alpha}{n}$ we have $\Lambda^*_{\alpha'}(\theta) \geq \Lambda^*_{\alpha}(\theta)$ for $\alpha' \in (\alpha, \alpha + \varepsilon)$ as required. Thus, for the remainder of the argument of this case, we assume that $\lambda_n < \Lambda^*_{\alpha}(\theta)$ for all n.

As we argued in Case 1.2, when $\Lambda^*_{\alpha}(\theta) < 1$, the cross partial at $(\alpha, \Lambda^*_{\alpha}(\theta))$ satisfies $\frac{\partial^2 g(\alpha, \Lambda^*_{\alpha}(\theta))}{\partial\hat{\lambda}\partial\hat{\alpha}} > 0$. Consequently, because the cross partial is continuous (by assumption), there exists an $\varepsilon > 0$ such that for all (α', λ') in an ε-ball $||(\alpha', \lambda') - (\alpha, \Lambda^*_{\alpha}(\theta))|| < \varepsilon$ around $(\alpha, \Lambda^*_{\alpha}(\theta))$, the cross-partial also satisfies $\frac{\partial^2 g(\alpha', \lambda')}{\partial\hat{\lambda}\partial\hat{\alpha}} > 0$.

Now take an $\alpha' > \alpha$ such that $||(\alpha', \Lambda^*_{\alpha'}(\theta)) - (\alpha, \Lambda^*_{\alpha}(\theta))|| < \varepsilon$ and $\Lambda^*_{\alpha'}(\theta) < \Lambda^*_{\alpha}(\theta)$. Such an α' exists because $\lim_{n\to\infty} \lambda_n = \Lambda^*_{\alpha}(\theta)$ and $\lambda_n > \Lambda^*_{\alpha}(\theta)$ for all n. For this α', the optimality of $\Lambda^*_{\alpha'}(\theta)$ implies that

$$\begin{aligned}
&g(\alpha', \Lambda^*_{\alpha'}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)} \geq g(\alpha', \Lambda^*_{\alpha}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha}(\theta)} \\
\implies &g(\alpha', \Lambda^*_{\alpha'}(\theta)) - \int_{\alpha}^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_{\alpha'}(\theta))}{\partial\hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)} \geq g(\alpha', \Lambda^*_{\alpha}(\theta)) - \int_{\alpha}^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_{\alpha'}(\theta))}{\partial\hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_{\alpha}(\theta)} \\
\implies &g(\alpha', \Lambda^*_{\alpha'}(\theta)) - \int_{\alpha}^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_{\alpha'}(\theta))}{\partial\hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)} > g(\alpha', \Lambda^*_{\alpha}(\theta)) - \int_{\alpha}^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_{\alpha}(\theta))}{\partial\hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_{\alpha}(\theta)} \\
\implies &g(\alpha, \Lambda^*_{\alpha'}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)} > g(\alpha, \Lambda^*_{\alpha}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha}(\theta)}.
\end{aligned}$$

The third inequality is a consequence of the fact that $\frac{\partial g(\alpha'',\Lambda^*_{\alpha'}(\theta))}{\partial\hat{\alpha}} < \frac{\partial g(\alpha'',\Lambda^*_{\alpha}(\theta))}{\partial\hat{\alpha}}$ for all $\alpha'' \in [\alpha', \alpha]$ because $\frac{\partial^2 g(\alpha'',\lambda'')}{\partial\hat{\lambda}\partial\hat{\alpha}} > 0$ for all $\alpha'' \in [\alpha, \alpha']$ and $\lambda'' \in [\Lambda^*_\alpha(\theta), \Lambda^*_{\alpha'}(\theta)]$ since these values of (α'', λ'') lie in the ε-ball around $(\alpha, \Lambda^*_\alpha(\theta))$.

This contradicts $\Lambda^*_\alpha(\theta)$ being optimal at α and thus it is not possible to have $\lim_{n\to\infty} \lambda_n = \Lambda^*_\alpha(\theta)$ and $\lambda_n < \Lambda^*_\alpha(\theta)$ for all n. This completes the argument for this case.

Case 2.2: $\lim_{n\to\infty} \lambda_n < \Lambda^*_\alpha(\theta)$.

Let $\tilde{\lambda} = \lim_{n\to\infty} \lambda_n < \Lambda^*_\alpha(\theta)$. First observe that this implies

$$g(\alpha, \Lambda^*_\alpha(\theta)) - \frac{\gamma}{\Lambda^*_\alpha(\theta)} = g(\alpha, \tilde{\lambda}) - \frac{\gamma}{\tilde{\lambda}} \implies g(\alpha, \Lambda^*_\alpha(\theta)) < g(\alpha, \tilde{\lambda}).$$

The inequality follows from the fact that $\frac{\gamma}{\tilde{\lambda}} > \frac{\gamma}{\Lambda^*_\alpha(\theta)}$. The equality follows from the following argument. Suppose $g(\alpha, \Lambda^*_\alpha(\theta)) - \frac{\gamma}{\Lambda^*_\alpha(\theta)} > g(\alpha, \tilde{\lambda}) - \frac{\gamma}{\tilde{\lambda}}$ then, by the continuity of $g(\cdot, \cdot)$, there exists an $\varepsilon > 0$ such that $g(\alpha', \Lambda^*_\alpha(\theta)) - \frac{\gamma}{\Lambda^*_\alpha(\theta)} > g(\alpha', \lambda') - \frac{\gamma}{\lambda'}$ for all (α', λ') in $||(\alpha', \lambda') - (\alpha, \tilde{\lambda})|| < \varepsilon$. But this contradicts $\tilde{\lambda} = \lim_{n\to\infty} \lambda_n$.

Moreover, $V^*_{g,\alpha}(\theta) > 0$ implies that $g(\alpha, \Lambda^*_\alpha(\theta)) > 0$. Taken together, we get

$$\frac{\partial g(\alpha, \Lambda^*_\alpha(\theta))}{\partial\hat{\alpha}} = \frac{1}{A(\Lambda^*_\alpha(\theta); \alpha)} \frac{\partial A(\Lambda^*_\alpha(\theta); \alpha)}{\partial\hat{\alpha}} g(\alpha, \Lambda^*_\alpha(\theta)) > \frac{1}{A(\tilde{\lambda}; \alpha)} \frac{\partial A(\tilde{\lambda}; \alpha)}{\partial\hat{\alpha}} g(\alpha, \tilde{\lambda}) = \frac{\partial g(\alpha, \tilde{\lambda})}{\partial\hat{\alpha}}$$

because $0 > \frac{1}{A(\Lambda^*_\alpha(\theta);\alpha)} \frac{\partial A(\Lambda^*_\alpha(\theta);\alpha)}{\partial\hat{\alpha}} > \frac{1}{A(\tilde{\lambda};\alpha)} \frac{\partial A(\tilde{\lambda};\alpha)}{\partial\hat{\alpha}}$ by assumption.

Now, since $\frac{\partial g(\cdot,\cdot)}{\partial\hat{\alpha}}$ is continuous (by assumption), there is an $\varepsilon > 0$ such that

$$\frac{\partial g(\alpha'', \lambda'')}{\partial\hat{\alpha}} > \frac{\partial g(\alpha', \lambda')}{\partial\hat{\alpha}}$$

for all $||(\alpha'', \lambda'') - (\alpha, \Lambda^*_\alpha(\theta))|| < \varepsilon$ and all $||(\alpha', \lambda') - (\alpha, \tilde{\lambda})|| < \varepsilon$.

So take an $\alpha' > \alpha$ such that $||(\alpha', \Lambda^*_{\alpha'}(\theta))) - (\alpha, \tilde{\lambda})|| < \varepsilon$. Such an α' must exist since $\tilde{\lambda} = \lim_{n\to\infty} \lambda_n$. Note that, for this α', we also have $||(\alpha', \Lambda^*_\alpha(\theta)) - (\alpha, \Lambda^*_\alpha(\theta))|| < \varepsilon$. Therefore,

$$\frac{\partial g(\alpha', \Lambda^*_\alpha(\theta))}{\partial\hat{\alpha}} > \frac{\partial g(\alpha', \Lambda^*_{\alpha'}(\theta))}{\partial\hat{\alpha}}$$

and consequently for all $\alpha'' \in (\alpha, \alpha')$, we also have

$$\frac{\partial g(\alpha'', \Lambda^*_\alpha(\theta))}{\partial\hat{\alpha}} > \frac{\partial g(\alpha'', \Lambda^*_{\alpha'}(\theta))}{\partial\hat{\alpha}}.$$

This inequality combined with the optimality of $\Lambda^*_\alpha(\theta)$ for α implies that

$$\begin{aligned} & g(\alpha, \Lambda^*_\alpha(\theta)) - \frac{\gamma}{\Lambda^*_\alpha(\theta)} \geq g(\alpha, \Lambda^*_{\alpha'}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)} \\ \implies & g(\alpha, \Lambda^*_\alpha(\theta)) + \int_\alpha^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_\alpha(\theta))}{\partial\hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_\alpha(\theta)} \geq g(\alpha, \Lambda^*_{\alpha'}(\theta)) + \int_\alpha^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_\alpha(\theta))}{\partial\hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)} \end{aligned}$$

$$\implies g(\alpha, \Lambda^*_\alpha(\theta)) + \int_\alpha^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_\alpha(\theta))}{\partial \hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_\alpha(\theta)} > g(\alpha, \Lambda^*_{\alpha'}(\theta)) + \int_\alpha^{\alpha'} \frac{\partial g(\hat{\alpha}, \Lambda^*_{\alpha'}(\theta))}{\partial \hat{\alpha}} d\hat{\alpha} - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)}$$

$$\implies g(\alpha', \Lambda^*_\alpha(\theta)) - \frac{\gamma}{\Lambda^*_\alpha(\theta)} > g(\alpha', \Lambda^*_{\alpha'}(\theta)) - \frac{\gamma}{\Lambda^*_{\alpha'}(\theta)}.$$

This violates the optimality of $\Lambda^*_{\alpha'}(\theta)$ for α' and therefore $\lim_{n\to\infty} \lambda_n < \Lambda^*_\alpha(\theta)$ cannot arise. This completes the proof of the result. □

Proposition 7. *Suppose the attention function $A(\hat{\lambda}; \hat{\alpha}) = A(\hat{\lambda})^{\hat{\alpha}}$ is parametrized by a constant $\hat{\alpha} > 0$. Denote the optimal mechanism for a given $\hat{\alpha}$ as $(V^*_{g,\hat{\alpha}}, \Lambda^*_{\hat{\alpha}})$. Then for every $\theta \in \Theta$ and every $0 < \alpha' < \alpha$ with $V^*_{g,\alpha}(\theta) > 0$, we have $V^*_{g,\alpha'}(\theta) > 0$ and $\Lambda^*_\alpha(\theta) \geq \Lambda^*_{\alpha'}(\theta)$.*

Proof. Observe that

$$A(\hat{\lambda}; \hat{\alpha}) = A(\hat{\lambda})^{\hat{\alpha}} \implies \frac{\partial A(\hat{\lambda}; \hat{\alpha})}{\partial \hat{\alpha}} = \ln(A(\hat{\lambda}))A(\hat{\lambda})^{\hat{\alpha}} \implies \frac{1}{A(\hat{\lambda}; \hat{\alpha})} \frac{\partial A(\hat{\lambda}; \hat{\alpha})}{\partial \hat{\alpha}} = \ln(A(\hat{\lambda})).$$

This implies that the result is an immediate consequence of Proposition A.1 because $A(\hat{\lambda}; \hat{\alpha})$ is strictly decreasing in $\hat{\alpha}$ and $\frac{1}{A(\hat{\lambda};\hat{\alpha})} \frac{\partial A(\hat{\lambda};\hat{\alpha})}{\partial \hat{\alpha}}$ is increasing in $\hat{\lambda}$. □

Proposition 8. *The optimal mechanism (V^*_g, Λ^*) must satisfy one of the following two properties.*

1. *Perfect certification is optimal: $\Lambda^*(\theta) = 1$ for all $\theta \in \Theta$ and consequently, $V^*_g(\theta) > 0$ if, and only if, $\phi(\theta) > \gamma$.*
2. *Enforced perfect certification reduces content diversity: $\Lambda^*(\theta) < 1$ for some $\theta \in \Theta$ with $\phi(\theta) > \gamma$ and $V^*_g(\theta) > 0$ for some $\theta \in \Theta$ with $\phi(\theta) < \gamma$.*

Proof. An exhaustive list of possibilities is Cases 1 and 2 in the statement of the Proposition plus the following two cases.

Case 3: $\Lambda^*(\theta) = 1$ for all θ such that $\phi(\theta) \geq \gamma$ and $V^*_g(\theta) > 0$ for some θ such that $\phi(\theta) \leq \gamma$.

Case 4: $\Lambda^*(\theta) < 1$ for some θ such that $\phi(\theta) > \gamma$ and $V^*_g(\theta) = 0$ for all θ such that $\phi(\theta) < \gamma$.

We show that Cases 3 and 4 cannot occur. Denote the contribution to profits by the type θ under the optimal mechanism by $\pi(\theta) = R(\phi(\theta), \Lambda^*(\theta))V^*_g(\theta) - c(V^*_g(\theta))$. Similarly denote the analogous contribution to profits under perfect certification by $\pi^{s,1}(\theta) = R(\phi(\theta), 1)V^{s,1}_g(\theta) - c(V^{s,1}_g(\theta))$. These are both continuous by the theorem of the maximum. They are increasing since $R(\phi(\theta), \lambda)$ is increasing in θ.

Suppose Case 3 could arise. This requires that $\Lambda^*(\theta) = 1$ for all θ with $\phi(\theta) \geq \gamma$, so $\pi(\theta) = \pi^{s,1}(\theta) = 0$ for $\phi(\theta) = \gamma$, and therefore $\pi(\theta) = V^*_g(\theta) = 0$ for $\phi(\theta) \leq \gamma$. This contradicts $V^*_g(\theta) > 0$ for some θ such that $\phi(\theta) \leq \gamma$ as required in Case 3.

Finally, suppose Case 4 could arise. This requires $\pi(\theta) = 0$ for θ such that $\phi(\theta) < \gamma$, so $\pi(\theta) = 0$ for θ such that $\phi(\theta) = \gamma$ by continuity. Case 4 requires further that $\lim_{\theta \downarrow \phi^{-1}(\gamma)} \Lambda^*(\theta) < 1$ and $\lim_{\theta \downarrow \phi^{-1}(\gamma)} R(\phi(\theta), \Lambda^*(\theta)) = 0$. To see the latter, note that, if $\lim_{\theta \downarrow \phi^{-1}(\gamma)} R(\phi(\theta), \Lambda^*(\theta)) > 0$, this implies that $\lim_{\theta \downarrow \phi^{-1}(\gamma)} \pi(\theta) > 0$ which violates the continuity of π. But $R(\phi^{-1}(\gamma), 1) = 0$ and therefore we must have $\Lambda^*(\phi^{-1}(\gamma)) = 1$. This provides the requisite contradiction since $\lim_{\theta \downarrow \phi^{-1}(\gamma)} \Lambda^*(\theta) < 1$ and $\Lambda^*(\theta)$ is nondecreasing. □

Proposition 9. *Suppose $A(\lambda) = \lambda^\alpha$ where $\alpha \in (0,1)$ and $c(v_g) = \frac{v_g^\sigma}{\sigma}$ where $\sigma > 1$. Then, for all $\theta \in \Theta$, the optimal mechanism (V_g^*, Λ^*) satisfies*

$$\lim_{\gamma \downarrow 0} \Lambda^*(\theta) = 0 \quad \textit{and}$$

$$\lim_{\gamma \downarrow 0} A(\Lambda^*(\theta)) V_g^*(\theta) = \begin{cases} 0 & \textit{if } \alpha \in \left(\frac{1}{\sigma}, 1\right), \\ \alpha^{\frac{1}{\sigma - 1}} & \textit{if } \alpha = \frac{1}{\sigma}, \\ \infty & \textit{if } \alpha \in \left(0, \frac{1}{\sigma}\right). \end{cases}$$

Consequently, there exists a $\underline{\gamma} > 0$, such that for all $\gamma < \underline{\gamma}$ enforced perfect certification leads to higher (lower) total engagement than the optimal mechanism when $\alpha < (>) \frac{1}{\sigma}$.

Proof. For clarity in the proof, we denote the optimal mechanism as $(V_{g,\gamma}^*, \Lambda_\gamma^*)$ and the true effective virtual value as R_γ to make the dependence on γ explicit.

We first argue that $\lim_{\gamma \downarrow 0} \Lambda_\gamma^*(\theta) = 0$ for all $\theta \in \Theta$. For contradiction, suppose there exists a $\theta \in \Theta$ and a $\lambda_l > 0$ such that for every $\varepsilon > 0$, there exists a $\gamma \in (0, \varepsilon)$ for which $\Lambda_\gamma^*(\theta) \geq \lambda_l$. For such a γ, observe that

$$R_\gamma(\phi(\theta), \Lambda_\gamma^*(\theta)) = \max_{\lambda \in [\lambda_l, 1]} \left\{ \left(\phi(\theta) + \frac{1 - \lambda}{\lambda} \right) A(\lambda) - \frac{\gamma}{\lambda} \right\} < \max_{\lambda \in [\lambda_l, 1]} \left\{ \left(\phi(\theta) + \frac{1 - \lambda}{\lambda} \right) A(\lambda) \right\}$$

and so the value of $R(\phi(\theta), \Lambda_\gamma^*(\theta))$ is bounded from above by the right term (that does not depend on γ).

But suppose, instead of the optimal quality Λ_γ^*, the platform chooses quality $\lambda = \gamma$. Then,

$$R_\gamma(\phi(\theta), \gamma) = (\phi(\theta) - 1)\gamma^\alpha + \gamma^{\alpha - 1} - 1.$$

Now since $\alpha \in (0,1)$ and $\gamma < \varepsilon$, if we take ε sufficiently small, then

$$R_\gamma(\phi(\theta), \gamma) > \max_{\lambda \in [\lambda_l, 1]} \left\{ \left(\phi(\theta) + \frac{1 - \lambda}{\lambda} \right) A(\lambda) \right\} > R_\gamma(\phi(\theta), \Lambda_\gamma^*(\theta))$$

because $R_\gamma(\phi(\theta), \Lambda_\gamma^*(\theta))$ is bounded and $\gamma^\alpha \to 0$, $\gamma^{\alpha - 1} \to \infty$, and so $R_\gamma(\phi(\theta), \gamma) \to \infty$ as $\gamma \downarrow 0$. This provides the requisite contradiction and implies that $\lim_{\gamma \downarrow 0} \Lambda_\gamma^*(\theta) = 0$ for all $\theta \in \Theta$.

Now we derive the limit of the engagement $A(\Lambda^*_\gamma(\theta))V^*_{g,\gamma}(\theta)$ from the optimal mechanism as $\gamma \downarrow 0$. First observe that, for every $\theta \in \Theta$, the above argument implies that $R_\gamma(\phi(\theta), \Lambda^*_\gamma(\theta)) > 0$ for sufficiently small $\gamma > 0$. Thus, for sufficiently small γ, the views $V^*_{g,\gamma}(\theta) > 0$ and are given by the equation

$$c'(V^*_{g,\gamma}(\theta)) = R_\gamma(\phi(\theta), \Lambda^*_\gamma(\theta)) \iff V^*_{g,\gamma}(\theta) = \left(\left(\phi(\theta) + \frac{1-\Lambda^*_\gamma(\theta)}{\Lambda^*_\gamma(\theta)}\right)\Lambda^*_\gamma(\theta)^\alpha - \frac{\gamma}{\Lambda^*_\gamma(\theta)}\right)^{\frac{1}{\sigma-1}}.$$

Thus, engagement from the optimal mechanism is given by

$$A(\Lambda^*_\gamma(\theta))V^*_{g,\gamma}(\theta) = \Lambda^*_\gamma(\theta)^\alpha \left(\left(\phi(\theta) + \frac{1-\Lambda^*_\gamma(\theta)}{\Lambda^*_\gamma(\theta)}\right)\Lambda^*_\gamma(\theta)^\alpha - \frac{\gamma}{\Lambda^*_\gamma(\theta)}\right)^{\frac{1}{\sigma-1}}.$$

Since $\lim_{\gamma\downarrow 0} \Lambda^*_\gamma(\theta) = 0$, the optimal quality $\Lambda^*_\gamma(\theta)$ must be less than 1 for sufficiently small $\gamma > 0$. Thus, for such small values of γ, it must satisfy the first-order condition

$$\alpha(\phi(\theta)-1)\Lambda^*_\gamma(\theta)^{\alpha-1} + (\alpha-1)\Lambda^*_\gamma(\theta)^{\alpha-2} + \frac{\gamma}{\Lambda^*_\gamma(\theta)^2} = 0.$$

This can be rewritten as

$$\alpha(\phi(\theta)-1)\Lambda^*_\gamma(\theta)^{\alpha} + (\alpha-1)\Lambda^*_\gamma(\theta)^{\alpha-1} = -\frac{\gamma}{\Lambda^*_\gamma(\theta)}.$$

We use this equation to replace the term $-\frac{\gamma}{\Lambda^*_\gamma(\theta)}$ in the expression for the engagement $A(\Lambda^*_\gamma(\theta))V^*_{g,\gamma}(\theta)$ to get

$$\begin{aligned}
A(\Lambda^*_\gamma(\theta))V^*_{g,\gamma}(\theta) &= \Lambda^*_\gamma(\theta)^\alpha \left(\left(\phi(\theta) + \frac{1-\Lambda^*_\gamma(\theta)}{\Lambda^*_\gamma(\theta)}\right)\Lambda^*_\gamma(\theta)^\alpha + \alpha(\phi(\theta)-1)\Lambda^*_\gamma(\theta)^\alpha + (\alpha-1)\Lambda^*_\gamma(\theta)^{\alpha-1}\right)^{\frac{1}{\sigma-1}} \\
&= \Lambda^*_\gamma(\theta)^\alpha \left((\phi(\theta)-1)\Lambda^*_\gamma(\theta)^\alpha + \Lambda^*_\gamma(\theta)^{\alpha-1} + \alpha(\phi(\theta)-1)\Lambda^*_\gamma(\theta)^\alpha + (\alpha-1)\Lambda^*_\gamma(\theta)^{\alpha-1}\right)^{\frac{1}{\sigma-1}} \\
&= \Lambda^*_\gamma(\theta)^\alpha \left((\phi(\theta)-1)(1+\alpha)\Lambda^*_\gamma(\theta)^\alpha + \alpha\Lambda^*_\gamma(\theta)^{\alpha-1}\right)^{\frac{1}{\sigma-1}} \\
&= \left((\phi(\theta)-1)(1+\alpha)\Lambda^*_\gamma(\theta)^{\sigma\alpha} + \alpha\Lambda^*_\gamma(\theta)^{\sigma\alpha-1}\right)^{\frac{1}{\sigma-1}}.
\end{aligned}$$

As $\gamma \downarrow 0$, the optimal quality $\Lambda^*_\gamma(\theta) \to 0$ and therefore, the first term goes to zero, since $\alpha\sigma > 0$. The second term goes to zero if $\alpha > \frac{1}{\sigma}$, to infinity if $\alpha < \frac{1}{\sigma}$, and to α if $\alpha = \frac{1}{\sigma}$. This completes the proof. □

Proposition 10. *Let $A_{\hat{b}}(\lambda) = A(\max\{\lambda(1+\hat{b}) - \hat{b}, 0\})$ with $\hat{b} > 0$ and $A_{\hat{z}}(\lambda) = A(\min\{\lambda + z, 1\})$ with $0 < \hat{z} < A^{-1}(\gamma)$ where $A(\lambda)$ is an attention function such that $\frac{1}{A(\lambda)}\frac{\partial A(\lambda)}{\partial \lambda}$ is strictly decreasing.*

Then, for every $\theta \in \Theta$ and every $0 < b' < b$ with $V^*_{g,b}(\theta) > 0$, we have $V^*_{g,b'}(\theta) > 0$ and $\Lambda^*_b(\theta) \geq \Lambda^*_{b'}(\theta)$. For every $\theta \in \Theta$ and every $0 < z < z' < A^{-1}(\gamma)$ with $V^*_{g,z}(\theta) > 0$, we have $V^*_{g,z'}(\theta) > 0$ and $\Lambda^*_z(\theta) \geq \Lambda^*_{z'}(\theta)$.

Proof. To prove this result, we will invoke the argument in the proof of Proposition A.1. We cannot apply that result directly since both A_b and A_z do not satisfy some of the assumptions—$A(0) = 0$, $A(1) = 1$ and A increasing—that we imposed on attention functions. That said, the argument in the proof can be applied with minor adjustments. Similar to (12) in the proof of Proposition A.1, we employ shorthand notation

$$g(\hat{b}, \hat{\lambda}) := \left(\phi(\theta) + \frac{1-\hat{\lambda}}{\hat{\lambda}}\right) A_{\hat{b}}(\hat{\lambda})$$

with $g(\hat{z}, \hat{\lambda})$ defined analogously (with $\hat{z}$ replacing $\hat{b}$). We deliberately overload the notation for simplicity since the meaning is clear from the first argument.

First, we consider the case of a b and a θ such that $V^*_{g,b}(\theta) > 0$. Observe that, for any $b' < b$, we have

$$0 < g(b, \Lambda^*_b(\theta)) - \frac{\gamma}{\Lambda^*_b(\theta)} \leq g(b', \Lambda^*_b(\theta)) - \frac{\gamma}{\Lambda^*_b(\theta)} \leq \max_{\lambda \in [0,1]} \left\{ g(b', \lambda) - \frac{\gamma}{\lambda} \right\} = g(b', \Lambda^*_{b'}(\theta)) - \frac{\gamma}{\Lambda^*_{b'}(\theta)}$$

where the first inequality follows from the fact that $V^*_{g,b}(\theta) > 0$ and the second inequality from the fact that $g(\hat{b}, \hat{\lambda})$ is decreasing in $\hat{b}$. Therefore, $V^*_{g,b'}(\theta) > 0$.

We now argue that for all b, there exists an $\varepsilon > 0$ such that $\Lambda^*_b(\theta) \geq \Lambda^*_{b'}(\theta)$ for $b' \in (b - \varepsilon, b)$ and $\Lambda^*_b(\theta) \leq \Lambda^*_{b'}(\theta)$ for $b' \in (b, b + \varepsilon)$. This suffices to establish the required monotonicity.

First, suppose $\Lambda^*_b(\theta) = 1$, then $\Lambda^*_b(\theta) \geq \Lambda^*_{b'}(\theta)$ for $b' < b$. For all $b' > b$, the optimality of $\Lambda^*_b(\theta) = 1$ for b implies that

$$g(b, \Lambda^*_b(\theta)) - \frac{\gamma}{\Lambda^*_b(\theta)} = \phi(\theta) - \gamma \geq g(b, \Lambda^*_{b'}(\theta)) - \frac{\gamma}{\Lambda^*_{b'}(\theta)}.$$

Suppose, for contradiction, that $\Lambda^*_{b'}(\theta) < 1$ for some $b' > b$.

Now $V^*_{g,b}(\theta) > 0$ implies that $\phi(\theta) - \gamma > 0$ which, in turn, implies that $\phi(\theta) > 0$. By assumption, $A_b(\Lambda^*_{b'}(\theta)) > A_{b'}(\Lambda^*_{b'}(\theta))$ and this combined with the fact that $\phi(\theta) > 0$ implies $g(b, \Lambda^*_{b'}(\theta)) > g(b', \Lambda^*_{b'}(\theta))$. Consequently,

$$\phi(\theta) - \gamma \geq g(b, \Lambda^*_{b'}(\theta)) - \frac{\gamma}{\Lambda^*_{b'}(\theta)} \implies \phi(\theta) - \gamma > g(b', \Lambda^*_{b'}(\theta)) - \frac{\gamma}{\Lambda^*_{b'}(\theta)}$$

violating the optimality of $\Lambda^*_{b'}(\theta)$. To summarize, if $\Lambda^*_b(\theta) = 1$, then we must have $\Lambda^*_{b'}(\theta) = 1$ for all $b' > b$ as required.

So consider a b and a θ such that $V^*_{g,b}(\theta) > 0$ and $\Lambda^*_b(\theta) < 1$. Since $V^*_{g,b}(\theta) > 0$, we must have $\Lambda^*_b(\theta) > \frac{b}{1+b}$. But then, following the identical argument as in the proof of Proposition A.1, there exists an $\varepsilon > 0$ such that $\Lambda^*_b(\theta) \geq \Lambda^*_{b'}(\theta)$ for $b' \in (b-\varepsilon, b)$ and $\Lambda^*_b(\theta) \leq \Lambda^*_{b'}(\theta)$ for $b' \in (b, b+\varepsilon)$. This is because there is a neighborhood around $\Lambda^*_b(\theta)$ such that A_b satisfies all the assumptions in the statement of Proposition A.1 (with b playing the role of $\hat{\alpha}$). To see this, observe that $A_b(\lambda)$ is locally strictly decreasing in b and

$$\frac{1}{A_b(\lambda)} \frac{\partial A_b(\lambda)}{\partial \hat{b}} = (\lambda - 1) \frac{A'((1+b)\lambda - b)}{A((1+b)\lambda - b)} < 0$$

is locally increasing in λ since $\lambda < 1$ and $\frac{A'(\cdot)}{A(\cdot)} > 0$ is assumed to be strictly decreasing. This completes the monotonicity argument for b.

Now consider the case of a z and a θ such that $V^*_{g,z}(\theta) > 0$. We will argue that there exists an $\varepsilon > 0$ such that $\Lambda^*_z(\theta) \leq \Lambda^*_{z'}(\theta)$ for $z' \in (z-\varepsilon, z)$ and $\Lambda^*_z(\theta) \geq \Lambda^*_{z'}(\theta)$ for $z' \in (z, z+\varepsilon)$. This suffices to establish the required monotonicity.

We begin by arguing that $\Lambda^*_z(\theta) \leq 1 - z$. Suppose, for contradiction, that $\Lambda^*_z(\theta) > 1 - z$ and so $A_z(\Lambda^*_z(\theta)) = 1$. Then,

$$g(z, \Lambda^*_z(\theta)) - \frac{\gamma}{\Lambda^*_z(\theta)} = \phi(\theta) - 1 + \frac{1-\gamma}{\Lambda^*_z(\theta)} < \phi(\theta) - 1 + \frac{1-\gamma}{1-z} = g(z, 1-z) - \frac{\gamma}{1-z}$$

which contradicts the optimality of $\Lambda^*_z(\theta)$.

So first suppose that $\Lambda^*_z(\theta) = 1 - z$. Then, for $z' > z$, we

$$\Lambda^*_z(\theta) = 1 - z \geq 1 - z' \geq \Lambda^*_{z'}(\theta)$$

as required. For all $z' < z$, the optimality of $\Lambda^*_z(\theta) = 1 - z$ for z implies that

$$g(z, \Lambda^*_z(\theta)) - \frac{\gamma}{\Lambda^*_z(\theta)} = \phi(\theta) + \frac{z-\gamma}{1-z} \geq g(z, \Lambda^*_{z'}(\theta)) - \frac{\gamma}{\Lambda^*_{z'}(\theta)}.$$

Suppose, for contradiction, that $\Lambda^*_{z'}(\theta) < 1 - z$ for some $z' < z$.

Now $V^*_{g,z}(\theta) > 0$ implies that $\phi(\theta) + \frac{z-\gamma}{1-z} > 0$ which, in turn, implies that $\phi(\theta) + \frac{z}{1-z} > 0$. Moreover, $\phi(\theta) + \frac{z}{1-z} > 0$ implies that $\phi(\theta) + \frac{1-\Lambda^*_{z'}(\theta)}{\Lambda^*_{z'}(\theta)} > 0$ because we assumed $\Lambda^*_{z'}(\theta) < 1 - z$. This combined with the fact that $A_z(\Lambda^*_{z'}(\theta)) > A_{z'}(\Lambda^*_{z'}(\theta))$ (by assumption) implies $g(z, \Lambda^*_{z'}(\theta)) > g(z', \Lambda^*_{z'}(\theta))$. Consequently,

$$\phi(\theta) + \frac{z-\gamma}{1-z} \geq g(z, \Lambda^*_{z'}(\theta)) - \frac{\gamma}{\Lambda^*_{z'}(\theta)} \implies \phi(\theta) + \frac{z-\gamma}{1-z} > g(z', \Lambda^*_{z'}(\theta)) - \frac{\gamma}{\Lambda^*_{z'}(\theta)}$$

violating the optimality of $\Lambda^*_{z'}(\theta)$. To summarize, if $\Lambda^*_z(\theta) = 1 - z$, then we must have $\Lambda^*_{z'}(\theta) \geq 1 - z$ for all $z' < z$ as required.

So consider a z and a θ such that $\Lambda_z^*(\theta) < 1 - z$ and therefore $A_z(\Lambda_z^*(\theta)) < 1$. We first note that $A_z(\Lambda_z^*(\theta)) > 0$. This is because

$$\left(\phi(\theta) + \frac{1-\lambda}{\lambda}\right) A_z(\lambda) - \frac{\gamma}{\lambda} = (\phi(\theta) - 1)A_z(\lambda) + \frac{A_z(\lambda) - \gamma}{\lambda} \quad \longrightarrow \quad -\infty$$

as $\lambda \to 0$ because $A_z(0) = A(z) < \gamma$ by assumption.

Therefore we are left to consider the case where $\Lambda_z^*(\theta) \in (0, 1-z)$. Once again, for this interior case, we can use the identical argument as in the proof of Proposition A.1. To do so, we redefine $A(\lambda; \hat{\alpha}) = A(\min\{\lambda + 1 - \hat{\alpha}, 1\})$ where $\hat{\alpha} \in (1 - A^{-1}(\gamma), 1)$. Take $\alpha = 1 - z$ and observe that $A(\lambda; \alpha) = A_z(\alpha) = A(\lambda + z)$; therefore $\Lambda_\alpha^*(\theta) = \Lambda_z^*(\theta) \in (0, 1-z)$. Observe that there is a neighborhood around $\Lambda_\alpha^*(\theta)$ such that $A(\lambda; \alpha)$ satisfies all the assumptions in the statement of Proposition A.1. To see this, observe that $A_\alpha(\lambda)$ is locally strictly decreasing in α and

$$\frac{1}{A_\alpha(\lambda)} \frac{\partial A_\alpha(\lambda)}{\partial \hat{\alpha}} = -\frac{A'(\lambda + 1 - \alpha)}{A(\lambda + 1 - \alpha)} < 0$$

is locally increasing in λ since $\frac{A'(\cdot)}{A(\cdot)} > 0$ is assumed to be strictly decreasing.

Therefore, the argument in the proof of Proposition A.1 implies that there exists an $\varepsilon' > 0$ such that $\Lambda_\alpha^*(\theta) \geq \Lambda_{\alpha'}^*(\theta)$ for $\alpha' \in (\alpha - \varepsilon', \alpha)$ and $\Lambda_\alpha^*(\theta) \leq \Lambda_{\alpha'}^*(\theta)$ for $\alpha' \in (\alpha, \alpha + \varepsilon')$. Consequently, there also exists an $\varepsilon > 0$ such that $\Lambda_z^*(\theta) \leq \Lambda_{z'}^*(\theta)$ for $z' \in (z - \varepsilon, z)$ and $\Lambda_z^*(\theta) \geq \Lambda_{z'}^*(\theta)$ for $z' \in (z, z + \varepsilon)$. This establishes that $\Lambda_z^*(\theta) \geq \Lambda_{z'}^*(\theta)$ for all $z' > z$ as required to complete the proof of this result. □

APPENDIX B. THE RELATIONSHIP BETWEEN ENGAGEMENT AND CONSUMER WELFARE

Here we substantiate the claim in Sections 3.1 and 4.3 that maximizing engagement is equivalent to maximizing consumer welfare.

Proposition B.1. *Suppose $A(\lambda) = \lambda^\alpha$. Then, the consumer welfare from v_g good content views assigned a certificate of quality λ is proportional to $A(\lambda)v_g$.*

Proof. We calculate the consumer welfare from v_g good content views when the content is marked with a certificate of quality λ. This, of course, implies that there are $v_b = \frac{1-\lambda}{\lambda} v_g$ bad content views assigned to the same certificate.

Recall that consumers pay a cost q drawn from a distribution A of reading content. If content is read, the consumer receives a payoff of 1 if the content is good and 0 if it is bad. Thus, only consumers with cost $q \leq \lambda$ read the content and thus consumer welfare is given by

$$v_g \int_0^\lambda (1 - q)\, dA(q) - v_b \int_0^\lambda q dA(q).$$

This expression can be rewritten as

$$\begin{aligned} v_g \int_0^{\lambda} (1-q)\, dA(q) - v_b \int_0^{\lambda} q dA(q) &= v_g \int_0^{\lambda} \left((1-q) - \frac{(1-\lambda)}{\lambda} q \right) dA(q) \\ &= v_g \int_0^{\lambda} \left(1 - \frac{1}{\lambda} q \right) dA(q) \\ &= v_g \left(A(\lambda) - \frac{1}{\lambda} \int_0^{\lambda} q dA(q) \right). \end{aligned}$$

For $A(\lambda) = \lambda^{\alpha}$, this simplifies to $\frac{1}{\alpha+1} v_g A(\lambda)$; that is, welfare is proportional to engagement as required. □

Appendix C. Two certificates analysis

In this section, we provide a full characterization of the platform's optimal mechanism when restricted to two certificates.

Proposition C.1. *Consider the version of the platform's problem (2) where $\Lambda : \Theta \to \{\underline{\lambda}, \overline{\lambda}\}$ and $V_g : \Theta \to \mathbb{R}_+$.*

There is an optimal mechanism $(V_g^{bin}, \Lambda^{bin})$ given by

$$\hat{\theta} = \begin{cases} \min\left\{ \theta \in \Theta \;\middle|\; \left(\phi(\theta) + \frac{1-\overline{\lambda}}{\overline{\lambda}} \right) A(\overline{\lambda}) - \frac{\gamma}{\overline{\lambda}} \geq \left(\phi(\theta) + \frac{1-\underline{\lambda}}{\underline{\lambda}} \right) A(\underline{\lambda}) - \frac{\gamma}{\underline{\lambda}} \right\} & \textit{if the set is non-empty,} \\ \overline{\theta} & \textit{otherwise,} \end{cases}$$

$$\Lambda^{bin}(\theta) = \begin{cases} \overline{\lambda} & \textit{if } \theta > \hat{\theta} \textit{ or } \theta = \hat{\theta} < \overline{\theta}, \\ \underline{\lambda} & \textit{if } \theta < \hat{\theta} \textit{ or } \theta = \hat{\theta} = \overline{\theta}, \end{cases} \qquad \textit{and}$$

$$V_g^{bin}(\theta) = c'^{-1} \left(\max\left\{ \left[\phi(\theta) + \frac{1-\Lambda^{bin}(\theta)}{\Lambda^{bin}(\theta)} \right] A(\Lambda^{bin}(\theta)) - \frac{\gamma}{\Lambda^{bin}(\theta)}, 0 \right\} \right).$$

Proof. We maximize the objective function pointwise and show that the mechanism we obtain satisfies the necessary monotonicity properties to satisfy the incentive compatibility constraints.

First, observe that, if

$$\left(\phi(\theta) + \frac{1-\overline{\lambda}}{\overline{\lambda}} \right) A(\overline{\lambda}) - \frac{\gamma}{\overline{\lambda}} \geq \left(\phi(\theta) + \frac{1-\underline{\lambda}}{\underline{\lambda}} \right) A(\underline{\lambda}) - \frac{\gamma}{\underline{\lambda}} \tag{13}$$

then, for any $V_g(\theta) \in \mathbb{R}_+$, the value of the objective function satisfies

$$\left[\left(\phi(\theta) + \frac{1-\overline{\lambda}}{\overline{\lambda}} \right) A(\overline{\lambda}) - \frac{\gamma}{\overline{\lambda}} \right] V_g(\theta) - c(V_g(\theta)) \geq \left[\left(\phi(\theta) + \frac{1-\underline{\lambda}}{\underline{\lambda}} \right) A(\underline{\lambda}) - \frac{\gamma}{\underline{\lambda}} \right] V_g(\theta) - c(V_g(\theta))$$

and vice versa when inequality (13) is reversed.

Consequently, there is a pointwise optimum that satisfies

$$\Lambda^{bin}(\theta) \in \operatorname*{argmax}_{\lambda \in \{\underline{\lambda}, \overline{\lambda}\}} \left\{ \left(\phi(\theta) + \frac{1-\lambda}{\lambda} \right) A(\lambda) - \frac{\gamma}{\lambda} \right\}.$$

For any $\theta \in \Theta$ at which both $\underline{\lambda}$ and $\overline{\lambda}$ are maximizers, we set $\Lambda^{bin}(\theta) = \overline{\lambda}$.

Now note that, if (13) holds for some θ, it also holds for all $\theta' > \theta$. This is because $A(\overline{\lambda}) > A(\underline{\lambda})$ and $\phi(\cdot)$ is nondecreasing. Consequently, Λ^{bin} must take the cutoff form

$$\Lambda^{bin}(\theta) = \begin{cases} \overline{\lambda} & \text{when } \theta > \hat{\theta}, \\ \underline{\lambda} & \text{when } \theta < \hat{\theta}. \end{cases}$$

as in the statement of the proposition and thus Λ^{bin} is nondecreasing.

The function V_g^{bin} as defined in the statement of the proposition is the solution to

$$V_g^{bin}(\theta) = \operatorname*{argmax}_{v_g \in \mathbb{R}_+} \left\{ \left[\left(\phi(\theta) + \frac{1-\Lambda^{bin}(\theta)}{\Lambda^{bin}(\theta)} \right) A(\Lambda^{bin}(\theta)) - \frac{\gamma}{\Lambda^{bin}(\theta)} \right] v_g - c(v_g) \right\} \tag{14}$$

and the exact expression is obtained from the first-order condition.

Now observe that, because ϕ is nondecreasing, the function

$$\left(\phi(\theta) + \frac{1-\Lambda^{bin}(\theta)}{\Lambda^{bin}(\theta)} \right) A(\Lambda^{bin}(\theta)) - \frac{\gamma}{\Lambda^{bin}(\theta)} = \max_{\lambda \in \{\underline{\lambda}, \overline{\lambda}\}} \left\{ \left(\phi(\theta) + \frac{1-\lambda}{\lambda} \right) A(\lambda) - \frac{\gamma}{\lambda} \right\}$$

is nondecreasing because it is the maximum of two nondecreasing functions. This, in turn, implies from (14), that V^{bin} is nondecreasing. Therefore $A(\Lambda^{bin}(\cdot))V_g^{bin}(\cdot)$ is nondecreasing and consequently the pointwise solution is incentive compatible as required. This completes the proof. □

References

ACEMOGLU, D., A. OZDAGLAR, AND J. SIDERIUS (2024): "A model of online misinformation," *Review of Economic Studies*, 91(6), 3117–3159.

ALI, S. N., N. HAGHPANAH, X. LIN, AND R. SIEGEL (2022): "How to sell hard information," *Quarterly Journal of Economics*, 137(1), 619–678.

ARIDOR, G., R. JIMÉNEZ-DURÁN, R. LEVY, AND L. SONG (forthcoming): "The Economics of Social Media," *Journal of Economic Literature*.

ARMSTRONG, M., AND J. ZHOU (2011): "Paying for prominence," *Economic Journal*, 121(556), F368–F395.

ATHEY, S., AND G. ELLISON (2011): "Position auctions with consumer search," *Quarterly Journal of Economics*, 126(3), 1213–1270.

BAR-ISAAC, H., AND S. SHELEGIA (2025): *Monetizing steering*. Working Paper.

BHARGAVA, H. K. (2022): "The Creator Economy: Managing Ecosystem Supply, Revenue Sharing, and Platform Design," *Management Science*, 68(7), 5233–5251.

BÖHME, E. (2016): "Second-degree price discrimination on two-sided markets," *Review of Network Economics*, 15(2), 91–115.

BOUVARD, M., AND R. LEVY (2018): "Two-sided reputation in certification markets," *Management Science*, 64(10), 4755–4774.

BURGUET, R., R. CAMINAL, AND M. ELLMAN (2015): "In Google we trust?," *International Journal of Industrial Organization*, 39, 44–55.

BURSZTYN, L., B. R. HANDEL, R. JIMENEZ, AND C. ROTH (2023): "When product markets become collective traps: The case of social media," Discussion paper, National Bureau of Economic Research.

CHEN, Y., AND C. HE (2011): "Paid placement: Advertising and search on the internet," *Economic Journal*, 121(556), F309–F328.

CHEN, Y., F. LI, AND M. PREUSS (2025): "Algorithmic Attention and Content Creation on Social Media Platforms," Available at SSRN: `https://ssrn.com/abstract=5182754`.

CHOI, J. P., D.-S. JEON, AND B.-C. KIM (2015): "Net neutrality, business models, and internet interconnection," *American Economic Journal: Microeconomics*, 7(3), 104–141.

COMPETITION AND MARKETS AUTHORITY (2022): "Auditing algorithms: the existing landscape, role of regulators and future outlook," accessed on June 7, 2024 at https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook.

CORTS, K. S. (2013): "Prohibitions on false and unsubstantiated claims: Inducing the acquisition and revelation of information through competition policy," *Journal of Law and Economics*, 56(2), 453–486.

——— (2014): "Finite optimal penalties for false advertising," *Journal of Industrial Economics*, 62(4), 661–681.

DE CORNIERE, A., AND G. TAYLOR (2014): "Integration and search engine bias," *RAND Journal of Economics*, 45(3), 576–597.

——— (2019): "A model of biased intermediation," *RAND Journal of Economics*, 50(4), 854–882.

DENECKERE, R. J., AND P. MCAFEE (1996): "Damaged goods," *Journal of Economics and Management Strategy*, 5(2), 149–174.

DRANOVE, D., AND G. Z. JIN (2010): "Quality disclosure and certification: Theory and practice," *Journal of economic literature*, 48(4), 935–963.

EDELMAN, B., M. OSTROVSKY, AND M. SCHWARZ (2007): "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords," *American Economic Review*, 97(1), 242–259.

ERSHOV, D., AND M. MITCHELL (forthcoming): "The Effects of Advertising Disclosure Regulations on Social Media: Evidence From Instagram," *RAND Journal of Economics*, forthcoming.

FAINMESSER, I. P., AND A. GALEOTTI (2021): "The Market for Online Influence," *American Economic Journal: Microeconomics*, 13(3), 1–28.

GHOSH, A., AND P. MCAFEE (2011): "Incentivizing high-quality user-generated content," in *Proceedings of the 20th International Conference on World Wide Web (WWW)*, pp. 137–146. ACM.

GLAESER, E. L., AND G. UJHELYI (2010): "Regulating misinformation," *Journal of Public Economics*, 94(3-4), 247–257.

GOLDFARB, A., AND C. TUCKER (2019): "Digital economics," *Journal of Economic Literature*, 57(1), 3–43.

HAGIU, A., AND B. JULLIEN (2014): "Search diversion and platform competition," *International Journal of Industrial Organization*, 33, 48–60.

ICHIHASHI, S., AND A. SMOLIN (2023): "Buyer-Optimal Algorithmic Consumption," SSRN Working Paper, September 21, 2023.

INDERST, R., AND M. OTTAVIANI (2012): "Competition through Commissions and Kickbacks," *American Economic Review*, 102(2), 780–809.

JEON, D.-S., B.-C. KIM, AND D. MENICUCCI (2022): "Second-degree price discrimination by a two-sided monopoly platform," *American Economic Journal: Microeconomics*, 14(2), 322–369.

JEON, D.-S., AND S. LOVO (2013): "Credit rating industry: A helicopter tour of stylized facts and recent theories," *International Journal of Industrial Organization*, 31(5), 643–651.

JOHNSON, J. P., A. RHODES, AND M. WILDENBEEST (2023): "Platform Design When Sellers Use Pricing Algorithms," *Econometrica*, 91(5), 1841–1879.

KOMINERS, S. D., AND J. M. SHAPIRO (2024): "Content Moderation with Opaque Policies," Working Paper w32156, National Bureau of Economic Research (NBER), Available at SSRN: https://ssrn.com/abstract=4731065.

LIZZERI, A. (1999): "Information revelation and certification intermediaries," *RAND Journal of Economics*, pp. 214–231.

MADIO, L., AND M. QUINN (2024): "Content Moderation and Advertising in Social Media Platforms," Working Paper 11169, CESifo Working Paper, Available at SSRN: https://ssrn.com/abstract=4875546 or http://dx.doi.org/10.2139/ssrn.4875546.

MITCHELL, M. (2021): "Free ad (vice): internet influencers and disclosure regulation," *RAND Journal of Economics*, 52(1), 3–21.

MUSSA, M., AND S. ROSEN (1978): "Monopoly and product quality," *Journal of Economic theory*, 18(2), 301–317.

QIAN, K., AND S. JAIN (2024): "Digital Content Creation: An Analysis of the Impact of Recommendation Systems," *Management Science*, 70(12), 8668–8684.

RHODES, D., AND J. WILSON (2018): "False Advertising," *RAND Journal of Economics*, 49(4), 1011–1039.

SRINIVASAN, K. (2023): "Paying Attention," Discussion paper, Mimeo.

WHITE, L. J. (2010): "Markets: The credit rating agencies," *Journal of Economic Perspectives*, 24(2), 211–226.

ZOU, T., Y. WU, AND M. SARVARY (2025): "Designing Recommendation Systems on Content Platforms: Trading off Quality and Variety," *Available at SSRN*.