ISSN 1936-5349 (print)

ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

Holding Platforms Liable

Xinyu Hua & Kathryn E. Spier

Updated Discussion Paper

11/2024

Harvard Law School

Cambridge, MA 02138

Holding Platforms Liable*

Xinyu Hua† Kathryn E. Spier‡
HKUST Harvard University

November 15, 2024

Abstract

Should platforms be liable for harms suffered by users? A platform enables interactions between firms and users. Harmful firms impose larger costs on users than safe firms. If firms have deep pockets and are fully liable for harms, platform liability is unnecessary. If firms have limited liability, holding platforms liable for residual harm increases platforms' incentives to raise interaction prices and invest in auditing to deter, detect, and block harmful firms. The social desirability and optimal level of platform liability depend on whether interactions require user consent, the degree to which users internalize harms, and the observability of platform effort.

*We would like to thank the Co-editor, three anonymous referees, Gary Biglaiser, Luís Cabral, Jay Pil Choi, James Dana, Andrei Hagiu, Bård Harstad, Ginger Jin, Yassine Lefouili, Hong Luo, Bentley MacLeod, Sarit Markovich, Haggai Porat, Urs Schweizer, Emil Temnyalov, Marshall Van Alstyne, Rory Van Loo, Julian Wright and seminar audiences at Boston University, Chinese University of Hong Kong, Columbia University, Fudan University, Georgetown University, Harvard University, the Kellogg School at Northwestern University, Asia Pacific Industrial Organization Conference (APIOC 2021), International Industrial Organization Conference (IIOC 2022), Economics of Platforms Seminar (TSE 2022), the 2022 Asia Meeting of the Econometric Society, American Law and Economics Association Conference (ALEA 2022), Society for Institutional & Organizational Economics Conference (SIOE 2022), JRC-TSE Workshop on Liability in the Digital Economy (2022), Society for the Advancement of Economic Theory Conference (SAET 2022). We also thank the support from the Hong Kong Research Grants Council (GRF Grant Number: 16500722).

†Hong Kong University of Science and Technology. xyhua@ust.hk.

‡Harvard Law School and NBER. kspier@law.harvard.edu.

1 Introduction

Online platforms are ubiquitous in the modern world. We connect with friends on Facebook, shop for products on Amazon, and search online for jobs, information, and entertainment. While the economic and social benefits created by platforms are undeniable, the costs and hazards for users are very real too. For example, platform users run the risk that their personal data and privacy will be compromised. Users of social networking sites and search engines may be misled by fraudulent advertisements and misinformation. Consumers who shop online run the risk of purchasing counterfeit, defective, or dangerous goods. Should internet platforms like Facebook and Amazon be liable for the harms suffered by users?

In the United States, platforms enjoy relatively broad immunity from lawsuits brought by users.[1] Section 230 of the Communications Decency Act shields platforms from liability for the digital content created by their participants.[2] In the EU, under the Digital Services Act (DSA), platforms may avoid liability for illegal content posted by users, assuming they are not aware of it.[3] This immunity is being challenged in legislatures and the courts. In an early class-action lawsuit, users sued Google for the financial losses that they suffered from being duped by an unscrupulous advertiser into purchasing unwanted cell phone services.[4] The FTC has been investigating how "platforms screen for misleading ads for scams and fraudulent and counterfeit products" and, "in 2022 alone, consumers reported losing more than $1.2 billion to fraud that started on social media, more than any other contact method."[5] Proposed federal legislation in the U.S. would hold platforms liable if they fail to protect users.[6]

Marketplace platforms have largely avoided responsibility for defective products and services sold by third-party vendors. In 2019 the Fourth Circuit held that Amazon.com is not a traditional seller and therefore not subject to strict tort liability.[7] The following

[1] See generally Spier and Van Loo (2025).

[2] See *Force v. Facebook, Inc.*, 934 F.3d 53 (2d Cir. 2019).

[3] Regulation (EU) 2022/2065. Exemption from liability is a core concept of the EU's e-Commerce Directive. See Buiten et al. (2020).

[4] See *Goddard v. Google, Inc.*, 640 F. Supp. 2d 1193 (N.D. Cal. 2009). The court dismissed the lawsuit, holding that the action was barred under Section 230.

[5] https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-issues-orders-social-media-video-streaming-platforms-regarding-efforts-address-surge-advertising

[6] The bipartisan Internet PACT Act is one example. https://www.schatz.senate.gov/news/press-releases/schatz-thune-reintroduce-legislation-to-strengthen-rules-transparency-for-online-content-moderation-hold-internet-companies-accountable

[7] See *Erie Ins. Co. v. Amazon.com, Inc.*, 925 F.3d 135 (4th Cir. 2019); *State Farm Fire & Cas. Co. v. Amazon.com, Inc.*, 835 F. App'x 213 (9th Cir. 2020), and *Great N. Ins. Co. v. Amazon.com, Inc.*,

year, a California court found that Amazon could be held strictly liable for a defective laptop battery that was sold by third-party vendors but "Fulfilled by Amazon."[8] Then, in 2021, Amazon was held strictly liable for harms caused by a defective hoverboard that was shipped directly to the consumer by an overseas third-party vendor. Although Amazon did not fulfill the hoverboard order, the court opined that Amazon was "instrumental" in its sale and that "Amazon is well situated to take cost-effective measures to minimize the social costs of accidents."[9] In short, the law is far from settled.

This paper presents a formal model of a two-sided platform with two kinds of participants, "firms" and "users." The platform enables interactions between the firms and users, and charges the firms a fixed price per interaction. There are two types of firms: harmful and safe. The harmful firms enjoy higher gross benefits per interaction but impose larger costs on the users. Interactions between harmful firms and users are socially inefficient (the costs exceed the benefits). In an ideal world, the harmful firms are deterred from joining the platform. If the harmful firms remain undeterred, however, the platform plays an instrumental role in reducing social costs. The platform has the ability to prevent harmful interactions by raising the interaction price or by investing resources to detect and block the harmful firms.[10]

In our baseline model, users are effectively *bystanders* of the firms. By joining the platform, the users consent to subsequent firm-user interactions. Such settings include social and professional networking platforms such as Facebook and LinkedIn where the users enjoy same-side network benefits from sharing content with each other and the firms pay the platform to access user data or to engage in influential activities (e.g., advertising). In these settings, bad actors may post fraudulent advertisements[11] and use consumer data for "identity theft, phishing, fraud, and other harmful purposes."[12] Absent liability, the harmful firms have a strong incentive to join the platform and the platform has an insufficient incentive to detect and block them. Holding the firms and the platform jointly liable gets them to internalize the negative externalities on the user-bystanders.

524 F. Supp. 3d 852 (N.D. Ill. 2021).

[8]See *Bolger v. Amazon.com, LLC*, 53 Cal. App. 5th 431 (2020).

[9]See *Loomis v. Amazon.com, LLC*, 63 Cal. App. 5th 466 (2021).

[10]Platforms can and do invest resources to vet and block participants. LinkedIn uses automatic and manual investigations to remove scams; Amazon employs machine learning scientists, software developers and expert investigators to detect fraud. See Van Loo (2020a, 2020b) and Spier and Van Loo (2025).

[11]Fraudulent business and job-opportunity postings on LinkedIn and other platforms have proliferated, with reported harms topping $367 million in 2022. See https://consumer.ftc.gov/consumer-alerts/2023/04/you-got-job.

[12]See *United States of America v. Facebook Inc.*, Case 1:19-cv-02184, Complaint for Civil Penalties, Injunction, and Other Relief (Filed 07/24/19). See also Farooqi et al. (2020).

If the firms have deep pockets, and must pay in full for the harms they cause, then platform liability is unwarranted. Holding the firms fully liable deters the harmful firms. Platform liability is socially desirable when the firms are *judgment proof* and immune from liability.[13] First, if the platform is held liable, the platform will raise the interaction price for the firms to reflect the platform's future liability costs. If the harmful firms are "marginal" (i.e., the harmful firms have a lower willingness to pay than the safe firms) then the higher interaction price deters the harmful firms from joining the platform. Second, if the harmful firms are "inframarginal" and undeterrable, the platform will invest resources to detect and block the harmful firms from interacting with users. Interestingly, the optimal level of platform liability may be less than the residual harm (i.e. total harm minus firm liability), as the platform does not fully internalize firms' surplus and large liability could lead to excessive auditing.

Next, we extend the baseline model to settings where interactions are market transactions that require the users' consent. Relevant settings include retail platforms like eBay and Amazon where participants enjoy cross-side benefits from the sale of goods and services. As in the baseline model there are two types of sellers, harmful and safe. The harmful sellers have lower production costs but cause harms more frequently. In these settings, users have the option, but not the obligation, to interact with the firms. A user's willingness to pay for a product will depend on their expectations about product risks. The risk of harmful products depresses the price that consumers are willing to pay and, by extension, depresses the revenues that the platform can generate.

In general, the socially-optimal platform liability is *lower* for retail platforms than in the baseline model (e.g., for social media platforms). The right level of liability hinges on several factors, including the transparency of the platform's enforcement activities and the degree to which users internalize the social harm. If users observe the platform's choice of auditing effort and fully internalize the harms, then platform liability is unnecessary and may in fact reduce social welfare. However, if users do not observe the platform's efforts directly, or do not fully internalize the future harms, then platform liability plays a valuable role. Interestingly, platform liability and firm liability may be *complements* in the retail setting. In the baseline model, platform liability and firm liability are *substitutes*.

[13] Shavell (1986) provides the first rigorous treatment of the judgment proof problem. Injurers with limited assets tend to engage in risky activities too frequently and take too little care.

1.1 Related Literature

Our paper is related to the law-and-economics literature on products liability. Products liability may be socially desirable if consumers misperceive product risks (Spence, 1977; Epple and Raviv, 1978; Polinsky and Rogerson, 1983) or if consumers are not able to observe product safety at the time of purchase (Simon, 1981; Daughety and Reinganum, 1995).[14] Building on Spence (1975), Hua and Spier (2020) emphasize the particular importance of firm liability when consumers are heterogeneous so the marginal consumer's preferences are not representative of the average consumer.

Our paper is also related to the literature about *extending* liability to parties who are not directly responsible for the victim's harms. Hay and Spier (2005) examine whether manufacturers should be held liable if a consumer, while using the product, harms somebody else. If consumers are judgment proof, then extending liability to the manufacturer can help the market to internalize the harms. Pitchford (1995) explores the desirability of extending liability to an injurer's lenders and Dari Mattiacci and Parisi (2003) consider vicarious liability where liability is extended to the injurer's employer.[15] Arlen and MacLeod (2005a) show that holding managed care organizations liable for medical malpractice by their physicians can raise the physicians' incentives to take care. Our model investigates the design of platform liability when the platform can audit and block harmful participants.

There is a vast literature on multi-sided platforms. The early studies (e.g., Caillaud and Jullien, 2003; Rochet and Tirole, 2003, 2006; Amstrong, 2006; and Weyl, 2010) have identified how cross-side externalities affect platform pricing schemes and users' participation incentives.[16] Some recent studies pay attention to non-pricing strategies, including seller exclusion (Hagiu, 2009), information management (Jullien and Pavan, 2019; Choi and Mukherjee, 2020), and control right allocation (Hagiu and Wright, 2015, 2018). Teh (2022) makes the fundamental point that retail platforms cannot be trusted to act in the public interest when designing access rules for outside sellers. In particular, if a platform imposes a minimum quality standard on sellers, consumers' search costs

[14] See also Daughety and Reinganum (1995, 2006, 2008a and b), Arlen and Macleod (2003), Wickelgren (2006), Chen and Hua (2012, 2017), Choi and Spier (2014).

[15] See also Boyer and Laffont (1997) and Che and Spier (2008) on lender liability, Kraakman (1986) on gatekeepers, Hamdani (2002) on internet service providers, Hamdani (2003) on accountants and lawyers, Van Loo (2020a) on big technology, and Grimmelmann and Zhang (2023) on content moderation.

[16] See also Dukes and Gal-Or (2003), Hagiu (2006), Armstrong and Wright (2007), Nocke et al. (2007), Galeotti and Moraga-Gonzalez (2009), Hagiu (2009), White and Weyl (2010), Gomes (2014), Belleflamme and Peitz (2019), Karle et al. (2020), Tan and Zhou (2021).

become lower, which intensifies seller competition. The platform may host either too many sellers or too few, depending on the fee structure. Choi and Jeon (2023) show that ad-funded platforms will tend to distort their advertising policies to favor either consumers or advertisers, to the detriment of social welfare. Their analysis provides a rationale for public policies that constrain platforms' market power. Our study complements this literature by considering the platform's incentive to detect, block, and deter bad actors and the instrumental role that platform liability plays in making platforms safer.

There is a small but growing literature on platform liability. The policy papers by Buiten et al. (2020) and Lefouili and Madio (2022) discuss informally whether platforms should bear liability for harms caused by participants. A few working papers study copyright infringement where the victims are the rights holders. De Chiara et al. (2021) examine hosting platforms' incentives to filter infringing materials when the rights holders are not platform participants. They show that strict liability can be desirable, but do not consider pricing mechanisms. Jeon et al. (2022) consider negligence-based liability when the rights holders are themselves platform participants and compete with infringing firms. Removing infringing products makes the platform less competitive but stimulates innovation by the rights holders. If the innovation effect is sufficiently strong, platform liability harms consumers.

A few recent papers explore retail settings with firm moral hazard where sellers invest to reduce product defects. In Zennyo (2023), holding sellers strictly liable for consumer harm solves the firms' moral hazard problem but exacerbates a double-marginalization problem. Shifting liability away from the firms and towards the platform can reduce the monopoly deadweight loss. In Yasui (2022), firms have long-run incentives and platform liability can interfere with the reputation mechanism. If the platform is held strictly liable for consumer harm, then a consumer's willingness to pay is less sensitive to their perceptions of product safety. In their frameworks, platform liability often reduces product safety. By contrast, we show that that platform liability raises product safety. When held liable for user harms, the platform has an incentive to raise the interaction price to deter bad actors and invest resources to detect and block them.

Our paper is organized as follows. Section 2 presents the baseline model where firm-user interactions do not require the users' explicit consent. Section 3 generalizes the baseline model to retail settings where users have the option but not the obligation to transact with firms. Section 4 discusses several extensions, including alternative pricing structures, false positives, litigation costs, and platform competition. Section 5 offers

concluding thoughts. Proofs are in the appendix.

2 The Baseline Model

Consider a two-sided platform (P) with two kinds of participants, firms (S) and users. Firms and users are small, have outside options of zero, and the mass of each is normalized to unity.

The platform provides two goods. First, the platform provides a quasi-public good that gives each user a private benefit $v > 0$, which we assume is the same for all users.[17] Second, the platform provides opportunities for the firms and the users to interact. The benefits and costs of these interactions depend on the firms' type, $i \in \{b, g\}$, where λ is the mass of type b and $1 - \lambda$ is the mass of type g in the firm population.[18] The firms privately observe their types. The type-b firms have higher interaction benefits, $\alpha_b > \alpha_g$, but impose higher losses on users, $\theta_b d > \theta_g d$ where $\theta_i \in [0, 1]$ is the probability of harm and $d > 0$ is the level of harm per firm-user interaction. We assume in this section that the platform users themselves suffer harm, but the analysis would be the same if the victims include bystanders or society at large.[19]

The platform charges the firms a price p per interaction and allows users to join the platform for free.[20] Importantly, in this section, firm-user interactions do not require the users' consent (as on social platforms). By joining the platform, the users agree to allow the firms to access their data, show advertisements, and the like. This is, in effect, the price that the users must pay to use the platform and enjoy the quasi-public good. We will relax the assumption in Section 3, where the users who join the platform can avoid firm-user interactions (e.g., by declining to purchase in a retail setting).

The platform has the capability to detect and block the type-b firms. We will refer to the platform's efforts to detect type-b firms as auditing. Specifically, by spending effort

[17] Online Appendix B5 in our working paper (Hua and Spier, 2023) considers the more general setting with heterogeneous users where some join the platform and others do not. Platform liability can still be beneficial and has the added benefit of stimulating user participation.

[18] λ is exogenous, so there is firm adverse selection. Online Appendix B6 presents a moral hazard extension where λ is endogenous.

[19] For example, U.S. Prosecutors allege that Russian entities used paid advertising on social media sites to interfere with the 2016 U.S. presidential elections. See *United States v. Internet Research Agency, et al.* (D.D.C. Feb. 16, 2018).

[20] This pricing strategy can be very profitable for the platform in strategic environments with strong network effects and our assumption is also aligned with other papers in the platform literature (Hagiu and Wright, 2015, 2018; and Karle et al., 2020). Armstrong and Wright (2007), Choi and Jeon (2023), and Gans (2022) justify non-negative prices on adverse selection and moral hazard grounds.

$e \in [0, 1)$ per firm, the platform can detect type-b firms with probability e and block them from interacting with users (e.g. removing their advertisements or making the content invisible to users).[21] The cost of effort $c(e)$ increases in e and satisfies $c(0) = 0, c''(e) > 0, c'(0) = 0$, and $c'(e) \to \infty$ as $e \to 1$. We assume that the platform's effort e is publicly observed.

Suppose that both types of firm seek to join the platform. Given audit intensity e, the number of firms that remain on the platform is $\lambda(1 - e) + (1 - \lambda)$. Since there is a unit mass of consumers, this is also the number of firm-user interactions. This may be interpreted as the volume of (infinitesimally small) interactions per consumer, assuming that each retained firm interacts with each and every consumer. Alternatively, one may interpret $\lambda(1 - e) + (1 - \lambda)$ as the probability of an exclusive match between a user and a randomly selected firm.

The platform operates in a legal environment where harmed users may sue the platform and the firms for monetary damages. If a user suffers harm d, the court orders the firm and the platform to pay damages w_s and w_p, respectively, to the user. We will assume that $w_s, w_p \geq 0$ and $w = w_s + w_p \leq d$ so the total damage award does not exceed the harm suffered by the user.[22] In practice, third-party vendors are often liquidity-constrained or "judgment proof" and cannot be held fully accountable for the harm that they cause. Thus, firm liability w_s may be limited. For simplicity, there are no litigation costs or other transaction costs associated with using the court system.

If the type-b firms seek to join the platform and the platform takes audit intensity e, each retained type-i firm's profit is $\alpha_i - \theta_i w_s - p$, the platform's profit is

$$\Pi(e) = (1 - e)\lambda(p - \theta_b w_p) + (1 - \lambda)(p - \theta_g w_p) - c(e), \tag{1}$$

and a user's expected surplus is $v - [(1 - e)\lambda\theta_b + (1 - \lambda)\theta_g](d - w_s - w_p)$. The user's surplus from platform participation reflects the expected uncompensated harm.

[21] We rule out the possibility for the platform to charge a higher price or assign other penalties to the detected type-b firms. Such price discrimination would allow the platform to extract more surplus and therefore raise its incentive to keep type-b firms. In this case, higher platform liability is needed to restore the platform's incentive to deter harmful firms.

[22] Our main results remain valid if punitive damage awards ($w > d$) are feasible but not too large. If the total damage award is very large, the platform would not be active.

In the following analysis, we assume

$$\begin{aligned} A0 &: v - [\lambda\theta_b + (1-\lambda)\theta_g]d > 0; \\ A1 &: \alpha_g - \theta_g d > 0 > \alpha_b - \theta_b d; \\ A2 &: \alpha_g - (\lambda\theta_b + (1-\lambda)\theta_g)d > 0. \end{aligned}$$

A0 implies that the users' benefit from the quasi-public good is sufficiently high that the users would join the platform even if the type-b firms join the platform and there is no liability. A1 implies that it is socially efficient (inefficient) for the type-g (type-b) firms to join the platform. A2 guarantees that the platform always gets non-negative profits and implies that the net social benefit from interactions is positive even if both types of firms join the platform and interact with users. These assumptions are not essential for the main insights, but simplify the analysis.

The timing of the game is as follows.

1. Platform creates the quasi-public good for users, commits to effort level $e \in [0, 1)$, and sets interaction price p for the firms. Effort e and price p are publicly observed.

2. Firms privately learn their types $i \in \{b, g\}$ and firms and users decide whether to join the platform.

3. Platform audits and blocks any detected type-b firms.

4. Firms interact with users and interaction benefit α_i and harm $\theta_i d$ are realized.

5. Harmed users sue for monetary damages and receive compensation w_s and w_p from the responsible firm and platform, respectively.

The equilibrium concept is perfect Bayesian Nash equilibrium. Our social welfare concept is the aggregate value captured by all players: the platform, the firms (both type-b and type-g), and the users. The platform, firms, and users are weighed equally in the social welfare function.

We now present two social welfare benchmarks. First, in an ideal world, the type-b firms are fully deterred. The type-b firms do not even attempt to join the platform or interact with users. Auditing is unnecessary and social welfare is:

$$v + (1-\lambda)(\alpha_g - \theta_g d). \tag{2}$$

Next, suppose that the type-b firms are undeterred and seek to join the platform. This is less than ideal, as costly auditing is now necessary to detect and block the type-b firms. Social welfare is:

$$S(e) = v + \lambda(1-e)(\alpha_b - \theta_b d) + (1-\lambda)(\alpha_g - \theta_g d) - c(e). \tag{3}$$

The socially-optimal auditing effort $e^{**} > 0$ satisfies

$$-\lambda(\alpha_b - \theta_b d) - c'(e^{**}) = 0. \tag{4}$$

At the optimum, the marginal cost of auditing, $c'(e^{**})$, equals the marginal benefit of blocking type-b firms from interacting with users, $-\lambda(\alpha_b - \theta_b d)$. Note that $e^{**} \in (0,1)$ so some type-b firms remain on the platform.

2.1 Equilibrium Analysis

In this subsection, we characterize the platform's pricing and auditing strategies, p and e, given the assignment of liability, w_s and w_p. Assumption A0 implies that users always join the platform. A type-i firm will seek to join the platform when their expected surplus per interaction is non-negative,

$$\alpha_i - \theta_i w_s - p \geq 0, \tag{5}$$

where α_i is the firm's interaction benefit, $\theta_i w_s$ is the firm's expected liability, and p is the interaction price. Depending on the level of firm liability, w_s, type-b firms may have higher or lower surplus than type-g firms. The two types have the same surplus when

$$w_s = \widehat{w} = \frac{\alpha_b - \alpha_g}{\theta_b - \theta_g} < d. \tag{6}$$

The threshold $\widehat{w}$ defined in (6) is critical for understanding the impact of platform liability on the interaction price and audit intensity. When the firms are sufficiently judgment-proof, $w_s < \widehat{w}$, then the type-g firms are "marginal." If the type-g firms are indifferent about joining the platform then the type-b firms strictly prefer to join. When $w_s = \widehat{w}$, then the two types have the same surplus. If the type-g firms join the platform, type-b firms would join too. Auditing is necessary to detect and block the type-b firms.

When the firms are only moderately judgment proof, $w_s > \widehat{w}$, then the type-b firms

are marginal. If the type-b firms are indifferent about joining the platform then the type-g firms strictly prefer to join. In this setting, the platform has the ability, but may not have the incentive, to deter the type-b firms from joining by raising the interaction price p.

To summarize, the platform has two possible mechanisms to reduce the harm to users: the price per interaction p and the audit intensity e. In principle, the pricing mechanism is privately and socially more efficient than the costly auditing mechanism. However, the pricing mechanism is infeasible when the firm's liability is below a threshold, $w_s \leq \widehat{w}$.[23]

We now characterize the equilibrium for $w_s \leq \widehat{w}$ and $w_s > \widehat{w}$.

Case 1: $w_s \leq \widehat{w}$. Suppose that firm liability is below the threshold, so type-g firms are marginal. We show in the appendix that the platform will set the interaction price to extract the type-g firms' surplus,

$$p^* = \alpha_g - \theta_g w_s. \tag{7}$$

The type-b firms seek to join the platform. Using the definition of $\widehat{w}$ in (6), the type-b firms' surplus per interaction is $\alpha_b - \theta_b w_s - p^* = (\theta_b - \theta_g)(\widehat{w} - w_s) \geq 0$. As firm liability w_s grows, the type-b firms' surplus falls.

We now explore the platform's incentive to audit and block the type-b firms. A necessary and sufficient condition for the firm to audit, $e^* > 0$, is that the platform's profit associated with each retained type-b firm is negative, $p^* - \theta_b w_p < 0$. Using the formula for $\widehat{w}$ in (6) and p^* in (7), and letting $w = w_s + w_p$ be the joint liability of the firm and platform, $e^* > 0$ if and only if

$$(\alpha_b - \theta_b d) + \theta_b(d - w) - (\theta_b - \theta_g)(\widehat{w} - w_s) < 0. \tag{8}$$

The first term on the left-hand side of (8) is the social loss associated with each retained type-b firm and the second term is the uncompensated harm to the users. The sum of these two terms, $\alpha_b - \theta_b w$, is the joint platform-firm surplus associated with each retained type-b firm. The third term in (8) is the surplus captured by the type-b firm.

Next, we explore how the private and social incentives for auditing diverge when $e^* > 0$. Using the definition of $S(e)$ in (3), $\widehat{w}$ in (6), and p^* in (7), the platform's profit

[23] Our analysis can be extended to consider continuous types of firms. Suppose that the probability of harm θ follows a distribution on $[0, 1]$ and a type-θ firm's interaction benefit is $\alpha(\theta) = \beta_0 + \beta\theta$, where $\beta_0 + \beta < d$. The social value $\beta_0 + (\beta - d)\theta$ decreases in θ while firm surplus $\beta_0 + (\beta - w_s)\theta$ increases in θ if $w_s \leq \beta$ and decreases in θ if otherwise. If $w_s \leq \beta$, the pricing mechanism cannot deter the more harmful firms and the auditing effort (to detect and block firms with higher θ) is valuable.

function in (1) may be rewritten as:

$$\Pi(e) = S(e) - (1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) + [(1-e)\lambda\theta_b + (1-\lambda)\theta_g](d-w) - v. \quad (9)$$

The platform's auditing effort $e^* > 0$ satisfies

$$\Pi'(e^*) = S'(e^*) + \lambda(\theta_b - \theta_g)(\widehat{w} - w_s) - \lambda\theta_b(d-w) = 0. \quad (10)$$

The first-order condition in (10) underscores that the platform's private incentive to invest in auditing may be either socially excessive or socially insufficient. First, when the platform increases e and blocks type-b firms, the blocked firms lose their surplus, $\lambda(\theta_H - \theta_L)(\widehat{w} - w_s)$. Auditing imposes a *negative externality* on the type-b firms. Second, when the platform blocks type-b firms, the user-bystanders' uncompensated loss is reduced by $\lambda\theta_b(d-w)$. Auditing confers a *positive externality* on the user-bystanders. Because there are two offsetting effects, the platform's effort, e^*, may be larger than or smaller than the socially optimal level, e^{**}, as shown in the following lemma.

Lemma 1. *Suppose $w_s \leq \widehat{w}$. The platform sets $p^* = \alpha_g - \theta_g w_s$ and attracts the type-b firms. Let $r_b(w_s) \equiv (\theta_b - \theta_g)(\widehat{w} - w_s)$ denote the type-b firms' surplus per interaction.*

1. *If $\alpha_b - \theta_b w \geq r_b(w_s)$ then the platform does not audit, $e^* = 0 < e^{**}$.*
2. *If $\alpha_b - \theta_b w < r_b(w_s)$ then $e^* > 0$. The platform's auditing efforts e^* increase with firm and platform liability, $de^*/dw_s > 0$ and $de^*/dw_p > 0$.*
   - *(a) If $\theta_b(d-w) > r_b(w_s)$ then $0 < e^* < e^{**}$.*
   - *(b) If $\theta_b(d-w) = r_b(w_s)$ then $0 < e^* = e^{**}$.*
   - *(c) If $\theta_b(d-w) < r_b(w_s)$ then $0 < e^{**} < e^*$.*

To summarize, the platform's incentive to audit and block the type-b firms is stronger when w_p and w_s are larger. This private incentive is socially insufficient when the joint liability for the platform and firms is small (as in case 2(a)) but socially excessive if the joint liability is large (as in case 2(c)).

These insights are consistent with the recent literature on platform governance. In a model of an ad-funded platform with homogeneous advertisers, Choi and Jeon (2023) show that when the membership fee is fixed at zero then the platform's policies will be biased in favor of the advertisers. In a retail setting where the platform controls the number of

homogeneous third-party vendors, Teh (2022) shows that a platform may set a socially excessive (or insufficient) quality standard if it cares more (or less) about transaction volumes but less (or more) about firm profits. This distortion can lead to too few or too many firms on the platform. In both papers, the platform's private governance choice can be systematically biased against (or in favor of) users.

Case 2: $w_s > \widehat{w}$. Now suppose that firm liability is above the threshold, so the type-b firms are marginal. The platform's profit-maximizing strategy is to either charge $p = \alpha_g - \theta_g w_s$ and deter the b-types from joining the platform or charge $p = \alpha_b - \theta_b w_s < \alpha_g - \theta_g w_s$ and attract both types. Notably, if the platform chooses the latter strategy, then it will not invest in auditing, $e^* = 0$.

The platform will charge the low price and attract the type-b firms if and only if

$$\alpha_b - \theta_b w_s - [\lambda\theta_b + (1-\lambda)\theta_g]w_p > (1-\lambda)(\alpha_g - \theta_g w_s - \theta_g w_p).$$

Using the definition of $\widehat{w}$ in equation (6) this condition becomes:

$$\lambda(\alpha_b - \theta_b w) > (1-\lambda)(\theta_b - \theta_g)(w_s - \widehat{w}). \tag{11}$$

The left-hand side is the joint platform-firm surplus of attracting the type-b firms on the platform: the fraction λ of type-b firms multiplied by the interaction benefit α_b minus the joint liability $\theta_b(w_s + w_p)$. The expression on the right-hand side is the surplus captured by the inframarginal type-g firms. The platform has incentives to deter the type-b firms if and only if the joint benefit of attracting the type-b firms is less than the type-g firms' surplus. This is summarized in the following Lemma.

Lemma 2. *Suppose $w_s > \widehat{w}$. Let $r_g(w_s) \equiv (\theta_b - \theta_g)(w_s - \widehat{w})$ denote the type-g firm's surplus per interaction.*

1. *If $\lambda(\alpha_b - \theta_b w) > (1-\lambda) r_g(w_s)$ then the platform sets $p^* = \alpha_b - \theta_b w_s$, attracts the type-b firms, and does not audit, $e^* = 0 < e^{**}$.*
2. *If $\lambda(\alpha_b - \theta_b w) \leq (1-\lambda) r_g(w_s)$ then the platform sets $p^* = \alpha_g - \theta_g w_s$ and deters the type-b firms.*

Lemma 2 implies that the platform's private incentive to deter type-b firms is insufficient when the joint liability for the platform and firms is not large. This possible distortion is consistent with the observations in the literature on platform governance.

However, the previous studies (Teh, 2022; Choi and Jeon, 2023) focus on homogeneous firms and do not consider the possibility of using the pricing mechanism to deter firms.

2.2 Platform Liability

This subsection explores the social desirability and optimal design of platform liability when interactions do not require users' consent, taking the level of firm liability w_s as fixed. Note that platform liability is designed to supplement, not replace, firm liability. We begin by presenting a benchmark where the platform is not liable for the harm.

Proposition 1. *(Firm-Only Liability.) Suppose that the platform is not liable for harm to users, $w_p = 0$. There exists a unique threshold $\widetilde{w} = \widetilde{w}(\lambda) \in [\widehat{w}, d)$, where $\widetilde{w}(\lambda)$ weakly increases in the number of type-b firms, λ.*

1. *If $w_s < \widetilde{w}$ then the platform attracts the type-b firms and does not invest in auditing, $e^* = 0 < e^{**}$.*

2. *If $w_s \geq \widetilde{w}$ then the platform sets $p^* = \alpha_g - \theta_g w_s$ and deters the type-b firms.*

Proposition 1 establishes that platform liability is unnecessary when the firms themselves are held sufficiently liable for harm to the users (case 2 in Proposition 1). In this case, the joint platform-firm surplus of including the type-b firms is low, so the platform has incentives to deter them by charging a high price. However, when the firms are more judgment proof and the platform faces no liability (case 1 in Proposition 1), the private and social incentives diverge. The platform attracts the type-b firms and does not invest in costly auditing. In such cases, platform liability can be socially desirable, as shown in the next proposition.

Proposition 2. *(Optimal Platform Liability.) The socially-optimal platform liability, w_p^*, is as follows:*

1. *If $w_s \leq \widehat{w}$ then $w_p^* = d - w_s - \left(1 - \frac{\theta_g}{\theta_b}\right)(\widehat{w} - w_s) \in (0, d - w_s]$. The platform attracts the type-b firms. The platform's auditing incentives are socially optimal, $e^* = e^{**}$.*

2. *If $w_s \in (\widehat{w}, \widetilde{w})$ then there exists a threshold $\underline{w}_p > 0$ such that, under any $w_p^* \in [\underline{w}_p, d - w_s]$, the platform deters the type-b firms.*

3. *If $w_s \geq \widetilde{w}$ then platform liability is unnecessary. Under any $w_p^* \in [0, d - w_s]$, the platform deters the type-b firms.*

Proposition 2 describes how platform liability can be designed to increase social welfare. In case 1, firm liability is below the threshold ($w_s \leq \widehat{w}$) and the type-g firms are marginal. From Proposition 1 we know that firm-only liability fails to deter the type-b firms and gives the platform no incentive to audit and block the type-b firms. Imposing liability on the platform motivates the platform to take auditing effort. If $w_s < \widehat{w}$ and the platform was held responsible for the full residual harm, $w_p = d - w_s$, then the platform would *overinvest* in auditing. Therefore the auditing effort is efficient when the platform bears some but not all of the residual damage, $w_p^* \in (0, d - w_s)$. If $w_s = \widehat{w}$, then the auditing effort is efficient when the platform bears full residual liability.

In case 2, the firms' liability is in an intermediate range and the type-b firms are marginal. According to Proposition 1, without platform liability, the platform would attract the type-b firms since the joint platform-firm benefit of including the type-b firms is larger than the type-g firms' surplus. Since the firms' surplus is independent of w_p while the joint benefit of keeping the type-b firms decreases in w_p, the social planner can motivate the platform to raise the price and thus deter the type-b firms by imposing residual liability on the platform, $w_p^* = d - w_s$.

Finally, in case 3, platform liability is unnecessary when firm liability is sufficiently high. As in Proposition 1, the deterrence outcome is obtained without platform liability.

Proposition 2 also implies that, if $w_s \leq \widehat{w}$, the optimal platform liability decreases in w_s. From the social planner's perspective, platform liability and firm liability are (imperfect) substitutes. To see the intuition, note that, in case 1 of Proposition 2, w_p^* satisfies

$$(\theta_b - \theta_g)(\widehat{w} - w_s) = \theta_b(d - w_s - w_p^*). \tag{12}$$

The left-hand side of (12) is each retained type-b firm's surplus, while the right-hand side is the uncompensated losses that are caused by the type-b firm. When firm liability w_s rises, both sides fall. However, the drop in the firms' surplus on the left-hand side is smaller. Holding w_p fixed, the platform would invest *too much* in auditing. To prevent excessive auditing, platform liability w_p must fall.

Corollary 1. *Suppose $w_s \leq \widehat{w}$. Firm liability and platform liability are substitutes: If firm liability w_s increases then the optimal platform liability w_p^* decreases.*

2.3 Discussion

This section investigated the need for platform liability when users are effectively bystanders and interactions do not require the users' consent. If firms have deep pockets and can compensate the users for the harm, then platform liability is unwarranted. If firms are judgment proof or can evade liability in other ways, then holding the platform liable for some or all of the residual harm motivates the platform to raise the interaction price or invest resources in auditing, which deters or blocks risky firms.

Some of the assumptions made in this section were stronger than necessary. First, we assumed that users fully understood and internalized the potential risks. Users were willing to join the platform despite the potential harm. Second, we assumed that users could observe the platform's auditing effort e, so the probability of harm was common knowledge. Both of these assumptions could be easily relaxed. Assumption A0 guarantees that users join the platform despite the risk of harm.

By contrast, the assumption that user consent is not required for interactions was critical for our results. The next section relaxes this key assumption by giving users the power to block interactions with firms. Platform liability is arguably less useful when users can look after themselves. However, as we will see, platform liability plays an instrumental role in improving safety when (1) users do not internalize all the harm (e.g., they underestimate risks or some of the harm is externalized to bystanders or society) and/or (2) users cannot observe the platform's auditing effort (i.e., platform moral hazard).

3 Retail Platforms

We now assume that interactions between firms and users are *market transactions* that require the users' consent. Users who join the platform have the option, but not the obligation, to purchase products. This extension captures retail platforms such as Amazon where the firms are third-party sellers of a product or service and the users are consumers. The third-party sellers, especially those without existing reputations, may have incentives to sell products that have low costs but may harm consumers. This problem is particularly severe when the third-party vendors are judgment-proof, and cannot be held accountable for the injuries that their products cause.

The basic setup is isomorphic to the baseline model. As in the baseline model, there are two types of firm, b and g. The type-i firm produces a good or service which causes accidents with probability θ_i. The unsafe products are cheaper to produce and cause

harm more frequently, $\theta_b > \theta_g$. We normalize the type-b firm's production cost as 0 and the type-g firm's production cost as $k > 0$. A user-consumer's gross value from the good is α_0. Letting $\alpha_b = \alpha_0$ and $\alpha_g = \alpha_0 - k$, the social value from an interaction is $\alpha_i - \theta_i d$ (as in the baseline model). In stage 4, the firm-sellers are randomly matched with the user-consumers and propose price t. If a user accepts the price offer t, then the user pays t to the firm and the firm pays p to the platform.

We assume that a user considers fraction $\delta \in [0, 1]$ of the social harm d, when making their purchase decision. $\delta < 1$ is realistic in various settings of interest. First, some users (including children) put insufficient weight on the risks that will be borne by their future selves, i.e., negative internalities (Allcott and Sunstein, 2015; Johnen and Somogyi, 2024). They may be unaware of product risks, underestimate potential damages (Spence, 1977; Epple and Raviv, 1978; Polinsky and Rogerson, 1983), or exhibit present bias due to hyperbolic discounting (Laibson, 1997; O'Donoghue and Rabin, 1999). Second, dangerous products often harm bystanders as well as consumers (i.e., negative externalities). When a defective hoverboard battery causes a residential fire, the consumer's family, roommates, and neighbors are victims too. When a consumer purchases a counterfeit or pirated product instead of legal versions, the owner of the intellectual property may be harmed too.

We first explore the case where users observe the platform's choice of auditing effort e when they make their purchase decisions, and then consider the case where the platform's choice of auditing effort e is not observed by users.

3.1 Observable Effort

In this subsection, we assume that users observe the platform's auditing effort e. Although users cannot distinguish safer products from harmful ones, they can and do form correct beliefs about the proportion of harmful firms in the market. As we will see, if users fully internalize the social harm ($\delta = 1$) then platform liability has no effect on the platform's profit or on social welfare. However, if users do not internalize all of the social harm ($\delta < 1$) then platform liability can play an instrumental role in assuring public safety.

If the type-b firms seek to join the platform and the platform invests e in auditing, the conditional probability of harm per interaction is

$$E(\theta|e) = \frac{(1-e)\lambda\theta_b + (1-\lambda)\theta_g}{(1-e)\lambda + (1-\lambda)}, \tag{13}$$

which is a decreasing function of e. We let $\theta^{**} = E(\theta|e^{**})$ be the probability of harm when auditing is socially optimal ($e = e^{**}$) and let $\theta^0 = E(\theta|0) = \lambda\theta_b + (1-\lambda)\theta_g$ be the probability of harm when the platform does not audit ($e = 0$).

There is no separating equilibrium where the type-b and type-g firms charge different prices and have positive sales. If such a separating equilibrium existed, the firms charging the low price would have incentives to mimic the firms charging the high price. In any pooling equilibrium where both types of firm seek to join the platform and offer the same t, the type-b firm's surplus is $t-\theta_b w_s - p$ and the type-g firm's surplus is $t-(\theta_g w_s + k) - p$. The two types have equal surplus when $w_s = \widehat{w} = \frac{\alpha_b - \alpha_g}{\theta_b - \theta_g}$ as defined in (6) above.

In equilibrium, given the auditing effort e^r, the retail price t^r cannot be larger than the users' maximum willingness to pay, $\alpha_0 - \theta^r\delta(d-w)$, where $\theta^r = E(\theta|e^r)$. However, there can be multiple pooling equilibria. Any price $t \in (\alpha_0 - \theta_b\delta(d-w), \alpha_0 - \theta^r\delta(d-w)]$ can be supported if users hold the off-equilibrium belief that any firm charging a different price is the type-b firm. As shown in the appendix, the platform's profit is maximized in the equilibrium with

$$t^r = \alpha_0 - \theta^r\delta(d-w), \tag{14}$$

so the user's expected surplus is zero. No firm has an incentive to raise its price, as otherwise the users would not buy from the firm. In the following, we will focus on this equilibrium.

Case 1: $w_s \leq \widehat{w}$. Since the type-g firms are marginal, the platform sets p^r to extract rents from the type-g firms, $p^r = t^r - (\theta_g w_s + k)$. Using (14) and $\alpha_g = \alpha_0 - k$,

$$p^r = \alpha_g - \theta_g w_s - \theta^r\delta(d-w). \tag{15}$$

Comparing p^r to its counterpart p^* in equation (7) in the baseline model reveals an important difference: the interaction price paid by the firms reflects the user-consumers' expected uncompensated harm, $\theta^r\delta(d-w)$.

Substituting p^r from (15), $S(e)$ from (3), and $\widehat{w}$ from (6) into (1) gives

$$\begin{aligned}\Pi(e) = S(e) - v - (1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) \\ + [(1-e)\lambda\theta_b + (1-\lambda)\theta_g](1-\delta)(d-w).\end{aligned} \tag{16}$$

The platform's profits $\Pi(e)$ diverge from social welfare $S(e)$ for two reasons. First, the platform does not internalize each retained type-b firm's surplus, $(\theta_b - \theta_g)(\widehat{w} - w_s)$. Second,

for any $w < d$, the platform does not fully internalize the losses that are not anticipated or considered by the users, $(1-\delta)(d-w)$. Note that the platform does fully consider the losses that are anticipated by the users, that is, $\delta(d-w)$.

If the firm's equilibrium auditing effort is positive, then $e^r > 0$ satisfies

$$\Pi'(e^r) = S'(e^r) + \lambda(\theta_b - \theta_g)(\widehat{w} - w_s) - \lambda\theta_b(1-\delta)(d-w) = 0. \tag{17}$$

The platform's auditing incentive is socially insufficient (or excessive) if and only if the type-b firms' surplus is smaller (or larger) than the harm not anticipated by the users.

Let us compare the two extreme cases, $\delta = 0$ and $\delta = 1$. If $\delta = 0$ then users totally fail to internalize product risks when making purchase decisions. This may be because users are unaware of the risks, are completely myopic, or extremely impulsive and disregard to harms to their future selves (internalities). Alternatively, this may be because the harms are borne by people other than the users (externalities). If $\delta = 0$, then equation (17) is the same as equation (10) in the baseline model and all of our earlier results apply.

At the other extreme, if $\delta = 1$ then users are fully aware of the risks and internalize all of the social harm when making their purchase decisions. In this case, the third term in equation (17) is zero. This has two notable implications. First, platform liability has no effect on the platform's choice of auditing, e^r, or indeed on the platform's profit. Intuitively, the risk to users is fully neutralized by the price mechanism. Second, the platform's private incentive to audit exceeds the social incentive. This happens because the platform does not internalize the firms' surplus when choosing its effort.

Case 2: $w_s > \widehat{w}$. As in the baseline model, the platform will either set a high price and deter the marginal type-b firms or set a low price and accommodate them.

Suppose that the platform sets a high price and deters the marginal type-b firms. Since users observe the price and anticipate that type-b firms are deterred, the retail price reflects the type-g risks only, $t^r = \alpha_0 - \theta_g\delta(d-w)$. The platform charges the firms an interaction price to extract the type-g firm's surplus, $p^r = t^r - (\theta_g w_s + k)$ or $p^r = \alpha_g - \theta_g w_s - \theta_g\delta(d-w)$.

Suppose instead that the platform sets a low price, accommodates the type-b firms, and sets $e = 0$. The platform would have no incentive to audit and block the type-b firms in this case. This is by revealed preference, as the platform can easily deter the type-b firms by raising the price. Since users anticipate the average probability of harm $\theta^0 = E(\theta|0)$, the retail price is $t^r = \alpha_0 - \theta^0\delta(d-w)$. The platform charges the firms an

interaction price $p^r = t^r - \theta_b w_s$ or $p^r = \alpha_b - \theta_b w_s - \theta^0 \delta(d - w)$.

The platform will charge the low price and attract the type-b firms if

$$\alpha_b - \theta_b w_s - \theta^0 \delta(d - w) - \theta^0 w_p > (1 - \lambda)(\alpha_g - \theta_g w_s - \theta_g \delta(d - w) - \theta_g w_p).$$

Using the definition of $\widehat{w}$ in (6), this condition becomes:

$$\lambda[\alpha_b - \theta_b w - \theta_b \delta(d - w)] > (1 - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w}). \tag{18}$$

The left-hand side is the incremental revenue from attracting the type-b firms on the platform. Different from equation (11) in the baseline model, the joint surplus of the platform and the type-b firm includes a new term, $\theta_b \delta(d-w)$, which is the user's expected uncompensated harm from the type-b firm. The right-hand side is the surplus captured by the inframarginal type-g firms.

Consider the two extreme cases, $\delta = 0$ and $\delta = 1$. If users totally fail to internalize product risks, $\delta = 0$, the platform's incentives are compromised. As in the baseline model, the platform may have insufficient incentives to raise the price and deter the harmful firms. If users are fully aware of the risks and internalize all of the social harm, $\delta = 1$, platform liability is unnecessary. Even without liability, since $\alpha_b - \theta_b d < 0$, (18) would not hold, that is, the platform would charge a high price and deter the type-b firms.

The next proposition characterizes the socially-optimal platform liability rule.

Proposition 3. *(Retail Platform with Observable Effort.) Suppose users observe e. There exists $\widetilde{w}^r(\delta) \in [\widehat{w}, d)$, which is weakly decreasing in δ. The socially-optimal platform liability, w_p^r, satisfies:*

1. *If $w_s \leq \widehat{w}$ the platform attracts the type-b firms. There exists a threshold δ^r with $\delta^r \in (\frac{\theta_g}{\theta_b}, 1)$ if $w_s < \widehat{w}$ and $\delta^r = 1$ if $w_s = \widehat{w}$.*

   (a) *If $\delta \leq \delta^r$, $w_p^r = d - w_s - \left(\frac{\theta_b - \theta_g}{\theta_b - \delta\theta_b}\right)(\widehat{w} - w_s)$. The platform's auditing incentives are socially optimal, $e^r = e^{**}$.*

   (b) *If $\delta > \delta^r$, $w_p^r = 0$. The platform's auditing incentives are excessive, $e^r > e^{**}$.*

2. *If $w_s \in (\widehat{w}, \widetilde{w}^r(\delta))$ there exists $\underline{w}_p^r > 0$ such that, under any $w_p^r \in [\underline{w}_p^r, d - w_s]$, the platform deters the type-b firms.*

3. *If $w_s \geq \widetilde{w}^r(\delta)$ platform liability is unnecessary. Under any $w_p^r \in [0, d - w_s]$, the platform deters the type-b firms.*

Proposition 3 establishes that platform liability can be socially beneficial when δ is not too large, so users fail to internalize the true social harm. In case 2, holding the platform liable gets the platform to raise interaction price and deter the type-b firms. In cases 1(a), holding the platform partially responsible (when $w_s < \widehat{w}$) or fully responsible (when $w_s = \widehat{w}$) for the residual harm leads to the socially-optimal auditing effort.

When δ is large, so users internalize a large percentage of the social harm, platform liability is either unnecessary or socially inefficient. Notice that, when δ is large, case 2 does not exist. Then for any $w_s > \widehat{w}$, as shown in case 3, the platform has a strong incentive to charge a high price and deter the type-b firms to stimulate user demand. In this case, platform liability is unnecessary. In case 1(b), even absent platform liability, the platform's auditing incentive exceeds the social incentive. Platform liability would make matters even worse by exacerbating the over-investment problem.

These observations indicate that, when firm-user interactions require users' consent and firms are very judgment proof, courts should impose some of the residual liability on the platform if and only if a large fraction of the harm is on bystanders or if users bear the harm but tend to be myopic (e.g. children). The optimal level of residual liability imposed on the platform depends on how much money firms can pay. These conditions and information can be verified by courts in practice.

Recall that, in the baseline model where interactions do not require the users' consent, from the social planner's perspective, platform liability and firm liability are substitutes (see Corollary 1). By contrast, when interactions require the users' consent, Proposition 3 implies that, if $w_s \leq \widehat{w}$ and δ is not too large, the optimal platform liability can increase or decrease in w_s, that is, platform liability and firm liability can be complements or substitutes. To see the intuition, note that, in case 1(a) of Proposition 3, w_p^r satisfies

$$(\theta_b - \theta_g)(\widehat{w} - w_s) = (\theta_b - \delta\theta_b)(d - w_s - w_p^r). \tag{19}$$

The left-hand side of (19) is each retained type-b firm's surplus, while the right-hand side is the uncompensated losses that are caused by the type-b firm but not anticipated or considered by the users.

When firm liability w_s rises marginally, the firms' surplus decreases by $\theta_b - \theta_g$ while the users' unanticipated loss drops by $\theta_b - \delta\theta_b$. If δ is large (i.e., the users internalize a large fraction of the harm), the former effect dominates. In this case, holding w_p fixed, the platform would invest *too little* in auditing. Hence, platform liability w_p should be raised.

However, if δ is small, the drop in the firms' surplus on the left-hand side is smaller. In this case, holding w_p fixed, the platform would invest *too much* in auditing. To prevent excessive auditing, platform liability w_p must fall.

Corollary 2. *Suppose $w_s \leq \widehat{w}$ and $\delta \leq \delta^r \in (\frac{\theta_g}{\theta_b}, 1)$. If $\delta < \frac{\theta_g}{\theta_b}$ then firm and platform liability are substitutes: If firm liability w_s increases then the optimal platform liability w_p^r decreases. It $\delta \geq \frac{\theta_g}{\theta_b}$ then firm and platform liability are complements: If firm liability w_s increases then the optimal platform liability w_p^r increases.*

3.2 Unobservable Effort

In this subsection, we assume that users do not observe the platform's auditing effort but form the belief about the average safety, which is correct in equilibrium. In practice, the inner workings of platforms, including the technology and systems for improving safety for users, are often less than transparent. The Digital Markets Act in the European Union and the INFORM Consumers Act proposed in the US contain many disclosure requirements, which reflects lawmakers' concerns about the lack of transparency on platform safety and effort.[24] As we will see, the scope for platform liability to improve social welfare is larger when users do not observe the platform's choice of auditing effort.

To begin, note that if $w_s > \widehat{w}$, the analysis is exactly the same as in the previous subsection. In this case, the type-b firms are marginal and can be easily deterred through the price mechanism. The observability of platform effort is not pertinent in this case. We shall therefore focus on the case with $w_s \leq \widehat{w}$. In this case, the platform needs to exert effort to detect and block the type-b firms.

The characterization of the pooling equilibrium when $w_s \leq \widehat{w}$ is similar to the previous subsection but with some notable differences. Suppose that consumers believe that the platform's effort is e^{rn} and the corresponding probability of harm is $\theta^{rn} = E(\theta|e^{rn})$. The platform's profit function may be written as

$$\begin{aligned}\Pi(e) = S(e) - v - (1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) \\ + [(1-e)\lambda\theta_b + (1-\lambda)\theta_g](1-\delta)(d-w). \\ + [(1-e)\lambda(\theta_b - \theta^{rn}) + (1-\lambda)(\theta_g - \theta^{rn})]\delta(d-w). \quad (20)\end{aligned}$$

[24]The Senators proposing the Act opined that "Consumers often do not even know the identity of the businesses that sell them goods. Unfortunately, online marketplaces like Amazon are going to great lengths to keep it that way." See https://www.cassidy.senate.gov/newsroom/press-releases/icymi-cassidy-durbin-pen-op-ed-on-online-marketplace-transparency/

The first two lines are the same as equation (16). The platform does not consider each retained type-b firm's surplus, $(\theta_b - \theta_g)(\widehat{w} - w_s)$ or the uncompensated losses that are not anticipated by the users along the equilibrium path, $(1-\delta)(d-w)$. However, now that the users cannot observe e, the platform's off-the-equilibrium-path choice of auditing may diverge from the users' expectations. The expression in the third line of (20) is unique to the unobservable-effort setting, and represents the users' *unanticipated loss or gain* when the platform deviates and invests $e \neq e^{rn}$.

When choosing auditing effort, the platform maximizes the profits in (20) taking users' belief θ^{rn} as given. If the equilibrium auditing effort is positive, then $e^{rn} > 0$ satisfies

$$\Pi'(e^{rn}) = S'(e^{rn}) + \lambda(\theta_b - \theta_g)(\widehat{w} - w_s) - \lambda(\theta_b - \delta\theta^{rn})(d-w) = 0. \tag{21}$$

Comparing (21) to (17) reveals that the platform's auditing incentive is weaker than in the scenario with observable effort. Again, the platform's auditing incentive can be socially insufficient or excessive.

Recall that $\theta^{**} = E(\theta|e^{**})$ is the probability of harm when auditing is socially optimal. Using (21), we have the following result (the proof is similar to Proposition 3 and omitted).

Proposition 4. *(Retail Platform with Unobservable Effort.) Suppose users cannot observe e.*

1. *If $w_s \leq \widehat{w}$ the platform attracts the type-b firms. The socially-optimal platform liability is $w_p^{rn} = d - w_s - \left(\frac{\theta_b - \theta_g}{\theta_b - \delta\theta^{**}}\right)(\widehat{w} - w_s) \in (0, d - w_s]$. The platform's auditing incentives are socially optimal, $e^{rn} = e^{**}$.*

2. *If $w_s > \widehat{w}$ the socially-optimal platform liability and audit intensity are the same as in Proposition 3 (observable effort).*

According to Proposition 4, if $w_s \leq \widehat{w}$, the optimal platform liability is a decreasing function of δ.[25] If users are more sophisticated and internalize the potential risks from product use, then the need for platform liability is smaller. However, it is important to note that platform liability is socially beneficial even if the users fully understand and internalize the harms, $\delta = 1$. When the auditing effort is not directly observed, platform liability helps to keep the platform's incentives in check.

[25] As in Corollary 2 with observable effort, platform liability and firm liability are substitutes when δ is small and complements when δ is large.

3.3 Discussion

This section investigated the need for platform liability when the interactions require the users' consent (as on retail platforms). Comparing the liability rules in Propositions 3 and 4 to Proposition 2 in the baseline model underscores an important point. When interactions are market transactions that require users' consent, the platform has stronger incentives to assure higher product safety to stimulate demand. If users observe the platform's auditing effort AND fully internalize social harm, platform liability is unnecessary and can even reduce social welfare.

However, if users do not fully internalize social harm and/or do not observe the auditing effort, the platform's incentive for safety can still fall short. As in the baseline model, holding the platform liable for some or all of the residual harm motivates the platform to raise the interaction price or invest resources in auditing, which deters or blocks risky firms. Since the platform has incentives to stimulate demand, the socially-optimal platform liability is (weakly) smaller than that in the baseline model.

The next corollary ranks the socially-optimal liability rules in the baseline model, w_p^*, retail model with observable effort, w_p^r, and retail model with unobservable effort, w_p^{rn}.

Corollary 3. *Suppose $\delta \in (0, 1]$. The socially-optimal level of platform liability is (weakly) lower when interactions require user consent.*

1. *If $w_s < \widehat{w}$ then the socially-optimal platform liability is less-than-full and satisfies $0 \leq w_p^r < w_p^{rn} < w_p^* < d - w_s$.*

2. *If $w_s = \widehat{w}$ then the socially-optimal platform liability satisfies $w_p^r = w_p^{rn} = w_p^* = d - w_s$.*

3. *If $w_s > \widehat{w}$ then the lowest socially-optimal platform liability satisfies $0 \leq \underline{w}_p^r = \underline{w}_p^{rn} \leq \underline{w}_p < d - w_s$.*

Finally, although the analysis in this section considered homogeneous users with unit demand, the insights extend to heterogeneous users. The appendix provides an illustrative example where firms have the same surplus and the platform's auditing effort is observable. Without liability, the platform's incentives diverge from the social planner's. As before, the platform's incentives to detect and block harmful firms may be insufficient (if δ is small). But a new effect emerges. Holding user participation fixed, inframarginal users enjoy higher network benefits from firm interactions and therefore value auditing

less than marginal users. Since the platform focuses on marginal (rather than average) users, auditing can be socially excessive.[26] Moreover, absent liability, the level of user participation may be too low (due to the classic monopoly distortion) or too high (if δ is small). When δ is small, platform liability raises social welfare by enhancing the platform's auditing incentive but also by (possibly) preventing socially inefficient transactions.[27]

4 Further Extensions

Alternative Pricing Structure. Our analysis assumed that the platform monetized its activities through an interaction price paid by the firms. The results would be unaffected if only the firms that are retained by the platform pay a lump-sum membership fee. With additional instruments, such as a non-refundable application fee or bond, the platform's ability to deter risky firms would be enhanced.[28] However, when the firms are very judgment proof ($w_s \leq \widehat{w}$), Proposition 1 implies that, absent platform liability, the platform would prefer to accommodate the type-b firms. Platform liability raises the platform's incentive to deter the harmful firms by charging a non-refundable fee. Note that, even if the non-refundable fee deters the type-b firms, the platform still needs to commit to some auditing effort, as otherwise the type-b firms would deviate to join the platform.[29]

False Positives. Our analysis assumed that the platform did not erroneously block the type-g firms. If there are false positives, the platform has weaker incentives to invest in auditing than in the baseline model, since the platform loses revenue when it excludes the type-g firms. And the platform's incentives are even weaker relative to the social incentives, because the platform does not account for the positive externality that excluding the type-g firms confers on the platform users. It follows that the optimal platform liability is (weakly) larger when there are false positives, as shown by Online Appendix B2.

[26] This effect is similar to Spence (1975). There, a monopolist would overinvest (underinvest) in quality if the marginal consumer has a higher (lower) willingness to pay for quality than the average consumer.

[27] When users fully internalize the social harm, $\delta = 1$, platform liability has no effect on the outcome.

[28] Similarly, if firms incur opportunity costs when joining the platform, the platform's ability to deter risky firms is enhanced too. In this scenario, platform liability can still be valuable in motivating the platform to exert auditing effort.

[29] Suppose that the platform charges a non-refundable application fee y and sets the interaction price $p = 0$. The type-b firm's surplus is $(1-e)(\alpha_b - \theta_b w_s) - y$, while the type-$g$ firm's surplus is $\alpha_g - \theta_g w_s - y$. When $w_s \leq \widehat{w}$, to deter the type-b firms but attract the type-g firms, the platform sets $\alpha_g - \theta_g w_s \geq y > (1-e)(\alpha_b - \theta_b w_s)$, which implies $e > 0$. See Online Appendix B1.

Litigation Costs. The implications of litigation costs for the design of optimal platform liability is nuanced. On the one hand, when the type-g firms are marginal, litigation costs reduce the type-b firms' surplus and raise the users' uncompensated harm, as compared to the baseline model. These effects make the platform's auditing incentives even weaker relative to the social incentives. Moreover, litigation costs may discourage victims from bringing meritorious claims. Thus, if litigation is more costly, a higher level of platform liability may be necessary to encourage plaintiffs (and their lawyers) to sue and raise the platform's auditing incentives, as shown by Online Appendix B3. On the other hand, when the type-b firms are marginal, litigation costs raise the platform's incentives to deter these harmful firms, so that platform liability can be lower than in the baseline model. Furthermore, insofar as the costs of litigation exceed the benefits of improved platform incentives, a lower level of liability, or indeed the elimination of liability altogether, may be warranted.

Platform Competition. Our analysis can be extended to consider platform competition. Suppose that there are two competing platforms providing differentiated services. Users can participate on only one of the platforms (i.e., single-homing), while the firms can participate on both platforms (i.e., multi-homing). If the type-g firms are marginal, absent platform liability, the platform's auditing incentives diverge from the social incentives: On one hand, taking the allocation of users as fixed, the platforms underinvest in auditing because they do not consider the benefit of auditing for users; on the other hand, when competition is fierce, the platforms may have excessive incentives for auditing. Using a Hotelling model of platform competition, Online Appendix B4 shows that platform liability can mitigate these distortions, though the socially-optimal platform liability can be lower than that for a monopoly platform.[30]

5 Conclusion

Should platforms be held liable for the harms suffered by platform participants? This question is of practical as well as academic interest. Platforms in the United States and abroad face lax regulatory oversight from public enforcement agencies and are largely immune from private litigation. We explored the social desirability of platform liability

[30] If the type-b firms are marginal, then competition raises the platforms' incentives to deter the harmful firms by charging high prices, relative to the baseline model. In this case, platform liability is socially beneficial if the platforms are sufficiently differentiated but unnecessary if otherwise.

in a two-sided platform model where firms impose cross-side harms on users.

The model, while very simple, underscores several key insights. First, if firms have sufficiently deep pockets and are held fully accountable for the harms they cause, then platform liability is unwarranted. Holding the firms (and only the firms) liable deters the harmful firms from joining the platform and interacting with users. Second, if firms are judgment proof and they do not need users' consent for interactions (as on social platforms), imposing residual liability on the platform can be socially desirable. With platform liability, the platform has an incentive to (1) raise the interaction price to deter the harmful firms and (2) invest resources to detect and block the harmful firms from interacting with users. To prevent overinvestment in auditing, the residual liability assigned to the platform may be partial instead of full and depend on how much money firms can pay, which is observed by courts in practice.

Third, the justification for platform liability is weaker when interactions are market transactions that require users' consent (as on retail platforms). The transaction price paid by the users, and accordingly the interaction price paid by the firms to the platform, reflects the users' expected harm. This raises the platform's incentive to deter or block the harmful firms, even absent liability. However, if users do not internalize all the harm or they cannot observe the platform's auditing effort, then platform liability provides additional incentives for the platform to reduce the social harm. A lower level of platform liability may be appropriate in market settings.

Our basic argument for holding platforms liable is valid regardless of the accuracy of the platforms' screening technologies and moderation efforts. First, the lack of effort by some platforms could reflect the weak incentives provided by the legal, economic, and political systems. Platforms may even have "perverse incentives" to reduce their control of online activities, similar to the potential distortion caused by vicarious liability on organizations.[31] Second, our model shows that platform liability may be socially desirable even if auditing is very costly or *completely ineffective* at detecting bad actors. Although platforms would not engage in auditing in this case, liability would motivate platforms to use the price mechanism to deter bad actors.

Although internet platforms provided the motivation for this paper, our insights apply more broadly. Our analysis provides an economic rationale for holding traditional news-

[31] In the EU, safe-harbor provisions create "perverse incentives for platforms not to monitor online activity." See Lefouili and Madio (2022, p.322). Similarly, under vicarious liability, organizations may eschew control over agents (e.g., by using subcontractors) to avoid tort liability. See Arlen and MacLeod (2005b).

papers liable for harmful advertising content and for holding bricks-and-mortar retailers liable for the harm caused by defective products if advertisers and producers are judgment proof. However, we believe that the insights are particularly salient for online platforms. First, the harmful participants on platforms are frequently small and judgment proof with insufficient incentives to curtail their harmful activities. Second, the big tech giants have the data and technologies to detect and block participants that are more likely to harm others.

A natural question is whether private solutions and industry self-regulation can assure platform safety and render platform liability unnecessary. To attract and retain users, platforms have a long run interest in establishing feedback and recommendation systems that provide valuable information about sellers.[32] Furthermore, many platforms have established internal systems to resolve buyer-seller disputes (Hui et al., 2016) and to protect user data (Jullien et al., 2020; Perdikakis, 2024). In practice, however, feedback systems are often plagued by inaccuracies and biases fueled by the users' fear of retaliation and harassment when leaving negative reviews (Bolton et al., 2013), strategic review manipulation by sellers (Mayzlin et al., 2014), and the market for fake reviews. The systems to protect user data or resolve buyer-seller disputes may not be fully effective due to limited transparency.[33]

Another open question is whether platform regulation would be more or less effective than civil liability in reducing social harm. For example, there is active debate over whether platforms should be treated as common carriers (Rahman, 2018; Volokh, 2021) and subject to regulations to ensure public safety.[34] Specifically, given the diversity of platform business models and the rapidly changing surveillance technologies and market conditions, it would be difficult for regulators to set uniform safety standards.[35] Platforms, especially big tech platforms, have the relevant information to weigh the social costs

[32]Tadelis (2016) describes this as a central feature of digital platform business models, and offers a thoughtful discussion of the limits and biases in peer-to-peer feedback mechanisms.

[33]The long-run interest of platforms may also diverge from the interest of society. For example, there is some empirical evidence about platforms' incentives to steer users to their own products or high-margin products (Aguiar and Waldfogel, 2021; Farronato et al., 2023), consistent with the theoretical predictions (de Corniere and Taylor, 2019; Bourreau and Gaudim, 2022; Hagiu et al., 2022).

[34]Common carriers, including telephone companies, mail carriers, and transportation systems have a duty to serve the public and may not generally exclude users (15 U.S. Code §375). GA Code §46-9-132 (2020) states that "a common carrier of passengers is bound to exercise extraordinary diligence." See also California Civil Code §2100.

[35]eBay's 2022 Transparency Report states: "regulatory regimes or technology mandates that are 'one size fits all' can actually serve to limit the tools, resources and partnerships necessary to combat bad actors."https://static.ebayinc.com/assets/Uploads/Documents/eBay-2022-Global-Transparency-Report.pdf

and benefits. Liability has the advantage of harnessing the information and expertise of platforms, giving them a financial incentive to use their discretion for the greater good.

References

[1] Aguiar, Luis and Joel Waldfogel, "Platforms, Power, and Promotion: Evidence from Spotify Playlists," *Journal of Industrial Economics*, Vol. 69 (2021), pp. 653-691.

[2] Allcott, Hunt and Cass R. Sunstein, "Regulating Internalities," *Journal of Policy Analysis and Management*, Vol. 34 (2015), pp. 698-705.

[3] Arlen, Jennifer, and W. Bentley MacLeod, "Malpractice Liability for Physicians and Managed Care Organizations," *New York University Law Review*, Vol. 78 (2003), pp. 1929-2006.

[4] Arlen, Jennifer, and W. Bentley MacLeod, "Torts, Expertise, and Authority: Liability of Physicians and Managed Care Organization," *RAND Journal of Economics*, Vol. 36 (2005a), pp. 494-515.

[5] Arlen, Jennifer, and W. Bentley MacLeod, "Beyond Master-Servant: A Critique of Vicarious Liability," *Exploring Tort Law*, Edited by M. Stuart Madden, Cambridge University Press (2005b).

[6] Armstrong, Mark, "Competition in Two-sided Markets," *RAND Journal of Economics*, Vol. 37 (2006), pp. 668-691.

[7] Armstrong, Mark and Julian Wright, "Two-Sided Markets, Competitive Bottlenecks, and Exclusive Contracts," *Economic Theory*, Vol. 32 (2007), pp. 353-380.

[8] Belleflamme, Paul and Martin Peitz, "Managing Competition on a Two-Sided Platform," *Journal of Economics & Management Strategy*, Vol. 28 (2019), pp. 5-22.

[9] Bolton, Gary, Ben Greiner, and Axel Ockenfels, "Engineering Trust: Reciprocity in the Production of Reputation Information," *Management science* Vol. 59 (2013), pp. 265-285.

[10] Boyer, Marcel, and Jean-Jacques Laffont, "Environmental Risk and Bank Liability," *European Economic Review*, Vol. 41 (1997), pp. 1427-1459.

[11] Buiten, Miriam C., Alexandre de Streel, and Martin Peitz, "Rethinking Liability Rules for Online Hosting Platforms," *International Journal of Law and Information Technology*, Vol. 28 (2020), pp. 139-166.

[12] Bourreau, Marc, and Germain Gaudin, "Streaming Platform and Strategic Recommendation Bias," *Journal of Economics & Management Strategy* Vol. 31 (2022), pp. 25-47.

[13] Caillaud, Bernard, and Bruno Jullien, "Chicken & Egg: Competition among Intermediation Service Providers," *RAND Journal of Economics*, Vol. 34 (2003), pp. 309-328.

[14] Che, Yeon-Koo, and Kathryn E. Spier, "Strategic Judgment Proofing," *RAND Journal of Economics*, Vol. 39 (2008), pp. 926-948.

[15] Chen, Yongmin and Xinyu Hua, "Ex ante Investment, Ex post Remedies, and Product Liability," *International Economic Review*, Vol 53 (2012), pp. 845-866.

[16] Chen, Yongmin and Xinyu Hua, "Competition, Product Safety, and Product Liability," *Journal of Law, Economics, & Organization*, Vol. 33 (2017), pp. 237-267.

[17] Choi, Albert, and Kathryn E. Spier, "Should Consumers Be Permitted to Waive Products Liability? Product Safety, Private Contracts, and Adverse Selection," *Journal of Law, Economics, & Organization*, Vol. 30 (2014), pp. 734-766.

[18] Choi, Jay Pil, and Arijit Mukherjee, "Optimal Certification Policy, Entry, and Investment in the Presence of Public Signals," *RAND Journal of Economics*, Vol. 51 (2020), pp. 989-1013.

[19] Choi, Jay Pil, and Doh-Shin Jeon, "A Leverage Theory of Tying in Two-sided Markets with Nonnegative Price Constraints," *Americal Economic Journal: Microeconoimcs*, Vol. 13 (2021), pp. 283-337.

[20] Choi, Jay Pil, and Doh-Shin Jeon, "Platform Design Biases in Ad-funded Two-sided Markets," *RAND Journal of Economics*, Vol. 54 (2023), pp. 240-267.

[21] Daughety, Andrew F., and Jennifer F. Reinganum, "Product Safety: Liability, R&D, and Signaling," *American Economic Review,* Vol. 85 (1995), pp. 1187-1206.

[22] Daughety, Andrew F. and Jennifer F. Reinganum, "Market, Torts, and Social Inefficiency," *RAND Journal of Economics*, Vol. 37 (2006), pp. 300-323.

[23] Daughety, Andrew F., and Jennifer F. Reinganum, "Communicating Quality: a Unified Model of Disclosure and Signaling," *RAND Journal of Economics*, Vol. 39 (2008b), pp. 973-989.

[24] Daughety, Andrew F., and Jennifer F. Reinganum, "Imperfect Competition and Quality Signaling," *RAND Journal of Economics,* Vol. 39 (2008a), pp. 163-183.

[25] Dari Mattiacci, Giuseppe, and Francesco Parisi, "The Cost of Delegated Control: Vicarious Liability, Secondary Liability and Mandatory Insurance," *International Review of Law and Economics*, Vol. 23 (2003), pp. 453-475.

[26] De Chiara, Alessandro, Ester Manna, Antoni Rubi-Puig and Adrian Segura-Moreiras, "Efficient Copyright Filters for Online Hosting Platforms," (2021), working paper.

[27] De Corniere, Alexandre, and Greg Taylor, "A Model of Biased Intermediation," *The RAND Journal of Economics*, Vol. 50 (2019), pp. 854-882.

[28] Dukes, Anthony, and Esther Gal-Or, "Negotiations and Exclusivity Contracts for Advertising," *Management Science*, Vol. 22 (2003), pp. 222-245.

[29] Epple, Dennis, and Artur Raviv, "Product Safety: Liability Rules, Market Structure, and Imperfect Information," *American Economic Review*, Vol. 68 (1978), pp. 80-95.

[30] Farooqi, Shehroze, Maaz Musa, Zubair Shafiq, and Fareed Zaffar, "Canary-Trap: Detecting Data Misuse by Third-party Apps on Online Social Networks," arXiv:2006.15794v1 [cs.CY], (2020), https://arxiv.org/pdf/2006.15794.pdf.

[31] Farronato, Chiara, Andrey Fradkin, Alexander MacKay, "Self-Preferencing at Amazon: Evidence from Search Rankings," *AEA Papers and Proceedings*, Vol. 113 (2023), pp. 239-243.

[32] Galeotti, Andrea, and Jose Luis Moraga-Gonzalez, "Platform Intermediation in a Market for Differentiated Products," *European Economic Review*, Vol. 53 (2009), pp. 417-428.

[33] Gans, Joshua S., "The Specialness of Zero," *Journal of Law and Economics*, Vol. 65 (2022), pp. 157-176.

[34] Gomes, Renato, "Optimal Auction Design in Two-Sided Markets," *RAND Journal of Economics*, Vol. 45 (2014), pp. 248-272.

[35] Grimmelmann, James, and Pengfei Zhang, "An Economic Model of Intermediary Liability," *Berkeley Technology Law Journal*, Vol. 38 (2023).

[36] Hagiu, Andrei, "Pricing and Commitment by Two-Sided Platforms," *RAND Journal of Economics*, Vol. 37 (2006), pp. 720-737.

[37] Hagiu, Andrei, "Quantity vs. Quality and Exclusion by Two-Sided Platforms," (2009), working paper.

[38] Hagiu, Andrei, and Julian Wright, "Marketplace or Reseller?" *Management Science*, Vol. 61 (2015), pp. 184-203.

[39] Hagiu, Andrei, and Julian Wright, "Controlling vs. Enabling," *Management Science*, Vol. 65 (2018), pp. 577-595.

[40] Hagiu, Andrei, Tat-How Teh, and Julian Wright, "Should platforms be Allowed to Sell on Their Own Marketplaces?" *RAND Journal of Economics*, Vol. 53 (2022), pp. 297-327.

[41] Hamdani, Assaf, "Who is Liable for Cyberwrongs?" *Cornell Law Review*, Vol. 87 (2002), pp. 901-957.

[42] Hamdani, Assaf, "Gatekeeper Liability," *Southern California Law Review*, Vol. 77 (2003), pp. 53-122.

[43] Hay, Bruce, and Kathryn E. Spier, "Manufacturer Liability for Harms Caused by Consumers to Others," *American Economic Review*, Vol.95 (2005), pp. 1700-1711.

[44] Hua, Xinyu and Kathryn E. Spier, "Product Safety, Contracts, and Liability," *RAND Journal of Economics*, Vol. 51 (2020), pp. 233-259.

[45] Hua, Xinyu and Kathryn E. Spier, "Holding Platforms Liable," (2023) working paper, https://ssrn.com/abstract=3985066.

[46] Hui, Xiang, Maryam Saeedi, Zeqian Shen, and Neel Sundaresan, "Reputation and Regulations: Evidence From eBay." *Management Science*, Vol. 62,(2016). pp. 3604-3616.

[47] Jeon, Doh-Shin, Yassine Lefouili, and Leonardo Madio, "Platform Liability and Innovation," working paper, 2022.

[48] Johnen, Johannes, and Robert Somogyi, "Deceptive Features on Platforms," *The Economic Journal*, Vol. 134 (2024), pp. 2470-2493.

[49] Jullien, Bruno, and Alessandro Pavan, "Information Management and Pricing in Platform Markets," *Review of Economic Studies*, Vol. 86 (2019), pp. 1666-1703.

[50] Jullien, Bruno, Yassine Lefouili, Michael H. Riordan, "Privacy Protection, Security, and Consumer Retention," CEPR discussion paper, 2020.

[51] Kraakman, Reinier H., "Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy," *Journal of Law, Economics, & Organization*, Vol. 2 (1986), pp. 53-104.

[52] Karle, Heiko, Martin Peitz, and Markus Reisinger, "Segmentation versus Agglomeration: Competition between Platforms with Competitive Sellers," *Journal of Political Economy*, Vol. 128 (2020), pp. 2329-2374.

[53] Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, Vol. 112 (1997), pp. 443-478.

[54] Lefouili, Yassine, and Leonardo Madio, "The Economics of Platform Liability," *European Journal of Law and Economics*, Vol. 53 (2022), pp. 319-351.

[55] Mayzlin, Dina, Yaniv Dover, and Judith Chevalier, "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review* Vol. 104 (2014), pp. 2421-2455.

[56] Nocke, Volker, Martin Peitz, and Konrad Stahl, "Platform Ownership," *Journal of the European Economic Association*, Vol. 5 (2007), pp. 1130-1160.

[57] O'Donoghue, Ted and Matthew Rabin, "Doing It Now or Later," *American Economic Review*, Vol. 89 (1999), pp. 103-124.

[58] Perdikakis, Manos, "Reputation and the Provision of Data Security," working paper, 2024.

[59] Pitchford, Rohan, "How Liable Should a Lender Be? The Case of Judgment-Proof Firms and Environment Risk," *American Economic Review*, Vol. 85 (1995), pp. 1171-1186.

[60] Polinsky, A. Mitchell and William P. Rogerson, "Products Liability, Consumer Misperceptions, and Market Power." *Bell Journal of Economics*, Vol. 14 (1983), pp. 581-589.

[61] Rahman, K. Sabeel, "Regulating Informational Infrastructure: Internet Platforms as New Public Utilities." *Georgetown Law and Technology Review*, (2018), pp. 234-251.

[62] Rochet, Jean-Charles, and Jean Tirole, "Platform Competition in Two-Sided Markets," *Journal of the European Economic Association*, Vol. 1 (2003), pp. 990-1029.

[63] Rochet, Jean-Charles, and Jean Tirole, "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, Vol. 37 (2006), pp. 645-667.

[64] Shavell, Steven, "The Judgment Proof Problem," *International Review of Law and Economics*, Vol. 6 (1986), pp. 45-58.

[65] Simon, Marilyn J., "Imperfect Information, Costly Litigation, and Product Quality," *Bell Journal of Economics*, Vol. 12 (1981), pp. 171-184.

[66] Spence, A. Michael, "Consumer Misperceptions, Product Failure, and Producer Liability," *Review of Economic Studies*, Vol. 44 (1977), pp. 561-572.

[67] Spence, A. Michael, "Monopoly, Quality, and Reputation," *Bell Journal of Economics*, Vol. 6 (1975), pp. 417-429.

[68] Spier, Kathryn E. and Rory Van Loo. "Foundations for Platform Liability," *Notre Dame Law Review*, forthcoming (2025).

[69] Tadelis, Steven, "Reputation and Feedback Systems in Online Platform Markets," *Annual Review of Economics*, Vol. 8 (2016), pp. 321-340.

[70] Tan, Guofu and Junjie Zhou, "The Effects of Competition and Entry in Multi-sided Markets," *Review of Economic Studies*, Vol. 88 (2021), pp. 1002-1030.

[71] Teh, Tat-How, "Platform Governance," *American Economic Journal: Microeconomics*, Vol. 14 (2022), pp. 213-254.

[72] Van Loo, Rory, "The New Gatekeepers: Private Firms as Public Enforcers," *Virginia Law Review*, Vol. 106 (2020a), pp. 467-522.

[73] Van Loo, Rory, "The Revival of Respondeat Superior and Evolution of Gatekeeper Liability," *Georgetown Law Journal*, Vol. 109 (2020b), pp. 141-189.

[74] Volokh, Eugene, "Treating Social Media Platforms as Common Carriers?" *Journal of Free Speech Law*, Vol. 1 (2021), pp. 377-462.

[75] Weyl, Glen, "A Price Theory of Multi-Sided Platforms," *American Economic Review*, Vol. 100 (2010), pp. 1642-1672.

[76] White, Alexander, and Glen Weyl, "Imperfect Platform Competition: A General Framework," (2010), working paper.

[77] Wickelgren, Abraham L., "The Inefficiency of Contractually Based Liability with Rational Consumers," *Journal of Law, Economics, & Organization*, Vol. 22 (2006), pp. 168-183.

[78] Yasui, Yuta, "Platform Liability for Third-party Defective Products," (2022), working paper.

[79] Zennyo, Yusuke, "Should Platforms be Held Liable for Defective Third-party Goods?" (2023), working paper.

Appendix

Proof of Lemma 1. We first show that, if $w_s \leq \widehat{w}$, the platform does not find it profitable to deter the type-g firms and retain the type-b firms. If the platform deters the type-g firms by setting a high price $p_b = \alpha_b - \theta_b w_s$, its profit is

$$\Pi_b(e) = \lambda(1-e)(\alpha_b - \theta_b w) - c(e),$$

where $w = w_s + w_p$. If the platform charges $p_g = \alpha_g - \theta_g w_s$ then both types join and profits are $\Pi(e)$ in (1). Consider two scenarios.

First, suppose $w > \frac{\alpha_b}{\theta_b}$. Then $\Pi_b(e) < 0$ for any e. Assumption A2 implies $\Pi(0) > 0$.

Second, suppose $w \leq \frac{\alpha_b}{\theta_b}$ so $\alpha_b - \theta_b w \geq 0$. If the platform sets $p_b = \alpha_b - \theta_b w_s$ and deters the type-g firms then $e = 0$ and $\Pi_b(0) = \lambda(\alpha_b - \theta_b w)$. If the platform sets $p_g = \alpha_g - \theta_g w_s$ and attracts the type-g firms then $\Pi(0) = \alpha_g - \theta_g w - \lambda(\theta_b - \theta_g)w_p$. We have

$$\begin{aligned}
\Pi(0) - \Pi_b(0) &= \alpha_g - \lambda\alpha_b - (1-\lambda)\theta_g w + \lambda(\theta_b - \theta_g)w_s \\
&\geq \alpha_g - \lambda\alpha_b - (1-\lambda)\theta_g \frac{\alpha_b}{\theta_b} \\
&= \alpha_g - (\lambda\theta_b + (1-\lambda)\theta_g)\frac{\alpha_b}{\theta_b} \\
&> 0,
\end{aligned}$$

where the first inequality holds given $w \leq \frac{\alpha_b}{\theta_b}$ and the second inequality follows from Assumption A2. Therefore, the platform would not deter the type-g firms.

Now we prove the lemma. (8) implies $e^* > 0$ if and only if $(\alpha_b - \theta_b w) - (\theta_b - \theta_g)(\widehat{w} - w_s) < 0$. This gives the condition for cases 1 and 2. Totally differentiating (10), and using the fact the social welfare function is concave, gives $de^*/dw_s = -\lambda\theta_g/S''(e) > 0$ and $de^*/dw_p = -\lambda\theta_b/S''(e) > 0$. Equation (10) implies $e^* > e^{**}$ if and only if $\lambda r_b(w_s) - \lambda\theta_b(d - w) > 0$. This gives the condition for subcases 2(a), 2(b) and 2(c).

Proof of Proposition 1. Note that $\widehat{w} < d < \frac{\alpha_g}{\theta_g}$ by Assumption A1. Suppose $w_p = 0$ and $w_s \leq \widehat{w}$. From Lemma 1, $e^* = 0$ if and only if

$$\alpha_b - \theta_b w_s > (\theta_b - \theta_g)(\widehat{w} - w_s).$$

Substituting for $\widehat{w}$ from (6),

$$\alpha_b - \theta_b w_s > (\alpha_b - \alpha_g) - (\theta_b - \theta_g)w_s,$$

which is equivalent to $w_s < \frac{\alpha_g}{\theta_g}$. Since $w_s \leq \widehat{w} < \frac{\alpha_g}{\theta_g}$ we have $e^* = 0$.

Suppose $w_s > \widehat{w}$. There are two possible scenarios. First, if $\theta_g/\theta_b < \alpha_g/\alpha_b$, then setting $w_p = 0$ in Lemma 2 and rearranging terms gives a threshold value $\frac{\alpha_b - \alpha_g + \lambda\alpha_g}{\theta_b - \theta_g + \lambda\theta_g} \in (\widehat{w}, \frac{\alpha_b}{\theta_b})$, which increases in λ. When $w_s < \frac{\alpha_b - \alpha_g + \lambda\alpha_g}{\theta_b - \theta_g + \lambda\theta_g}$, the platform sets $p^* = \alpha_b - \theta_b w_s$, and attracts the type-$b$ firms; when $w_s \geq \frac{\alpha_b - \alpha_g + \lambda\alpha_g}{\theta_b - \theta_g + \lambda\theta_g}$, the platform sets $p^* = \alpha_g - \theta_g w_s$ and deters the type-b firms. Second, if $\theta_g/\theta_b \geq \alpha_g/\alpha_b$, then $\widehat{w} \geq \frac{\alpha_b - \alpha_g + \lambda\alpha_g}{\theta_b - \theta_g + \lambda\theta_g} \geq \frac{\alpha_b}{\theta_b}$. Note that, if $w_s > \frac{\alpha_b}{\theta_b}$, the type-$b$ firms have no incentive to join the platform. Hence, the platform sets $p^* = \alpha_g - \theta_g w_s$ and the type-b firms do not join the platform. The two scenarios can be combined by defining $\widetilde{w}(\lambda) = \max\left\{\frac{\alpha_b - \alpha_g + \lambda\alpha_g}{\theta_b - \theta_g + \lambda\theta_g}, \widehat{w}\right\}$.

Proof of Proposition 2. Suppose $w_s \leq \widehat{w}$, so the type-g firms are marginal. From equation (10) we have $e^* = e^{**}$ if and only if $w_p = w_p^* = d - w_s - \left(1 - \frac{\theta_g}{\theta_b}\right)(\widehat{w} - w_s)$. Note that $w_p^* \in (0, d - w_s)$ if $w_s < \widehat{w}$ and $w_p^* = d - w_s$ if $w_s = \widehat{w}$.

Suppose $w_s \in (\widehat{w}, \widetilde{w})$. From Proposition 1, if $w_p = 0$, the platform sets $p = \alpha_b - \theta_b w_s$, and attracts the type-$b$ firms. Lemma 2 implies that the platform would deter the type-b firms if $\lambda(\alpha_b - \theta_b w) \leq (1-\lambda) r_g(w_s)$. Note that $\lambda(\alpha_b - \theta_b w)$ decreases in w_p and the firms' surplus $(1-\lambda) r_g(w_s)$ is independent of w_p. Setting $\lambda(\alpha_b - \theta_b w) = (1-\lambda) r_g(w_s)$ gives the lower bound $\underline{w}_p$:

$$\underline{w}_p = \frac{\alpha_b}{\theta_b} - w_s - \frac{1-\lambda}{\lambda}\left(1 - \frac{\theta_g}{\theta_b}\right)(w_s - \widehat{w}) > 0.$$

For any $w_p^* \geq \underline{w}_p$, the platform deters the type-b firms.

Suppose $w_s \geq \widetilde{w}$. Proposition 1 implies that even if $w_p = 0$ the platform sets $p^* = \alpha_g - \theta_g w_s$ and deters type-b firms. Platform liability is unnecessary.

Proof of Proposition 3. We first prove a claim.

Claim 1*: Suppose $w_s \leq \widehat{w}$. The platform sets $p^r = \alpha_g - \theta_g w_s - \theta^r \delta(d - w)$ and attracts the type-b firms, where $\theta^r = E(\theta | e^r)$. Let $r_b(w_s) = (\theta_b - \theta_g)(\widehat{w} - w_s)$ and $\theta^0 = \lambda\theta_b + (1-\lambda)\theta_g$.*

1. *If* $(\alpha_b - \theta_b d) + (\theta_b - \theta^0\delta)(d - w) \geq r_b(w_s)$ *then* $e^r = 0 < e^{**}$.
2. *If* $(\alpha_b - \theta_b d) + (\theta_b - \theta^0\delta)(d - w) < r_b(w_s)$ *then* $e^r > 0$.
   (a) *If* $\theta_b(1-\delta)(d-w) > r_b(w_s)$ *then* $0 < e^r < e^{**}$.
   (b) *If* $\theta_b(1-\delta)(d-w) = r_b(w_s)$ *then* $0 < e^r = e^{**}$.
   (c) *If* $\theta_b(1-\delta)(d-w) < r_b(w_s)$ *then* $0 < e^{**} < e^r$.

Proof of Claim 1: To begin, we construct values $\{e^r, p^r, t^r\}$ that maximize the platform's profits subject to the participation constraints of the users and type-g firms (as the type-g firm is marginal). Then, we verify that this is an equilibrium of the game.

$$\max_{\{e,p,t\}} \quad \Phi(e,p) = (1-e)\lambda(p-\theta_b w_p) + (1-\lambda)(p-\theta_g w_p) - c(e) \tag{22}$$

subject to

$$\alpha_0 - t - E(\theta|e)\delta(d - w_s - w_p) \geq 0 \tag{23}$$

$$t - (\theta_g w_s + k) - p \geq 0. \tag{24}$$

(23) and (24) are the user's and type-g firm's participation constraints, respectively.

The type-g firm's participation constraint (24) must bind. If not, the platform would increase the price p which would increase the platform's profits in (22). Since (24) binds, $p = t - (\theta_g w_s + k)$ and we can rewrite the optimand (22) as

$$(1-e)\lambda(t - (\theta_g w_s + k) - \theta_b w_p) + (1-\lambda)(t - (\theta_g w_s + k) - \theta_g w_p) - c(e).$$

Next, we show that the user's participation constraint (23) binds. If not, the platform would increase t and its profits would rise. Since (23) and (24) bind, we have

$$p = \alpha_0 - E(\theta|e)\delta(d - w_s - w_p) - (\theta_g w_s + k).$$

Since $\alpha_g = \alpha_0 - k$ and $w = w_s + w_p$ the solution to the optimization problem is:

$$e^r = \arg\max_{e \geq 0} \Phi(e, p^r) \tag{25}$$

$$t^r = \alpha_0 - E(\theta|e^r)\delta(d - w) \tag{26}$$

$$p^r = \alpha_g - \theta_g w_s - E(\theta|e^r)\delta(d - w). \tag{27}$$

We now verify that $\{e^r, p^r, t^r\}$ defined in (25), (26), and (27) is an equilibrium of the game. Suppose that the platform charges p^r in (27) and chooses e^r in (25). The firms and users observe e^r and therefore believe that the probability of harm is $\theta^r = E(\theta|e^r)$. The users are (just) willing to pay t^r in (26) and the type-g firms are (just) willing to pay p^r in (27). Anticipating that the users and the firms all participate, the platform exerts effort e^r in (25). Therefore $\{e^r, p^r, t^r\}$ is an equilibrium.

Next, we verify that Assumption A2 guarantees that the platform's profits are positive.

Note that the platform is better off if the users believe that the product is safer. Users would pay a higher price t so the platform could charge firms a higher price p. It is sufficient to show that the platform's profits are positive when $e = 0$ so $E(\theta|0) = \theta^0$. In this scenario, $t = \alpha_0 - \theta^0\delta(d - w)$ from (23). The type-g firms are willing to pay $p = \alpha_g - \theta_g w_s - \theta^0\delta(d - w)$ from (24). The platform's profits can be rewritten as

$$\begin{aligned}\Pi(0) &= \alpha_g - \theta_g w_s - \theta^0\delta(d - w) - \theta^0 w_p \\ &\geq \alpha_g - \theta^0 d + \lambda(\theta_b - \theta_g)w_s \\ &\geq 0,\end{aligned}$$

where the first inequality follows from $\delta \leq 1$ and the second from Assumption $A2$.

We now show that the condition in case 1 of the claim is necessary and sufficient for a corner solution, $e^r = 0$. We first show the condition is necessary. If $e^r = 0$ then $E(\theta|0) = \theta^0$. Since the user's participation constraint (23) binds we have $t^r = \alpha_0 - \theta^0\delta(d-w)$; since the type-$g$ firm's participation constraint (24) binds we have $p^r = \alpha_g - \theta_g w_s - \theta^0\delta(d-w)$. Finally, for $e^r = 0$ to be optimal for the platform we need $\partial\Phi(e, p^r)/\partial e \leq 0$ or equivalently $p^r - \theta_b w_p \geq 0$. Substituting p^r, this condition becomes

$$\alpha_g - \theta_g w_s - \theta^0\delta(d - w) - \theta_b w_p \geq 0.$$

Adding $r_b(w_s) = (\theta_b - \theta_g)(\widehat{w} - w_s)$ on both sides and rearranging terms lead to

$$(\alpha_b - \theta_b d) + (\theta_b - \theta^0\delta)(d - w) \geq r_b(w_s).$$

This confirms that the condition in case 1 is necessary.

Next, we show that the condition in case 1 is sufficient for $e^r = 0$. Suppose the condition holds and $e^r > 0$. Since $E(\theta|e^r) < \theta^0$, $t^r > \alpha_0 - \theta^0\delta(d - w)$ and $p^r > \alpha_g - \theta_g w_s - \theta^0\delta(d - w) > \theta_b w_p$. So, the platform does not audit, $e^r = 0$, a contradiction.

Now consider case 2. The condition implies $p^r - \theta_b w_p < 0$ so the platform is losing money from each type-b transaction. The equilibrium effort $e^r > 0$ and $\theta^r = E(\theta|e^r)$. The platform charges $p^r = \alpha_g - \theta_g w_s - \theta^r\delta(d-w)$ and users pay $t^r = \alpha_0 - \theta^r\delta(d-w)$. Condition (17) implies that $e^{**} < e^r$ if and only if $\theta_b(1 - \delta)(d - w) < (\theta_b - \theta_g)(\widehat{w} - w_s)$. Totally differentiating (17) and using the fact that the welfare function is concave, $de^r/dw_p > 0$.

We now proceed to prove Proposition 3.

(1) Suppose $w_s \leq \widehat{w}$, so the type-g firm is marginal. From Claim 1 case 2b, we have

$e^r = e^{**}$ if and only if

$$(\theta_b - \theta_g)(\widehat{w} - w_s) - \theta_b(1-\delta)(d-w) = 0.$$

When $w_s = \widehat{w}$, it is socially efficient to set $w = d$, or equivalently, $w_p^r = d - w_s$.

Consider $w_s < \widehat{w}$. Substituting $w = w_p + w_s$, we have $w_p = d - w_s - \left(\frac{\theta_b - \theta_g}{\theta_b - \delta\theta_b}\right)(\widehat{w} - w_s)$, which decreases in δ and goes to $-\infty$ when $\delta \to 1$. Moreover, when $\delta = \frac{\theta_g}{\theta_b}$,

$$d - w_s - \left(\frac{\theta_b - \theta_g}{\theta_b - \delta\theta_b}\right)(\widehat{w} - w_s) = d - \widehat{w} > 0.$$

Hence, there exists a unique threshold $\delta^r \in (\frac{\theta_g}{\theta_b}, 1)$ such that

$$d - w_s - \left(\frac{\theta_b - \theta_g}{\theta_b - \delta^r\theta_b}\right)(\widehat{w} - w_s) = 0.$$

If $\delta \leq \delta^r$, $w_p^r = d - w_s - \left(\frac{\theta_b - \theta_g}{\theta_b - \delta\theta_b}\right)(\widehat{w} - w_s)$ leads to $e^r = e^{**}$. If $\delta > \delta^r$, for any w_p, we have

$$(\theta_b - \theta_g)(\widehat{w} - w_s) - \theta_b(1-\delta)(d-w) > 0,$$

which, together with (17), implies that $S'(e^r) < 0$, that is, $e^r > e^{**}$. If $\delta < 1$, Claim 1 shows that e^r increases w_p, so it is socially optimal to have $w_p^r = 0$. If $\delta = 1$, (17) implies that e^r is indepedent of w_p.

(2) Suppose $w_s > \widehat{w}$. As shown in the text, the platform will charge the low price $p^r = \alpha_b - \theta_b w_s - \theta^0\delta(d-w)$ and attract the type-b firms if and only if

$$\lambda[\alpha_b - \theta_b w - \theta_b\delta(d-w)] > (1-\lambda)(\theta_b - \theta_g)(w_s - \widehat{w}), \tag{28}$$

where the left-hand side decreases in w_p if $\delta < 1$ and is independent of w_p if $\delta = 1$. If $\delta = 1$, the left-hand side becomes $\lambda(\alpha_b - \theta_b d) < 0$, so the platform will charge the high price $p^r = \alpha_g - \theta_g w_s - \theta_g\delta(d-w)$ and deter the type-b firms.

Now, consider the scenario with $\delta < 1$. Suppose $w_p = 0$. Then (28) can be written as

$$w_s < \frac{\alpha_b - \alpha_g + \lambda\alpha_g - \lambda\theta_b\delta d}{\theta_b - \theta_g + \lambda\theta_g - \lambda\theta_b\delta},$$

where the right-hand side can be shown to be decreasing in δ, less than $\widehat{w}$ if $\delta = 1$ or if $\theta_g/\theta_b > \alpha_g/\alpha_b$, and greater than $\widehat{w}$ if $\delta = 0$ and $\theta_g/\theta_b < \alpha_g/\alpha_b$.

Consider two cases.

(2.1) Suppose that $\theta_g/\theta_b \geq \alpha_g/\alpha_b$. Define $\widetilde{w}^r(\delta) = \widehat{w}$. The earlier analysis implies that, given any δ and $w_s > \widehat{w} = \widetilde{w}^r(\delta)$, (28) does not hold. The platform always charges $p^r = \alpha_g - \theta_g w_s - \theta_g \delta(d-w)$ and deter the type-b firms. Platform liability is unnecessary.

(2.2) Suppose that $\theta_g/\theta_b < \alpha_g/\alpha_b$. The earlier analysis implies that there exists a unique threshold $\delta^R \in (0,1)$ such that

$$\frac{\alpha_b - \alpha_b + \lambda\alpha_b - \lambda\theta_b\delta^R d}{\theta_b - \theta_g + \lambda\theta_g - \lambda\theta_b\delta^R} = \widehat{w}.$$

If $\delta \geq \delta^R$, then $w_s > \widehat{w} \geq \frac{\alpha_b - \alpha_g + \lambda\alpha_g - \lambda\theta_b\delta d}{\theta_b - \theta_g + \lambda\theta_g - \lambda\theta_b\delta}$. Define $\widetilde{w}^r(\delta) = \widehat{w}$. Condition (28) does not hold with $w_p = 0$. Since the left-hand side of (28) decreases in w_p while the right-hand side is independent of w_p, condition (28) does not hold for any w_p. Hence, the platform always charges $p^r = \alpha_g - \theta_g w_s - \theta_g\delta(d-w)$ and deter the type-b firms. Platform liability is unnecessary.

If $\delta < \delta^R$, then $\widehat{w} < \frac{\alpha_b - \alpha_g + \lambda\alpha_g - \lambda\theta_b\delta d}{\theta_b - \theta_g + \lambda\theta_g - \lambda\theta_b\delta}$. It can also be verified that $\frac{\alpha_b - \alpha_g + \lambda\alpha_g - \lambda\theta_b\delta d}{\theta_b - \theta_g + \lambda\theta_g - \lambda\theta_b\delta} < \frac{\alpha_b}{\theta_b}$. Define $\widetilde{w}^r(\delta) = \frac{\alpha_b - \alpha_g + \lambda\alpha_g - \lambda\theta_b\delta d}{\theta_b - \theta_g + \lambda\theta_g - \lambda\theta_b\delta}$. When $w_s \geq \widetilde{w}^r(\delta)$, under any w_p, the platform charges $p^r = \alpha_g - \theta_g w_s - \theta_g\delta(d-w)$ and deter the type-b firms, so platform liability is unnecessary. When $w_s < \widetilde{w}^r(\delta)$, condition (28) holds with $w_p = 0$, that is, the platform attracts the type-b firms. Setting $\lambda[\alpha_b - \theta_b w - \theta_b\delta(d-w)] = (1-\lambda)r_g(w_s)$ gives the lower bound $\underline{w}_p^r$:

$$\underline{w}_p^r = \frac{\alpha_b - \theta_b\delta d}{\theta_b} - w_s - \frac{1-\lambda}{\lambda}\Big(\frac{\theta_b - \theta_g}{\theta_b(1-\delta)}\Big)(w_s - \widehat{w}) > 0.$$

Thus, when $w_s < \widetilde{w}^r(\delta)$, for any $w_p^r \geq \underline{w}_p^r$, the platform deters the type-b firms.

A Simple Extension on Retail Platforms with Heterogeneous Users

Suppose that users' values from transactions are drawn from density $f(\alpha)$ for $\alpha \in [0, \overline{\alpha}]$, with cumulative density $F(\alpha)$ and $\frac{1-F(\alpha)}{f(\alpha)}$ decreasing in α. For simplicity, assume that $\theta_g d < \overline{\alpha} < \theta_b d$, $k = 0$ and $w_s = 0$. That is, transactions with type-g firms are socially efficient when α is sufficiently high, while transactions with type-b firms are always inefficient. Moreover, the firms are completely judgment-proof. Thus, the two types of firms have the same surplus if they can stay on the platform. If type-g firms join the platform, type-b firms will join too. Users observe the platform's auditing effort e.

If the platform only charges a per-interaction price p, the firms would set a transaction price $t > p$, which leads to double marginalization. To avoid this distortion, we assume that the platform can charge a listing fee p_0 for retained firms and a per-interaction price

p. (Note that the analysis in Section 3 remains the same under this alternative pricing structure.) Recall that $\theta^r = \frac{(1-e)\lambda\theta_b+(1-\lambda)\theta_g}{(1-e)\lambda+(1-\lambda)}$ is the average risk. It can be shown that the firm would set the per-interaction price equal to its expected liability cost, $p = \theta^r w_p$, and choose p_0 to extract the firms' surplus.

After being matched with a firm, a user will purchase from the firm if and only if

$$\alpha \geq t + \theta^r \delta(d - w_p).$$

Given $p = \theta^r w_p$, a firm chooses t to maximize

$$(t - \theta^r w_p)[1 - F(t + \theta^r \delta(d - w_p))].$$

Therefore, the equilibrium transaction price t^r satisfies

$$t^r - \theta^r w_p - \frac{1 - F(t^r + \theta^r \delta(d - w_p))}{f(t^r + \theta^r \delta(d - w_p))} = 0.$$

Accordingly, the marginal user's value α^r satisfies

$$\alpha^r - \frac{1 - F(\alpha^r)}{f(\alpha^r)} = \theta^r w_p + \theta^r \delta(d - w_p), \tag{29}$$

which implies that α^r decreases in e, that is, more auditing stimulates user participation. Importantly, the social value from the marginal user's transaction, $\alpha^r - \theta^r d$, can be positive or negative. In other words, user participation may be socially insufficient (due to the classic monopoly distortion) or excessive (if w_p and δ are small).

The platform chooses the listing fee p_0 to extract the firms' surplus, that is, $p_0 = \int_{\alpha^r}^{\overline{\alpha}} (t^r - p) dF(\alpha)$. Given $p = \theta^r w_p$, the platform's profit can be written as

$$\begin{aligned} \Pi(e, \alpha^r) &= \int_{\alpha^r}^{\overline{\alpha}} [(1-e)\lambda + (1-\lambda)](t^r - \theta^r w_p) dF(\alpha) - c(e) \\ &= \int_{\alpha^r}^{\overline{\alpha}} [(1-e)\lambda + (1-\lambda)] \frac{1 - F(\alpha^r)}{f(\alpha^r)} dF(\alpha) - c(e) \end{aligned} \tag{30}$$

Social welfare is

$$S(e, \alpha^r) = \int_{\alpha^r}^{\overline{\alpha}} [(1-e)\lambda + (1-\lambda)](\alpha - \theta^r d) dF(\alpha) - c(e), \tag{31}$$

It can be shown that

$$\begin{aligned}\frac{d\Pi(e,\alpha^r)}{de} &= \frac{dS(e,\alpha^r)}{de} \\ &\quad -\lambda\theta_b(1-\delta)(d-w_p)[1-F(\alpha^r)] - [(1-e)\lambda\theta_b + (1-\lambda)\theta_g](1-\delta)(d-w_p)f(\alpha^r)\frac{d\alpha^r}{de} \\ &\quad +\lambda\int_{\alpha^r}^{\overline{\alpha}}(\alpha-\alpha^r)dF(\alpha) + [(1-e)\lambda+(1-\lambda)][1-F(\alpha^r)]\frac{d\alpha^r}{de}. \qquad (32)\end{aligned}$$

In general, the platform's auditing effort can be socially insufficient or excessive. The second line in (32) shows that the platform does not consider the impact of auditing (and the corresponding change in user participation) on harm not anticipated by users. The last line in (32) reflects that the platform does not internalize the impact of auditing (and the corresponding change in user participation) on inframarginal users' surplus.

We now consider two cases. First, if $\delta = 1$, then (29) and (30) imply that α^r and e^r are independent of w_p. Platform liability has no effect on welfare if the users fully anticipate or internalize the harm.

Second, suppose $\delta = 0$ and $\underline{\alpha} < \theta^0 d$, where $\underline{\alpha} - \frac{1-F(\underline{\alpha})}{f(\underline{\alpha})} = 0$. We will show that a marginal increase in platform liability from $w_p = 0$ raises social welfare. To see this, note that, if $w_p = 0$ and $\delta = 0$, the firm does not take any auditing effort ($e^r = 0$) and (29) implies $\alpha^r = \underline{\alpha}$ and $\frac{d\alpha^r}{de} = 0$. Hence, (32) becomes

$$\frac{d\Pi(e^r,\alpha^r)}{de} = \frac{dS(e^r,\alpha^r)}{de} + \lambda\int_{\alpha^r}^{\overline{\alpha}}(\alpha-\alpha^r-\theta_b d)dF(\alpha),$$

which, together with the assumption $\alpha < \theta_b d$ for any α, implies $\frac{dS}{de} > 0$. The equilibrium auditing effort is socially insufficient. Moreover, since $\alpha^r = \underline{\alpha} < \theta^0 d$, user participation is socially excessive, that is, $\frac{\partial S}{\partial \alpha^r} > 0$. Note that $\alpha^r > \underline{\alpha}$ for any $w_p > 0$, which implies $\frac{\partial \alpha^r}{\partial w_p} > 0$ at $w_p = 0$. Differentiating the social welfare function (30) with respect to w_p,

$$\frac{dS}{dw_p} = \frac{dS}{de}\frac{de^r}{dw_p} + \frac{\partial S}{\partial \alpha^r}\frac{\partial \alpha^r}{\partial w_p}. \qquad (33)$$

Since $\frac{dS}{de} > 0$, $\frac{\partial S}{\partial \alpha^r} > 0$, $\frac{\partial \alpha^r}{\partial w_p} > 0$, and $\frac{de^r}{dw_p} \geq 0$ given $e^r = 0$, we have $\frac{dS}{dw_p} > 0$. A marginal increase in platform liability raises social welfare. Platform liability not only motivates the platform to take auditing effort, but can also prevent some socially inefficient transactions. By continuity, platform liability can be beneficial if δ is positive but small.

Holding Platforms Liable

Xinyu Hua* Kathryn E. Spier†
HKUST Harvard University

November 15, 2024

Online Appendix B

This appendix contains the analysis of six additional extensions to the baseline model: (B1) alternative pricing structure with non-refundable fees, (B2) false positives, (B3) litigation costs, (B4) competing platforms, (B5) user participation, and (B6) firm moral hazard.

B1. Alternative Pricing Structure

Our baseline model assumed that the platform could only charge an interaction price to the firms. In this extension, assume that the platform can use two-part tariffs: a non-refundable application fee y and an interaction price p. We will show that platform liability can still be socially beneficial.

When $w_s > \widehat{w}$, the type-b firms are marginal and the platform can – but may not have incentives – to deter them by charging a high interaction price and setting $y = 0$. The analysis is the same as in the baseline model. Therefore, in this extension, we focus on the case with $w_s \leq \widehat{w}$.

Given $w_s \leq \widehat{w}$, the type-g firms are marginal and the platfrom sets $p + y = a_g - \theta_g w_s$. If a type-$b$ firm seeks to join the platform, its expected surplus is

$$\begin{aligned} & (1-e)(a_b - \theta_b w_s - p) - y \\ = \; & (1-e)(\theta_b - \theta_g)(\widehat{w} - w_s) - ey, \end{aligned}$$

*Hong Kong University of Science and Technology. xyhua@ust.hk.
†Harvard Law School and NBER. kspier@law.harvard.edu.

which decreases in e and equals $(\theta_b - \theta_g)(\widehat{w} - w_s)$ when $e = 0$.

Similar to the analysis in the baseline model, the platform will accommodate the type-b firms by setting $y = 0$ and $e = 0$ if and only if the joint benefit for the platform and firms is larger than the type-b firms' surplus.

$$a_b - \theta_b(w_s + w_p) \geq (\theta_b - \theta_g)(\widehat{w} - w_s). \tag{1}$$

Absent platform liability ($w_p = 0$), as shown in Section 2, the above condition holds given $w_s \leq \widehat{w}$. Therefore, if $w_p = 0$, the platform would accommodate the type-b firms.

If $w_p = d - w_s$, given $a_b - \theta_b d < 0$, condition (1) does not hold. Thus, when w_p is sufficiently large, the platform has incentives to block or deter the type-b firms. Note that the type-b firms can be fully deterred if and only if

$$y > \frac{(1-e)(\theta_b - \theta_g)(\widehat{w} - w_s)}{e}. \tag{2}$$

If the platform sets $y \leq \frac{(1-e)(\theta_b-\theta_g)(\widehat{w}-w_s)}{e}$, then the type-$b$ firms seek to join the platform and the analysis of the equilibrium is the same as in the baseline model.

If the platform sets $y > \frac{(1-e)(\theta_b-\theta_g)(\widehat{w}-w_s)}{e}$, then the type-$b$ firms do not join the platform. However, the platform still needs to commit to some auditing effort, because condition (2) cannot hold when e is arbitrarily close to 0. Since $y = a_g - \theta_g w_s - p$ and the right-hand side of (2) decreases in e, to fully deter the type-b firms and minimize the auditing cost, the platform would set $p = 0$, $y = a_g - \theta_g w_s$, and e larger than but arbitrarily close to $\underline{e}$, where $\underline{e}$ satisfies

$$a_g - \theta_g w_s = \frac{(1-\underline{e})(\theta_b - \theta_g)(\widehat{w} - w_s)}{\underline{e}},$$

or, equivalently,

$$\underline{e} = 1 - \frac{a_g - \theta_g w_s}{a_b - \theta_b w_s} > 0.$$

In general, $\underline{e}$ can be larger or smaller than e^{**}, which is the socially optimal auditing effort in the baseline model (when the type-b firms cannot be deterred by the pricing mechanism). If $\underline{e} < e^{**}$, it is socially optimal to deter the type-b firms by using a high non-refundable application fee. Imposing large platform liability (for example, $w_p = d - w_s$) motivates the platform to do so.

Proposition 5. *(Non-Refundable Fees) Suppose $w_s \leq \widehat{w}$ and $\underline{e} < e^{**}$. If $w_p = 0$, the*

platform accomodates the type-b firms by choosing $y = 0$, $p = a_g - \theta_g w_s$, *and* $e = 0$. *If* $w_p = d - w_s$, *the platform deters the type-b firms by choosing* $y = a_g - \theta_g w_s$, $p = 0$, *and* $e = \underline{e} + \varepsilon$ *with arbitrarily small* $\varepsilon > 0$.

B2. False Positives (Type-I Errors)

Now we extend the baseline model by considering false positives. Suppose that the auditing effort of the platform may erroneously block the type-g firms with probability ϕe, where $\phi < 1$. If the type-b firms seek to join the platform, social welfare is:

$$S(e) = v + \lambda(1-e)(\alpha_b - \theta_b d) + (1-\lambda)(1-\phi e)(\alpha_g - \theta_g d) - c(e). \tag{3}$$

The socially optimal auditing effort $\widetilde{e}^{**}$ (if it is positive) satisfies

$$-\lambda(\alpha_b - \theta_b d) - \phi(1-\lambda)(\alpha_g - \theta_g d) - c'(\widetilde{e}^{**}) = 0. \tag{4}$$

When $w_s > \widehat{w}$, the type-b firms are marginal and the platform would not take auditing effort. There is no type-I error. The analysis is the same as in the baseline model.

When $w_s \leq \widehat{w}$, the type-g firms are marginal. The platform sets the interaction price $p^f = \alpha_g - \theta_g w_s$, and its profits can be written as

$$\Pi(e) = S(e) - (1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) + [(1-e)\lambda\theta_b + (1-\lambda)(1-\phi e)\theta_g](d-w) - v.$$

Denote the equilibrium auditing effort by e^f. If $e^f > 0$, the first-order condition is

$$\Pi'(e^f) = S'(e^f) + \lambda(\theta_b - \theta_g)(\widehat{w} - w_s) - [\lambda\theta_b + (1-\lambda)\phi\theta_g](d-w) = 0. \tag{5}$$

Note that the users' (marginal) uncompensated harm, $[\lambda\theta_b + (1-\lambda)\phi\theta_g](d-w)$, is larger than that in the baseline model, while the firms' surplus, $\lambda(\theta_b - \theta_g)(\widehat{w} - w_s)$, remains the same. Thus, the platform's incentives for auditing are weaker than in the baseline model. Hence, the optimal platform liability becomes larger as shown below (the proof is similar to that in the baseline model and omitted).

Proposition 6. *(False Positives.) The socially-optimal platform liability for harm to users,* w_p^f, *is as follows:*

1. *If* $w_s \leq \widehat{w}$ *then* $w_p^f = d - w_s - \frac{\lambda(\theta_b - \theta_g)}{\lambda\theta_b + (1-\lambda)\phi\theta_g}(\widehat{w} - w_s) \geq w_p^*$. *The platform attracts the type-b firms and its auditing incentives are socially efficient,* $e^f = \widetilde{e}^{**}$.

2. *If* $w_s \in (\widehat{w}, \widetilde{w})$ *then there exists a threshold* $\underline{w}_p > 0$ *such that, under any* $w_p^f \in [\underline{w}_p, d - w_s]$, *the platform deters the type-b firms.*

3. *If* $w_s \geq \widetilde{w}$ *then platform liability is unnecessary. Under any* $w_p^f \in [0, d - w_s]$, *the platform deters the type-b firms.*

B3. Litigation Costs

We now extend the baseline model by considering litigation costs. When a user gets harmed by a firm and files a lawsuit, the litigation costs are z_p, z_s, z_u, respectively for the platform, the firm, and the user. Denote $z = z_p + z_s + z_u$. Assume that $z_u \leq w_s + w_p$ and $\alpha_g - \theta_g d - z > 0$.[1] So, litigation is credible and it is efficient to have interactions between the type-g firms and users. If the type-b firms seek to join the platform, social welfare is

$$S(e) = v + \lambda(1-e)(\alpha_b - \theta_b(d+z)) + (1-\lambda)(\alpha_g - \theta_g(d+z)) - c(e).$$

The socially optimal auditing effort $\overline{e}^{**} > 0$ satisfies

$$-\lambda(\alpha_b - \theta_b(d+z)) - c'(\overline{e}^{**}) = 0.$$

The two types of firms have the same surplus when:

$$w_s + z_s = \widehat{w} = \frac{\alpha_b - \alpha_g}{\theta_b - \theta_g}. \tag{6}$$

Case 1: $w_s + z_s \leq \widehat{w}$. The platform sets $p^z = \alpha_g - \theta_g(w_s + z_s)$ to extract the type-g firms' surplus. The platform chooses $e > 0$ if and only if $p^z - \theta_b(w_p + z_p) < 0$, which can be rewritten as

$$\alpha_b - \theta_b(w + z_p + z_s) - (\theta_b - \theta_g)(\widehat{w} - w_s - z_s) < 0.$$

[1] We also assume that z is lower than the benefit of improved platform incentives.

The platform's profits can be written as

$$\Pi(e) = S(e) - (1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s - z_s) + [(1-e)\lambda\theta_b + (1-\lambda)\theta_g](d + z_u - w) - v.$$

Denote the equilibrium auditing effort as e^z. If $e^z > 0$, the first-order condition is

$$\Pi'(e^z) = S'(e^z) + \lambda(\theta_b - \theta_g)(\widehat{w} - w_s - z_s) - \lambda\theta_b(d + z_u - w) = 0. \tag{7}$$

The users' uncompensated loss caused by the type-b firms, $\lambda\theta_b(d + z_u - w)$, increases in z_u; and the firms' surplus, $\lambda(\theta_b - \theta_g)(\widehat{w} - w_s - z_s)$, decreases in z_s. Therefore, as compared to the baseline model, the platform's auditing incentives are even weaker relative to the social incentives. Moreover, condition (7) implies that $e^z = \overline{e}^{**}$ if and only if $w_p^z = d + z_u - w_s - (1 - \frac{\theta_g}{\theta_b})(\widehat{w} - w_s - z_s) \geq w_p^*$.

Case 2: $w_s + z_s > \widehat{w}$. The platform's profit-maximizing strategy is to either charge $p = \alpha_g - \theta_g(w_s + z_s)$ and deter the type-b firms from joining the platform or charge $p = \alpha_b - \theta_b(w_s + z_s)$ and attract both types. The platform will charge $p = \alpha_b - \theta_b(w_s + z_s)$ and attract the type-b firms if

$$\lambda(\alpha_b - \theta_b(w + z_s + z_p)) > (1-\lambda)(\theta_b - \theta_g)(w_s + z_s - \widehat{w}), \tag{8}$$

which is less likely to hold when z_s or z_p is larger. That is, the platform is more likely to deter the type-b firms when the litigation costs for the platform or the firms are larger. This also implies that the platform has stronger incentives to deter the type-b firms than in the baseline model.

Similar to the analysis in the baseline model, we can characterize the optimal platform liability.

Proposition 7. *(Litigation Costs) There exists a threshold $\widetilde{w}^z \in (\widehat{w}, d)$. The socially-optimal platform liability for harm to users, w_p^z, is as follows:*

1. *If $w_s + z_s \leq \widehat{w}$ then $w_p^z = d + z_u - w_s - (1 - \frac{\theta_g}{\theta_b})(\widehat{w} - w_s - z_s) \geq w_p^*$. The platform attracts the type-b firms and its auditing incentives are socially efficient, $e^z = \overline{e}^{**}$.*

2. *If $w_s + z_s \in (\widehat{w}, \widetilde{w}^z)$ then there exists a threshold $\underline{w}_p^z \in (0, \underline{w}_p)$ such that, under any $w_p^z \in [\underline{w}_p^z, d - w_s]$, the platform deters the type-b firms.*

3. *If $w_s + z_s \geq \widetilde{w}^z$ then platform liability is unnecessary. Under any $w_p^z \in [0, d - w_s]$, the platform deters the type-b firms.*

When $w_s + z_s \leq \widehat{w}$, as shown earlier, the platform's auditing incentives are even weaker relative to the social incentives, as compared to the baseline model. Hence, the optimal platform liability is larger than that in the baseline model, $w_p^z \geq w_p^*$, where the inequality holds strictly if $z_b > 0$ or $w_s + z_s < \widehat{w}$.

When $w_s + z_s \in (\widehat{w}, \widetilde{w}^z)$, with litigation costs, the platform has stronger incentives to deter the type-b firms than in the baseline model. Hence, the lowest platform liability that motivates the platform to deter the type-b firms is smaller than that in the baseline model, $\underline{w}_p^z < \underline{w}_p$.

B4. Platform Competition

Now consider two competing platforms, Platform 1 and Platform 2. Users are distributed symmetrically on a Hotelling line with density $f^c(x) = f^c(1 - x) > 0$ on $x \in [0, 1]$, Platform 1 is located at $x = 0$ while Platform 2 is located at $x = 1$. A user at location $x \in [0, 1]$ receives consumption value $v - \tau x$ if they join Platform 1 but $v - \tau(1 - x)$ if they join Platform 2, where $\tau \geq 0$ reflects the level of differentiation. Assume that v is sufficiently large and τ is not too large such that the market is fully covered. The firms can join both platforms, while each user only joins one platform.[2] Thus, the platforms compete for users but not for firms.

In stage 1, the platforms simultaneously set interaction prices p_j and commit to their audit intensities e_j, $j = 1, 2$. Suppose that the auditing effort is per interaction and the users observe auditing effort before deciding which platform to join.[3] The timing and the other assumptions are otherwise identical to the baseline model. We shall focus on the symmetric equilibrium where $p_1 = p_2$ and $e_1 = e_2$ and, accordingly, each platform serves half of the users. We will show that platform liability can be socially beneficial in this competitive environment.

Case 1: $w_s \leq \widehat{w}$. The platforms set $p^c = \alpha_g - \theta_g w_s$, which attracts the type-b firms.

[2] In practice, many users choose single-homing due to switching costs or same-side network effects.

[3] The results hold qualitatively if auditing costs are per firm and the platforms are sufficiently differentiated (i.e., τ is not too small). With per firm auditing costs, it would be socially efficient to have two platforms if τ is large but efficient to have one platform if τ is small, due to large economies of scale in auditing.

Denote the location of the indifferent user as $\widehat{x}$. If $\widehat{x} \in [0,1]$, then it satisfies

$$\begin{aligned} & v - \tau\widehat{x} - [\lambda(1-e_1)\theta_b + (1-\lambda)\theta_g](d-w) \\ = \; & v - \tau(1-\widehat{x}) - [\lambda(1-e_2)\theta_b + (1-\lambda)\theta_g](d-w), \end{aligned}$$

or equivalently,

$$\widehat{x} = \frac{1}{2} + \frac{\lambda(e_1 - e_2)\theta_b(d-w)}{2\tau}.$$

If $w_p = d - w_s$ then $\widehat{x} = \frac{1}{2}$. The users are fully compensated for any harm. Similar to the analysis in the baseline model, the platforms over-invest in auditing.

If $w_p < d - w_s$, given e_2, Platform 1 can attract all the users ($\widehat{x} = 1$) by choosing $e_1 \geq \overline{e}_1$, where

$$\overline{e}_1 = e_2 + \frac{\tau}{\lambda\theta_b(d-w)}.$$

When $\tau \to 0$, $\overline{e}_1 \to e_2$, so Platform 1 would raise its auditing effort slightly to attract all the users as long as its profit is positive. When $\tau \to \infty$, $\overline{e}_1 \to \infty$, so Platform 1 would not be able to capture the whole market. Hence, there exist two thresholds $\underline{\tau}$ and $\overline{\tau}$, with $0 < \underline{\tau} \leq \overline{\tau}$, such that both platforms get positive profits if $\tau > \overline{\tau}$ while they get zero profits if $\tau < \underline{\tau}$. We consider these two cases separately.

First, suppose $\tau > \overline{\tau}$. In this case, competition is not fierce and $\widehat{x} \in (0,1)$. Platform 1 chooses e_1 to maximize its profit

$$F^c(\widehat{x})[(1-e_1)\lambda(p^c - \theta_b w_p) + (1-\lambda)(p^c - \theta_g w_p) - c(e_1)],$$

where $F^c(\widehat{x})$ is the number of users choosing Platform 1. The profit-maximizing auditing effort by Platform 1, e_1^c (if it is positive), satisfies

$$\begin{aligned} 0 \;=\; & -F^c(\widehat{x})[\lambda(p^c - \theta_b w_p) + c'(e_1^c)] \\ & + f^c(\widehat{x})\frac{\lambda\theta_b(d-w)}{2\tau}[(1-e_1^c)\lambda(p^c - \theta_b w_p) + (1-\lambda)(p^c - \theta_g w_p) - c(e_1^c]. \end{aligned} \tag{9}$$

In the symmetric equilibrium with $F^c(\widehat{x}) = \frac{1}{2}$ and $e_1^c = e_2^c = e^c$, this can be rewritten as

$$\begin{aligned} 0 \;=\; & \frac{1}{2}S'(e^c) + \frac{1}{2}[\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) - \lambda\theta_b(d-w)] \\ & + f^c(\widehat{x})\frac{\lambda\theta_b(d-w)}{2\tau}[(1-e^c)\lambda(p^c - \theta_b w_p) + (1-\lambda)(p^c - \theta_g w_p) - c(e^c)], \end{aligned} \tag{10}$$

where the last term captures the competition effect. If $w_p > w_p^*$, as shown in the baseline model, the second term on the right-hand side of (10) is positive while the last term is non-negative, so the platforms over-invest in auditing, $e^c > e^{**}$. If $w_p = w_p^*$, the second term becomes 0 while the last term is positive if $e^c = e^{**}$, so the platforms over-invest in auditing, $e^c > e^{**}$. Finally, if $w_p = 0$ and $\tau \to \infty$, similar to the analysis in the baseline model, $e^c \to 0$. By continuity, there exists a unique threshold $\widehat{\tau} \geq \overline{\tau}$ such that $e^c < e^{**}$ if $\tau > \widehat{\tau}$ and $w_p = 0$. These observations imply that, given $\tau > \widehat{\tau}$, there exists $\widehat{w}_p \in (0, w_p^*)$ under which $e^c = e^{**}$. Hence, the optimal platform liability is $w_p^c = \widehat{w}_p < w_p^*$, which motivates the platform to choose the socially efficient auditing effort. Competition raises the platforms' auditing incentives, so that the optimal platform liability is less than in the baseline model.

Next, suppose $\tau < \underline{\tau}$. Given fierce competition, the platforms invest to the point where profits are dissipated,

$$(1 - e^c)\lambda(p^c - \theta_b w_p) + (1 - \lambda)(p^c - \theta_g w_p) - c(e^c) = 0. \tag{11}$$

If $w_p = 0$ then platform safety is socially excessive, $e^c > e^{**}$. Absent platform liability, the platforms take too much auditing effort. Equation (11) also implies $\frac{de^c}{dw_p} < 0$. Therefore, if $\tau < \underline{\tau}$, platform liability mitigates the over-investment problem and raises social welfare.

Case 2: $w_s > \widehat{w}$. In this case, the type-b firms are marginal. The platforms have a choice: they can either charge the firms $p = \alpha_g - \theta_g w_s$ and deter the type-b firms or charge the firms $p = \alpha_b - \theta_b w_s < \alpha_g - \theta_g w_s$ and attract both types. As shown in the baseline model, when $w_s \geq \widetilde{w} > \widehat{w}$, a monopoly platform has incentives to charge the high price and deter the type-b firms. With competition, a platform can attract more users by deterring the type-b firms, because the users observe the prices and prefer to join a safer platform. Therefore, given $w_s \geq \widetilde{w}$, both platforms deter the type-b firms. As in the baseline model, platform liability is unnecessary.

Now suppose $w_s \in (\widehat{w}, \widetilde{w})$. If $w_p = d - w_s$, the users would be fully compensated for any harm and therefore each platform attracts half of the users. Each platform charges the high price and deter the type-b firms if

$$\frac{1}{2}(1 - \lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) > \frac{1}{2}[\alpha_b - \theta_b w_s - (\lambda\theta_b + (1 - \lambda)\theta_g)w_p],$$

which holds given $\alpha_b - \theta_b d < 0$. Hence, imposing full residual liability on the platforms gets the platforms to raise the interaction price and deter the type-b firms.

We now show that platform liability is necessary when $w_s \in (\widehat{w}, \widetilde{w})$ and τ is sufficiently large. Suppose to the contrary that, under $w_p = 0$, the platforms charge $p = \alpha_g - \theta_g w_s$ and deter the type-b firms. Each platform's profit is $(1-\lambda)(\alpha_g - \theta_g w_s)/2$. If Platform 1 deviates to $p = \alpha_b - \theta_b w_s$, the indifferent user's location $\widehat{x}$ satisfies

$$\tau\widehat{x} + [\lambda\theta_b + (1-\lambda)\theta_g](d - w_s) = \tau(1-\widehat{x}) + (1-\lambda)\theta_g(d - w_s),$$

that is,

$$\widehat{x} = \frac{1}{2} - \frac{\lambda\theta_b(d - w_s)}{2\tau}.$$

Accordingly, Platform 1's profit from deviation is

$$F^c\Big(\max\Big\{0, \frac{1}{2} - \frac{\lambda\theta_b(d - w_s)}{2\tau}\Big\}\Big)(\alpha_b - \theta_b w_s), \tag{12}$$

which goes to 0 when $\tau \to 0$ and goes to $(\alpha_b - \theta_b w_s)/2$ when $\tau \to \infty$. Note that $(1-\lambda)(\alpha_g - \theta_g w_s) < (\alpha_b - \theta_b w_s)$ given $w_s \in (\widehat{w}, \widetilde{w})$. Hence, there exists a threshold $\widetilde{\tau} > 0$ such that, absent platform liability, both platforms deter the type-b firms if and only if $\tau \leq \widetilde{\tau}$. If $\tau > \widetilde{\tau}$, platform liability is socially desired. If $\tau \leq \widetilde{\tau}$, platform liability is unnecessary. Since the price that the platforms charge is observed by users, and the platforms are not highly differentiated, the users will prefer to join a platform that completely deters the harmful type-b firms.

Proposition 8. *(Platform Competition with Observable Effort.) The socially-optimal liability for the competing platforms, w_p^c, is as follows.*

1. *If $w_s \leq \widehat{w}$, there exist $\widehat{\tau}$ and $\underline{\tau}$ with $0 < \underline{\tau} \leq \widehat{\tau}$: when $\tau > \widehat{\tau}$, $w_p^c \in (0, w_p^*)$ motivates the platforms to choose the socially efficient auditing effort; when $\tau < \underline{\tau}$, $w_p^c > 0$ mitigates the over-investment problem and raises social welfare.*

2. *If $w_s \in (\widehat{w}, \widetilde{w})$, there exists $\widetilde{\tau} > 0$: when $\tau > \widetilde{\tau}$, $w_p^c = d - w_s$ motivates the platforms to deter the type-b firms; when $\tau \leq \widetilde{\tau}$, platform liability is unnecessary and the platforms deter the type-b firms under any $w_p^c \in [0, d - w_s]$.*

3. *If $w_s > \widetilde{w}$, platform liability is unnecessary. Under any $w_p^c \in [0, d - w_s]$, the platforms deter the type-b firms.*

B5. User Participation

Suppose that the users' valuations of the quasi-public good are drawn from density $f^u(v) > 0$ for $v \in [0, \infty)$, with cumulative density $F^u(v)$.[4] As in the baseline model, the platform charges the firms price p per interaction and takes auditing effort e per firm. The users have the option to join the platform for free.[5]

Assumption A2 implies that it is socially efficient for all users to participate and assumption A1 implies that it is socially inefficient for the type-b firms to participate. As in the baseline model, full deterrence of the type-b firms may not be possible. If the type-b firms seek to join the platform, social welfare is

$$S(e, \widehat{v}) = \int_{\widehat{v}}^{\infty} [v + \lambda(1-e)(\alpha_b - \theta_b d) + (1-\lambda)(\alpha_g - \theta_g d)] f^u(v) dv - c(e), \tag{13}$$

where $\widehat{v}$ is the value of the marginal user,

$$\widehat{v}(e, w) = (\lambda(1-e)\theta_b + (1-\lambda)\theta_g)(d - w). \tag{14}$$

Notice that $\widehat{v}(e, w)$ is decreasing in e and w for all $d - w > 0$: higher levels of effort and liability stimulate user participation. Holding e constant, the users view w as a "rebate" for joining the platform. Therefore, the social planner would like to set $w = d$ (that is, $w_p = d - w_s$), so that all the users participate. Given full participation by the users, the socially efficient auditing effort is e^{**}, the same as in the baseline model.

Case 1: $w_s \leq \widehat{w}$. The type-g firms are marginal and the platform charges $p^u = \alpha_g - \theta_g w_s$. The platform's profit function can be written as:

$$\Pi(e, \widehat{v}) = S(e, \widehat{v}) + \int_{\widehat{v}}^{\infty} \big\{ -(1-e)\lambda(\theta_b - \theta_g)(\widehat{w} - w_s) + ((1-e)\lambda\theta_b + (1-\lambda)\theta_g)(d - w) - v \big\} f^u(v) dv, \tag{15}$$

[4] This framework is equivalent to the model where users decide how much time (T) to spend on the platform. The user's marginal value decreases in T. At each moment, the user is randomly matched with a firm and may be harmed. Intuitively, when platform liability increases and/or the platform raises audit intensity, the user spends more time.

[5] The platform might also charge a membership fee $m \geq 0$ to each user. However, it can be shown that $m = 0$ in equilibrium if $\alpha_g - (\lambda\theta_b + (1-\lambda)\theta_g)d$ is sufficiently large (that is, if cross-side network effects are strong). We maintain the assumption that $\alpha_g - (\lambda\theta_b + (1-\lambda)\theta_g)d$ is sufficiently large such that the platform does not charge the users.

Since users observe the auditing effort, the platform's effort (if it is positive) satisfies

$$\frac{d\Pi(e^u,\widehat{v})}{de} = \frac{dS(e^u,\widehat{v})}{de} + \int_{\widehat{v}}^{\infty} [\lambda(\theta_b-\theta_g)(\widehat{w}-w_s) - \lambda\theta_b(d-w)] f^u(v)dv \\ - \lambda\theta_b(d-w)[\lambda(1-e^u)(\theta_b-\theta_g)(\widehat{w}-w_s)]f^u(\widehat{v}) = 0 \quad (16)$$

where $\widehat{v} \equiv \widehat{v}(e,w)$.

When $w_s = \widehat{w}$, $\frac{d\Pi(e^u,\widehat{v})}{de} = \frac{dS(e^u,\widehat{v})}{de}$ if and only if $w_p^u = d - w_s$. Therefore, imposing full residual liability on the platform motivates the platform to choose $e^u = e^{**}$ and attracts all the users to join the platform.

When $w_s < \widehat{w}$, the last term on the right-hand side of equation (16) is negative. Moreover, if $w_p \leq w_p^*$, where $w_p^* \in (0, d-w_s)$ is the optimal platform liability in the baseline model, then the second term on the right-hand side of equation (16) is non-positive. Therefore, $\frac{dS(e^u,\widehat{v})}{de} > 0$, that is, the platform's auditing incentive is socially insufficient. The social planner chooses w_p to maximize social welfare:

$$\frac{dS(e^u,\widehat{v})}{dw_p} = \frac{dS(e^u,\widehat{v})}{de}\frac{de^u}{dw_p} + \frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}, \quad (17)$$

where $\frac{\partial \widehat{v}}{\partial w_p} = -(\lambda(1-e^u)\theta_b + (1-\lambda)\theta_g) < 0$. Since $\frac{\partial S(\cdot)}{\partial \widehat{v}} < 0$, the last term in (17), $\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}$, is non-negative. Intuitively, given the auditing effort, platform liability stimulates user participation and therefore raises social welfare. Moreover, as shown earlier, $\frac{dS(e^u,\widehat{v})}{de} > 0$ if $w_p \leq w_p^*$. Hence, as long as $\frac{de^u}{dw_p} > 0$, it is socially optimal to set $w_p^u > w_p^*$.

Case 2: $w_s > \widehat{w}$. In this case, type-b firms are marginal. First, suppose $w_s \geq \widetilde{w}$, where $\widetilde{w}$ is defined in Section 2 (the baseline model) . As shown in Section 2, the platform charges $p^u = \alpha_g - \theta_g w_s$, which deters all of the type-b firms. Anticipating that the type-b firms are fully deterred, the users participate if $v \geq (1-\lambda)\theta_g(d-w)$. Hence, all the users participate when $w_p = d - w_s$. Second, suppose $w_s \in (\widehat{w}, \widetilde{w})$. As shown in Section 2, given $w_p \geq \underline{w}_p$, the platform charges $p^u = \alpha_g - \theta_g w_s$, which deters all of the type-b firms. Again, setting $w_p = d - w_s$ attracts all the users.

Proposition 9. *(User Participation with Observable Effort) The socially-optimal platform liability for harm to users, w_p^u, is as follows:*

1. *If $w_s < \widehat{w}$, then $w_p^u > w_p^*$ as long as $\frac{de^u}{dw_p} > 0$. The platform's auditing effort is not socially optimal.*

2. *If $w_s = \widehat{w}$, then $w_p^u = d - w_s$. The platform chooses the socially optimal auditing effort $e^u = e^{**}$ and all users participate.*

3. *If $w_s > \widehat{w}$, then $w_p^u = d - w_s$. The platform deters the type-b firms and all users participate.*

As in the baseline model, platform liability motivates the platform to take auditing effort or set high interaction prices to block or deter risky firms. When users are heterogeneous, platform liability has the additional benefit in stimulating user participation.

Example: Uniform Distribution. In case 1 of Proposition 9, the socially optimal platform liability is larger than that in the baseline model as long as the equilibrium auditing effort increases in w_p. Recall that, in the baseline model, the equilibrium effort always increases in w_p. However, in this extension, the equilibrium effort may increase or decrease in w_p. For illustration, suppose that v follows the uniform distribution on $[0, \overline{v}]$. Then with observable effort, the platform's effort (if it is positive) satisfies

$$\begin{aligned}\frac{d\Pi(e^u, \widehat{v})}{de} &= -c'(e^u) - \lambda(\alpha_g - \theta_g w_s - \theta_b w_p)\Big[1 - \frac{\widehat{v}}{\overline{v}}\Big] \\ &\quad +\lambda\theta_b(d-w)\Big[\lambda(1-e^u)(\alpha_g - \theta_g w_s - \theta_b w_p) + (1-\lambda)(\alpha_g - \theta_g w)\Big]\frac{1}{\overline{v}} \\ &= 0,\end{aligned}$$

which implies

$$\begin{aligned}\frac{d^2\Pi(e^u, \widehat{v})}{dedw_p} &= \frac{\lambda}{\overline{v}}\Big\{\overline{v} - (\lambda(1-e^u)\theta_b \\ &\quad +(1-\lambda)\theta_g)\Big[(1+\beta)\theta_b(d-w) + \alpha_g - \theta_g w_s - \theta_b w_p\Big] \\ &\quad -\theta_b\Big[(1-e^u)\lambda(\alpha_g - \theta_g w_s - \theta_b w_p) + (1-\lambda)(\alpha_g - \theta_g w)\Big]\Big\}.\end{aligned}$$

If $\overline{v}$ is very small and $w_p = 0$ then $\frac{d^2\Pi(e^u, \widehat{v})}{dedw_p} < 0$ and, accordingly, $\frac{de^u}{dw_p} < 0$. By contrast, if $\overline{v}$ is sufficiently large then for any $w_p \leq w_p^*$ we have $\frac{d^2\Pi(e^u, \widehat{v})}{dedw_p} > 0$ and, accordingly, $\frac{de^u}{dw_p} > 0$. Intuitively, given the participation threshold, an increase in platform liability raises the marginal profit from auditing effort; at the same time, the increase in platform liability decreases the participation threshold, which in turn reduces the marginal profit from auditing effort. The former effect dominates when $\overline{v}$ is sufficiently large.

To summarize, even if the heterogeneous users observe the auditing effort and choose whether to join the platform or not, platform liability can be socially desired. The optimal

platform liability is (weakly) larger than in the baseline model, as long as the equilibrium effort increases in w_p, which holds when v follows the uniform distribution on $[0, \overline{v}]$ with sufficiently large $\overline{v}$.

B6. Firm Moral Hazard

The baseline model assumes that the firms' types are exogenously given. Platform liability can still be socially beneficial if the firms' types are endogenous and the firms can take effort to improve safety. In this section, suppose all the firms are identical ex ante but may become either the type-g or type-b ex post. If a firm takes (unobservable) care with cost $c > 0$, the probability of becoming type-b is λ. If the firm does not take care, the probability of being type-b rises to $\widehat{\lambda} > \lambda$. The platform commits to its price p before the firms decide to take care or not. The firms privately learn their realized types and decide whether to join the platform.

For simplicity, we maintain the following assumption

$$c < (\widehat{\lambda} - \lambda)(\alpha_g - \theta_g d) + \lambda(\alpha_b - \theta_b d). \tag{18}$$

Assumption (18) leads to several implications.

First, since $\alpha_b - \theta_b d < 0$, $c < (\widehat{\lambda} - \lambda)(\alpha_g - \theta_g d)$. If the type-$b$ firms never join the platform, it is socially efficient for the (ex ante identical) firms to invest c.

Second, Assumption (18) implies

$$c < (\widehat{\lambda} - \lambda)[(\alpha_g - \theta_g d) - (\alpha_b - \theta_b d)] = (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(d - \widehat{w}).$$

Even if both types join the platform, it is efficient for the firms to invest c.

Finally, Assumption (18) implies

$$\lambda(\alpha_b - \theta_b d) + (1 - \lambda)(\alpha_g - \theta_g d) - c > (1 - \widehat{\lambda})(\alpha_g - \theta_g d),$$

that is, social welfare is larger if all the firms invest c and join the platform than if no firm invests and only the type-g firms join the platform.

In the first-best benchmark, all the firms invest c ex ante and only the type-g firms join the platform. Given c, there exists $w^m \in (\widehat{w}, d)$ such that, if and only if $w_s > w^m$,

$$c < (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w}).$$

Case 1: $w_s \leq \widehat{w}$. The type-g firms are marginal. The platform charges $p = \alpha_g - \theta_g w_s$. Since the type-$g$ firms do not have any surplus, ex ante the firms have no incentive to take care. As in the baseline model, $w_p^m = w_p^* \in (0, d - w_s]$ motivates the platform to choose the socially optimal auditing effort.

Case 2: $w_s > \widehat{w}$. The type-b firms are marginal. Consider three scenarios.

Case 2.1: $w_s > \frac{\alpha_b}{\theta_b}$. Then the type-$b$ firms would never join the platform. The platform either charges $p_g = \alpha_g - \theta_g w_s$, under which the firms would not invest c, or charges p_0, where

$$p_0 = \alpha_g - \theta_g w_s - c/(\widehat{\lambda} - \lambda) > 0,$$

under which the firms would invest c. Social welfare is larger if the platform charges p_0. The platform's profit under p_g is

$$\Pi^g = (1 - \widehat{\lambda})(\alpha_g - \theta_g w_s - \theta_g w_p);$$

while its profit under p_0 is

$$\Pi^0 = (1 - \lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1 - \lambda)/(\widehat{\lambda} - \lambda).$$

The profit difference,

$$\Pi^0 - \Pi^g = (\widehat{\lambda} - \lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1 - \lambda)/(\widehat{\lambda} - \lambda),$$

decreases in w_p. That is, the platform has stronger incentives to charge p_0 if w_p is lower. When $c > \frac{(\widehat{\lambda}-\lambda)^2}{(1-\lambda)}(\alpha_g - \theta_g w_s)$, then the platform never charges p_0, so platform liability is unnecessary. When $c \leq \frac{(\widehat{\lambda}-\lambda)^2}{(1-\lambda)}(\alpha_g - \theta_g w_s)$, then $\Pi^0 - \Pi^g \geq 0$ if $w_p = 0$ but may become negative if w_p is large, so it is optimal to set $w_p = 0$.

Case 2.2: $w_s \in (w^m, \frac{\alpha_b}{\theta_b})$**.** Given $w_s < \frac{\alpha_b}{\theta_b}$, the type-$b$ firms may have incentives to join the platform. Moreover, given $w_s > w^m$, we have $c < (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w})$, which implies $p_0 > p_b = \alpha_b - \theta_b w_s > 0$. If the platform charges p_g, the firms would not invest c and the platform's profit is

$$\Pi^g = (1 - \widehat{\lambda})(\alpha_g - \theta_g w_s - \theta_g w_p).$$

If the platform charges p_b, the type-g firms' surplus is $(\theta_b - \theta_g)(w_s - \widehat{w})$. Since $c <$

$(\widehat{\lambda}-\lambda)(\theta_b-\theta_g)(w_s-\widehat{w})$, the firms would invest c and always join the platform. Then the platform's profit is

$$\Pi^b = \lambda(\alpha_b - \theta_b w_s - \theta_b w_p) + (1-\lambda)(\alpha_b - \theta_b w_s - \theta_g w_p).$$

If the platform charges p_0, the firms would invest c but the type-b firms would not join the platform. Then the platform's profit becomes

$$\Pi^0 = (1-\lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1-\lambda)/(\widehat{\lambda}-\lambda).$$

Note that

$$\Pi^0 - \Pi^b = (1-\lambda)(\theta_b-\theta_g)(w_s-\widehat{w}) - \lambda(\alpha_b - \theta_b w_s - \theta_b w_p) - c(1-\lambda)/(\widehat{\lambda}-\lambda)$$

increases in w_p, while

$$\Pi^0 - \Pi^g = (\widehat{\lambda}-\lambda)(\alpha_g - \theta_g w_s - \theta_g w_p) - c(1-\lambda)/(\widehat{\lambda}-\lambda)$$

decreases in w_p. It can be verified that, when $w_s = w^m$, $\Pi^0 - \Pi^b \geq 0$ if and only if $w_p \geq (\alpha_b - \theta_b w_s)/\theta_b > 0$, and $\Pi^0 - \Pi^g \geq 0$ if $w_p = (\alpha_b - \theta_b w_s)/\theta_b$ and

$$\Big(1 - \frac{\widehat{\lambda}-\lambda}{1-\lambda}\Big)(\alpha_g - \theta_g w_s) \leq \Big(1 - \frac{\theta_g(\widehat{\lambda}-\lambda)}{\theta_b(1-\lambda)}\Big)(\alpha_b - \theta_b w_s),$$

which holds if θ_g is close to 0 and $\widehat{\lambda}$ is close to 1. Moreover, given $w_s \in (w^m, \frac{\alpha_b}{\theta_b})$, if there exists $w_p > 0$ under which $\Pi^0 - \Pi^b \geq 0$ and $\Pi^0 - \Pi^g \geq 0$, then for any $w_s' = w_s + \varepsilon$ with arbitrarily small $\varepsilon > 0$, $\Pi^0 - \Pi^b \geq 0$ and $\Pi^0 - \Pi^g \geq 0$ if platform liability is set at $w_p' = w_p - \varepsilon > 0$. Hence, there exists a unique threshold $\overline{w} \in [w^m, \frac{\alpha_b}{\theta_b}]$ such that, given $w_s \in (w^m, \overline{w})$, only under a non-empty set of $w_p > 0$, the platform charges p_0 and the first-best outcome is achieved.[6] That is, if $w_s \in (w^m, \overline{w})$, platform liability is socially desired.

If $w_s = \overline{w}$, $\Pi^0 - \Pi^b \geq 0$ and $\Pi^0 - \Pi^g \geq 0$ only under $w_p = 0$, so it is optimal to set $w_p = 0$. If $w_s \in (\overline{w}, \frac{\alpha_b}{\theta_b})$, the platform never charges p_0. Since it is efficient for all the firms to invest c and the profit difference $\Pi^b - \Pi^g$ decreases in w_p, it is optimal to set $w_p = 0$, under which the platform charges p_b and the firms invest c.

[6]Note that $\overline{w}$ may equal w^m or $\frac{\alpha_b}{\theta_b}$ under certain parameter values.

Case 2.3: $w_s \in (\widehat{w}, w^m)$. Given $w_s < w^m$, we have $c > (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w})$, which implies $p_0 < p_b$. If the platform charges p_g, the firms would not invest c and the platform's profit is

$$\Pi^g = (1 - \widehat{\lambda})(\alpha_g - \theta_g w_s - \theta_g w_p).$$

If the platform charges p_b, the type-g firms' surplus is $(\theta_b - \theta_g)(w_s - \widehat{w})$. Since $c > (\widehat{\lambda} - \lambda)(\theta_b - \theta_g)(w_s - \widehat{w})$, the firms would not invest c but always join the platform. The platform's profit is

$$\Pi^b = \widehat{\lambda}(\alpha_b - \theta_b w_s - \theta_b w_p) + (1 - \widehat{\lambda})(\alpha_b - \theta_b w_s - \theta_g w_p).$$

If the platform charges $p_0 < p_b$, the firms would invest c and join the platform, so the platform's profit becomes

$$\Pi^0 = \alpha_g - \theta_g w_s - c/(\widehat{\lambda} - \lambda) - [\lambda\theta_b + (1 - \lambda)\theta_g]w_p.$$

When $w_p = 0$, it can be verified that $\Pi^b > \Pi^g$ and $\Pi^b > \Pi^0$, that is, the platform would charge p_b and the firms do not invest c but join the platform. Similar to the analysis in the baseline model, with full residual liability ($w_p = d - w_s$), the platform's profit is larger under p_g than under p_b, so the platform may charge either p_0 or p_g. Under either price, social welfare is larger than under p_b. Hence, given $w_s \in (\widehat{w}, w^m)$, platform liability is socially desired.

Summarizing the above analysis, we have

Proposition 10. *(Firm Moral Hazard.) Suppose that firm liability is $w_s \in [0, d]$ and the firms can take effort with costs c. The socially-optimal liability, w_p^m, is as follows:*

1. *If $w_s \leq \widehat{w}$, it is optimal to set $w_p^m = w_p^* \in (0, d - w_s]$. The platform charges $p^m = \alpha_g - \theta_g w_s$ and takes auditing effort e^{**}. The firms do not invest c.*
2. *If $w_s \in (\widehat{w}, \overline{w})$, it is optimal to set $w_p^m > 0$. The firms invest c if $w_s \in (w^m, \overline{w})$.*
3. *If $w_s \geq \overline{w}$, either platform liability is unnecessary or it is optimal to set $w_p^m = 0$.*