Holding Platforms Liable*

Xinyu Hua† — HKUST
Kathryn E. Spier‡ — Harvard University

January 2, 2023

Abstract

Should platforms be held liable for the harms suffered by users? A two-sided platform enables interactions between firms and users. There are two types of firms: harmful and safe. The harmful firms impose larger costs on the users. If firms have deep pockets then platform liability is unwarranted. Holding the firms liable for user harms deters the harmful firms from joining the platform. If firms are judgment proof then platform liability plays an instrumental role in reducing social costs. With platform liability, the platform has an incentive to (1) raise the interaction price to deter harmful firms and (2) invest resources to detect and remove harmful firms from the platform. The residual liability assigned to the platform may be partial instead of full. The optimal level of platform liability depends on whether users are involuntary bystanders or voluntary consumers, the intensity of platform competition, and the impact on user participation.

*We would like to thank 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).

†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 may be misled by false information or harmed by cyberbullying and hate speech. 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, although this immunity is being challenged in legislatures and the courts.[1] Section 230 of the Communications Decency Act, enacted in 1996, shields platforms from liability for the digital content created by their participants.[2] Early proponents argued that the law was necessary to allow the internet to grow and flourish, but its application is controversial and many critics question the law's merits.[3] Proposed federal legislation, including the "Health Misinformation Act of 2021," would strip platforms of Section 230 protections if the platforms facilitate the spread of misinformation about public health emergencies.[4] In 2021, Zoom reportedly agreed to pay $85 million to settle a lawsuit alleging that Zoom shared users' personal data with third parties and failed to provide appropriate security measures.[5] Currently, the Supreme Court is poised to consider whether Google should be responsible for targeted recommendations made by their algorithms.[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 year, a California court found that Amazon could be held strictly liable for a defective

[1]See Buiten et al. (2020) for discussion of the European Commission's e-Commerce Directive. Hosting platforms in the EU may avoid liability for illegal content posted by users, assuming they are not aware of it, and are not responsible for monitoring the legality of the posted content.

[2]Section 230(c)(1) says that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." See "Why Hate Speech on the Internet Is a Never-Ending Problem." New York Times, August 6, 2019.

[3]See *Force v. Facebook, Inc.*, No. 18-397 (2d Cir. 2019). The court opined that Section 230 "should be construed broadly in favor of immunity."

[4]Senators Amy Klobuchar and Ben Ray Lujan introduced this bill to combat misinformation about COVID-19. See "Bill Targets Health Misinformation." *Wall Street Journal*, July 23, 2021.

[5]See "After Issue With Hackers, Zoom Agrees to Settle Lawsuit Over 'Zoombombing'." *New York Times*, August 2, 2021.

[6]See "Google Case Before High Court Could Reshape Internet Economy." *Wall Street Journal*, Oct 30, 2022.

[7]See *Erie Ins. Co. v. Amazon.com*, 2019 WL 2195146 (4th Cir. May 22, 2019). See also *State Farm Fire and Casualty Company v. Amazon.com, Inc.*, 2020 WL 6746745 (9th Cir. Nov. 17, 2020), and *Great Northern Insurance v. Amazon.com, Inc.*, 2021 WL 872949 (N.D. Ill. March 9, 2021).

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.[10] 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.[11] 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 either raising the interaction price or by investing resources to detect and remove the harmful firms from the platform.[12]

In our baseline model, the *users are bystanders* of the firms. 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). Platform users may be harmed by the firms when their private data is breached or when they are exposed to harmful advertising or misinformation. Absent liability the harmful firms have no incentive to leave the platform, and the platform has an insufficient incentive to detect and remove them. Holding the firms and the platform jointly liable gets them to internalize the negative externalities on the user-bystanders.

If the firms have deep pockets, and must pay in full for the harms they cause, then platform liability is unwarranted. Holding just the firms liable achieves the first-best outcome. 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

[8] See *Bolger v. Amazon.com, LLC*, 53 Cal.App.5th 431, 267 Cal.Rptr.3d 601 (2020). The court held that Amazon "is an integral part of the overall producing and marketing enterprise that should bear the cost of injuries resulting from defective products."

[9] See *Loomis v. Amazon.com LLC,* 2021 WL 1608878 (Cal. App. Ct. April 26, 2021).

[10] Consistent with the literature, we assume that the platform does not charge users. Section 3.1 extends the model to retail platforms where the consumers pay the firms and the firms pay the platform.

[11] The focus of this paper is cross-side harms. Similar issues arise when the injurers and victims are on the same side of the market. See Section 3.4.

[12] According to Amazon's post, "[Amazon's] proactive measures begin when a seller attempts to open an account. Our new seller account vetting includes a number of verifications and uses proprietary machine learning technology that stops bad actors before they can register or list a single product in our store." See https://www.aboutamazon.com/news/company-news/product-safety-and-compliance-in-our-store.

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

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 remove the harmful firms from the platform.[14] Interestingly, the optimal level of platform liability may be partial instead of full, as full liability could lead to excessive auditing by the platform.[15]

We then extend the baseline model to settings where *users are customers* of the firms, so interactions require the users' consent. Relevant settings include online marketplaces 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. The consumers are sophisticated and their willingness-to-pay reflects their rational 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. If the harmful firms are marginal, then platform liability is unwarranted. Since consumers are willing to pay more for safer products, the platform has a private incentive to raise the interaction price to deter the harmful firms from joining the platform. If the harmful firms are inframarginal, however, then partial platform liability gives the platform an appropriate incentive to audit and remove the harmful firms.[16] Since the platform internalizes the average harm to consumers, the socially-optimal platform liability is *lower* than in the baseline model (e.g., for social media platforms).[17]

Next, we extend the baseline model to consider two *competing platforms*. The users are bystanders and can participate on both platforms (i.e., multi-homing), while the firms can only participate on one of the platforms (i.e., single-homing). If the harmful firms are marginal then competition reduces the platforms' incentives to deter the harmful firms by charging high prices, relative to the baseline model. Therefore the socially optimal platform liability is (weakly) *higher* than that in the baseline monopoly model. If the harmful firms are inframarginal, holding the platforms partially liable for the residual harms motivates them to make the socially efficient auditing effort. In this case, since competition reduces the price-cost margins from serving the harmful firms, the competing platforms have stronger incentives for auditing than the monopoly platform. Thus, the socially optimal platform liability is *lower* than that in the baseline model. These observations suggest that policies encouraging platform competition should be complemented by changes in platform liability.

Finally, we extend the model to consider *user participation* when users have heterogeneous valuations. As in the baseline model, platform liability motivates the platform

[14]If the firms are very judgment proof and can evade liability, then the harmful firms are inframarginal (i.e. the harmful firms have a strictly higher willingness to pay than the safe firms). If the firms are moderately judgment proof, then the harmful firms are "marginal."

[15]If the firms are completely judgment proof, then the safe firms are marginal and the harmful firms get information rents. When choosing its audit intensity, the platform does not take into account the lost rents when the harmful firms are removed from the platform.

[16]As in our baseline model, full liability would lead to excessive auditing by the platform.

[17]We also show that platform liability and firm liability may be *complements* in the retail setting. In the benchmark model, platform liability and firm liability are *substitutes*.

to raise the interaction price or take auditing effort, which deters or removes the harmful firms. In addition, platform liability stimulates user participation. This happens for two reasons. First, users anticipate that the platform's auditing incentives are improved and that the platform is safer. Second, users view the larger damage award as a "rebate" for joining the platform. Because of the *user-participation effect*, the optimal platform liability is higher than in the baseline model.[18]

Our paper is related to the law-and-economics literature on products liability where firms are held liable for the product-related harms suffered by consumers. 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).[19] Building on Spence (1975), Hua and Spier (2020) emphasize the particular importance of firm liability when consumers are heterogeneous so the marginal buyer'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 (third party bystanders). If consumers are judgment proof and cannot be held accountable for the harms they cause, then extending liability to the manufacturer can help the market to internalize the harms.[20] Pitchford (1995) explores the desirability of extending liability to an injurer's lenders[21] and Dari Mattiacci and Parisi (2003) consider vicarious liability where liability is extended to the injurer's employer.[22] Arlen and MacLeod (2005) show that holding managed care organizations liable for medical malpractice by their physicians can raise the physicians' incentives to take care. Our model, which has not been previously studied, investigates the design of platform liability when the platform can audit and remove harmful participants.[23]

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'

[18]Other salient factors, including litigation costs, court errors, and alternative pricing structures, are also discussed. Online Appendix B presents a formal analysis of litigation costs and court errors.

[19]See also Simon (1981), Daughety and Reinganum (1995, 1997, 2006, 2008a and b, 2014), Arlen and Macleod (2003), Wickelgren (2006), Chen and Hua (2012, 2017), Choi and Spier (2014).

[20]Brooks (2002), and Fu et al. (2018) investigate how legal responsibility affects firms' choice between vertical integration and outsourcing.

[21]See also Boyer and Laffont (1997) and Che and Spier (2008). Bebchuk and Fried (1996) argue informally for raising the priority of tort victims in bankruptcy above debt claims gives the debtholders an incentive to better monitor the borrower.

[22]There are related legal studies. See Kraakman (1986) for a general taxonomy of gatekeeper enforcement strategies, Hamdani (2002) for liability on internet service providers, Hamdani (2003) on accountants and lawyers, and Van Loo (2020a) on big technology.

[23]Our paper is also related to the studies comparing joint and several liability (JSL) to several liability (SL) for harms caused by multiple defendants (e.g., see Landes and Posner, 1980; Carvell et al., 2012). With JSL, the victim may recover full damages from a single deep-pocketed defendant. With SL, the victim's recovery from each defendant is limited by the defendant's share of responsibility.

participation incentives. The literature also examines the impact of seller competition[24] or the impact of platform competition on pricing.[25] Some recent studies pay attention to non-pricing strategies, including seller exclusion (Hagiu, 2009), information management (Julien and Pavan, 2019; Choi and Mukherjee, 2020), control right allocation (Hagiu and Wright, 2015, 2018), and platform governance (Teh, 2022). A few policy papers (Buiten et al., 2020; Lefouili and Madio, 2022) discuss informally whether platforms should bear liability for harms caused by participants. Our paper contributes to the literature by investigating the effects of platform liability on platform pricing and auditing incentives, as well as their welfare implications.

Our paper is organized as follows. Section 2 presents the baseline model where users are bystanders to firms on a monopoly platform. This section explores the impact of liability on the platform's pricing and auditing as well as social welfare. Section 3 examines several alternative settings, including a retail platform where the firms are sellers and the users are consumers, two competing platforms, and heterogeneous users who make participation decisions. Section 4 provides concluding thoughts. The proofs are in the appendix.

2 The Model

Consider a two-sided platform (P) with two kinds of participants, firms (S) and users (B). The platform is a monopolist and necessary for interactions between firms and users.[26] 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.[27] Second, the platform provides opportunities for the firms and the users to interact. The platform charges the firms a price p per interaction. Google, for example, currently enjoys a market share of more than 92% of the search engine market. There are approximately seven billion free Google searches conducted by users every day. Google monetizes the quasi-pubic good by selling online advertising to businesses through real-time auctions.[28]

We assume that interactions between firms and users do not require the users' consent and so the users are effectively "bystanders."[29] The benefits and costs of these interactions depend on the firms' type, $i \in \{H, L\}$, where λ is the mass of type H and $1-\lambda$ is the mass

[24]See Nocke et al. (2007), Galeotti and Moraga-Gonzalez (2009), Hagiu (2009), Gomes (2014), Belleflamme and Peitz (2019).

[25]See Dukes and Gal-Or (2003), Hagiu (2006), Armstrong and Wright (2007), White and Weyl (2010), Karle et al. (2020), Tan and Zhou (2021).

[26]Section 3.2 extends the analysis to consider platform competition.

[27]This assumption is made for simplicity. Section 3.3 considers heterogeneous users with endogenous participation.

[28]Over 80% of Google's revenues in 2020 came from selling ads. See Google's annual report. Google's expertise in collecting and analyzing troves of user data increases the firms' willingness to participate in these auctions. Similarly, most of Facebook's revenue comes from advertising.

[29]Section 3.1 extends the analysis to retail platforms where interactions require the users' consent.

of type L in the firm population.[30] The H-type firms have higher interaction benefits, $\alpha_H > \alpha_L$, but impose higher interaction losses on users, $\theta_H d > \theta_L d$ where $\theta_i \in [0, 1]$ is the probability of harm and $d > 0$ is the level of harm per firm-user interaction.[31] The firms privately observe their types.

This general specification is aligned with a variety of economic settings. First, platform users may be harmed when their personal data is compromised. Prominent examples include the breach of Facebook user data by consulting firm Cambridge Analytica.[32] Some of the firms participating in Google's auctions allegedly collect and store so-called "bidstream data" on users, which they subsequently sell to third parties (including hedge funds and political campaigns).[33] Second, users often bear direct harms from fake news[34] and fraudulent (or simply unwanted) advertising.[35] It has been estimated that displayed advertising accounts for a large share of the data costs for mobile telephone plan users in the United States.[36] Third, our specification is also aligned with retail platforms when user-consumers consent to transactions but are unaware that the products and services are potentially dangerous.[37] Finally, although our focus is on harms to platform participants themselves, our insights also apply to harms to third parties who are external to the platform.[38]

We assume that the platform has the capability to detect and block the H-type firms. We will refer to the platform's efforts to detect the H-types as auditing. By virtue of their scale, data, and technological sophistication, platforms like Google may be in a good position to root out harmful platform participants.[39] Specifically, by spending

[30]For simplicity, λ is taken as exogenous. One may endogenize λ by allowing firms to invest resources to increase the likelihood being safe. As shown below, if the firms are very judgment proof then the H-types earn information rents. It follows that the firms' incentive to invest in safety would be insufficient.

[31] If $\alpha_H < \alpha_L$ then the H-types are marginal for all liability rules and auditing is unnecessary. The threshold $\widehat{w}$ defined in (5) below is identically equal to zero, and all of our results apply.

[32]The user data was allegedly used for political purposes. Facebook paid a \$5 billion fine.

[33]See "Wyden FTC Letter" https://www.wyden.senate.gov/imo/media/doc/073120 20Wyden 20Cassidy 20Led 20FTC 20Investigation 20letter.pdf

[34]Sensational content is a key driver of viewer attention and clicks, and it is easier and cheaper to fabricate a sensational story than to identify a true one. Top fake news proprietors reportedly earned \$10,000 to \$30,000 per month working with major advertising networks (e.g., Google AdSense). See Sydell (2016). Google kicked 200 publishers off of AdSense following the 2016 presidential election. Google attributed heightened removals to improvements in detection technology. See Townsend (2017).

[35]Facebook has settled lawsuits alleging that they failed to block scam advertisements. See "Facebook Hit With UK Copyright Suit Over Fraudulent Ads," Law360.com, October 8, 2021. In a lawsuit brought against Google, a user clicked on a fraudulent advertisement that took her to a website where she was unknowingly charged. See *Goddard v. Google, Inc.*, 640 F. Supp. 2d 1193 (N.D. Cal. Jul. 30, 2009).

[36]A recent study by Enders Analysis places the share at 18-79 percent. Other costs include those of blocking unwanted advertising. See https://www.techdirt.com/articles/20160317/09274333934/why-are-people-using-ad-blockers-ads-can-eat-up-to-79-mobile-data-allotments.shtml.

[37]If consumers are naïve and unaware, the harm that they suffer is effectively "externalized" on their future selves. Thus, the consumers' future selves are effectively bystanders. Firms that sell low-quality goods have lower production costs and a higher willingness to pay per interaction.

[38]Example include harms to copyright holders when illegal material is posted on Facebook or YouTube, and harms to branded products when counterfeits are sold on Amazon.

[39]See Van Loo (2020a, 2020b) for additional examples and relevant case law.

effort $e \in [0,1)$ per firm, the platform can detect H-type firms with probability e and block them from interacting with users.[40] We assume that the cost of effort $c(e)$ satisfies $c(0) = 0, c'(e) > 0, c''(e) > 0, c'(0) = 0$, and $c'(e) \to \infty$ as $e \to 1$. The effort level e is neither observable nor contractible.[41] Thus, there is a potential moral hazard problem associated with auditing.[42]

Suppose that both types of firms 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.[43] 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.[44] For simplicity, there are no litigation costs or other transaction costs associated with using the court system.[45] There may be practical and legal limits on firm and platform liability. Third-party vendors are often liquidity-constrained or "judgment proof" and cannot be held fully accountable for the harm that they cause and platforms may enjoy immunity as well. Thus, in practice, liability is often limited.

In the following analysis, we assume

$$
\begin{aligned}
A0 &: \; v - [\lambda\theta_H + (1-\lambda)\theta_L]d > 0; \\
A1 &: \; \alpha_L - \theta_L d > 0 > \alpha_H - \theta_H d; \\
A2 &: \; \alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)d > 0.
\end{aligned}
$$

A0 implies that the users' benefit from the quasi-public good is sufficiently high that the users would join the platform even if the H-type firms join the platform and there is no liability.[46] A1 implies that it is socially efficient (inefficient) for the L-type (H-type) firms

[40] If the platform takes auditing effort per interaction instead of per firm, the analysis remains the same as long as the number of users is fixed.

[41] The results in the baseline model would be unchanged if the platform could commit to e (as users are bystanders). If e is observable and user participation is endogeneous, the platform may take auditing effort even absent liability.

[42] We abstract away from the possibility that, after detecting the H-type firms, the platform might retain these firms and charge them a higher price. Such price discrimination would reduce social welfare, creating an additional reason for increasing platform liability.

[43] This interpretation is aligned with platform models with non-exclusive matching including Armstrong (2006) and Weyl (2010).

[44] 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.

[45] Section 3.4 discusses the impact of litigation costs.

[46] Similar results are obtained in a model where users have heterogeneous valuations and some users do not join the platform. See Section 3.3. Note that if users were naïve or unaware of product risks then they would participate for any $v > 0$.

to join the platform.[47] A2 guarantees that the platform always gets non-negative profits and implies that it is socially efficient for both types to join the platform on average. These assumptions are not essential for the main insights, but simplify the analysis.

The timing of the game is as follows.

1. The platform creates the quasi-public good for users and sets the interaction price p for the firms. The price p is publicly observed.
2. Firms privately learn their types $i \in \{H, L\}$ and decide whether to join the platform.
3. The platform chooses $e \in [0, 1)$ to audit firms on the platform and removes any detected H-type firms. The audit intensity e is not publicly observed.
4. Firms interact with the users and the 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.

We will maintain the assumption that the platform, firms, and users are sophisticated and understand the risks of interacting on the platform. 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 H-types and L-types), and the users.

We now present two social welfare benchmarks.

First-Best Benchmark. The first-best outcome is achieved if the socially-harmful H-type firms do not join the platform or interact with users. Auditing is unnecessary (as there are no H-types to be detected and removed). Social welfare is:

$$v + (1 - \lambda)(\alpha_L - \theta_L d). \tag{1}$$

Second-Best Benchmark. Suppose that the H-type firms join the platform. Auditing is necessary to detect and remove the H-types. Social welfare is:

$$S(e) = v + \lambda(1 - e)(\alpha_H - \theta_H d) + (1 - \lambda)(\alpha_L - \theta_L d) - c(e). \tag{2}$$

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

$$-\lambda(\alpha_H - \theta_H d) - c'(e^{**}) = 0. \tag{3}$$

At the optimum, the marginal cost of auditing, $c'(e^{**})$, equals the marginal benefit of blocking H-type firms from interacting with users, $-\lambda(\alpha_H - \theta_H d)$. Note that $e^{**} \in (0, 1)$ so some H-types remain on the platform in this second-best world.

[47] In our model, society is better off when the monopolist excludes the H-type firms. Given our assumptions, there is no social loss from monopoly pricing. In a more general model, platform liability could exacerbate the monopoly pricing problem (as would a Pigouvian tax).

Our analysis proceeds in two steps. First, we characterize the platform's pricing and auditing strategies, p and e, given the assignment of liability, w_s and w_p. Second, we explore the socially-optimal platform liability rule.

2.1 Equilibrium Analysis

A type-i firm will seek to join the platform when their expected profit per interaction is non-negative,

$$\alpha_i - \theta_i w_s - p \geq 0, \tag{4}$$

where α_i is the firm's interaction benefit, $\theta_i w_s$ is the firm's expected liability, and p is the price paid to the platform. Note that depending on the level of firm liability, w_s, the H-type may have higher or lower rents than the L-type. If $w_s = 0$ then the H-type firms have higher rents than the L-type firms (since $\alpha_H > \alpha_L$). If $w_s = d$ then the H-type firms have lower rents than the L-type firms (see Assumption A1). The rents of the two types are equal when

$$w_s = \widehat{w} = \frac{\alpha_H - \alpha_L}{\theta_H - \theta_L} < d. \tag{5}$$

The threshold $\widehat{w}$ defined in (5) is critical for understanding the impact of platform liability on the interaction price and audit intensity. If the firms are sufficiently judgment-proof, $w_s < \widehat{w}$, then the L-type firms are "marginal." If the L-types are indifferent about joining the platform then the H-types strictly prefer to join.[48] Auditing is necessary to detect and remove the H-type firms. In this setting, we will see that a higher level of platform liability creates a stronger incentive for the platform to audit the firms and remove the harmful H-types from the platform.

If the firms are only moderately judgment proof, $w_s > \widehat{w}$, then the H-type firms are marginal. If the H-types are indifferent about joining the platform then the L-types strictly prefer to join. In this setting, the platform can deter the socially-harmful H-types from joining the platform by raising the interaction price p; the platform need not engage in costly auditing. A higher level of platform liability gives the platform a stronger incentive to raise the interaction price to deter the harmful H-types from joining the platform.

We now characterize the equilibrium for $w_s \leq \widehat{w}$ and $w_s > \widehat{w}$ and present the results in two lemmas.

Case 1: $\boldsymbol{w_s \leq \widehat{w}}$. Suppose that firm liability is below the threshold, so the L-type firms are marginal. The platform sets the interaction price to extract the L-type firms' rent,[49]

$$p^* = \alpha_L - \theta_L w_s. \tag{6}$$

[48]If $w_s = \widehat{w}$, then the two types have the same rents. If the L-type firms join the platform, the H-types would join too.

[49]If $w_s < \widehat{w}$, the platform will choose between a low price $p_L = \alpha_L - \theta_L w_s$ where both types seek to join the platform and a high price $p_H = \alpha_H - \theta_H w_s$ where only the H-type firms seek to join. Assumption A2 guarantees that the platform does not find it profitable to deter the L-types and retain the H-types.

The H-types seek to join the platform. Using the definition of $\widehat{w}$ in (5) and the formula for p^* in (6), the H-type firms' rent per interaction is $\alpha_H - \theta_H w_s - p^* = (\theta_H - \theta_L)(\widehat{w} - w_s) \geq 0$. Notice that as firm liability w_s grows, the H-type's information rent falls. In the limit when $w_s \to \widehat{w}$ the H-type's rent approaches zero.

We now explore the platform's incentive to audit and remove the H-type firms. The platform's aggregate profits are:

$$\Pi(e) = (1-e)\lambda(p^* - \theta_H w_p) + (1-\lambda)(p^* - \theta_L w_p) - c(e). \tag{7}$$

A necessary and sufficient condition for the firm to audit, $e^* > 0$, is that the platform's profit associated with each retained H-type is negative, $p^* - \theta_H w_p < 0$. Using the formula for $\widehat{w}$ in (5) and p^* in (6), and letting $w = w_s + w_p$ be the joint liability of the firm and platform, $e^* > 0$ if and only if

$$(\alpha_H - \theta_H d) + \theta_H(d - w) - (\theta_H - \theta_L)(\widehat{w} - w_s) < 0. \tag{8}$$

The first term on the left-hand side of (8) is the social loss associated with each retained H-type and the second term is the uncompensated harm to the users. The sum of these two terms, $\alpha_H - \theta_H w$, is the joint platform-firm surplus associated with each retained H-type. The third term in (8) is the information rent captured by the H-type firm. If condition (8) holds, then the platform loses money on each retained H-type firm and so the platform invests $e^* > 0$ and removes detected H-types from the platform. If (8) does not hold then the platform has no incentive to audit and remove the H-types from the platform, $e^* = 0$.

We now explore how the private and social incentives for auditing diverge when $e^* > 0$. Using the definition of $S(e)$ in (2), $\widehat{w}$ in (5), and p^* in (6) the platform's profit function in (7) above may be rewritten as:

$$\begin{aligned}\Pi(e) = S(e) &- (1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) \\ &+ [(1-e)\lambda\theta_H + (1-\lambda)\theta_L](d - w) - v. \end{aligned} \tag{9}$$

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

$$\Pi'(e^*) = S'(e^*) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - \lambda\theta_H(d - w) = 0. \tag{10}$$

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 removes H-types from the platform, the removed H-types lose their information rents, $\lambda(\theta_H - \theta_L)(\widehat{w} - w_s)$. Auditing imposes a *negative externality* on the H-type firms. Second, when the platform removes H-types, the user-bystanders' uncompensated loss is reduced by $\lambda\theta_H(d - w)$. 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^{**}.

These basic insights are summarized in the following lemma.

Lemma 1. *Suppose $w_s \leq \widehat{w}$. The platform sets $p^* = \alpha_L - \theta_L w_s$ and attracts the H-type firms. Let $r_H(w_s) \equiv (\theta_H - \theta_L)(\widehat{w} - w_s)$ denote the H-types' information rents per interaction.*

1. *If $\alpha_H - \theta_H w \geq r_H(w_s)$ then the platform does not audit, $e^* = 0 < e^{**}$.*
2. *If $\alpha_H - \theta_H w < r_H(w_s)$ 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 $\theta_H(d - w) > r_H(w_s)$ then $0 < e^* < e^{**}$.*
   - *(b) If $\theta_H(d - w) = r_H(w_s)$ then $0 < e^* = e^{**}$.*
   - *(c) If $\theta_H(d - w) < r_H(w_s)$ then $0 < e^{**} < e^*$.*

To summarize, when firm liability is below the threshold, $w_s \leq \widehat{w}$, the H-type firms cannot be deterred from joining the platform by the interaction price p. In case 1 of Lemma 1, the joint liability $w = w_s + w_p$ is small and the platform makes money on each H-type interaction. In this case, the platform takes no effort to audit, $e^* = 0$. The platform is enabling the H-type firms and profiting from their socially harmful activities.

In case 2 of Lemma 1, the joint liability $w = w_s + w_p$ is larger and the platform loses money on each H-type interaction. The platform therefore has incentives to audit and remove the H-types, $e^* > 0$. The platform's incentives to audit are stronger when w_p and w_s are larger. Intuitively, when platform liability w_p rises, the platform's cost of keeping H-types on the platform rises. When firm liability w_s rises, the interaction price that the platform can charge falls, reducing the platform's benefit of retaining the H-type firms.

Finally, and importantly, Lemma 1 establishes that the platform's incentive to audit and remove the H-types may be socially insufficient or socially excessive. In case 2(a) the level of joint liability is small and the platform's investment in auditing is suboptimal, $e^* < e^{**}$. The platform is not taking into account the positive impact that their investments have on the user-bystanders. In case 2(c) when the level of joint liability is large, then the platform is overly aggressive in its auditing efforts, $e^* > e^{**}$. The reason is that the platform is not taking into account the negative impact that their audit imposes on the H-type firms.

Case 2: $\boldsymbol{w_s > \widehat{w}}$. Now suppose that firm liability is above the threshold, so the H-type firms are marginal. The platform's profit-maximizing strategy is to either charge $p_L = \alpha_L - \theta_L w_s$ and deter the H-types from joining the platform or charge $p_H = \alpha_H - \theta_H w_s < p_L$ and attract both types. Notably, if the platform chooses the latter strategy and attracts the H-type firms then the platform will not invest in auditing, $e^* = 0$.[50]

The platform will charge p_H and attract the H-types (instead of charging p_L and deterring the H-types) if

$$\lambda(p_H - \theta_H w_p) + (1 - \lambda)(p_H - \theta_L w_p) > (1 - \lambda)(p_L - \theta_L w_p).$$

[50] Attracting the H-types and exerting auditing effort $e > 0$ is a dominated strategy, since the platform can deter the H-types by charging a higher price.

Substituting the formulas for p_H and p_L and using the definition of $\widehat{w}$ in equation (5) this condition becomes:

$$\lambda(\alpha_H - \theta_H w) > (1-\lambda)(\theta_H - \theta_L)(w_s - \widehat{w}). \tag{11}$$

The left-hand side is the joint value of attracting the H-type firms on the platform: the fraction λ of H-types multiplied by the interaction benefit α_H minus the joint liability $\theta_H(w_s + w_p)$. The expression on the right-hand side is the information rent captured by the inframarginal L-types.

We have the following result.

Lemma 2. *Suppose $w_s > \widehat{w}$. Let $r_L(w_s) \equiv (\theta_H - \theta_L)(w_s - \widehat{w})$ denote the L-type firm's information rents per interaction.*

1. *If $\lambda(\alpha_H - \theta_H w) > (1-\lambda) r_L(w_s)$ then the platform sets $p^* = \alpha_H - \theta_H w_s$, attracts the H-type firms, and does not audit, $e^* = 0 < e^{**}$.*

2. *If $\lambda(\alpha_H - \theta_H w) \leq (1-\lambda) r_L(w_s)$ then the platform sets $p^* = \alpha_L - \theta_L w_s$ and deters the H-type firms.*

If firm liability is above the threshold, $w_s > \widehat{w}$, then the H-type firms are marginal. The platform has the power to deter the H-type firms by raising the price. In case 1 of Lemma 2, the joint benefit of including the H-types is larger than the information rents captured by the L-type firms. In this case, the platform charges a low price, $p^* = \alpha_H - \theta_H w_s$, welcomes the H-types on the platform and takes no steps to detect or remove them. In case 2 of Lemma 2, the joint benefit of including the H-types is smaller than the L-types' information rents. In this case, the platform raises the price to $p^* = \alpha_L - \theta_L w_s$, and deter the H-types from joining the platform.

2.2 Platform Liability

This subsection explores the social desirability and optimal design of platform liability for harm to user-bystanders, taking the level of firm liability w_s as fixed.

We begin by presenting a benchmark where the platform is not liable for the harm, $w_p = 0$. In this benchmark, the platform has no incentive to engage in costly auditing to detect and remove harmful firms. However, the H-type firms may be deterred from participating on the platform if firm liability w_s and/or the interaction price p is large such that $\alpha_H - \theta_H w_s \leq p$.

Proposition 1. *(Firm-Only Liability.) Suppose that the platform is not liable for harm to users, $w_p = 0$, and firm liability is $w_s \in (0, d]$. There exists a unique threshold $\widetilde{w} = \widetilde{w}(\lambda) \in \left[\widehat{w}, \frac{\alpha_H}{\theta_H}\right)$, where $\widetilde{w}(\lambda)$ weakly increases in the number of H-types, λ.*[51]

[51] If $\theta_L/\theta_H \geq \alpha_L/\alpha_H$ then $\widetilde{w}(\lambda) = \widehat{w}$ for all λ.

1. *If $w_s \leq \widehat{w}$ then the platform sets $p^* = \alpha_L - \theta_L w_s$, attracts the H-type firms, and does not invest in auditing, $e^* = 0 < e^{**}$. The platform's auditing incentives are socially insufficient.*

2. *If $w_s \in (\widehat{w}, \widetilde{w})$ then the platform sets $p^* = \alpha_H - \theta_H w_s$, attracts the H-type firms, and does not invest in auditing, $e^* = 0 < e^{**}$. The platform's auditing incentives are socially insufficient.*

3. *If $w_s \geq \widetilde{w}$ then the platform sets $p^* = \alpha_L - \theta_L w_s$ and deters the H-type firms. The first-best outcome is achieved.*

Proposition 1 describes the market outcome when the firms, and only the firms, are liable for the harm to user-bystanders. In case 1, since $w_s \leq \widehat{w}$ the L-types are marginal. The platform cannot deter the H-types without excluding the L-types. So the platform attracts the H-type firms and does not invest in costly auditing to detect and remove them. This is obviously a socially undesirable outcome.

If firm liability is above the threshold, $w_s > \widehat{w}$, then the H-types are marginal. Increasing w_s reduces the joint value of attracting the H-types and therefore motivates the platform to deter them. In case 2, $w_s \in (\widehat{w}, \widetilde{w})$ and the platform charges $p^* = \alpha_H - \theta_H w_s$ and attracts the H-types. In case 3, $w_s \geq \widetilde{w}$ and the platform charges a higher price $p^* = \alpha_L - \theta_L w_s$ and deters the H-types.

Should platforms be held liable for the harm suffered by users? Proposition 1 establishes that platform liability is unnecessary when the firms themselves are held sufficiently liable for harm to bystanders, $w_s \geq \widetilde{w}$. However, when the firms are very judgment proof ($w_s < \widetilde{w}$) and the platform faces no liability, the private and social incentives diverge. The next proposition characterizes the optimal platform liability rule, w_p^*.

Proposition 2. *(Optimal Platform Liability.) Suppose firm liability is $w_s \in (0, d]$. The socially-optimal platform liability for harm to users, w_p^*, is as follows:*

1. *If $w_s \leq \widehat{w}$ then $w_p^* = d - w_s - \left(1 - \frac{\theta_L}{\theta_H}\right)(\widehat{w} - w_s) \in (0, d - w_s]$ achieves the second-best outcome. The platform sets $p^* = \alpha_L - \theta_L w_s$ and attracts the H-type firms. The platform's auditing incentives are socially efficient, $e^* = e^{**}$.*

2. *If $w_s \in (\widehat{w}, \widetilde{w})$ then there exists a threshold $\underline{w}_p > 0$ where any $w_p^* \in [\underline{w}_p, d - w_s]$ achieves the first-best outcome. The platform sets $p^* = \alpha_L - \theta_L w_s$ and deters the H-type firms.*

3. *If $w_s \geq \widetilde{w}$ then platform liability is unnecessary. Any $w_p^* \in [0, d - w_s]$ achieves the first-best outcome. The platform sets $p^* = \alpha_L - \theta_L w_s$ and deters the H-type firms.*

Proposition 2 describes how platform liability can be designed to maximize social welfare. Recall that the platform has two possible mechanisms to reduce the harm to users: the price per interaction p and the audit intensity e. If feasible, the pricing mechanism is privately and socially more efficient than the auditing mechanism, as the pricing mechanism can deter the H-types without the need for costly audits. The pricing mechanism is feasible if and only if firm liability is above a threshold, $w_s > \widehat{w}$.

In case 1, firm liability is below the threshold ($w_s \leq \widehat{w}$) and the L-type firms are marginal. From Proposition 1 we know that firm-only liability fails to deter the H-types and gives the platform no incentive to audit and remove the H-types. Imposing liability on the platform motivates the platform to take the socially efficient 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 second-best outcome is achieved 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 second-best outcome is achieved when the platform bears full residual liability, $w_p^* = d - w_s$.

Note that, in case 1, the optimal platform liability, w_p^*, decreases in w_s. From the social planner's perspective, platform liability and firm liability are *substitutes*. Intuitively, when firm liability (w_s) is larger, the H-type firms get less rent, which reduces the platform's auditing incentives; at the same time, the uncompensated harm for users becomes lower and the firms are less willing to pay, which raises the platform's auditing incentives. In equilibrium, the second effect dominates, so the increase in w_s leads to more auditing. To prevent excessive auditing, it is efficient to reduce platform liability.

In case 2, the firms' liability is in an intermediate range and the H-type firms are marginal. According to Proposition 1, without platform liability, the platform would charge p_H and attract the H-type firms since the joint value of including the H-types (for the platform and the firms) is larger than the L-type firms' rents. Since the firms' rent is independent of w_p while the joint value of keeping the H-types decreases in w_p, the social planner can motivate the platform to raise the price and thus deter the H-types by imposing residual liability on the platform, $w_p^* = d - w_s$. The first-best outcome is obtained.

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

This section investigated the need for platform liability when the firms that participate on the platform cause harm to user-bystanders. Our analysis has important implications for the design of liability rules. First, if firms have deep pockets and can compensate the user-bystanders for the harm that they cause, then platform liability is unwarranted. Placing liability on the firms themselves is socially optimal, as it solves the problem of negative externalities in the user-bystanders. Firms that pose excessive risks to users are deterred from participating on the platform by the threat of future litigation.

Second, if firms are judgment proof or can evade liability in other ways, then platform liability is socially desirable. Holding the platform liable for some or all of the residual harm has two potential benefits. First, the platform may raise the price that it charges to the firms, which will help to deter firms that pose excessive risks to users. Second, the platform will invest resources to detect and remove risky firms from the platform. Interestingly, we show that the socially-optimal level of platform liability may be less than full. When the firms have very limited resources, then holding the platform fully responsible for the residual harm would lead the platform to overinvest in auditing.

3 Extensions

The baseline model considers a monopoly platform where interactions between firms and users do not require the users' consent and all the users participate. In this section, we examine several important extensions, including retail platforms with consensual market transactions, platform competition, and heterogeneous users who make participation decisions. We will show that platform liability can still increase social welfare, though the optimal level of platform liability may be different from that in the baseline model.

3.1 Retail Platforms

We now extend the analysis to consider a retail platform where the firms are the sellers of a product or service and the users are sophisticated consumers. Interactions between the firms and the users are *market transactions* that require the users' consent. We will show that the optimal platform liability is (weakly) lower than in the baseline model.

This extension has many practical applications. Most of the products that are bought and sold through Amazon are manufactured and distributed by third-party vendors. Even relatively straightforward products like computer chargers and lightbulbs are of varying quality and safety. The third-party vendors, especially those without existing reputations, would 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. Extending liability to Amazon gives the platform the incentive to monitor third-party vendors and block dangerous products from reaching the marketplace.

As in the baseline model, there are two types of firm, H and L. The type-i firm produces a good or service at cost c_i which causes accidents with probability θ_i. The unsafe products are cheaper to produce, $c_H < c_L$, and cause harm more frequently, $\theta_H > \theta_L$. A user-consumer's gross value from the good is α_0. Letting $\alpha_i = \alpha_0 - c_i$, the net interaction value 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 the user accepts the price offer t then the user pays t to the firm, and the firm pays p to the platform.[52]

The users' willingness to transact with the firms depends on their beliefs about product safety. Users do not observe the safety of the product directly, or the auditing efforts of the platform, but are sophisticated and form beliefs that are, in equilibrium, correct.[53]

[52] The results would be the same if the firms pay the platform a percentage of their gross revenue.

[53] Our model may be adapted to consider naïve consumers. If a consumer is unaware of product risks, then each transaction imposes a negative externality on the consumer's future self. Since the consumer's future self is essentially a non-consenting "bystander" to the transaction, the analysis of the baseline model and all of its implications apply. The case for holding retail platforms liable would be stronger.

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

$$E(\theta|e) = \frac{(1-e)\lambda\theta_H + (1-\lambda)\theta_L}{(1-e)\lambda + (1-\lambda)}, \tag{12}$$

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 $\theta^0 = E(\theta|0) = \lambda\theta_H + (1-\lambda)\theta_L$ be the probability of harm when the platform does not audit ($e = 0$).[54] If a user believes that the platform invests e^r in auditing, then the expected probability of harm from an "average" transaction is $\theta^r = E(\theta|e^r)$. Note that, if all the H-types are deterred, then the expected probability of harm is $\theta^r = \theta_L$.

There is no separating equilibrium where the H-types and L-types charge different prices and have positive sales. If such a separating equilibrium existed, users would have correct beliefs about the firms' types. Since $\alpha_H - \theta_H d < 0$, jointly-beneficial transactions between users and H-types cannot occur.[55] In any pooling equilibrium where both types of firm seek to join the platform and offer the same t, the type-i firm's surplus is $t - (\theta_i w_s + c_i) - p$ and the two types have equal surplus when $w_s = \widehat{w}$ as defined in (5) in the baseline model.[56]

Given the users' belief of e^r, in equilibrium the retail price t^r cannot be larger than the users' maximum expected willingness to pay. We will construct perfect Bayesian equilibria with

$$t^r = \alpha_0 - \theta^r(d-w), \tag{13}$$

so consumer surplus is zero.[57] $\theta^r(d-w)$ is the users' expected uncompensated harm. The consumers believe that any firm charging a different price would have at least the average probability of harm, θ^r. No firm has an incentive to raise its price, as otherwise the users would not buy from the firm.

Case 1: $\boldsymbol{w_s \leq \widehat{w}}$. Since the L-type firms are marginal, the platform sets p^r to extract rents from the L-type firms, $p^r = t^r - (\theta_L w_s + c_L)$.[58] Using (13) and $\alpha_L = \alpha_0 - c_L$,

$$p^r = \alpha_L - \theta_L w_s - \theta^r(d-w). \tag{14}$$

[54] e^{**} is defined in equation (3).

[55] Assumption A2 implies that even if the platform does not audit at all, the gross profit for the L-type firms (before paying p to the platform) is positive. Thus, this assumption guarantees that an equilibrium exists for all assignments of liability, w_s and w_p. It is possible to have a separating equilibrium where the platform deters all the H-types through the pricing mechanism.

[56] They have equal surplus if $t-(\theta_H w_s + c_H) - p = t - (\theta_L w_s + c_L) - p$. Substituting $c_i = \alpha_0 - \alpha_i$ and rearranging gives $w_s = \widehat{w} = (\alpha_H - \alpha_L)/(\theta_H - \theta_L)$.

[57] This equilibrium maximizes the platform's profits. See the proof of Proposition 3. Other equilibria may exist: Any price $t \in (\alpha_0 - \theta_H(d-w), \alpha_0 - \theta^r(d-w))$ can be an equilibrium if the users hold the off-equilibrium belief that any firm charging a different price would be the H-type. However, in such equilibria, firms are playing a dominated strategy: their profits would be higher if they raise the prices.

[58] See the proof in the Appendix.

Comparing p^r to its counterpart p^* (see (6)) in the baseline model reveals an important difference: the interaction price paid by the firms (14) reflects the user-consumers' expected uncompensated harm, $\theta^r(d-w)$.

We now explore the platform's auditing incentives. Substituting p^r from (14), $S(e)$ from (2), and $\widehat{w}$ from (5) into (7) gives the platform's profit function

$$\Pi(e) = S(e) - v - (1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) \\ + [(1-e)\lambda(\theta_H - \theta^r) + (1-\lambda)(\theta_L - \theta^r)](d-w). \quad (15)$$

The platform's profits $\Pi(e)$ diverge from social welfare $S(e)$ for two reasons.[59] First, the platform does not internalize the information rents that are enjoyed by each retained H-type firm, $(\theta_H - \theta_L)(\widehat{w} - w_s)$. Second, the platform does not internalize the users' unanticipated losses or gains (relative to their expectations).[60] The expression in the second line of (15) represents the user's *unanticipated loss or gain* when the platform deviates and invests $e \neq e^r$.[61]

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

$$\Pi'(e^r) = S'(e^r) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - \lambda(\theta_H - \theta^r)(d-w) = 0 \quad (16)$$

where $w = w_s + w_p$. Note that the platform's auditing incentive may be socially insufficient ($e^r < e^{**}$) or excessive ($e^r > e^{**}$). The incentive is insufficient (or excessive) if and only if the H-type firms' rent, $\lambda(\theta_H - \theta_L)(\widehat{w} - w_s)$, is smaller (or larger) than the users' loss relative to their expectations $\lambda(\theta_H - \theta^{**})(d-w)$ where θ^{**} is the probability of harm if the auditing effort is socially efficient ($e = e^{**}$).

Case 2: $\boldsymbol{w_s > \widehat{w}}$. Suppose that the platform sets a high price and deters the marginal H-type firms. Since consumers rationally anticipate that H-types are deterred, $\theta^r = \theta_L$, the retail price is $t = \alpha_0 - \theta_L(d-w)$. The platform charges the firms a transaction price $p = t - (\theta_L w_s + c_L)$ or $p = \alpha_L - \theta_L(d - w_p)$. The platform's profit is $(1-\lambda)(p - \theta_L w_p)$ or

$$(1-\lambda)(\alpha_L - \theta_L d). \quad (17)$$

If the platform deters the H-type firms, the platform extracts all of the social surplus associated with the transactions between users and the L-type firms.

Now suppose that the platform sets a low price and accommodates the H-type firms.[62] The platform's profits would be strictly lower in this case. To see why, observe that the

[59] If the platform could commit to e then they would internalize the users' losses and gains. With commitment, the platform's auditing incentives may be socially excessive but not socially insufficient.

[60] Since the users cannot observe e, the platform's off-the-equilibrium-path choice of auditing may diverge from the users' expectations. If $e < e^r$ ($e > e^r$) then the users experience an unanticipated loss (gain) and expression in the second line of (15) is negative (positive).

[61] If $e = e^r$ then this term equals zero.

[62] As in the previous section where users were bystanders, the platform would have no incentive to audit and remove the H-types from the platform. This is by revealed preference, as it could deter the H-types by raising the price.

incremental social benefit of attracting the H-type firms is negative, $\lambda(\alpha_H - \theta_H d) < 0$. If the platform accommodates the H-types, then the consumers, firms, and platform are jointly worse off. In equilibrium, the consumers are compensated for purchasing the less safe products and the L-type firms capture rents. Therefore the platform's incremental profit from attracting the H-types is unambiguously negative.[63]

Proposition 3. *(Retail Platform.) Suppose firm liability is $w_s \in (0, d]$. Let $\theta^{**} = E(\theta|e^{**})$. The socially-optimal platform liability for harm to user-consumers, w_p^r, is as follows:*

1. *If $w_s \leq \widehat{w}$ then $w_p^r = d - w_s - \left(\frac{\theta_H - \theta_L}{\theta_H - \theta^{**}}\right)(\widehat{w} - w_s) \in (0, d - w_s]$ achieves the second-best outcome. The platform sets $p^r = \alpha_L - \theta_L w_s - \theta^{**}(d - w_s)$ and attracts the H-type firms. The platform's auditing incentives are socially efficient, $e^r = e^{**}$.*

2. *If $w_s > \widehat{w}$ then platform liability is unnecessary. Any $w_p^r \in [0, d - w_s]$ achieves the first-best outcome. The platform sets $p^r = \alpha_L - \theta_L(d - w_p^r)$ and deters the H-type firms.*

Comparing Proposition 3 to Proposition 2 in the baseline model reveals both similarities and differences. As in the baseline model, if $w_s = \widehat{w}$, then the second-best outcome is achieved when the platform bears full residual liability, $w_p^r = d - w_s$. If $w_s < \widehat{w}$, it is socially efficient to have the platform bear some but not all the residual damage, $w_p^r < d - w_s$. If the platform was responsible for the residual harm then the platform would overinvest in auditing. However, if $w_s < \widehat{w}$, the optimal platform liability is smaller than in the baseline model, because interactions require users' consent and the platform has stronger incentives to assure higher product safety to stimulate demand.

Moreover, in contrast to the baseline model, if $w_s \leq \widehat{w}$, the optimal platform liability, w_p^r, increases in w_s. From the social planner's perspective, platform liability and firm liability are *complements*. In Proposition 2 where the users are bystanders, platform liability and firm liability are *substitutes*. We now develop intuition for this fundamental difference.

When users are *bystanders*, liability encourages the platform to internalize the externalities imposed on the firms and users. In Proposition 2, w_p^* satisfies

$$(\theta_H - \theta_L)(\widehat{w} - w_s) = \theta_H(d - w_s - w_p^*). \tag{18}$$

The left-hand side are the rents enjoyed by the H-type firms and the right-hand side are the users' *uncompensated harm caused by the H-types.* When firm liability w_s rises, both sides fall. However, the drop in the firms' rent on the left is smaller than the drop in the users' uncompensated harm on the right. Holding w_p fixed, the platform would invest too much in auditing. To prevent excessive auditing, platform liability w_p must fall. This is why firm liability and platform liability were substitutes in the baseline model.

[63] In the baseline model of Section 2 where the users are bystanders, given $w_s > \widehat{w}$, the platform may (inefficiently) attract the H-type firms if the joint value for the platform and firms is larger than the firms' rent.

By contrast, when users are *consumers*, the retail price t^r paid by the users to the firms (and the price p^r paid by the firms to the platform) reflects the users' beliefs of the probability of harm. In Proposition 3, when the users are consumers, w_p^r satisfies

$$(\theta_H - \theta_L)(\widehat{w} - w_s) = (\theta_H - \theta^{**})(d - w_s - w_p^r). \tag{19}$$

Now the right-hand side reflects the users' *uncompensated harm beyond their expectations.* As in the baseline model, when firm liability w_s rises, both sides fall. However, the drop in the firms' rent on the left is *bigger* than the drop in the users' uncompensated harm (beyond their expectations) on the right. Holding w_p fixed, the platform would invest *too little* in auditing. To restore the efficient incentives for auditing, platform liability should be raised. This is why platform liability and firm liability are complements in the retail platform extension.

Corollary 1. *Suppose $w_s \leq \widehat{w}$. When the users are bystanders, the optimal platform liability decreases in w_s; when the users are consumers, the optimal platform liability increases in w_s.*

As a remark, the analysis above assumed that the platform removed discovered H-types from the platform. What would happen if the platform is required to disclose the audit results to the consumers, and the consumers decide for themselves whether to interact with the known H-types? Absent platform liability ($w_p = 0$), a rational consumer would decline to interact with a known H-type ex post.[64] Although ex post efficiency would be obtained without platform liability, the platform would have insufficient incentives to audit the sellers ex ante.[65] At the other extreme, with full platform liability ($w_p = d$), a rational consumer would interact with a known H-type.[66] That is, disclosure would not deter harmful interactions. These observations underscore the importance of granting retail platforms the discretion to remove bad actors rather than relying on disclosure alone.[67]

3.2 Platform Competition

We now extend our baseline model (where users are bystanders) by considering two competing platforms, Platform 1 and Platform 2. Users can join both platforms, but each

[64] The joint surplus for a consumer and an H-type firm from their transaction is $\alpha_H - \theta_H(d - w_p) - p^r$. If $w_p = 0$ then the joint surplus is negative, $\alpha_H - \theta_H d - p^r < 0$.

[65] If consumers are naïve and underestimate product risks then the platform's incentive to audit and disclose negative information would be further diluted. Recent empirical work by Culotta et al. (2022) shows that Airbnb may limit the flow of negative safety reviews.

[66] If $w_p = d$ then the consumer and seller's joint surplus is positive, $\alpha_H - p^r > 0$. The accident losses are externalized on the platform.

[67] In some settings, consumers can take pre- and post-sale precautions to mitigate the harm. A shopper can read the product reviews posted by others before purchase and take further precautions after receiving the item. The optimal design of platform liability must strike a balance between creating incentives for the platform to detect and remove harmful products and creating incentives for consumers to be prudent.

firm can only join one platform.[68] Thus, the platforms compete for firms but not for users.[69] We will show that platform liability can still be socially beneficial and the optimal platform liability may be higher or lower relative to the baseline model.

Denote the platforms' prices and auditing efforts as p_j and e_j, $j = 1, 2$. In stage 1, the platforms set their prices simultaneously. In contrast to the baseline model, imposing punitive damages may be necessary to implement the first-best outcome. We therefore allow $w = w_s + w_p$ to be higher or lower than d. The timing and the other assumptions are otherwise identical to the baseline model

Case 1: $\boldsymbol{w_s \leq \widehat{w}}$. In this case, the L-type firms are marginal. We show in the appendix that, as long as both platforms are active, the platforms get zero profits, charge the same interaction price $p_1 = p_2$, and take the same auditing effort $e_1 = e_2$ in any (symmetric or asymmetric) equilibrium.[70] Without loss of generality, we focus on the symmetric perfect Bayesian equilibria where each platform attracts half of the firms. The equilibrium price p^c and auditing effort e^c (if it is an interior solution) satisfy

$$(1 - e^c)\lambda(p^c - \theta_H w_p) + (1 - \lambda)(p^c - \theta_L w_p) - c(e^c) = 0, \tag{20}$$

$$-\lambda(p^c - \theta_H w_p) - c'(e^c) = 0. \tag{21}$$

If there is no platform liability, $w_p = 0$, then (20) and (21) imply that the platforms charge $p^c = 0$ and do not waste resources auditing the firms, $e^c = 0$. If there is platform liability, $w_p > 0$, then the platforms will engage in costly auditing, $e^c > 0$. To see why, suppose to the contrary that $w_p > 0$ and the platforms do not audit, $e^c = 0$. Then $c(e^c) = 0$ and the zero-profit condition (20) implies $p^c - \theta_H w_p < 0 < p^c - \theta_L w_p$. Since the platforms are losing money on each retained H-type, condition (21) implies that the platforms would invest $e^c > 0$ to detect and remove the H-types, a contradiction. So, if platforms are liable, $w_p > 0$, the platforms invest in auditing, $e^c > 0$.

Interestingly, platform competition increases the platforms' auditing incentives relative to the monopoly benchmark. The reason is that since the platforms compete to serve the L-type firms, the equilibrium interaction price is lower than in the baseline monopoly model, $p^c < p^* = \alpha_L - \theta_L w_s$. This implies that the price-cost margins from serving the H-type firms is lower, too.[71] Competing platforms will naturally spend more resources to detect and remove the harmful H-type firms than a monopoly platform: if $w_p > 0$ then $e^c > e^*$. Accordingly, the optimal platform liability is lower than before. We also show in the appendix that, similar to the observation in the baseline model, the competing platforms' auditing incentives can be socially insufficient or excessive.[72]

[68] In practice, many firms choose single-homing due to fixed costs or reputation concerns.

[69] In some applications, users may join only one platform due to switching costs. If a certain proportion of users are single-homing, then the platforms would compete for these users, which would raise their incentives to deter or remove the H-type firms. Accordingly, the optimal platform liability can be lower.

[70] See the proof of Proposition 4.

[71] Condition (21) confirms that $de^c/dp^c < 0$. Therefore, since $p^c < p^*$ we have $e^c > e^*$.

[72] See the proof of Proposition 4.

Case 2: $\boldsymbol{w_s > \widehat{w}}$. In this case, the H-type firms are marginal. If there is no platform liability, $w_p = 0$, the platforms compete the price down to $p_1 = p_2 = 0$. If $\alpha_H - \theta_H w_s < 0$ then the H-types do not join the platform; the first-best outcome is obtained without platform liability. However, if $\alpha_H - \theta_H w_s \geq 0$ then the H-types do join the platform, a socially-undesirable outcome. In this environment, platform liability operates as a Pigouvian tax to get the platforms to raise the interaction price and deter the harmful firms. If $\theta_L > 0$, the social planner can set $w_p \in (\frac{\alpha_H - \theta_H w_s}{\theta_L}, \frac{\alpha_L - \theta_L w_s}{\theta_L}]$, under which the equilibrium price is $p_1 = p_2 = \theta_L w_p \in (\alpha_H - \theta_H w_s, \alpha_L - \theta_L w_s]$. This deters the H-types from joining the platform and restores social optimality. The L-types are willing to join the platform because $\alpha_L - \theta_L w_s - \theta_L w_p \geq 0$.[73]

Proposition 4. *(Platform Competition.) Suppose that firm liability is $w_s \in (0, d]$. The socially-optimal liability for the competing platforms, w_p^c, is as follows:*

1. *If $w_s \leq \widehat{w}$ then there exists a unique $w_p^c < w_p^*$ that achieves the second-best outcome. The platforms set $p_1 = p_2 = p^c < p^*$ and attract the H-type firms. The platforms' auditing incentives are socially efficient, $e^c = e^{**}$.*

2. *If $w_s \in (\widehat{w}, \frac{\alpha_H}{\theta_H}]$ and $\theta_L > 0$, then any $w_p^c \in (\frac{\alpha_H - \theta_H w_s}{\theta_L}, \frac{\alpha_L - \theta_L w_s}{\theta_L}]$ achieves the first-best outcome. The platforms set $p_1 = p_2 = \theta_L w_p^c$ and deter the H-type firms.*

3. *If $w_s > \frac{\alpha_H}{\theta_H}$ then platform liability is unnecessary. Any $w_p^c \in [0, \frac{\alpha_L - \theta_L w_s}{\theta_L}]$ achieves the first-best outcome. The platforms set $p_1 = p_2 = \theta_L w_p^c$ and deter the H-type firms.*

Comparing Proposition 4 to Proposition 2 reveals how competition changes the socially-optimal level of platform liability.

If the firms are very judgment proof, $w_s \leq \widehat{w}$, then the socially-optimal level of platform liability with platform competition is lower than in the baseline model of monopoly, $w_p^c < w_p^*$. As in the baseline model, platform liability encourages the platforms to detect and remove the H-type firms from the platform. However, since the equilibrium interaction price is lower with competition, $p^c < p^*$, the platform's surplus from retaining the harmful H-types, $p^c - \theta_H w_p$ is lower too. Therefore the platforms' incentives to detect and remove the harmful H-type firms is stronger with competition. As a result, the optimal platform liability is lower when platforms are more competitive.

If the firms are modestly judgment proof, $w_s > \widehat{w}$, then the socially-optimal level of platform liability with platform competition is (weakly) higher than in the baseline model, $w_p^c \geq w_p^*$. In case 2 in Proposition 4, since $\alpha_H - \theta_H w_s \geq 0$, the H-type firms would participate if the interaction price is zero. Platform liability gets the platforms to raise the interaction price. If $\theta_L > 0$, to raise the price sufficiently to deter the H-types, platform liability must satisfy $w_p^c > \frac{\alpha_H - \theta_H w_s}{\theta_L}$, which is larger than the minimum liability

[73] If $\theta_L = 0$, there is no equilibrium with the first-best outcome being achieved. To see this, note that, the first-best outcome would be achieved when $p_1 = p_2 > \alpha_H - \theta_H w_s$. However, since $\theta_L = 0$, the platforms do not have any cost on the equilibrium path, so they would keep reducing the prices.

necessary for deterrence in Proposition 2.[74] Intuitively, since platform competition lowers the equilibrium interaction price, platform liability must be larger with platform competition. Note that $\frac{\alpha_H - \theta_H w_s}{\theta_L}$ may be even larger than $d - w_s$, that is, punitive damages may be necessary to implement the first-best outcome.[75] In case 3 in Proposition 4, if $\alpha_H - \theta_H w_s < 0$, the H-types do not participate even if the interaction price is zero so platform liability is unnecessary.

These observations suggest that policies encouraging platform competition should be complemented by changes in platform liability. A report written by Cremer, et al. and published by the European Commission (2019) raised concerns about increased concentration in platform markets.[76] The anti-trust authorities in both the U.S. and the EU have initiated investigations and lawsuits against platforms. For example, the Federal Trade Commission in the U.S. filed a lawsuit against Facebook, asking the court to force it to sell WhatsApp and Instagram.[77] The potential changes in market competition would affect platforms' incentives to deter or remove harmful firms, which would call for changes in platform liability.

In fact, the Digital Services Act and Digital Markets Act proposed in 2020 by the European Commission try to achieve the two goals together: creating a safer digital space and establishing a level playing field (to foster innovation and competitiveness).[78] Holding platforms liable for user harm can improve safety in the digital space. However, our analysis implies that, if these policies increase platform competition, the socially optimal platform liability could be higher or lower (depending on the extent to which the harmful firms are judgment proof).

3.3 User Participation

Our baseline model assumed that the value of the quasi-public good v was the same for all users and sufficiently high so that all of the users joined the platform, regardless of their beliefs about platform safety. In this section, we extend the model by considering the participation decisions of heterogeneous users. We will show that platform liability has the additional effect of stimulating user participation and that the level of optimal platform liability can be higher than in the baseline model.

Suppose that the users' valuations of the quasi-public good are drawn from density $f(v) > 0$ for $v \in [0, \infty)$, with cumulative density $F(v)$. As in the baseline model, the platform can charge a price p per interaction to the firms, and take auditing effort e per firm. Although the users do not observe the platform's auditing effort, they observe w_s and w_p, and form the correct belief of e in equilibrium. Note that there are economies of

[74]As defined in the proof of Proposition 2, $\underline{w}_p = \frac{\alpha_H}{\theta_H} - w_s - \frac{1-\lambda}{\lambda}(1 - \frac{\theta_L}{\theta_H})(w_s - \widehat{w})$, which is strictly lower than $\frac{\alpha_H - \theta_H w_s}{\theta_L}$ as long as $w_s > \widehat{w}$.

[75]For example, $\frac{\alpha_H - \theta_H w_s}{\theta_L} > d - w_s$ if w_s is close to $\widehat{w}$.

[76]See https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

[77]See https://www.reuters.com/technology/us-ftc-says-court-should-allow-antitrust-lawsuit-against-facebook-go-forward-2021-11-17/

[78]See https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

scale in (per-firm) auditing, so that both the private and the socially optimal incentives for auditing depend on the users' participation rate.[79] The users have the option to join the platform for free.[80]

Assumption A2 implies that it is socially efficient for all the users to participate. Thus, the first-best outcome is achieved if the H-type firms do not join but all the users participate. As in the baseline model, full deterrence may not be possible. If the H-type firms seek to join the platform, then auditing is necessary to reduce the social harm. In the second-best benchmark, social welfare is

$$S(e, \widehat{v}) = \int_{\widehat{v}} [v + \lambda(1-e)(\alpha_H - \theta_H d) + (1-\lambda)(\alpha_L - \theta_L d)] f(v) dv - c(e), \tag{22}$$

where $\widehat{v}$ is the value of the marginal user and given by

$$\widehat{v}(e, w) = (\lambda(1-e)\theta_H + (1-\lambda)\theta_L)(d - w). \tag{23}$$

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 would be e^{**}, the same as in the baseline model.

Case 1: $\boldsymbol{w_s \leq \widehat{w}}$. In this case, the L-type firms are marginal and the platform charges $p^u = \alpha_L - \theta_L w_s$. The platform's profit function may be written as:

$$\begin{aligned}\Pi(e, \widehat{v}) = S(e, \widehat{v}) + \int_{\widehat{v}} \Big\{ &- (1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) \\ &+ ((1-e)\lambda\theta_H + (1-\lambda)\theta_L)(d - w) - v \Big\} f(v) dv, \end{aligned} \tag{24}$$

where $\widehat{v}$ is the marginal user defined in (23). Since the platform chooses its auditing effort ex post, given $\widehat{v}$, the platform's auditing effort e^u (if it is positive) satisfies[81]

$$\begin{aligned}\frac{\partial \Pi(e^u, \widehat{v})}{\partial e} = \frac{dS(e^u, \widehat{v})}{de} &+ \int_{\widehat{v}} [\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - \lambda\theta_H(d - w)] f(v) dv \\ &+ \lambda\theta_H(d - w)\frac{\partial S(e^u, \widehat{v})}{\partial \widehat{v}} = 0. \end{aligned} \tag{25}$$

[79]If auditing is per interaction instead of per firm, the results are similar. The analysis is available upon request.

[80]The platform might also charge a membership fee $m \geq 0$ to each user. However, we show in the appendix that the platform sets $m = 0$ in equilibrium if $\alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)d$ is sufficiently large. Consistent with the literature, when there are strong cross-side network effects, platforms find it optimal to charge only one side of participants and offer free services to the other side. We maintain the assumption that $\alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)d$ is sufficiently large such that the platform does not charge the users.

[81]See the proof of Proposition 5.

The platform's auditing incentives diverge from the social planner's in several important aspects. As in equation (10) in the baseline model, when the platform increases e, the removed H-types lose their information rents, $\lambda(\theta_H - \theta_L)(\widehat{w} - w_s)$ and the users' uncompensated loss is reduced by $\lambda\theta_H(d - w)$. If $w_p = w_p^*$ as defined in the baseline model, these two effects offset each other. The last term in (25) is the users' benefit of increased participation.

If $w_s < \widehat{w}$, the level of optimal platform liability is higher than in the baseline model, $w_p^u > w_p^*$.[82] To understand why, recall that the user's participation threshold $\widehat{v}(e, w)$ in (23) is a decreasing function of e and w. An increase in w_p stimulates user participation for two reasons. First, holding e fixed, when w_p increases users who participate receive a larger "rebate." Second, an increase in w_p leads the platform to increase its effort e. Note that the second-best outcome with $\widehat{v} = 0$ and $e = e^{**}$ cannot be achieved. To attract all the users we must have $w_p = d - w_s$, but this would motivate the platform to invest excessively in auditing.

If $w_s = \widehat{w}$, the level of optimal platform liability is the same as in the baseline model, $w_p^u = w_p^* = d - w_s$, which attracts all the users and motivates the platform to choose $e = e^{**}$.

Case 2: $\boldsymbol{w_s > \widehat{w}}$. Since the H-types are marginal, the first-best outcome may be obtained with sufficiently high platform liability. First, suppose $w_s \geq \widetilde{w}$, where $\widetilde{w}$ is defined in Proposition 1. As shown in Proposition 1, the platform charges $p^u = \alpha_L - \theta_L w_s$, which deters all the H-type firms. Anticipating that all the H-type firms are deterred, the users participate if $v \geq (1 - \lambda)\theta_L(d - w)$. In this case, the first-best outcome is achieved when $w_p = d - w_s$. Second, suppose $w_s \in (\widehat{w}, \widetilde{w})$. As shown in Proposition 2, given $w_p \geq \underline{w}_p$, the platform charges $p^u = \alpha_L - \theta_L w_s$, which deters all the H-type firms. Again, the first-best outcome is achieved when $w_p = d - w_s$.

Proposition 5. *(User Participation.) Suppose firm liability is $w_s \in (0, d]$. The socially-optimal platform liability for harm to users, w_p^u, is as follows:*

1. *If $w_s < \widehat{w}$ then $w_p^u > w_p^*$. The platform sets $p^u = \alpha_L - \theta_L w_s$. The second-best outcome is not achieved.*

2. *If $w_s = \widehat{w}$ then $w_p^u = d - w_s$ achieves the second-best outcome. The platform sets $p^u = \alpha_L - \theta_L w_s$ and chooses the efficient auditing effort $e^u = e^{**}$. All users participate.*

3. *If $w_s > \widehat{w}$ then $w_p^u = d - w_s$ achieves the first-best outcome. The platform sets $p^u = \alpha_L - \theta_L w_s$ and deters the H-type firms. All users participate.*

To summarize, as in the baseline model, platform liability can motivate the platform to take more auditing effort or raise the interaction price, which removes or deters the harmful firms. Additionally, platform liability stimulates user participation. So, the optimal level of platform liability is (weakly) higher than in the baseline model.

[82] See the proof of Proposition 5.

3.4 Other Extensions

Pricing Structure. Our analysis assumed a very simple pricing structure where the platform monetized its activities through an interaction price paid by the firms. Alternatively, we could have assumed that the firms pay a lump-sum membership fee.[83] Our results would be unaffected if the membership fee is paid by the firms that are retained by the platform. With additional instruments, such as a non-refundable application fee or bond, the platform's ability to deter risky firms would be enhanced and the platform could save resources on auditing. However, the H-types may still join. To see this, suppose that the firms are very judgment proof ($w_s \leq \widehat{w}$) so that the L-types are marginal. If the H-types do not join the platform, the platform would not take any auditing effort. But anticipating this, the H-types would deviate to join. In this case, there is no equilibrium where the H-types are fully deterred.[84] Therefore, platform liability can increase the platform's auditing incentives.

False Positives. Our analysis assumed that there were no "false positives." The auditing efforts of the platform did not erroneously remove the L-type firms. Several new insights emerge when the analysis is extended to include false positives. First, the second-best auditing effort is lower than in our baseline model (since it is socially efficient for L-types to remain on the platform). Second, the platform has weaker incentives to invest in auditing than in the baseline model (since the platform loses revenue when it excludes the L-types). Third, the platform's incentives are even weaker relative to the social incentives. When choosing its audit intensity, the platform does not account for the positive externality that excluding the L-types confers on the platform users. It follows that the optimal platform liability is (weakly) larger when there are false positives, compared to our baseline model.[85]

Litigation Costs. Our baseline model assumed that litigation was free. In reality, bringing a lawsuit is expensive and requires the services of a lawyer. The implications of litigation costs for the design of optimal platform liability is nuanced. On the one hand, when the L-type firms are marginal, litigation costs reduce the H-type firms' information rent 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. Without a meaningful threat of litigation, the platform has little incentive to deter and remove harmful firms. Thus, a higher level of liability may be necessary to encourage plaintiffs (and their lawyers) to sue and raise the platform's auditing incentives.[86] On the other hand, when the H-type firms are marginal, litigation costs raise the platform's

[83] In practice, many firms have budget constraints so that they could not make a large upfront payment when joining platforms.

[84] There can be two possible equilibria: One where the platform attracts the H-types as in the baseline model; the other (a mixed-strategy equilibrium) where the platform randomizes on auditing and the H-types randomize on participation.

[85] See Appendix B for the formal analysis.

[86] See Appendix B for the formal analysis.

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.

Same-Side Harms. Our baseline model considered a setting with cross-side harms: Firms on one side of the platform harmed the users on the other side of the platform. Our model also applies to same-side harms where some users may harm other users. Consider for example a social networking platform where most user-generated content is perfectly safe but some of it is socially harmful (e.g., misinformation and hate speech). Suppose further that the advertising revenue that the platform enjoys is proportional to the volume of shared content, both safe and harmful. If the users are judgment proof, and cannot be held accountable for the harmful content that they post, then holding the platform liable may make sense. Without platform liability, the platform has a financial incentive to facilitate the posting and sharing of all content, both safe and harmful; with platform liability, the platform has a financial incentive to detect and remove harmful content.

4 Conclusion

Should platforms be held liable for the harms suffered by platform participants? This question is of practical as well as academic interest. Platforms like Amazon, Google, and Facebook create considerable social value for their users but may also expose them to considerable risk. These and other platforms claim that they value their users' privacy and safety, are careful to protect their users' sensitive personal information, and spend considerable sums of money to monitor platform activity and block harmful actors from participating. But in reality, 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 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. If firms are judgment proof and immune from liability, however, then platform liability is 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 remove the harmful firms from the platform. The optimal level of platform liability depends on whether users are involuntary bystanders or voluntary consumers of the firms, the intensity of platform competition, and the impact on user participation. With appropriate incentives, platforms can play an important role in reducing social costs.

Our model abstracted from other salient factors. First, we did not consider ex ante incentives for innovation.[87] Section 230 of the Communications Decency Act was adopted to allow the internet to grow and flourish, and has been referred to as "the one line of federal code that has created more economic value in this country than any other."[88] Do Facebook, Google, and Amazon still require the protection of Section 230? Second, our model abstracted from reputation building and peer-to-peer reviews. Is platform liability a substitute or a complement for decentralized market mechanisms?[89] Third, we assumed that the platform could audit and remove participants from the platform. Should a platform that maintains tight control be held to a higher legal standard? What if the allocation of control rights is endogenous?[90]

Although internet platforms provided the motivation for this paper, our insights apply more broadly. Our analysis provides a strong economic rationale for holding traditional newspapers liable for harmful advertising content[91] and for holding bricks-and-mortar retailers liable for the harm caused by defective products.[92] Although our model is broadly applicable, we believe that the insights are particularly salient for online platforms including Facebook, Google, and Amazon. First, the harmful participants on these platforms are frequently small and judgment proof with insufficient incentives to curtail their harmful activities. Second, the big tech giants have the data and technology to detect and block participants that are more likely to harm others.[93] It is therefore ironic that the big internet platforms enjoy legal protections that are unavailable to traditional business models.

[87] Jeon et al. (2022) examine how negligence-based liability changes platforms' incentives to remove IP-infringing products, which in turn affects brand owners' innovation incentives.

[88] This quote is attributed to Michael Beckerman with the Internet Association, a lobbying organization that represents some of the largest Internet companies. See https://www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-facebook-google-is-about-to-change.

[89] Many platforms rely on a combination of screening and peer-to-peer feedback mechanisms. For example, Uber runs various background checks on its drivers, eliminates drivers based on negative reviews, and shares reviews with users. See Einav et al. (2016). See Tadelis (2016) for a thoughtful discussion of the limits and biases in peer-to-peer feedback mechanisms.

[90] Hagiu and Wright (2015 and 2018) examine the allocation of control rights between intermediaries and firms over transferable decisions such as marketing activities. Platform liability is not addressed.

[91] See *Braun v. Soldier of Fortune Magazine, Inc.*, 968 F.2d 1110 (1992). The court opined: "[T]he first Amendment permits a state to impose upon a publisher liability for compensatory damages for negligently publishing a commercial advertisement where the ad on its face, and without the need for investigation, makes it apparent that there is substantial danger of harm to the public."

[92] See *In re Mattel, Inc.*, 588 F. Supp. 2d 1111 (C.D. Cal. 2008). Some toy buyers brought suit against manufacturers and retailers (including Wal-Mart) for unsafe toys. See also Restatement (Third) of Torts (1998). "One engaged in the business of selling or otherwise distributing products who sells or distributes a defective product is subject to liability for harm to persons or property caused by the defect."

[93] See Van Loo (2020a, 2020b).

References

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

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

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

[4] 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.

[5] Bebchuk, Lucian and Jesse Fried, "The Uneasy Case for the Priority of Secured Claims in Bankruptcy," *Yale Law Journal*, Vol. 105 (1996), pp. 857-934.

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

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

[8] 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.

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

[10] Carvell, Daniel, Janet Currie, and W. Bentley MacLeod, "Accidental Death and the Rule of Joint and Several Liability," *RAND Journal of Economics*, Vol. 43 (2012), pp. 51-77.

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

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

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

[14] 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.

[15] 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.

[16] Culotta, Aron, Ginger Zhe Jin, Yidan Sun, and Liad Wagman, "Safety Reviews on Airbnb: An Information Tale," (2022), working paper.

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

[18] Daughety, Andrew F., and Jennifer F. Reinganum, "Everybody Out of the Pool: Products Liability, Punitive Damages, and Competition," *Journal of Law, Economics and Organization*, Vol. 13 (1997), pp. 410-432.

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

[20] 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.

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

[22] Daughety, Andrew F. and Jennifer F. Reinganum, "Cumulative Harm and Resilient Liability Rules for Product Markets," *Journal of Law, Economics, and Organization,* Vol. 30 (2014), pp. 371-400.

[23] 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.

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

[25] Einav, Liran, Chiara Farronato, and Jonathan Levin, "Peer-to-Peer Markets," *Annual Review of Economics*, Vol. 8 (2016), pp. 615-635.

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

[27] Fu, Qiang, Jie Gong, and Ivan Png, "Law, Social Responsibility, and Outsourcing," *International Journal of Industrial Organization*, Vol. 57 (2018), pp. 114-146.

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

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

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

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

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

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

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

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

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

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

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

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

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

[41] 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.

[42] Landes, William M., and Richard A. Posner, "Joint and Multiple Tortfeasors: An Economic Analysis," *Journal of Legal Studies*, Vol. 9 (1980), pp. 517-555.

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

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

[45] 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.

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

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

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

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

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

[51] Shavell, Steven, “Liability for Accidents,” Chapter 2, Handbook of Law and Economics (2007), Elsevier.

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

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

[54] Sydell, Laura, “We Tracked Down a Fake-News Creator in the Suburbs. Here’s What We Learned,” *National Public Radio*, November 23, 2016. https://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs

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

[56] 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.

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

[58] Townsend, Tess, "Google Has Banned 200 Publishers Since It Passed a New Policy Against Fake News," *CNBC*, January 25, 2017. https://www.cnbc.com/2017/01/25/google-has-banned-200-publishers-since-it-passed-a-new-policy-against-fake-news.html

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

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

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

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

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

Appendix A

Proof of Lemma 1. We first show that, if $w_s < \widehat{w}$, the platform does not find it profitable to deter the L-types and retain the H-types. If the platform deters the L-types by setting a high price $p_H = \alpha_H - \theta_H w_s$, its profit is

$$\Pi_H(e) = \lambda(1-e)(\alpha_H - \theta_H w) - c(e),$$

where $w = w_s + w_p$. As defined in the text, $\Pi(e)$ is the platform's profit when it charges $p_L = \alpha_L - \theta_L w_s$. Consider two scenarios.

First, suppose $w > \frac{\alpha_H}{\theta_H}$. Then $\Pi_H(e) < 0$ for any e. Assumption A2 implies $\Pi(0) > 0$, that is, the profit from attracting both types is larger than the profit from deterring the L-types.

Second, suppose $w \leq \frac{\alpha_H}{\theta_H}$. Since $\alpha_H - \theta_H w \geq 0$, the platform would not take any auditing effort and the optimal profit is $\Pi_H(0) = \lambda(\alpha_H - \theta_H w)$. We have

$$\begin{aligned}
\Pi(0) - \Pi_H(0) &= \lambda(\alpha_L - \theta_L w_S - \theta_H w_p) + (1-\lambda)(\alpha_L - \theta_L w_S - \theta_L w_p) \\
&\quad -\lambda(\alpha_H - \theta_H w) \\
&= \alpha_L - \lambda\alpha_H - (1-\lambda)\theta_L w + \lambda(\theta_H - \theta_L) w_s \\
&\geq \alpha_L - \lambda\alpha_H - (1-\lambda)\theta_L \frac{\alpha_H}{\theta_H} \\
&= \alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)\frac{\alpha_H}{\theta_H} \\
&> 0,
\end{aligned}$$

where the first inequality holds given $w \leq \frac{\alpha_H}{\theta_H}$ and the second inequality follows from Assumption A2. Therefore, the platform would not deter the L-types.

Now we prove the remaining results in the lemma. Using the definition of $r_H(w_s)$ in the lemma, (8) implies $e^* > 0$ if and only if $(\alpha_H - \theta_H w) - (\theta_H - \theta_L)(\widehat{w} - w_s) < 0$. This gives the condition for cases 1 and 2. Totally differentiating (10), and using the fact the social welfare function is concave, gives $de^*/dw_s = -\lambda\theta_L/S''(e) > 0$ and $de^*/dw_p = -\lambda\theta_H/S''(e) > 0$. When $e^* > 0$ (an interior solution), increasing the level of liability for either the firm or the platform increases the platform's auditing effort. Equation (10) implies $e^* > e^{**}$ if and only if $\lambda r_H(w_s) - \lambda\theta_H(d - w) > 0$. This gives the condition for subcases 2(a), 2(b) and 2(c).

Proof of Proposition 1. Note that $\widehat{w} < d < \frac{\alpha_L}{\theta_L}$ by Assumption A1. Suppose $w_p = 0$ and $w_s \leq \widehat{w}$. From Lemma 1, a necessary and sufficient condition for $e^* = 0$ is (8) or

$$\alpha_H - \theta_H w_s > (\theta_H - \theta_L)(\widehat{w} - w_s).$$

Substituting for $\widehat{w}$ from (5),

$$\alpha_H - \theta_H w_s > (\alpha_H - \alpha_L) - (\theta_H - \theta_L) w_s,$$

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

Suppose $w_s > \widehat{w}$. There are two possible scenarios. First, if $\theta_L/\theta_H < \alpha_L/\alpha_H$, then setting $w_p = 0$ in Lemma 2 and rearranging terms gives a threshold value $\widetilde{w}(\lambda) = \frac{\alpha_H-\alpha_L+\lambda\alpha_L}{\theta_H-\theta_L+\lambda\theta_L} \in \left(\widehat{w}, \frac{\alpha_H}{\theta_H}\right)$. Moreover, $\frac{d\widetilde{w}(\lambda)}{d\lambda} > 0$ given $\theta_L/\theta_H < \alpha_L/\alpha_H$. When $w_s < \widetilde{w}(\lambda)$, the platform sets $p^* = \alpha_H - \theta_H w_s$, and attracts the H-types; when $w_s \geq \widetilde{w}(\lambda)$, the platform sets $p^* = \alpha_L - \theta_L w_s$ and deters the H-types. Second, if $\theta_L/\theta_H \geq \alpha_L/\alpha_H$, then $\frac{\alpha_H-\alpha_L+\lambda\alpha_L}{\theta_H-\theta_L+\lambda\theta_L} \leq \widehat{w} < w_s$. In this scenario, Lemma 2 implies that the platform always sets $p^* = \alpha_L - \theta_L w_s$ and deters the H-types. The two scenarios can be combined by defining $\widetilde{w}(\lambda) = \max\{\frac{\alpha_H-\alpha_L+\lambda\alpha_L}{\theta_H-\theta_L+\lambda\theta_L}, \widehat{w}\}$.

Proof of Proposition 2. Suppose $w_s \leq \widehat{w}$, so the L-type is marginal. The platform cannot deter the H-types directly through the price, but can remove them through auditing. From equation (10) we have $e^* = e^{**}$ if and only if $w_p = w_p^* = d - w_s - \left(1 - \frac{\theta_L}{\theta_H}\right)(\widehat{w} - w_s)$. 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 $p = \alpha_H - \theta_H w_s$, and attracts the H-type firms. This is socially inefficient. Lemma 2 implies that the platform would deter the H-type if $\lambda(\alpha_H - \theta_H w) \leq (1-\lambda) r_L(w_s)$. $\lambda(\alpha_H - \theta_H w)$ decreases in w_p and the firms' rent $(1-\lambda) r_L(w_s)$ is independent of w_p. Setting $\lambda(\alpha_H - \theta_H w) = (1-\lambda) r_L(w_s)$ gives the lower bound $\underline{w}_p$:

$$\underline{w}_p = \tfrac{\alpha_H}{\theta_H} - w_s - \tfrac{1-\lambda}{\lambda}\left(1 - \tfrac{\theta_L}{\theta_H}\right)(w_s - \widehat{w}) > 0.$$

For any $w_p^* \geq \underline{w}_p$, the platform deters the H-types and the first-best outcome is obtained.

Suppose $w_s \geq \widetilde{w}$. Proposition 1 implies that even if $w_p = 0$ the platform sets $p^* = \alpha_L - \theta_L w_s$, deters H-type firms, and the first-best outcome is obtained. Platform liability is unnecessary. Any $w_p^* \in [0, d - w_s]$ achieves the first-best outcome.

Proof of Proposition 3. We prove two claims respectively for $w_s \leq \widehat{w}$ and $w_s > \widehat{w}$.

Claim 1: *Suppose $w_s \leq \widehat{w}$. The platform sets $p^r = \alpha_L - \theta_L w_s - \theta^r(d-w)$ and attracts the H-type firms where $\theta^r = E(\theta|e^r)$ are the equilibrium posterior beliefs. Let $\theta^{**} = E(\theta|e^{**})$, $\theta^0 = E(\theta|0)$, and $r_H(w_s) = (\theta_H - \theta_L)(\widehat{w} - w_s)$.*

1. *If $(\alpha_H - \theta_H d) + (\theta_H - \theta^0)(d - w) \geq r_H(w_s)$ then the platform does not audit, $e^r = 0 < e^{**}$.*

2. *If $(\alpha_H - \theta_H d) + (\theta_H - \theta^0)(d - w) < r_H(w_s)$ then $e^r > 0$. The platform's auditing effort decreases in firm liability $de^r/dw_s < 0$ and increases in platform liability $de^r/dw_p > 0$.*

   (a) *If $(\theta_H - \theta^{**})(d - w) > r_H(w_s)$ then $0 < e^r < e^{**}$.*

   (b) *If $(\theta_H - \theta^{**})(d - w) = r_H(w_s)$ then $0 < e^r = e^{**}$.*

   (c) *If $(\theta_H - \theta^{**})(d - w) < r_H(w_s)$ then $0 < e^{**} < e^r$.*

Proof of Claim 1: Since $w_s \leq \widehat{w}$, it is not possible for the platform to deter the H-types without deterring the L-types, too. If the L-type is willing to participate, then the H-type also prefers to participate.

To begin,we construct values $\{e^r, p^r, t^r\}$ that maximize the platform's profits subject to the platform's incentive compatibility constraint and the participation constraints of the consumers and the L-type firms (as the L-type firm is marginal). Then, we will verify that these values are an equilibrium of the game.

$$\max_{\{e,p,t\}} \quad \Phi(e,p) = (1-e)\lambda(p - \theta_H w_p) + (1-\lambda)(p - \theta_L w_p) - c(e) \tag{26}$$

subject to

$$e = \arg\max_{e' \geq 0} \Phi(e', p) \tag{27}$$

$$\alpha_0 - t - E(\theta|e)(d - w_s - w_p) \geq 0 \tag{28}$$

$$t - (\theta_L w_s + c_L) - p \geq 0. \tag{29}$$

(27) is the platform's incentive compatibility constraint, (28) is the consumer's participation constraint, and (29) is the L-type firm's participation constraint.[94]

The L-type's participation constraint (29) must bind. To see this, consider two cases. First, suppose that neither (28) nor (29) binds. Then the platform would increase the price p which would increase the platform's profits in (26) and maintain the consumer's participation constraint (28). Second, suppose that (28) binds while (29) does not. Again, the platform would increase the price p marginally. The direct effect of increasing p is that the platform's profits in (26) increase. Since $\partial^2\Phi(e,p)/\partial e \partial p = -\lambda < 0$, increasing p also (weakly) decreases the platform's effort e in (27), which in turn raises $E(\theta|e)$ and, since (28) binds, reduces t. However, since t is not in (26), the platform's profits still increase.

Since the L-type's constraint (29) binds, $p = t - (\theta_L w_s + c_L)$ and we can rewrite the optimand (26) as a function of e and t:

$$(1-e)\lambda(t - (\theta_L w_s + c_L) - \theta_H w_p) + (1-\lambda)(t - (\theta_L w_s + c_L) - \theta_L w_p) - c(e). \tag{30}$$

Next, we show that the consumer's participation constraint (28) binds. Suppose not. Then, the platform would increase t and its profits would rise. Since both participation constraints (28) and (29) bind, we have

$$p = \alpha_0 - E(\theta|e)(d - w_s - w_p) - (\theta_L w_s + c_L). \tag{31}$$

Since $\alpha_L = \alpha_0 - c_L$ and $w = w_s + w_p$ the solution to the platform's optimization problem is:

$$e^r = \arg\max_{e \geq 0} \Phi(e, p^r) \tag{32}$$

[94] The H-type's participation constraint is satisfied if (29) holds, and is therefore not included in the program.

$$t^r = \alpha_0 - E(\theta|e^r)(d-w) \tag{33}$$

$$p^r = \alpha_L - \theta_L w_s - E(\theta|e^r)(d-w). \tag{34}$$

We now verify that the values $\{e^r, p^r, t^r\}$ defined in (32), (33), and (34) are an equilibrium of the game. Suppose that the platform charges p^r in (34), and that the firms and consumers believe that the probability of harm is $\theta^r = E(\theta|e^r)$ where e^r defined in (32). The consumers are (just) willing to pay t^r in (33) and the L-type firms are (just) willing to pay p^r in (34). If the consumers and the firms all participate, the platform exerts effort e^r in (32). Therefore the equilibrium beliefs $\theta^r = E(\theta|e^r)$ are consistent.

Next, we verify that Assumption A2 guarantees that the platform's profits are positive. To do this, we will show that the platform's profits are positive even if consumers and the firms believed that the platform is not auditing at all, so $E(\theta|0) = \theta^0$.[95] In this scenario, the most that consumers would be willing to pay is $t = \alpha_0 - \theta^0(d-w)$ from (28). The most that the L-type firms would be willing to pay is $p = \alpha_L - \theta_L w_s - \theta^0(d-w)$ from (29). The platform's profits can be rewritten as

$$\Pi(0) = \alpha_L - \theta^0 d + \lambda(\theta_H - \theta_L)w_s.$$

Therefore, $\Pi(0) > 0$ for any $w_s \geq 0$ if Assumption $A2$ holds.[96]

We now show that the algebraic condition in case 1 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 consumer's participation constraint (28) binds we have $t^r = \alpha_0 - \theta^0(d-w)$; since the L-type firm's participation constraint (29) binds we have $p^r = \alpha_L - \theta_L w_s - \theta^0(d-w)$. Finally, for $e^r = 0$ to satisfy the platform's IC constraint (27) we need $\partial\Phi(e,p)/\partial e \leq 0$ or equivalently $p^r - \theta_H w_p \geq 0$. Substituting p^r, this condition becomes

$$\alpha_L - \theta_L w_s - \theta^0(d-w) - \theta_H w_p \geq 0. \tag{35}$$

Adding and subtracting terms this becomes

$$(\alpha_H - \theta_H d) - (\alpha_H - \alpha_L) - \theta_L w_s - \theta_H w_p + \theta_H w + (\theta_H - \theta^0)(d-w) \geq 0, \tag{36}$$

and rearranging this expression gives

$$(\alpha_H - \theta_H d) + (\theta_H - \theta^0)(d-w) \geq (\alpha_H - \alpha_L) - (\theta_H - \theta_L)w_s. \tag{37}$$

The right-hand side is $r_H(w_s)$. This confirms that the condition in case 1 is necessary.

[95] The platform is better off if the consumers believe that the product is safer. If consumers perceive the product to be safer, they will pay a higher price t for the product which means that the platform can charge the firms a higher price p.

[96] If $e = 1$ then $E(\theta|1) = \theta_L$. One can verify that $\Pi(1) > 0$ if and only if $\alpha_L - \theta_L d > \frac{c(1)}{1-\lambda}$. This condition is independent of w_s and w_p. It may hold even if $A2$ is not satisfied (that is, $\alpha_L - \theta_L d \leq \lambda(\theta_H - \theta_L)d$). When this condition holds, even if $A2$ is not satisfied, the platform may still be active. That is, $A2$ is a sufficient but not necessary condition for the platform to be active.

Next, we show that the condition in case 1 is sufficient. Suppose the condition holds and $e^r > 0$. Since $E(\theta|e^r) < \theta^0$, $t^r > \alpha_0 - \theta^0(d-w)$ and $p^r > \alpha_L - \theta_L w_s - \theta^0(d-w)$. Assumption A2 implies $p^r - \theta_H w_p > 0$, so the platform does not audit, $e^r = 0$.

Now consider case 2. The condition implies $p^r - \theta_H w_p < 0$ so the platform is losing money from each H-type transaction. The equilibrium effort $e^r > 0$ and consumers' equilibrium beliefs $\theta^r = E(\theta|e^r)$ satisfy equation (16). the platform charges $p^r = \alpha_L - \theta_L w_s - \theta^r(d-w)$ and consumers believe that the platform will exert effort e^r and are willing to pay $t^r = \alpha_0 - \theta^r(d-w)$. Condition (16) implies that $e^{**} < e^r$ if and only if $(\theta_H - \theta^{**})(d-w) < (\theta_H - \theta_L)(\widehat{w} - w_s)$. Totally differentiating condition (16) and using the fact that the welfare function is concave, we have $de^r/dw_s < 0$ and $de^r/dw_p > 0$.

Claim 2: *Suppose $w_s > \widehat{w}$. The platform sets $p^r = \alpha_L - \theta_L(d-w_p)$ and deters the H-type firms.*

Proof of Claim 2: Since $w_s > \widehat{w}$ the H-type firms are marginal. The platform can deter the H-types by charging a price that only the L-types would accept. The users' posterior beliefs are $\theta^r = \theta_L$, and so the firms charge the consumers $t^r = \alpha_0 - \theta_L(d-w)$. The platform's price extracts the L-type firm's surplus, $p^r = t^r - (\theta_L w_s + c_L)$. Therefore

$$p^r = \alpha_L - \theta_L w_s - \theta_L(d-w) = \alpha_L - \theta_L(d-w_p) \tag{38}$$

and the platform's profits are

$$(1-\lambda)(p^r - \theta_L w_p) = (1-\lambda)(\alpha_L - \theta_L d). \tag{39}$$

In other words, the platform extracts the full social surplus from the L-types.

If the platform chooses to attract the H-type firms, then the platform will not audit them. The users' posterior beliefs are the same as their priors, $\theta^0 = \lambda\theta_H + (1-\lambda)\theta_L$, and the firms charge the consumers $t^r = \alpha_0 - \theta^0(d-w)$. The platform's price extracts the marginal H-type firm's surplus, that is, $p^r = t^r - (\theta_H w_s + c_H)$ or

$$p^r = \alpha_H - \theta_H w_s - \theta^0(d-w). \tag{40}$$

The platform's profits are

$$\begin{aligned} p^r - \theta^0 w_p &= (1-\lambda)(\alpha_L - \theta_L d) + \lambda(\alpha_H - \theta_H d) + (1-\lambda)[\alpha_H - \alpha_L - (\theta_H - \theta_L)w_s] \\ &= (1-\lambda)(\alpha_L - \theta_L d) + \lambda(\alpha_H - \theta_H d) + (1-\lambda)(\theta_H - \theta_L)(\widehat{w} - w_s) \\ &< (1-\lambda)(\alpha_L - \theta_L d) \end{aligned}$$

where the inequality follows from Assumption A1 and $w_s > \widehat{w}$. Therefore, if $w_s > \widehat{w}$, the platform charges $p^r = \alpha_L - \theta_L(d-w_p)$ and deters the H-types.

We now proceed to proof Proposition 3. Suppose $w_s \leq \widehat{w}$, so the L-type is marginal. From Claim 1, we have $e^r = e^{**}$ if and only if

$$(\theta_H - \theta_L)(\widehat{w} - w_s) - (\theta_H - \theta^{**})(d-w) = 0. \tag{41}$$

Substituting that $w = w_p + w_s$ and isolating w_p on the left-hand side establishes the result. Suppose $w_s > \widehat{w}$. The results follow from Claim 2.

Proof of Proposition 4. We first show a claim for $w_s \leq \widehat{w}$.

Claim 3: *Suppose $w_s \leq \widehat{w}$. The platforms set $p^c < \alpha_L - \theta_L w_s = p^*$ and attract the H-type firms. If $w_p = 0$ then the platforms do not audit, $e^c = e^* = 0$. If $w_p > 0$ then $e^c > 0$, and $de^c/dw_p > 0$. There exists a unique threshold $\overline{w}_p \in (0, w_p^*)$.*

1. *If $0 < w_p \leq \overline{w}_p$ then $e^* < e^c \leq e^{**}$.*
2. *If $w_p \in (\overline{w}_p, w_p^*)$ then $e^* < e^{**} < e^c$.*
3. *If $w_p \geq w_p^*$ then $e^{**} \leq e^* < e^c$.*

Proof of Claim 3: We first show that the platforms receive zero profits in equilibrium. If one platform received positive profits while the other got no profit, the second platform would deviate and imitate the first one's strategies.

Suppose that both platforms got positive profits and Platform 2 attracted weakly more H-type firms than Platform 1. Since $w_s \leq \widehat{w}$, the L-type firms get (weakly) lower rents than the H-types. Thus, the L-type firms must be indifferent between joining the two platforms. But then Platform 2 would reduce its price marginally, which would steal all the L-types (and possibly the H-types) from Platform 1 and therefore weakly reduce the proportion of H-types on Platform 2. Note that, in this off-equilibrium path, Platform 2 may raise its auditing effort marginally. Since Platform 2 got positive profits when having more H-types, attracting more firms with a larger proportion of L-types would strictly raise its profits. Therefore, both platforms should receive zero profits in equilibrium.

Next, we show that, as long as both platforms are active, they charge the same price and take the same auditing effort in equilibrium. Note that, if the L-type firms strictly prefer joining one platform, then the H-types would join this platform too because they get (weakly) higher rents than the L-types. Therefore, as long as both platforms are active, they should get some L-types. That is, the L-types are indifferent between joining the two platforms. Since the L-types would never be removed, the platforms' prices must be the same. Furthermore, if the two platforms chose different auditing levels, the one with less auditing would attract all the H-types. However, since the platforms' prices are the same, the one attracting all the H-types would have greater incentives to take auditing effort, a contradiction. Therefore, the platforms take the same auditing effort.

To summarize, the above analysis suggests that in any equilibrium the platforms get zero profits, charge the same price, and take the same auditing effort.

Now we show that e^c increases in w_p for any $w_p < d - w_s$. Define $Z = p^c - \theta_H w_p$. Condition (20) can be re-written as

$$\Pi = (1 - e^c)\lambda Z + (1 - \lambda)[Z + (\theta_H - \theta_L)w_p] - c(e^c) = 0.$$

Differentiating with respect to w_p, and recognizing that e^c and Z are functions of w_p, this implies

$$\frac{d\Pi}{de^c}\frac{de^c}{dw_p} + \frac{d\Pi}{dZ}\frac{dZ}{dw_p} + (1-\lambda)(\theta_H - \theta_L) = 0.$$

Since $\frac{d\Pi}{de^c} = 0$ (the first-order condition), $\frac{d\Pi}{dZ} > 0$, and $(1-\lambda)(\theta_H - \theta_L) > 0$, we have $\frac{dZ}{dw_p} < 0$. Condition (21) may be written as $-\lambda Z - c'(e^c) = 0$. Differentiating this with respect to w_p gives $-\lambda\frac{dZ}{dw_p} - c''(e^c)\frac{de^c}{dw_p} = 0$. Finally, $\frac{dZ}{dw_p} < 0$ and $c''(e^c) > 0$ imply that $\frac{de^c}{dw_p} > 0$.

Finally, note that $p^c < p^* = \alpha_L - \theta_L w_s$. To see this, suppose that $e = 0$ and $p \geq \alpha_L - \theta_L w_s$. Then

$$\begin{aligned} \Pi(0) &\geq \alpha_L - \theta_L w_s - [\lambda\theta_H + (1-\lambda)\theta_L]w_p \\ &\geq \alpha_L - \theta_L w_s - [\lambda\theta_H + (1-\lambda)\theta_L](d - w_s) \\ &> 0 \end{aligned}$$

where the second inequality follows from $w_p \leq d - w_s$ and the last inequality holds given Assumption A2. Thus, condition (20) implies that $p^c < p^* = \alpha_L - \theta_L w_s$. And condition (21) then implies $e^c > e^*$ as long as $w_p > 0$.

Lemma 1 implies that, when there is a monopoly platform, $0 < e^* < e^{**}$ if $w_p < w_p^*$ and $0 < e^{**} = e^*$ if $w_p = w_p^*$. Note that $w_p^* \in (0, d - w_s)$ if $w_s < \widehat{w}$. Since e^c increases in w_p and $e^c > e^*$ if $w_p > 0$, there exists a unique value $\overline{w}_p \in (0, w_p^*)$ such that $e^c = e^{**}$ if and only if $w_p = \overline{w}_p$.

If $0 < w_p \leq \overline{w}_p$, then $e^c \leq e^{**}$, while $e^* < e^c$ as shown earlier. Therefore, under competition, the auditing intensity is closer to the socially efficient level, which raises welfare.

If $w_p \geq w_p^*$ then $e^{**} \leq e^* < e^c$. Therefore, competition exacerbates the distortion in auditing and reduces welfare.

We now proceed to proof Proposition 4.

If $w_s \leq \widehat{w}$, from Claim 3, we have $e^c = e^{**}$ if and only if $w_p^c = \overline{w}_p \in (0, w_p^*)$.

If $w_s \in (\widehat{w}, \frac{\alpha_H}{\theta_H})$, then $\alpha_H - \theta_H w_s < \alpha_L - \theta_L w_s$, that is, the H-types are marginal. Note that platform i always charges $p_i \geq \theta_L w_p$, $i = 1, 2$, as otherwise its profit would be negative. In any equilibrium where all the H-type firms are deterred (if it exists), we must have $p_i > \alpha_H - \theta_H w_s$, $i = 1, 2$. If $\theta_L > 0$ and $\theta_L w_p > \alpha_H - \theta_H w_s$, then $p_1 = p_2 = \theta_L w_p$, which deters the H-type firms. If $\theta_L = 0$ or $\theta_L w_p \leq \alpha_H - \theta_H w_s$, then the equilibrium with the H-type firms being deterred does not exist.

Finally, if $w_s \geq \frac{\alpha_H}{\theta_H}$, the H-type firms would never join and therefore the platforms compete to attract the L-type firms, which leads to $p_1 = p_2 = \theta_L w_p$.

Proof of Proposition 5. We start by showing that the platform sets $m = 0$ if $\alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)d$ is sufficiently large. To see this, first consider the scenario where the

L-type firms are marginal ($w_s \leq \widehat{w}$). Given the belief e and damage award $w = w_s + w_p$, a user will participate when

$$v \geq m + [\lambda(1-e)\theta_H + (1-\lambda)\theta_L](d-w).$$

The platform's equilibrium price charge to the firms is the same as in the baseline model (see Lemma 1). Thus, the platform's profits are

$$[1 - F(m + (\lambda(1-e)\theta_H + (1-\lambda)\theta_L)(d-w))][\widehat{\Pi}(e) + m] - c(e),$$

where $1 - F(\cdot)$ is the users' participation rate and

$$\widehat{\Pi}(e) = (1-e)\lambda(\alpha_L - \theta_L w_s - \theta_H w_p) + (1-\lambda)(\alpha_L - \theta_L w). \tag{42}$$

Since $\alpha_L - \theta_H d < 0$, when $e = 0$, $w_s = 0$ and $w_p = d$, $\widehat{\Pi}(e)$ achieves the lowest value

$$\alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)d,$$

which is positive by Assumption A2. Taking differentiation of the profit function with respect to m, we have

$$[1 - F(\cdot)] - f(\cdot)[\widehat{\Pi}(e) + m],$$

which is negative if $\widehat{\Pi}(e)$ is sufficiently large. Hence, if $\alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)d$ is sufficiently large, the platform would set $m = 0$.

Next, consider the scenario where the H-type firms are marginal ($w_s > \widehat{w}$). If the platform accommodates all the H-type firms, a user will participate when

$$v \geq m + [\lambda\theta_H + (1-\lambda)\theta_L](d-w).$$

If the platform deters all the H-types firms by charging a larger price, a user will participate when

$$v \geq m + (1-\lambda)\theta_L(d-w).$$

Similar to the earlier analysis, we can show that, if $\alpha_L - \theta_L d$ is sufficiently large, the platform would set $m = 0$.

In the remaining analysis, we maintain the assumption that $\alpha_L - (\lambda\theta_H + (1-\lambda)\theta_L)d$ is sufficiently large, which also implies $\alpha_L - \theta_L d$ is sufficiently large, such that the platform does not charge the users.

Now we prove condition (25), which highlights the potential divergence between the private and social incentives for auditing. Given w, (22) implies

$$\frac{dS(e,\widehat{v})}{de} = \frac{\partial S(e,\widehat{v})}{\partial e} - \frac{\partial S(e,\widehat{v})}{\partial \widehat{v}}\lambda\theta_H(d-w) = 0.$$

Using (24), if the equilibrium auditing effort is positive, then e^u satisfies

$$
\begin{aligned}
\frac{\partial \Pi(e^u,\widehat{v})}{\partial e} &= \frac{\partial S(e^u,\widehat{v})}{\partial e} + \int_{\widehat{v}} [\lambda(\theta_H-\theta_L)(\widehat{w}-w_s) - \lambda\theta_H(d-w)] f(v)dv \\
&= \frac{dS(e^u,\widehat{v})}{de} + \int_{\widehat{v}} [\lambda(\theta_H-\theta_L)(\widehat{w}-w_s) - \lambda\theta_H(d-w)] f(v)dv + \lambda\theta_H(d-w)\frac{\partial S(e^u,\widehat{v})}{\partial\widehat{v}} \\
&= 0.
\end{aligned}
$$

Next, we show that, if $w_s < \widehat{w}$, then $w_p^u > w_p^*$. Totally differentiating (22) with respect to w_p gives

$$
\frac{dS(e^u,\widehat{v}(\cdot))}{dw_p} = [\frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial e} - \frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial \widehat{v}}\lambda\theta_H(d-w)]\frac{de^u}{dw_p} + \frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}, \qquad (43)
$$

where $\frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial \widehat{v}} < 0$ and $\frac{\partial \widehat{v}}{\partial w_p} < 0$. Similar to the analysis in the baseline model, we can show that, given $\widehat{v}$, if $w_p \leq w_p^*$,

$$
\lambda(\theta_H-\theta_L)(\widehat{w}-w_s) < \lambda\theta_H(d-w), \qquad (44)
$$

which implies $\frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial e} \geq 0$. Moreover, if $e^u > 0$, it satisfies

$$
\frac{\partial \Pi(e^u,\widehat{v})}{\partial e} = -\int_{\widehat{v}} \lambda(\alpha_L - \theta_L w_s - \theta_H w_p) f(v)dv - c'(e^u) = 0, \qquad (45)
$$

which implies $\frac{de^u}{dw_p} > 0$.

Given the above observations, if $w_p \leq w_p^*$, we have

$$
\frac{dS(e^u,\widehat{v}(\cdot))}{dw_p} = [\frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial e} - \frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial \widehat{v}}\lambda\theta_H(d-w)]\frac{de^u}{dw_p} + \frac{\partial S(e^u,\widehat{v}(\cdot))}{\partial \widehat{v}}\frac{d\widehat{v}}{dw_p} > 0. \qquad (46)
$$

Therefore, if $w_s < \widehat{w}$, it is socially optimal to set $w_p > w_p^*$.

Finally, if $w_s = \widehat{w}$, $w_p^u = d - w_s$ achieves the second-best outcome. To see this, note that $w_p^u = d - w_s$ attracts all the users. As shown by Proposition **2,** if $w_s = \widehat{w}$ and all the users participate, imposing full residual liability on the platform motivates it to choose the socially efficient auditing effort, $e = e^{**}$.

Online Appendix B

This appendix contains the analysis of two additional extensions: false positives and litigation costs.

False Positives (Type-I Errors)

Suppose that the auditing effort of the platform may erroneously remove the L-type firms with probability δe, where $\delta < 1$. The first-best benchmark is the same as in the baseline model. For the second-best benchmark, suppose that the H-type firms seek to join the platform. Social welfare is:

$$S(e) = v + \lambda(1-e)(\alpha_H - \theta_H d) + (1-\lambda)(1-\delta e)(\alpha_L - \theta_L d) - c(e). \tag{47}$$

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

$$-\lambda(\alpha_H - \theta_H d) - \delta(1-\lambda)(\alpha_L - \theta_L d) - c'(\widetilde{e}^{**}) = 0. \tag{48}$$

When $w_s > \widehat{w}$, the H-type 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 L-type firms are marginal. The platform sets the interaction price $p^f = \alpha_L - \theta_L w_s$, and its profits can be written as

$$\begin{aligned}\Pi(e) = S(e) &- (1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) \\ &+ [(1-e)\lambda\theta_H + (1-\lambda)(1-\delta e)\theta_L](d-w) - v.\end{aligned}$$

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

$$\Pi'(e^f) = S'(e^f) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - [\lambda\theta_H + (1-\lambda)\delta\theta_L](d-w) = 0. \tag{49}$$

Note that the users' (marginal) uncompensated harm, $[\lambda\theta_H + (1-\lambda)\delta\theta_L](d-w)$, is larger than that in the baseline model, while the firms' information rent, $\lambda(\theta_H - \theta_L)(\widehat{w} - w_s)$, 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 therefore omitted).

Proposition 6. *(False Positives.) Suppose firm liability is $w_s \in (0, d]$. The socially-optimal platform liability for harm to users, w_p^f, is as follows:*

1. *If $w_s \leq \widehat{w}$ then $w_p^f = d - w_s - \frac{\lambda(\theta_H - \theta_L)}{\lambda\theta_H + (1-\lambda)\delta\theta_L}(\widehat{w} - w_s) \geq w_p^*$ achieves the second-best outcome and it increases in δ. The platform sets $p^f = \alpha_L - \theta_L w_s$ and attracts the H-type firms. The platform's auditing incentives are socially efficient, $e^f = \widetilde{e}^{**}$.*

2. *If $w_s \in (\widehat{w}, \widetilde{w})$ then there exists a threshold $\underline{w}_p > 0$ where any $w_p^f \in [\underline{w}_p, d - w_s]$ achieves the first-best outcome. The platform sets $p^f = \alpha_L - \theta_L w_s$ and deters the H-type firms.*

3. *If $w_s \geq \widetilde{w}$ then platform liability is unnecessary. Any $w_p^f \in [0, d - w_s]$ achieves the first-best outcome. The platform sets $p^f = \alpha_L - \theta_L w_s$ and deters the H-type firms.*

Litigation Costs

When a user gets harmed by a firm and files a lawsuit, the litigation costs are k_p, k_s, k_b, respectively for the platform, the firm, and the user. Denote $k = k_p + k_s + k_b$. Assume that $k_b \le w_s + w_p$ and $\alpha_L - \theta_L d - k > 0$.[97] So, litigation is credible and it is efficient to have interactions between the L-type firms and users. If the H-type firms seek to join the platform, social welfare is

$$S(e) = v + \lambda(1-e)(\alpha_H - \theta_H(d+k)) + (1-\lambda)(\alpha_L - \theta_L(d+k)) - c(e).$$

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

$$-\lambda(\alpha_H - \theta_H(d+k)) - c'(\overline{e}^{**}) = 0.$$

The two types of firms have the same rent when:

$$w_s + k_s = \widehat{w} = \frac{\alpha_H - \alpha_L}{\theta_H - \theta_L}. \tag{50}$$

Case 1: $w_s + k_s \le \widehat{w}$. The platform sets $p^k = \alpha_L - \theta_L(w_s + k_s)$ to extract the L-type firms' rent. The platform chooses $e > 0$ if and only if $p^k - \theta_H(w_p + k_p) < 0$, which can be rewritten as

$$\alpha_H - \theta_H(w + k_p + k_s) - (\theta_H - \theta_L)(\widehat{w} - w_s - k_s) < 0.$$

The platform's profits can be written as

$$\begin{aligned}\Pi(e) = S(e) - (1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s - k_s) \\ + [(1-e)\lambda\theta_H + (1-\lambda)\theta_L](d + k_b - w) - v.\end{aligned}$$

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

$$\Pi'(e^k) = S'(e^k) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s - k_s) - \lambda\theta_H(d + k_b - w) = 0. \tag{51}$$

The users' uncompensated loss caused by the H-types, $\lambda\theta_H(d + k_b - w)$, increases in k_b; and the firms' information rent, $\lambda(\theta_H - \theta_L)(\widehat{w} - w_s - k_s)$, decreases in k_s. Therefore, as compared to the baseline model, the platform's auditing incentives are even weaker relative to the social incentives. We can show the following results.

Lemma 3. *Suppose $w_s + k_s \le \widehat{w}$. The platform sets $p^k = \alpha_L - \theta_L(w_s + k_s)$ and attracts the H-types. Let $r_H^k(w_s) \equiv (\theta_H - \theta_L)(\widehat{w} - w_s - k_s)$ denote the H-types' information rents.*

1. *If $\alpha_H - \theta_H(w + k_p + k_s) \ge r_H^k(w_s)$ then the platform does not audit, $e^k = 0 < \overline{e}^{**}$.*
2. *If $\alpha_H - \theta_H(w + k_p + k_s) < r_H^k(w_s)$ then $e^k > 0$.*
   (a) *If $\theta_H(d + k_b - w) > r_H^k(w_s)$ then $0 < e^k < \overline{e}^{**}$.*

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

(b) *If* $\theta_H(d + k_b - w) = r_H^k(w_s)$ *then* $0 < e^k = \overline{e}^{**}$.

(c) *If* $\theta_H(d + k_b - w) < r_H^k(w_s)$ *then* $0 < \overline{e}^{**} < e^k$.

Case 2: $w_s + k_s > \widehat{w}$. The platform's profit-maximizing strategy is to either charge $p = \alpha_L - \theta_L(w_s + k_s)$ and deter the H-types from joining the platform or charge $p = \alpha_H - \theta_H(w_s + k_s)$ and attract both types. The platform will charge $p = \alpha_H - \theta_H(w_s + k_s)$ and attract the H-types if

$$\lambda(\alpha_H - \theta_H(w + k_s + k_p)) > (1 - \lambda)(\theta_H - \theta_L)(w_s + k_s - \widehat{w}), \tag{52}$$

which is less likely to hold when k_s or k_p is larger. That is, the platform is more likely to deter the H-type 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 H-types than in the baseline model.

Lemma 4. *Suppose* $w_s + k_s > \widehat{w}$. *Let* $r_L^k(w_s) = (\theta_H - \theta_L)(w_s + k_s - \widehat{w})$ *denote the L-type firm's information rents.*

1. *If* $\lambda(\alpha_H - \theta_H(w + k_s + k_p)) > (1 - \lambda)r_L^k(w_s)$ *then the platform sets* $p^k = \alpha_H - \theta_H(w_s + k_s)$*, attracts the* H*-type firms, and does not audit,* $e^k = 0 < \overline{e}^{**}$.

2. *If* $\lambda(\alpha_H - \theta_H(w + k_s + k_p)) \leq (1 - \lambda)r_L^k(w_s)$ *then the platform sets* $p^k = \alpha_L - \theta_L(w_s + k_s)$ *and deters the* H*-type firms.*

Define $\widetilde{w}^k = \frac{\alpha_H - \alpha_L + \lambda\alpha_L - \lambda\theta_H k_p}{\theta_H - \theta_L + \lambda\theta_L}$. Similar to the analysis in the baseline model, we can characterize the optimal platform liability.

Proposition 7. *(Litigation Costs) The socially-optimal platform liability for harm to users,* w_p^k*, is as follows:*

1. *If* $w_s + k_s \leq \widehat{w}$ *then* $w_p^k = d + k_b - w_s - (1 - \frac{\theta_L}{\theta_H})(\widehat{w} - w_s - k_s) \geq w_p^*$ *achieves the second-best outcome. The platform sets* $p^k = \alpha_L - \theta_L(w_s + k_s)$ *and attracts the* H*-type firms. The platform's auditing incentives are socially efficient,* $e^k = \overline{e}^{**}$.

2. *If* $w_s + k_s \in (\widehat{w}, \widetilde{w}^k)$ *then there exists a threshold* $\underline{w}_p^k \in (0, \underline{w}_p)$ *such that any* $w_p^k \in [\underline{w}_p^k, d - w_s]$ *achieves the first-best outcome. The platform sets* $p^k = \alpha_L - \theta_L(w_s + k_s)$ *and deters the* H*-type firms.*

3. *If* $w_s + k_s \geq \widetilde{w}^k$ *then platform liability is unnecessary. Any* $w_p^k \in [0, d - w_s]$ *achieves the first-best outcome. The platform sets* $p^k = \alpha_L - \theta_L(w_s + k_s)$ *and deters the* H*-type firms.*

When $w_s + k_s \leq \widehat{w}$, 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, $w_p^k \geq w_p^*$, where the inequality holds strictly if $k_b > 0$ or $w_s + k_s < \widehat{w}$.

When $w_s + k_s \in (\widehat{w}, \widetilde{w}^k)$, with litigation costs, the platform has stronger incentives to deter the H-types than in the baseline model. Hence, the lowest platform liability that implements the first-best outcome is smaller than that in the baseline model, $\underline{w}_p^k < \underline{w}_p$.