Holding Platforms Liable*

Xinyu Hua† HKUST

Kathryn E. Spier‡ Harvard University

October 16, 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 raise the interaction price to deter harmful firms and invest resources to detect and remove harmful firms from the platform. To prevent overinvestment in detection and removal, the residual liability assigned to the platform may be partial instead of full. The optimal level of platform liability depends on the impact on user participation, the intensity of platform competition, and whether users are involuntary bystanders or voluntary consumers.

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

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

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

1 Introduction

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

In the United States, platforms enjoy relatively broad immunity from lawsuits brought by users, 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] In 2019, Facebook paid $5 billion to settle charges that they failed to take adequate precautions to protect user data.[4] The FTC has also been investigating how "platforms screen for misleading ads for scams and fraudulent and counterfeit products" and, "in 2022 alone, consumers reported losing more than $1.2 billion to fraud that started on social media, more than any other contact method."[5] Proposed federal legislation would hold platforms liable if they fail to protect users.[6]

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

[1]See 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."Proponents hoped Section 230 would address the "perverse incentives" created by *Stratton Oakmont v. Prodigy Services Co.*, WL 323710 (N.Y. Sup. Ct. 1995). In that case, the court reasoned that since Prodigy exercised some editorial control, Prodigy should also assume liability for user content.

[3]See *Force v. Facebook, Inc.*, 934 F.3d 53 (2d Cir. 2019). The court opined that Section 230 "should be construed broadly in favor of immunity."

[4]https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook

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

[6]One recent example is the bipartisan "Internet Platform Accountability and Consumer Transparency (Internet PACT) Act." https://www.schatz.senate.gov/news/press-releases/schatz-thune-reintroduce-legislation-to-strengthen-rules-transparency-for-online-content-moderation-hold-internet-companies-accountable

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

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

This paper presents a formal model of a two-sided platform with two kinds of participants, "firms" and "users." The platform enables interactions between the firms and users, and charges the firms a fixed price per interaction.[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.

In practice, platforms can and do invest resources to vet participants, monitor their online behavior, and block participants who are more likely to harm others.[12] For example, Facebook parent Meta utilizes various privacy-enhancing technologies to protect users.[13] Amazon employs machine learning scientists, software developers and expert investigators, to fight against fraudulent sellers.[14] Google has been licensing its technologies and providing cloud-based services for other platforms to improve safety.[15] LinkedIn uses both automatic and manual investigations to remove scams, though fraudulent business and job-opportunity postings are still skyrocketing.[16]

In our baseline model, users are homogeneous and are effectively *bystanders* of the firms. By joining the platform, the users consent to subsequent firm-user interactions.

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

[8]See *Bolger v. Amazon.com, LLC*, 53 Cal. App. 5th 431 (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*, 63 Cal. App. 5th 466 (2021).

[10]Consistent with the literature, we assume that the platform does not charge users. Section 4.1 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 4.3.

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

[13]See https://about.fb.com/news/2022/12/meta-launches-new-content-moderation-tool/ and https://about.fb.com/news/2021/08/privacy-enhancing-technologies-and-ads/

[14]www.retailcustomerexperience.com/news/amazon-innovating-in-fight-against-fraudulent-sellers/.

[15]Twitter, however, has reportedly refused to pay the recent bill. See "Twitter is Refusing to Pay its Google Cloud Bills, Platformer Reports," *Reuters*, Jun 11, 2023.

[16]See https://about.linkedin.com/transparency/community-report and a recent FTC report at https://consumer.ftc.gov/consumer-alerts/2023/04/you-got-job. Also, see "Fake Job Scams Are Skyrocketing Online And They're Getting Harder to Detect," *Los Angeles Times*, Jan 12, 2023.

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.[17] First, if the platform is held liable, the platform will raise the interaction price for the firms to reflect the platform's future liability costs. If the harmful firms are "marginal" (i.e., the harmful firms have a lower willingness to pay than the safe firms) then the higher interaction price deters the harmful firms from joining the platform. Second, if the harmful firms are "inframarginal" and undeterrable, the platform will invest resources to detect and remove the harmful firms from the platform.[18] Interestingly, the optimal level of platform liability may be partial instead of full, as full liability could lead to excessive auditing by the platform.[19]

We then consider the more general setting with *heterogeneous* users where some join the platform and others do not. We show that platform liability has the added benefit of stimulating 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.

Next, we 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.

[17]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.

[18]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."

[19]If the firms are very 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.

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.[20] 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).[21]

Finally, we extend the baseline model to consider two *competing platforms*. The users are bystanders and can participate on only one of the platforms (i.e., single-homing), while the firms can participate on both platforms (i.e., multi-homing). If the harmful firms are inframarginal, in equilibrium the users hold the belief that the two platforms take the same auditing effort and allocate themselves equally between the platforms. The platforms' auditing incentives are similar to those in the baseline model, so that the socially-optimal platform liability remains the same. If the harmful firms are marginal then competition raises the platforms' incentives to deter the harmful firms by charging high prices, relative to the baseline model. In this case, platform liability is socially beneficial if the platforms are sufficiently differentiated but unnecessary if otherwise. These observations suggest that policies encouraging platform competition should be complemented by changes in platform liability.[22]

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).[23] 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.[24] Pitchford (1995) explores the desirability of

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

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

[22] Other salient factors, including firm moral hazard, same-side harm, alternative pricing structures, court errors, and litigation costs are also discussed. Online Appendix B presents a formal analysis of several extensions.

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

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

extending liability to an injurer's lenders[25] and Dari Mattiacci and Parisi (2003) consider vicarious liability where liability is extended to the injurer's employer.[26] Arlen and MacLeod (2005a) show that holding managed care organizations liable for medical malpractice by their physicians can raise the physicians' incentives to take care. Our model, which has not been previously studied, investigates the design of platform liability when the platform can audit and remove harmful participants.[27]

There is a vast literature on multi-sided platforms. The early studies (e.g., Caillaud and Jullien, 2003; Rochet and Tirole, 2003, 2006; Amstrong, 2006; and Weyl, 2010) have identified how cross-side externalities affect platform pricing schemes and users' participation incentives. The literature also examines the impact of seller competition[28] or the impact of platform competition on pricing.[29] 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).

There is a small but growing literature on platform liability. The policy papers by Buiten et al. (2020) and Lefouili and Madio (2022) discuss informally whether platforms should bear liability for harms caused by participants. A few working papers study copyright infringement and retail settings. De Chiara et al. (2021) examine hosting platforms' incentives to filter copyright-infringing materials. They focus on harms to copyright owners and do not consider platforms' pricing strategies. 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. Zennyo (2023) considers the impact of platform liability on sellers' efforts to improve product safety, without discussing platforms' screening or auditing actions. Yasui (2022) discusses sellers' incentives to maintain reputation and platforms' ex-post efforts to discover and announce potential safety risks after consumers purchase products from sellers. Our paper considers a broad array of platform types and investigates the effects of liability on platform pricing, incentives to block bad actors, and social welfare.

Our paper is organized as follows. Section 2 presents the baseline model where users are homogeneous bystanders of the firms. Section 3 generalizes the baseline model by considering heterogeneous users with endogenous participation. Section 4 examines sev-

[25]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.

[26]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.

[27]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.

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

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

eral extensions including a retail setting where the firms are sellers and the users are consumers and a setting with two competing platforms. Section 5 provides concluding thoughts. The proofs are in the appendix.

2 The Baseline 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. 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 $v > 0$. For simplicity, we first consider the special case where users are homogeneous and have the same v. Section 3 generalizes the analysis to include heterogeneous users with different valuations. Second, the platform provides opportunities for the firms and the users to interact.

We assume that interactions between firms and users do not require the users' consent and so the users are effectively "bystanders."[30] 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 of type L in the firm population.[31] 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.[32] The firms privately observe their types.

We assume that the platform charges the firms a price p per interaction and allows users to join the platform for free. This is broadly aligned with what we often observe in practice. Platforms such as Google and Facebook monetize quasi-pubic goods by selling online advertising to businesses and/or sharing user data and do not charge users for access. In theory, this pricing strategy can be very profitable for the platform in strategic environments with strong network effects.[33] Our assumption is also aligned with other papers in the platform literature.[34]

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,

[30]Section 4.1 extends the analysis to retail platforms where interactions require the users' consent.

[31]For simplicity, λ is taken as exogenous. One may endogenize λ by allowing firms to invest resources to increase the likelihood being safe. Section 4.3 discusses an extension with firm moral hazard problems.

[32] 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.

[33]Suppose that each user receives v only if a large number of users join the platform. A user's decision to join depends on the price and their expectations about the number of other users. Following Harsanyi and Selten (1988), to avoid coordination failure, the platform should set a sufficiently low price (or even zero price) for the users. The appendix provides an illustrative example of the coordination game.

[34]Armstrong (2006) shows that, with strong network effects, platforms have incentives to set negative prices. However, negative prices may be infeasible. Armstrong and Wright (2007) and Choi and Jeon (2021) justify non-negative prices on adverse selection and moral hazard grounds. Gans (2022) justifies this based on free disposal. See also Belleflamme and Peitz (2021).

and technological sophistication, platforms like Google may be in a good position to root out harmful platform participants.[35] Specifically, by spending effort $e \in [0, 1)$ per firm, the platform can detect H-type firms with probability e and block them from interacting with users.[36] 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 $e \to 1$.[37] In this baseline model, it will not matter whether the platform's effort level e is observable or not.[38]

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.[39] 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.[40] For simplicity, there are no litigation costs or other transaction costs associated with using the court system.[41] 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.[42] A1 implies that it is socially efficient (inefficient) for the L-type (H-type) firms

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

[36] The analysis is the same if the platform takes auditing effort per interaction instead of per firm.

[37] 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.

[38] Assumption A0 below guarantees that homogeneous users will join regardless of their beliefs. Observability is relevant if users have heterogeneous values, however. See the discussion in Section 3.

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

[40] 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.

[41] Section 4.3 discusses the impact of litigation costs.

[42] Similar results are obtained in a model where users have heterogeneous valuations and some users

to join the platform.[43] 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 firms and users 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.
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.

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 present two social welfare benchmarks below.

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. 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.

do not join the platform. See Section 3. Note that if users were naïve or unaware of product risks then they would participate for any $v > 0$.

[43] 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).

2.1 Motivating Examples

In our baseline model, bad actors on one side of a two-sided platform may harm users on the other side of the platform. In the following, we motivate the baseline model with three broad examples: fraudulent advertising, data misuse by technology partners, and the sale of harmful products. For each of these three settings, we will document the platform's financial incentives, the presence of bad actors, and the potential for user harm.

Advertisers. Many platforms rely on paid advertising as their main source of revenue.[44] Although most online advertising is benign, fraudulent and misleading ads abound. Victims of online scams are often left powerless and without recourse. In an early class-action lawsuit, users sued Google for the financial losses that they suffered from being duped by an unscrupulous advertiser into purchasing unwanted cell phone services, including ringtones.[45] A recent report estimated that Google earned $10 million from fake abortion clinics posting advertisements and aiming to stop women from having the procedure.[46] Fraudulent business and job-opportunity postings on LinkedIn and other platforms have also proliferated, with reported harms topping $367 million in 2022, 76% higher than the year before.[47]

The harm from fraudulent and misleading advertising can extend beyond platforms' direct users to society at large. In the Cambridge Analytica scandal, U.S. Prosecutors allege that Russian entities used fake and stolen online personas, including paid advertising on social media sites, to interfere with the 2016 U.S. presidential elections.[48] The charges against the Russian entities, who were effectively judgment proof, were subsequently dropped.[49] In another example, Facebook settled a defamation lawsuit brought by a well-known British journalist over misleading cryptocurrency advertisements claiming the journalist's endorsement.[50] Our baseline model applies to settings like these where the victims are bystanders (i.e., the harms are externalities).

Technology Partners. Platforms often share user data with technology partners, includ-

[44]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 firms' willingness to participate in auctions for advertising. Similarly, most of Facebook's revenue comes from advertising.

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

[46]See "Google Earned $10 Million by Allowing Misleading Anti-abortion Ads From Fake Clinics, Report Says," *CNN*, Jun 15, 2023.

[47]See https://consumer.ftc.gov/consumer-alerts/2023/04/you-got-job and "Too Good to Be True? The Fake Recruiters Targeting Jobseekers," *Financial Times*, June 12, 2023.

[48]See *United States v. Internet Research Agency, et al.* (D.D.C. Feb. 16, 2018). https://www.justice.gov/file/1035477/download. See also Report on the Select Committee on Intelligence at https://www.intelligence.senate.gov/sites/default/files/documents/Report Volume2.pdf

[49]The court noted that prosecuting defendants with "no presence in the United States and no exposure to meaningful punishment in the event of a conviction, promotes neither the interests of justice nor the nation's security." See Motion to Dismiss Concord Defendants, Case 1:18-cr-00032-DLF Document 381 Filed March 16, 2020.

[50]See "Facebook Settles Martin Lewis Lawsuit over Scam Ads," *Financial Times*, Jan 24, 2019.

ing app developers, who can use the data to improve their offerings and the experiences of platform users. When deciding whether to grant developers access to user data, platforms may consider the financial benefit (among other things). For example, in 2013, Facebook allegedly granted or denied access based on the developer spending at least $250,000 in mobile advertising.[51] In the Spring of 2023, Twitter, Reddit, and other platforms announced hefty charges for developers to access the platforms' API, leading some partners to reduce their data usage and others to terminate their contracts.[52] There is substantial evidence that technology partners violate their platform developer agreements[53] and sell data to others.[54] In the wrong hands, platform data can "be used for identity theft, phishing, fraud, and other harmful purposes."[55]

Third-Party Sellers. Retail and gaming platforms make money by sharing sales revenue with third-party sellers. While many consumers are sophisticated and can understand the risks that they face online,[56] others may be unaware of the risks or cannot meaningfully consent to transactions (e.g., children). Consumers who are naïve and unaware of the harms are, for all intents and purposes, bystanders. For example, some online games use "dark patterns" to exploit cognitive biases and to manipulate users into making online purchases. Trickery was central in the FTC cases against Apple, Google, and Amazon for in-app charges associated with "free" games for children.[57] The FTC complaint against Apple described third-party game Dragon Story "as 'sucker[ing] young children into spending huge amounts of money' without their parents' knowledge."[58] In-app sales are financially lucrative for third-party sellers and for the platform. Apple, for example, retained thirty percent of all revenue, including in-app sales.[59]

[51] "Facebook's enforcement of its policies, terms, and conditions, however, was inadequate and was influenced by the financial benefit that violator third-party app developers provided to Facebook." See *United States of America v. Facebook Inc.*, Case 1:19-cv-02184, Complaint for Civil Penalties, Injunction, and Other Relief (Filed 07/24/19). https://www.justice.gov/opa/press-release/file/1186506/download

[52] See "Reddit Wants to Get Paid for Helping to Teach Big AI Systems," *New York Times*, Apr 18, 2023.

[53] The "misuse of data shared with third-party apps on Facebook [includes] ransomware, spam, and targeted advertising." See Farooqi et al. (2020)

[54] App developer Aleksandr Kogan shared the personal information of 87 million Facebook users with Cambridge Analytica. Kogan allegedly received over $800,000 for the collaboration. See "Scholars Have Data on Millions of Facebook Users. Who Is Guarding It? " *New York Times*, May 6, 2018, and "How Trump Consultants Exploited the Facebook Data of Millions," *New York Times*, Mar 17, 2018.

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

[56] We consider sophisticated consumers in Section 4.1 on retail platforms.

[57] See "Bringing Dark Patterns to Light," FTC Staff Report, September 2022.

[58] https://www.ftc.gov/sites/default/files/documents/cases/140115applecmpt.pdf Similarly, when playing Air Penguins on Google Play, it was difficult for users to distinguish between virtual currency and real dollars. https://www.ftc.gov/system/files/documents/cases/140904googleplaycmpt.pdf

[59] The harm from sellers on retail platforms can also go beyond platforms' direct users. For example, counterfeits sold on Amazon harm brand owners, who are effectively bystanders in our baseline model.

2.2 Equilibrium Analysis

In this subsection, we characterize the platform's pricing and auditing strategies, p and e, given the assignment of liability, w_s and w_p. 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. 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.[60] Auditing is necessary to detect and remove the H-type firms.

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 has the ability — but may not have incentive — to deter the H-types from joining the platform by raising the interaction price p.

To summarize, the platform has two possible mechanisms to reduce the harm to users: the price per interaction p and the audit intensity e. In principle, the pricing mechanism is privately and socially more efficient than the auditing mechanism, as the pricing mechanism can deter the H-types without the need for costly audits. However, the pricing mechanism is infeasible when the firm's liability is below a threshold, $w_s \leq \widehat{w}$.

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

Case 1: $w_s \leq \widehat{w}$. Suppose that firm liability is below the threshold, so the L-type firms are marginal. The platform sets the interaction price to extract the L-type firms' rent,[61]

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

The H-types seek to join the platform. Using the definition of $\widehat{w}$ in (5), 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.

[60] 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.

[61] 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.

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.

Next, we 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) may be rewritten as:

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

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. The platform invests in auditing if and only if the joint platform-firm surplus is larger than the firms' information rent. Note that the platform's incentives to audit are stronger when w_p and w_s are larger. The platform's incentive to audit and remove the H-types is socially insufficient when the joint liability for the platform and firms is small (as in case 2(a)) but socially excessive if the joint liability is large (as in case 2(c)).

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, then it will not invest in auditing, $e^* = 0$.[62]

The platform will charge p_H and attract the H-types if and only if

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

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 platform-firm surplus 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. The platform has incentives to deter the H types if and only if the joint platform-firm surplus is less than the firms' information rents, as summarized in the following Lemma.

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.*

[62] 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.

2.3 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, $w_p = 0$.

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.*[63]

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.*

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 the users (case 3 in Proposition 1). In this case, the joint platform-firm surplus of including the H-types is low, so the platform has incentives to deter them by charging a high price. However, when the firms are more judgment proof and the platform faces no liability (cases 1 and 2 in Proposition 1), the private and social incentives diverge. The platform attracts the H-types and does not invest in costly auditing. In such cases, platform liability can be socially desirable, as shown in the next proposition.

Proposition 2. *(Optimal Platform Liability.) 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.*

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

Proposition 2 describes how platform liability can be designed to increase social welfare. In case 1, firm liability is below the threshold ($w_s \leq \widehat{w}$) and the 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 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.

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 attract the H-type firms since the joint platform-firm surplus of including the H-types is larger than the L-type firms' rents. Since the firms' rent is independent of w_p while the joint surplus 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, $w_p^* = d - w_s$.

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 homogeneous user-bystanders. If firms have deep pockets and can compensate the user-bystanders for the harm that they cause, then platform liability is unwarranted. 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. However, when the firms have very limited resources, large platform liability can cause social costs by leading to excessive auditing. In this case, the socially optimal level of platform liability may be less than the full residual harm.

3 Heterogeneous Users

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 generalize the model by considering heterogeneous users who choose whether to joint the platform or not. We will show that, as in the baseline model, platform liability motivates the platform to remove or deter the H-type firms. Moreover, platform liability has the additional effect of stimulating user participation, so that the optimal level of 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 $F(v)$.[64] As in the baseline model, the platform charges the firms price p per interaction and takes auditing effort e per firm. Note that there are economies of scale in (per-firm) auditing, so that both the private and the socially optimal incentives for auditing depend on the users' participation rate.[65] Users have the option to join the platform for free. As discussed in the baseline model, many platforms do not charge users in practice and this observation could emerge in equilibrium when there are strong same-side or cross-side network effects.[66]

We assume that the users cannot directly observe the platform's audit intensity, or equivalently, the platform chooses its audit intensity after the users make their participation decisions.[67] Although the users do not observe the platform's auditing effort e when making their participation decisions, they observe the liability rule, w_s and w_p, and form correct beliefs about e in equilibrium.

In practice, the public does not directly observe platforms' enforcement efforts or technologies used in improving platform safety. In the words of former Facebook employee and whistleblower Frances Haugen, "Facebook became a $1 trillion company by paying for its profits with our safety, including the safety of our children" and "almost no one outside of Facebook knows what happens inside Facebook."[68] The Digital Services Act in the European Union and the PACT Act recently proposed in the US contain many disclosure requirements,[69] which reflects lawmakers' concerns about the lack of transparency on platform safety and effort.[70]

Consider the first-best outcome. Assumption A2 implies that it is socially efficient for all users to participate and assumption A1 implies that it is socially inefficient for the H-type firms to participate. The first-best outcome is achieved if the H-type firms do

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

[65]If auditing is per interaction instead of per firm, the results are similar. Analysis available upon request.

[66]The platform might 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 (that is, if cross-side network effects are strong). In this section, 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.

[67]Recall that, in the baseline model, the observability of the platform's auditing effort e is irrelevant to the results. Observability is relevant in this section. See discussion below.

[68]Written Testimony of Frances Haugen for Congressional Hearing Regarding "Holding Big Tech Accountable," Dec. 1, 2021. https://docs.house.gov/meetings/IF/IF16/20211201/114268/HHRG-117-IF16-Wstate-HaugenF-20211201-U1.pdf. When Facebook, Amazon, Google, and Twitter downsized their safety teams and terminated some fact-checking projects in early 2023, it was hard for the public to understand the implications for platform safety. See "Tech Layoffs Ravage the Teams that Fight Online Misinformation and Hate Speech," *CNBC*, May 26, 2023.

[69]See https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package and https://www.schatz.senate.gov/news/press-releases/schatz-thune-reintroduce-legislation-to-strengthen-rules-transparency-for-online-content-moderation-hold-internet-companies-accountable

[70]However, Maroni (2023) questions the effectiveness of the disclosure requirements in the Digital Services Act and argues that they "cannot solve the problem of information asymmetry."

not join the platform and all the users join the platform.

Next, consider the second-best outcome. As in the baseline model, full deterrence of the H-types may not be possible. If the H-type firms seek to join the platform, then costly auditing is necessary to reduce the social harm. In the second-best benchmark, social welfare is

$$S(e,\widehat{v}) = \int_{\widehat{v}} [v + \lambda(1-e)(\alpha_H - \theta_H d) + (1-\lambda)(\alpha_L - \theta_L d)] f(v)dv - c(e), \tag{12}$$

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

$$\widehat{v}(e,w) = (\lambda(1-e)\theta_H + (1-\lambda)\theta_L)(d-w). \tag{13}$$

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

We now characterize the equilibrium and the optimal platform liability. As in the baseline model, the L-type firms are marginal if $w_s \leq \widehat{w}$, while the H-types are marginal if $w_s > \widehat{w}$. We consider each case in turn.

Case 1: $w_s \leq \widehat{w}$. 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:

$$\Pi(e,\widehat{v}) = S(e,\widehat{v}) + \int_{\widehat{v}} \big\{ -(1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) + ((1-e)\lambda\theta_H + (1-\lambda)\theta_L)(d-w) - v \big\} f(v)dv, \tag{14}$$

where $\widehat{v}$ is the marginal user defined in (13). Since the platform chooses its auditing effort ex post given $\widehat{v}$, the platform's auditing effort e^u (if it is positive) satisfies[71]

$$\frac{\partial \Pi(e^u,\widehat{v})}{\partial e} = \frac{dS(e^u,\widehat{v})}{de} + \int_{\widehat{v}} [\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - \lambda\theta_H(d-w)] f(v)dv + \lambda\theta_H(d-w)\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}} = 0, \tag{15}$$

where

$$\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}} = \frac{\partial \Pi(e^u,\widehat{v})}{\partial \widehat{v}} - \big[\lambda(1-e^u)(\theta_H - \theta_L)(\widehat{w} - w_s)\big] f(\widehat{v}). \tag{16}$$

Equation (15) shows that the platform's auditing incentives diverge from the social planner's. The first line of equation (15) is familiar. As in the baseline model, when the platform increases e, the removed H-types lose their information rents, $\lambda(\theta_H - \theta_L)(\widehat{w} - w_s)$

[71] See the proof of Proposition 3.

and the users' uncompensated loss is reduced by $\lambda\theta_H(d-w)$. If $w_p = w_p^*$ as defined in Proposition 2, these two effects offset each other. The last line of equation (15) identifies a new source of divergence: the platform does not consider the marginal social gain from increased user participation. This new divergence occurs *partly* due to the moral hazard problem that the platform cannot commit to its audit intensity and therefore does not consider the impact of increased participation on its profit (i.e., the first term on the right-hand side of Equation (16)). Importantly, another reason for this new divergence is that the platform does not consider the impact of increased user participation on the firms' information rents (i.e., the last term in Equation (16)).

The above observations imply that the level of optimal platform liability is higher than in the baseline model, $w_p^u > w_p^*$. To see this, recall that the user's participation threshold $\widehat{v}(e,w)$ in equation (13) is a decreasing function of e and w. An increase in platform liability 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 however that platform liability cannot achieve the second-best outcome. Attracting all the users to the platform, $\widehat{v} = 0$, would require fully-compensatory damages, $w_p = d - w_s$, but this would motivate the platform to invest excessively in auditing, $e > e^{**}$.[72]

Case 2: $w_s > \widehat{w}$. In this case, the H-type firms are marginal. As shown in the baseline model, the platform charges either $p_L = \alpha_L - \theta_L w_s$ or $p_H = \alpha_H - \theta_H w_s < p_L$. If the platform charges p_L, the H-types firms are deterred, and anticipating this, the users participate if $v \geq (1-\lambda)\theta_L(d-w)$. If instead, the platform charges p_H, the H-types firms join the platform, and anticipating this, the users participate if $v \geq [\lambda\theta_H + (1-\lambda)\theta_L](d-w)$. Under both scenarios, however, the users' participation incentives are socially insufficient if $w_p < d - w_s$.[73]

The first-best outcome may be obtained by holding the platform liable for full residual damage, $w_p = d - w_s$. First, since the users are fully compensated, they all participate. Second, as shown in Proposition 2, given full user participation and $w_p = d - w_s$, the platform charges $p_L = \alpha_L - \theta_L w_s$ and deters the H-type firms from joining the platform.

Proposition 3. *(Heterogenous Users.) 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.*[74]

[72] 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 $e = e^{**}$.

[73] The platform may choose either price in equilibrium, though it has stronger incentives to deter the H-type firms than in the baseline model.

[74] If the user-participation effect is not overly strong (e.g. $f(0)$ is not too large), it is optimal to assign partial (instead of full) residual liability to the platform. See proof of Proposition 3.

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, if the firms have deep pockets and can be held fully liable ($w_s = d$), platform liability is unnecessary. However, if the firms are judgment proof, 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. Note that, when the firms are very judgment proof (Case 1 in Proposition 3), the optimal platform liability leads to excessive auditing.

Remark on the Chilling Effects of Liability. Lawmakers and commentators have historically expressed concern that the burden of liability might chill economic activity. These concerns were part of the rhetoric for platform immunity to liability in the early years. Section 230 of the Communications Decency Act was adopted to allow the internet to grow and flourish.[75] To be sure, defending against frivolous lawsuits can be costly and distract managers from the core business.[76] However, our analysis shows that platform liability can stimulate user participation, both directly and indirectly.[77]

First, platform liability serves as a "rebate" to attract users. This effect is unique to the platform market. To see this, consider a non-platform market where a seller sells its product to consumers. Although products liability reduces consumers' uncompensated harm, it raises the seller's costs and leads the seller to raise the price of the product, which can neutralize the impacts on output.[78] By contrast, in a platform market with strong network effects, users have the option to join the platform for free. The platform does not adjust the price to fully reflect the users' uncompensated harm or the platform's liability costs and cross-side network benefits (i.e. revenue from the firms). Platform liability stimulates participation by reducing the "effective" price for users.

Second, platform liability can raise the audit intensity, which attracts users indirectly. For both platform and non-platform markets, when users cannot observe product safety or the platform's audit intensity, liability addresses the moral hazard problem and improves

[75]Section 230 and has been called "the one line of federal code that has created more economic value in this country than any other." See https://www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-facebook-google-is-about-to-change.

[76]Court errors and litigation costs are discussed in Section 4.3.

[77]Some empirical studies observe a positive correlation between liability and innovation. Viscusi and Moore (1993) observe that when products liability is low or moderate, raising liability encouraged firms' investments in innovation. Galasso and Luo (2017) identify a positive correlation between liability and innovation.

[78]If consumers have the same preference for product safety and can observe safety before purchase, then liability is irrelevant to output (Hamada, 1976).

safety. The increased safety reduces the joint costs for the platform and users (or the seller and consumers), thereby stimulating user participation. However, the moral hazard problem is not the only reason for the divergence between the platform's auditing incentive and the social incentive. As shown by equation (15), the divergence occurs also because the platform does not consider the benefit of auditing to the inframarginal users or the impact of increased participation on the firms' rents. Platform liability addresses these externalities and motivates the platform to raise audit intensity.

Remark on Observable Effort. Platform liability may be socially beneficial when users observe the platform's auditing effort e before making their participation decisions. In this setting, the platform's auditing incentives are stronger. Recall that in equation (15), when effort is not observable, the platform disregards the social benefit of increased participation (the last term). With observable effort and $w_s \leq \widehat{w}$, the platform's effort (if it is positive) satisfies:

$$\frac{d\Pi(e^u,\widehat{v})}{de} = \frac{dS(e^u,\widehat{v})}{de} + \int_{\widehat{v}} [\lambda(\theta_H-\theta_L)(\widehat{w}-w_s) - \lambda\theta_H(d-w)]f(v)dv \\ + \lambda\theta_H(d-w)\Big(\frac{\partial S(e^u,\widehat{v})}{\partial\widehat{v}} - \frac{\partial\Pi(e^u,\widehat{v})}{\partial\widehat{v}}\Big) = 0. \quad (17)$$

where

$$\frac{\partial S(e^u,\widehat{v})}{\partial\widehat{v}} - \frac{\partial\Pi(e^u,\widehat{v})}{\partial\widehat{v}} = -\big[\lambda(1-e^u)(\theta_H-\theta_L)(\widehat{w}-w_s)\big]f(\widehat{v}). \quad (18)$$

Comparing equation (17) to (15) reveals that the platform has stronger incentives to detect and remove bad actors when its effort is observable (compared to not observable). In equation (17), the platform captures part of the social gain from increased user participation.[79]

Although observability creates stronger incentives to improve platform safety, the platform's incentives are socially insufficient absent platform liability. This is true for two reasons. First, as in the baseline model, taking user participation as fixed, the platform does not fully consider the benefit of auditing for users.[80] Second, greater audit intensity stimulates participation, which generates additional revenue for the platform and additional rents for the firms. However, the platform does not consider the impact of increased participation on the firms' rents.[81]

To illustrate the necessity of platform liability, consider a special case of our model where the H-type firms do not capture equilibrium rents, $w_s = \widehat{w}$.[82] In this case, the platform captures the entire *marginal* social gain from increased participation. The last

[79]The other part of the social gain from increased user participation accrues to the H-type firms.

[80]See case 1 of Proposition 1. In the baseline model where participation is fixed, the platform doesn't consider the effect of auditing on firm rents, which is outweighed by the effect on users (absent platform liability). In other words, the second term on the right-hand side of equation (17) is negative.

[81]In other words, the last term of equation (17) is negative.

[82]For example, this would happen if the two firm types have the same gross willingness to pay, $\alpha_H = \alpha_L$, and are totally judgment-proof, $w_s = 0$.

term in equation (17) drops out and we are left with $d\Pi(\cdot)/de = dS(\cdot)/de - \int_v \lambda\theta_H(d - w)f(v)dv = 0$. Private and social incentives diverge because the platform does not consider the safety benefits that accrue to the participating users. Imposing full residual liability on the platform, $w_p = d - w_s$, aligns the private and social incentives. More broadly, with observable effort and $w_s < \widehat{w}$, the optimal level of platform liability can be higher than in the baseline model, $w_p^u > w_p^*$.[83]

4 Extensions

Our baseline model considers a monopoly platform where interactions between firms and users do not require the users' consent. In this section, we examine several extensions, including retail platforms with consensual market transactions, and platform competition.

4.1 Retail Platforms

We now extend the baseline model 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.[84]

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 effort of

[83]Suppose $w_s < \widehat{w}$ and $w_p = w_p^* < d - w_s$ as defined in Proposition 2. The second term in equation (17) is zero, and the last term is negative. A higher level of platform liability is necessary when the platform's equilibrium effort is an increasing function of platform liability, w_p. Online Appendix B1 provides sufficient conditions for the equilibrium effort to increase in w_p.

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

the platform, but are sophisticated and form beliefs that are, in equilibrium, correct.[85] 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

$$E(\theta|e) = \frac{(1-e)\lambda\theta_H + (1-\lambda)\theta_L}{(1-e)\lambda + (1-\lambda)}, \tag{19}$$

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 ($e = 0$).[86] 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.[87] 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.[88]

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

$$t^r = \alpha_0 - \theta^r(d-w), \tag{20}$$

so consumer surplus is zero.[89] $\theta^r(d-w)$ 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: $w_s \leq \widehat{w}$. Since the L-type firms are marginal, the platform sets p^r to extract

[85] 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. See discussion in Section 2.1.

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

[87] 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.

[88] 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 $w_s = \widehat{w} = (\alpha_H - \alpha_L)/(\theta_H - \theta_L)$.

[89] This equilibrium maximizes the platform's profits. See the proof of Proposition 4. 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.

rents from the L-type firms, $p^r = t^r - (\theta_L w_s + c_L)$. Using (20) and $\alpha_L = \alpha_0 - c_L$,

$$p^r = \alpha_L - \theta_L w_s - \theta^r(d - w). \tag{21}$$

Comparing p^r to its counterpart p^* (see (6)) in the baseline model reveals an important difference: the interaction price paid by the firms (21) reflects the user-consumers' expected uncompensated harm, $\theta^r(d-w)$.

We now explore the platform's auditing incentives. Substituting p^r from (21), $S(e)$ from (2), and $\widehat{w}$ from (5) into (7) gives the platform's profit function

$$\begin{aligned}\Pi(e) = S(e) - v - (1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s) \\ + [(1-e)\lambda(\theta_H - \theta^r) + (1-\lambda)(\theta_L - \theta^r)](d-w).\end{aligned} \tag{22}$$

The platform's profits $\Pi(e)$ diverge from social welfare $S(e)$ for two reasons. 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).[90] The expression in the second line of (22) represents the user's *unanticipated loss or gain* when the platform deviates and invests $e \neq e^r$.[91]

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

$$\Pi'(e^r) = S'(e^r) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - \lambda(\theta_H - \theta^r)(d-w) = 0 \tag{23}$$

where $w = w_s + w_p$. Note that the platform's auditing 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: $w_s > \widehat{w}$. 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

$$(1-\lambda)(\alpha_L - \theta_L d). \tag{24}$$

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.[92] The platform's profits would be strictly lower in this case. To see why, observe that the

[90] 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 (22) is negative (positive).

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

[92] 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.[93]

Proposition 4. *(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 4 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, 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 contrast, in Proposition 2 where the users are bystanders, w_p^*, increases in w_s, that is, 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

$$(\theta_H - \theta_L)(\widehat{w} - w_s) = \theta_H(d - w_s - w_p^*). \tag{25}$$

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.

[93] 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 4, when the users are consumers, w_p^r satisfies

$$(\theta_H - \theta_L)(\widehat{w} - w_s) = (\theta_H - \theta^{**})(d - w_s - w_p^r). \tag{26}$$

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.*

Remark on Disclosure. 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.[94] Although ex post efficiency would be obtained without platform liability, the platform would have insufficient incentives to audit the sellers ex ante.[95] At the other extreme, with full platform liability ($w_p = d$), a rational consumer would interact with a known H-type.[96] 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.[97]

4.2 Platform Competition

We now extend our baseline model (with user-bystanders) by considering two competing platforms, Platform 1 and Platform 2. Users are distributed symmetrically on a Hotelling

[94]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, $\alpha_H - \theta_H d - p^r < 0$.

[95]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.

[96]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.

[97]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.

line with density $g(x) = g(1-x) > 0$ on $x \in [0,1]$, Platform 1 is located at $x = 0$ while Platform 2 is located at $x = 1$. A user at location $x \in [0,1]$ receives consumption value $v - \tau x$ if they join Platform 1 but $v - \tau(1-x)$ if they join Platform 2, where $\tau \geq 0$ reflects the level of differentiation. Assume that v is sufficiently large such that the market is fully covered. The firms can join both platforms, while each user only joins one platform.[98] Thus, the platforms compete for users but not for firms.[99]

In stage 1, the platforms set their prices simultaneously. The timing and the other assumptions are otherwise identical to the baseline model. Denote the platforms' prices and auditing efforts as p_j and e_j, $j = 1, 2$. We shall focus on the symmetric equilibrium where $p_1 = p_2$ and $e_1 = e_2$ and, accordingly, each platform serves half of the users. We will show that platform liability can still be socially beneficial in this competitive environment.

Case 1: $w_s \leq \widehat{w}$. In this case, the L-type firms are marginal and the platforms set $p_1 = p_2 = \alpha_L - \theta_L w_s > 0$. Although the users do not observe the platforms' auditing efforts directly, they are sophisticated and form rational inferences in equilibrium. In the symmetric equilibrium, users hold the belief that the two platforms take the same auditing effort, $e_1 = e_2 = e^c$ and allocate themselves equally between the two platforms. Each platform's profit ($j = 1, 2$) can be written as

$$\Pi_j(e_j) = \frac{1}{2}\big\{S(e_j) - \lambda(1-e_j)(\theta_H - \theta_L)(\widehat{w} - w_s) + [\lambda(1-e_j)\theta_H + (1-\lambda)\theta_L](d-w)\big\} \quad (27)$$

and each platform's auditing effort e^c (if it is positive) satisfies

$$\Pi_j'(e^c) = \frac{1}{2}\big[S'(e^c) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - \lambda\theta_H(d-w)\big] = 0. \quad (28)$$

This is equivalent to equation (10) in the baseline model. Therefore the optimal platform liability is the same as in Proposition 2 in the baseline model, $w_p^c = w_p^* \in (0, d - w_s]$.

Case 2: $w_s > \widehat{w}$. In this case, the H-type firms are marginal. The platforms have a choice: they can either charge the firms $p_L = \alpha_L - \theta_L w_s$ and deter the H-types or charge the firms $p_H = \alpha_H - \theta_H w_s < p_L$ and attract both types. As shown in the baseline model, when $w_s \geq \widetilde{w} > \widehat{w}$, a platform's per-user profit by charging p_L is always larger than charging p_H. With competition, a platform can attract more users by charging p_L instead of p_H, because the users observe the prices and prefer to join a platform that deters the H-type firms. Therefore, given $w_s \geq \widetilde{w}$, both platforms charge p_L, which implements the first-best outcome. As in the baseline model, platform liability is unnecessary.

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

[99] In some applications, firms may join only one platform. In an earlier version of the paper, we considered an extension where the platforms would compete for firms. If firms are very judgment proof, competition reduces the platforms' profit margin and therefore raises their auditing incentives. In this case, the optimal platform liability is positive but less than that in the baseline model.

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

$$\frac{1}{2}(1-\lambda)(p_L - \theta_L w_p) > \frac{1}{2}[\lambda(p_H - \theta_H w_p) + (1-\lambda)(p_H - \theta_L w_p)],$$

which holds given $p_H < p_L$ and $p_H - \theta_H w_p = \alpha_H - \theta_H d < 0$. Hence, imposing full residual liability on the platforms gets the platforms to raise the interaction price and deter the H-type firms, implementing the first-best outcome.

We now show that platform liability is necessary when $w_s \in (\widehat{w}, \widetilde{w})$ and τ is sufficiently large. Suppose to the contrary that the first-best outcome is obtained with no platform liability, $w_p = 0$. If both platforms charge p_L, each platform's profit is $(1-\lambda)p_L/2$. If Platform 1 deviates to p_H, the indifferent user's location $\widehat{x}$ satisfies

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

that is,

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

Accordingly, Platform 1's profit from deviation is

$$G\Big(\max\Big\{0, \frac{1}{2} - \frac{\lambda\theta_H(d - w_s)}{2\tau}\Big\}\Big)p_H, \tag{29}$$

which goes to 0 when $\tau \to 0$ and goes to $p_H/2$ when $\tau \to \infty$. Note that $(1-\lambda)p_L < p_H$ given $w_s \in (\widehat{w}, \widetilde{w})$. Hence, there exists a threshold $\widetilde{\tau} > 0$ such that, absent platform liability, both platforms charge p_L if and only if $\tau \leq \widetilde{\tau}$. If $\tau > \widetilde{\tau}$, platform liability is socially desired and, in particular, imposing full residual liability on the platforms implements the first-best outcome.

If $\tau \leq \widetilde{\tau}$, platform liability is unnecessary. Since the price that the platforms charge is observed by users, and the platforms are not highly differentiated, the users will prefer to join a platform that charges p_L and completely deters the harmful H-types. In the unique symmetric Bayesian Nash equilibrium, the platforms charge the firms p_L, deter the harmful firms, and split the market.

Proposition 5. *(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 $w_p^c = w_p^*$ achieves the second-best outcome. The platforms set $p^c = \alpha_L - \theta_L w_s$ and attract the H-type firms. The platforms' auditing incentives are socially efficient.*

2. *If $w_s \in (\widehat{w}, \widetilde{w})$, there exists $\widetilde{\tau} > 0$: when $\tau \leq \widetilde{\tau}$, platform liability is unnecessary; when $\tau > \widetilde{\tau}$, $w_p^c = d - w_s$ achieves the first-best outcome. The platforms set $p^c = \alpha_L - \theta_L w_s$ and deter the H-type firms.*

3. *If $w_s \geq \widetilde{w}$, platform liability is unnecessary. The platforms set $p^c = \alpha_L - \theta_L w_s$ and deter the H-type firms.*

Comparing Proposition 5 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 is the same as for monopoly, $w_p^c = w_p^*$. As before, platform liability encourages the platforms to detect and remove the H-type firms from the platforms. If the firms are modestly judgment proof, $w_s \in (\widehat{w}, \widetilde{w})$, then platform liability is socially beneficial when the platforms are sufficiently differentiated (large τ) but unnecessary when platform competition is fierce (small τ). By contrast, in the baseline model, platform liability was necessary to induce the platform to raise the interaction price to deter the bad actors. Here, when competition is fierce, the market mechanism gives the platforms the incentive to raise their interaction prices and deter the bad actors from participating.

Regulators across the globe have been focusing efforts on increasing competition and reducing market power in platform markets. 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.[100] The Digital Services Act and Digital Markets Act in the European Union are geared towards establishing a level playing field (to foster innovation and competitiveness) and creating a safer digital space for users and others.[101] Our analysis shows that policies that encourage platform competition should be complemented by changes in platform liability. When bad actors are judgment proof and undeterred, then platform liability plays an important role of encouraging platforms to invest efficiently to protect users from harm.

4.3 Other Extensions

Firm Moral Hazard. In our baseline model and main extensions, platforms played an instrumental role in solving the adverse selection problem by detecting and removing bad actors from the platforms. As discussed in Section 2.1, adverse selection is empirically relevant: Bad actors, masquerading as legitimate firms, post fraudulent advertisements, steal user data, and sell counterfeit products. Moral hazard is also empirically relevant: Otherwise legitimate app developers may sell user data to others and manufacturers may cut corners to lower costs and raise profit margins. When firms are judgment proof, platform liability can play an instrumental role in solving moral hazard problems, too.

Our baseline model can be easily adapted to reflect a moral hazard problem. Suppose all the firms are identical ex ante but may become either the L-type or H-type ex post. A firm can take (unobservable) care at cost $k > 0$, which reduces the probability of

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

[101] See https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package. A report written by Cremer, et al. and published by the European Commission (2019) raised concerns about increased concentration in platform markets. See https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

becoming an H-type. If the firms are very judgment proof ($w_s \leq \widehat{w}$), then the H-types earn information rents. It follows that ex ante the firms have no incentive to take care and, as in the baseline model, platform liability raises the platform's auditing incentives.[102] When the firms are modestly judgment proof (w_s in a middle range), platform liability motivates the platform to raise the interaction price, which deters the H-type firms and, under certain conditions, motivates the firms to take ex ante effort.[103]

Same-Side Harms. The previous analysis considered a setting with cross-side harms: Firms on one side of the platform harmed the users on the other side of the platform. In practice, some users on platforms may harm other users. For example, some influencers on TikTok create videos that draw attention but may induce children to engage in dangerous activities; celebrities' endorsement of cryptocurrency may persuade investors to buy risky tokens.[104] These influencers can monetize user attention by collaborating with brands or sharing advertising revenue with platforms.[105]

Our model can be adapted to investigate such cases with same-side harms. Consider for example a social networking platform where most user-generated content is perfectly safe but some of it is socially harmful. 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 incentives to detect and remove harmful content.

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.[106] 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

[102] If the firms can escape liability, their incentives are diluted. This leads to an ex-post adverse selection problem.

[103] See online Appendix B2 for the formal analysis. When the firms have deep pockets (w_s is high), however, it can be optimal to *not* impose platform liability, as platform liability may lead to a very high interaction price, which removes the L-types' rent and therefore lowers the firms' incentives to take effort.

[104] See "TikTok's Viral Challenges Keep Luring Young Kids to Their Deaths," *Bloomberg*, Nov 30, 2022, and "Celebrities Who Endorsed Crypto, NFTs Land in Legal Crosshairs After Investor Losses," *Wall Street Journal*, Jan 30, 2023.

[105] YouTube and TikTok have programs sharing advertising revenue with influencers, while Facebook has a "Brand Collabs Manager" program that matches brands with influencers. The value of influencer marketing was estimated to be $16.4 billion in 2022. (https://influencermarketinghub.com/influencer-marketing/)

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

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.[107] 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 baseline model 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.[108]

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.[109] On the other hand, when the H-type firms are marginal, litigation costs raise the platform's incentives to deter these harmful firms, so that platform liability can be lower than in the baseline model. Furthermore, insofar as the costs of litigation exceed the benefits of improved platform incentives, a lower level of liability, or indeed the elimination of liability altogether, may be warranted.

5 Conclusion

Should platforms be held liable for the harms suffered by platform participants? This question is of practical as well as academic interest. Platforms in the United States and

[107] 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.

[108] See online Appendix B3 for the formal analysis.

[109] See online Appendix B4 for the formal analysis.

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. Moreover, platform liability can stimulate user participation when users are heterogeneous. To prevent overinvestment in detection and removal, the residual liability assigned to the platform may be partial instead of full. The optimal level of platform liability also depends on the intensity of platform competition, and whether users are involuntary bystanders or voluntary consumers of the firms. With appropriate incentives, platforms can play an important role in reducing social costs.[110]

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[111] and for holding bricks-and-mortar retailers liable for the harm caused by defective products.[112] However, we believe that the insights are particularly salient for online platforms including Facebook, Google, and Amazon. First, the harmful participants on platforms are frequently small and judgment proof with insufficient incentives to curtail their harmful activities. Second, the big tech giants have the data and technologies to detect and block participants that are more likely to harm others.

Our basic argument for holding platforms liable is valid regardless of the accuracy of the platforms' current screening technologies and existing moderation efforts. First, the lack of effort by some platforms could reflect the weak incentives provided by the legal, economic, and political systems.[113] Platforms may even have "perverse incentives" to

[110]Our model abstracted from reputation building and peer-to-peer reviews. Many platforms rely on a combination of screening and peer-to-peer feedback mechanisms. See Tadelis (2016) for a thoughtful discussion of the limits and biases in peer-to-peer feedback mechanisms. Future work can explore whether platform liability is a substitute or a complement for reputation building.

[111]See *Braun v. Soldier of Fortune Magazine, Inc.*, 968 F.2d 1110, 6 Fla. L. Weekly Fed. C 985 (11th Cir. 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."

[112]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."

[113]In the U.S., Section 230's "Good Samaritan" provision grants broad discretion to platforms to remove content that (in the platform's view) is socially harmful, but immunity from liability creates incentive misalignment.

reduce their control of online activities, similar to the potential distortion caused by vicarious liability on organizations.[114] Second, our model shows that platform liability may be socially desirable even if auditing is very costly or *completely ineffective* at detecting bad actors. Although platforms would not engage in auditing in this case, liability would force platforms to internalize the social harms and create an incentive for them to use the price mechanism to deter bad actors.

There is active debate over whether platforms may be treated as common carriers.[115] Common carriers, including telephone companies, mail carriers, and transportation systems (e.g., railroads and airlines) have a duty to serve the general public and may not generally exclude users.[116] Common carriers are, however, subject to regulations that ensure public safety and sometimes have discretion or even a duty to exclude parties that may cause harm to others. For example, under federal law, airlines must deny transport to passengers who refuse to be searched for weapons,[117] and airline pilots have "permissive removal" authority to deny service to passengers who appear nervous or potentially disruptive.[118] Although the Digital Millennium Copyright Act limits liability for internet service providers (ISPs), it also requires ISPs to terminate the accounts of repeat infringers.[119] In a lawsuit brought against Western Union, the court opined that the defendant was in fact *obligated* to discontinue service for illegal gambling communications.[120] Common carriers can be held liable if they fail to meet their duties[121] and, in many jurisdictions, the standard of care exceeds "reasonable care."[122]

This article advances the idea that liability can play an instrumental role making

[114]In the EU, platforms are not liable for harmful content if they are unaware of it, which "creates perverse incentives for platforms not to monitor online activity." See Lefouili and Madio (2022). Similarly, under vicarious liability, organizations may eschew control over agents (e.g. using subcontractors instead of employees) to avoid tort liability. See Arlen and MacLeod (2005b).

[115]See Rahman (2018) and Volokh (2021).

[116]See 15 U.S. Code §375. If platforms were considered common carriers, they would not be able to exclude users with certain political views. However, as newspapers, platforms can create value by having some discretion in selecting and/or recommending content and products to users. See Bhagwat (2022).

[117]See 49 U.S.C. §44902(a)(1). Similarly, in *Holton v. Boston E. R. Co.*, 303 Mass. 242 (1939), the court held that a street car company had a legal obligation to exclude passengers "who manifest a boisterous and belligerent attitude and threatened to assault persons with his reach."

[118]See 49 U.S.C. §44902(b). Other common carriers may have such discretion too. In *Occhino v. Northwestern Bell Tel. Co.*, 675 F.2d 220 (8th Cir. 1982), the court upheld the telephone company's decision to exclude a subscriber for repeatedly making "harassing" and "abusive" phone calls. See other examples in Sitaraman (2023).

[119]See 17 U.S.C. §512.

[120]*Hamilton v. Western Union Telegraph Co.*, 34 F. Supp. 928 (N.D. Ohio 1940). Similarly, 18 U.S.C. §1084(d) requires FCC-regulated common carriers to discontinue service when they are notified by a law enforcement agency about illegal gambling.

[121]In *BMG Rights Mgt. (US) LLC v. Cox Communs., Inc.*, 881 F.3d 293 (4th Cir. 2018), the appeals court agreed with the district court denying Cox a safe harbor defense under the Digital Millennium Copyright Act (DMCA), because the ISP failed to implement its policy to terminate infringers' accounts.

[122]For example, GA Code §46-9-132 (2020) states that "a common carrier of passengers is bound to exercise extraordinary diligence." Also, California Civil Code §2100 specifies that a common carrier must "use the utmost care and diligence for their safe carriage."

platforms safer for users and for society more broadly. An open question is whether civil liability is the best mechanism to accomplish these goals, or whether regulation would prove more effective. Social media and other platforms share similarities to common carriers and public utilities and so, by analogy, one could in principle regulate them in similar ways. Platform liability arguably has substantial advantages over regulation. Specifically, given the complexity and diversity of platforms, it would be difficult (and perhaps inadvisable) for regulators to set uniform safety standards.[123] Moreover, given the rapidly changing market conditions, regulators would be chasing a moving target. Platforms, especially big tech platforms, have the relevant information to weigh the social costs and benefits. Liability would give platforms financial incentives to use their discretion for the greater good.

[123]This view is shared by many platforms; eBay's 2022 Transparency Report states: "regulatory regimes or technology mandates that are 'one size fits all' can actually serve to limit the tools, resources and partnerships necessary to combat bad actors."

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 (2005a), pp. 494-515.

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

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

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

[5] 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.

[6] 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.

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

[8] Belleflamme, Paul and Martin Peitz, *The Economics of Platforms*, Cambridge University Press (2021).

[9] Bhagwat, Ashutosh, "Why Social Media Platforms Are Not Common Carriers?" *Journal of Free Speech Law*, Vol. 2 (2022), pp. 127-156.

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

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

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

[13] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[31] Galasso, Alberto, and Hong Luo, "Tort Reform and Innovation," *Journal of Law and Economics*, Vol. 60 (2017), pp. 385-412.

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

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

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

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

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

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

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

[39] Hamada, Koichi, "Liability Rules and Income Distribution in Product Liability," *American Economic Review*, Vol. 66 (1976), pp. 228-234.

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

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

[42] Harsanyi, John C., and Reinhard Selten, *A General Theory of Equilibrium Selection in Games*, MIT Press (1988).

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

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

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

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

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

[48] 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.

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

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

[51] Maroni, Marta, "Mediated Transparency: The Digital Services Act and the Legitimisation of Platform Power," (2023) working paper.

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

[53] 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.

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

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

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

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

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

[59] Sitaraman, Ganesh, "Deplatforming," *Yale Law Journal*, forthcoming.

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

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

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

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

[64] 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.

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

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

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

[68] Viscuci, W. Kip, and Michael J. Moore, "Product Liability, Research and Development, and Innovation," *Journal of Political Economy*, Vol. 101 (1993), pp. 161-184.

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

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

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

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

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

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

Appendix A

An Example of the Coordination Game. This example illustrates the idea that, given the same-side network effects, the platform finds it optimal to set a sufficiently small price (or even zero price) for the users.

Suppose that there are two potential users, 1 and 2, who independently choose whether to join the platform or not. Each user receives a private benefit, v, if and only if both users join the platform. In addition, when joining the platform, a user incurs costs, $x \leq v/2$, which can include entry costs, opportunity costs, and the expected harm caused by the firms on the platform. The platform charges the same membership fee, $m \geq 0$, to all users.

If both users join the platform, each user's net payoff is $v - x - m$. If only one user joins, this user's net payoff is $-x - m$, while the other user's payoff is 0. If neither user joins, each user receives 0. In this game, there are two pure-strategy Nash equilibria, one with both users joining and the other with neither joining (i.e., coordination failure). Applying the risk-dominance refinement by Harsanyi and Selten (1988), we can show that in equilibrium both users join the platform if $(v-x-m)+(-x-m) \geq 0$ but do not join if otherwise. That is, coordination failure can be avoided if $m \leq v/2 - x$. When $x = v/2$, the platform chooses $m = 0$.

Proof of Lemma 1. We first show that, if $w_s \leq \widehat{w}$, 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 $\widetilde{w}(\lambda) = \frac{\alpha_H - \alpha_L + \lambda\alpha_L}{\theta_H - \theta_L + \lambda\theta_L} \in \big(\widehat{w}, \frac{\alpha_H}{\theta_H}\big)$. 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 $\widetilde{w}(\lambda) = \max\big\{\frac{\alpha_H - \alpha_L + \lambda\alpha_L}{\theta_H - \theta_L + \lambda\theta_L}, \widehat{w}\big\}$.

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 $w_p = w_p^* = d - w_s - \big(1 - \frac{\theta_L}{\theta_H}\big)(\widehat{w} - w_s)$. 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 $\underline{w}_p$:

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

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 start by showing that the platform does not charge the users (i.e. $m = 0$) 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

$$\widehat{\Pi}(e) = (1-e)\lambda(\alpha_L - \theta_L w_s - \theta_H w_p) + (1-\lambda)(\alpha_L - \theta_L w). \tag{30}$$

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 (15), which highlights the potential divergence between the private and social incentives for auditing. Given w, (12) implies

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

Using (14), if the equilibrium auditing effort is positive, then e^u satisfies

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

Next, we show that, if $w_s<\widehat{w}$, then $w_p^u>w_p^*$. Totally differentiating (12) with respect to w_p gives

$$\frac{dS(e^u,\widehat{v})}{dw_p}=\Big[\frac{\partial S(e^u,\widehat{v})}{\partial e}-\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\lambda\theta_H(d-w)\Big]\frac{\partial e^u}{\partial w_p}+\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}, \tag{31}$$

where $\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}<0$ 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 $w_p\le w_p^*$,

$$\lambda(\theta_H-\theta_L)(\widehat{w}-w_s)<\lambda\theta_H(d-w), \tag{32}$$

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

$$\frac{\partial \Pi(e^u,\widehat{v})}{\partial e}=-\int_{\widehat{v}}\lambda(\alpha_L-\theta_L w_s-\theta_H w_p)f(v)dv-c'(e^u)=0, \tag{33}$$

which implies $\frac{\partial e^u}{\partial w_p}>0$.

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

$$\frac{dS(e^u,\widehat{v})}{dw_p}=\Big[\frac{\partial S(e^u,\widehat{v})}{\partial e}-\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\lambda\theta_H(d-w)\Big]\frac{\partial e^u}{\partial w_p}+\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}>0. \tag{34}$$

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

Moreover, if $w_p=d-w_s>w_p^*$, then all the users join the platform, that is, $\widehat{v}=0$. In this case, $\frac{\partial S(e^u,\widehat{v})}{\partial e}<0$ and $\frac{\partial e^u}{\partial w_p}>0$ are independent of $f(0)$. Note that the absolute value of $\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}$ is

$$\big[\lambda(1-e^u)(\alpha_H-\theta_H w)+(1-\lambda)(\alpha_L-\theta_L w)\big]f(0), \tag{35}$$

which decreases in $f(0)$. Hence, if $w_p=d-w_s$ and $f(0)$ is sufficiently small, then $\frac{dS(e^u,\widehat{v})}{dw_p}<0$, which implies $w_p^u<d-w_s$.

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, $e=e^{**}$.

Proof of Proposition 4. 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.

$$\max_{\{e,p,t\}} \quad \Phi(e,p) = (1-e)\lambda(p - \theta_H w_p) + (1-\lambda)(p - \theta_L w_p) - c(e) \tag{36}$$

subject to

$$e = \arg\max_{e' \geq 0} \Phi(e', p) \tag{37}$$

$$\alpha_0 - t - E(\theta|e)(d - w_s - w_p) \geq 0 \tag{38}$$

$$t - (\theta_L w_s + c_L) - p \geq 0. \tag{39}$$

(37) is the platform's incentive compatibility constraint, (38) is the consumer's participation constraint, and (39) is the L-type firm's participation constraint.[124]

The L-type's participation constraint (39) must bind. To see this, consider two cases. First, suppose that neither (38) nor (39) binds. Then the platform would increase the price p which would increase the platform's profits in (36) and maintain the consumer's participation constraint (38). Second, suppose that (38) binds while (39) does not. Again, the platform would increase the price p marginally. The direct effect of increasing p is

[124]The H-type's participation constraint is satisfied if (39) holds, and is therefore not included in the program.

that the platform's profits in (36) increase. Since $\partial^2\Phi(e,p)/\partial e\partial p = -\lambda < 0$, increasing p also (weakly) decreases the platform's effort e in (37), which in turn raises $E(\theta|e)$ and, since (38) binds, reduces t. However, since t is not in (36), the platform's profits still increase.

Since the L-type's constraint (39) binds, $p = t - (\theta_L w_s + c_L)$ and we can rewrite the optimand (36) as a function of e and t:

$$(1-e)\lambda(t - (\theta_L w_s + c_L) - \theta_H w_p) + (1-\lambda)(t - (\theta_L w_s + c_L) - \theta_L w_p) - c(e). \quad (40)$$

Next, we show that the consumer's participation constraint (38) binds. Suppose not. Then, the platform would increase t and its profits would rise. Since both participation constraints (38) and (39) bind, we have

$$p = \alpha_0 - E(\theta|e)(d - w_s - w_p) - (\theta_L w_s + c_L). \quad (41)$$

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

$$e^r = \arg\max_{e\geq 0} \Phi(e, p^r) \quad (42)$$

$$t^r = \alpha_0 - E(\theta|e^r)(d - w) \quad (43)$$

$$p^r = \alpha_L - \theta_L w_s - E(\theta|e^r)(d - w). \quad (44)$$

We now verify that the values $\{e^r, p^r, t^r\}$ defined in (42), (43), and (44) are an equilibrium of the game. Suppose that the platform charges p^r in (44), and that the firms and consumers believe that the probability of harm is $\theta^r = E(\theta|e^r)$ where e^r is defined in (42). The consumers are (just) willing to pay t^r in (43) and the L-type firms are (just) willing to pay p^r in (44). If the consumers and the firms all participate, the platform exerts effort e^r in (42). 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 $E(\theta|0) = \theta^0$.[125] In this scenario, the most that consumers would be willing to pay is $t = \alpha_0 - \theta^0(d - w)$ from (38). 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 (39). 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.[126]

[125]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.

[126]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.

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 (38) binds we have $t^r = \alpha_0 - \theta^0(d-w)$; since the L-type firm's participation constraint (39) 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 (37) 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

$$\alpha_L - \theta_L w_s - \theta^0(d-w) - \theta_H w_p \geq 0. \tag{45}$$

Adding and subtracting terms this becomes

$$(\alpha_H - \theta_H d) - (\alpha_H - \alpha_L) - \theta_L w_s - \theta_H w_p + \theta_H w + (\theta_H - \theta^0)(d-w) \geq 0, \tag{46}$$

and rearranging this expression gives

$$(\alpha_H - \theta_H d) + (\theta_H - \theta^0)(d-w) \geq (\alpha_H - \alpha_L) - (\theta_H - \theta_L)w_s. \tag{47}$$

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

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 (23). 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 (23) 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 (23) 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

$$p^r = \alpha_L - \theta_L w_s - \theta_L(d-w) = \alpha_L - \theta_L(d - w_p) \tag{48}$$

and the platform's profits are

$$(1-\lambda)(p^r - \theta_L w_p) = (1-\lambda)(\alpha_L - \theta_L d). \tag{49}$$

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

$$p^r = \alpha_H - \theta_H w_s - \theta^0(d - w). \tag{50}$$

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 4. Suppose $w_s \leq \widehat{w}$, so the L-type is marginal. From Claim 1, we have $e^r = e^{**}$ if and only if

$$(\theta_H - \theta_L)(\widehat{w} - w_s) - (\theta_H - \theta^{**})(d - w) = 0. \tag{51}$$

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.

Online Appendix B

This appendix contains the analysis of four additional extensions: heterogeneous users with observable effort, firm moral hazard, false positives and litigation costs.

B1. Heterogeneous Users with Observable Effort

Section 3 shows that platform liability can be socially desired when heterogenous users make participation decisions but do not observe the platform's auditing effort. Now we consider the setting where the platform can commit to its auditing effort before the users make participation decisions.

If $w_s > \widehat{w}$ then the analysis is the same as case 2 in Section 3. As shown in Section 3, if $w_s > \widehat{w}$, the platform would not take any auditing effort, and imposing full residual liability on the platform implements the first-best outcome. The following analysis examines case 1 where $w_s \leq \widehat{w}$.

When auditing effort is observable, equation (14) implies that the platform's effort (if it is positive) satisfies

$$\frac{d\Pi(e^u,\widehat{v})}{de} = \frac{dS(e^u,\widehat{v})}{de} + \int_{\widehat{v}} [\lambda(\theta_H-\theta_L)(\widehat{w}-w_s) - \lambda\theta_H(d-w)]f(v)dv \\ - \lambda\theta_H(d-w)[\lambda(1-e^u)(\theta_H-\theta_L)(\widehat{w}-w_s)]f(\widehat{v}) = 0 \quad (52)$$

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

When $w_s = \widehat{w}$, $\frac{d\Pi(e^u,\widehat{v})}{de} = \frac{dS(e^u,\widehat{v})}{de}$ if and only if $w_p^u = d - w_s$. Therefore, imposing full residual liability on the platform implements the second-best outcome: the platform chooses $e^u = e^{**}$ and all the users join the platform.

When $w_s < \widehat{w}$, the last term on the right-hand side of equation (52) is negative. Moreover, if $w_p \leq w_p^*$, where $w_p^* \in (0, d-w_s)$ is the optimal platform liability in Proposition 2 of the baseline model, then the second term on the right-hand side of equation (52) is non-positive. Therefore, $\frac{dS(e^u,\widehat{v})}{de} > 0$, that is, the platform's auditing incentive is socially insufficient. The social planner chooses w_p to maximize social welfare:

$$\frac{dS(e^u,\widehat{v})}{dw_p} = \frac{dS(e^u,\widehat{v})}{de}\frac{de^u}{dw_p} + \frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}, \quad (53)$$

where $\frac{\partial \widehat{v}}{\partial w_p} = -(\lambda(1-e^u)\theta_H + (1-\lambda)\theta_L) < 0$. Since $\frac{\partial S(\cdot)}{\partial \widehat{v}} < 0$, the last term in (53), $\frac{\partial S(e^u,\widehat{v})}{\partial \widehat{v}}\frac{\partial \widehat{v}}{\partial w_p}$, is non-negative. Intuitively, given auditing effort, platform liability stimulates user participation and therefore raises social welfare. Moreover, as shown earlier, $\frac{dS(e^u,\widehat{v})}{de} > 0$ if $w_p \leq w_p^*$. Hence, as long as $\frac{de^u}{dw_p} > 0$, it is socially optimal to set $w_p^u > w_p^*$.

Proposition 6. *(Heterogeneous Users with Observable Effort) Suppose firm liability is $w_s \leq \widehat{w}$ and the platform commits to its auditing effort. The socially-optimal platform liability for harm to users, w_p^u, is as follows:*

1. *If $w_s < \widehat{w}$, then $w_p^u > w_p^*$ as long as $\frac{de^u}{dw_p} > 0$. The platform 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 socially efficient auditing effort $e^u = e^{**}$. All users participate.*

Example: Uniform Distribution. Recall that, if the users cannot observe auditing effort, the equilibrium effort increases in w_p. However, with observable effort, the equilibrium effort may increase or decrease in w_p. For illustration, suppose that v follows the uniform distribution on $[0, \overline{v}]$. Then with observable effort, the platform's effort (if it is positive) satisfies

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

which implies

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

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

To summarize, even if the users observe the auditing effort, platform liability can be socially desired. The optimal platform liability is (weakly) larger than in the baseline model, as long as the equilibrium effort increases in w_p, which holds when v follows the uniform distribution on $[0, \overline{v}]$ with sufficiently large $\overline{v}$.

B2. Firm Moral Hazard

The baseline model assumes that the firms' types are exogenously given. Platform liability can still be socially beneficial if the firms' types are endogenous and the firms can take effort to improve safety. In this section, suppose all the firms are identical ex ante but

may become either the L-type or H-type ex post. If a firm takes (unobservable) care with cost $k > 0$, the probability of becoming an H-type is λ. If the firm does not take care, the probability of being an H-type rises to $\widehat{\lambda} > \lambda$. The platform commits to its price p before the firms decide to take care or not. The firms privately learn their realized types and decide whether to join the platform.

For simplicity, we maintain the following assumption

$$k < (\widehat{\lambda} - \lambda)(\alpha_L - \theta_L d) + \lambda(\alpha_H - \theta_H d). \tag{54}$$

Assumption (54) leads to several implications.

First, since $\alpha_H - \theta_H d < 0$, $k < (\widehat{\lambda} - \lambda)(\alpha_L - \theta_L d)$. If the H-types never join the platform, it is socially efficient for the (ex ante identical) firms to invest k.

Second, Assumption (54) implies

$$k < (\widehat{\lambda} - \lambda)[(\alpha_L - \theta_L d) - (\alpha_H - \theta_H d)] = (\widehat{\lambda} - \lambda)(\theta_H - \theta_L)(d - \widehat{w}).$$

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

Finally, Assumption (54) implies

$$\lambda(\alpha_H - \theta_H d) + (1 - \lambda)(\alpha_L - \theta_L d) - k > (1 - \widehat{\lambda})(\alpha_L - \theta_L d),$$

that is, social welfare is larger if all the firms invest k and join the platform than if no firm invests and only the L-types join the platform.

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

$$k < (\widehat{\lambda} - \lambda)(\theta_H - \theta_L)(w_s - \widehat{w}).$$

Case 1: $w_s \leq \widehat{w}$. The L-types are marginal. The platform charges $p = \alpha_L - \theta_L w_s$. Since the L-types do not receive any rent, ex ante the firms have no incentive to take care. As in the baseline model, $w_p^k = w_p^* \in (0, d - w_s]$ achieves the second-best outcome.

Case 2: $w_s > \widehat{w}$. The H-types are marginal. Consider three scenarios.

Case 2.1: $w_s > \frac{\alpha_H}{\theta_H}$. Then the H-types would never join the platform. The platform either charges $p_L = \alpha_L - \theta_L w_s$, under which the firms would not invest k, or charges p_0, where

$$p_0 = \alpha_L - \theta_L w_s - k/(\widehat{\lambda} - \lambda) > 0,$$

under which the firms would invest k. Social welfare is larger if the platform charges p_0. The paltform's profit under p_L is

$$\Pi^L = (1 - \widehat{\lambda})(\alpha_L - \theta_L w_s - \theta_L w_p);$$

while its profit under p_0 is

$$\Pi^0 = (1 - \lambda)(\alpha_L - \theta_L w_s - \theta_L w_p) - k(1 - \lambda)/(\widehat{\lambda} - \lambda).$$

The profit difference,

$$\Pi^0 - \Pi^L = (\widehat{\lambda} - \lambda)(\alpha_L - \theta_L w_s - \theta_L w_p) - k(1-\lambda)/(\widehat{\lambda} - \lambda),$$

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

Case 2.2: $w_s \in (w^k, \frac{\alpha_H}{\theta_H})$**.** Given $w_s < \frac{\alpha_H}{\theta_H}$, the H-types may have incentives to join the platform. Moreover, given $w_s > w^k$, we have $k < (\widehat{\lambda} - \lambda)(\theta_H - \theta_L)(w_s - \widehat{w})$, which implies $p_0 > p_H = \alpha_H - \theta_H w_s > 0$. If the platform charges p_L, the firms would not invest k and the platform's profit is

$$\Pi^L = (1 - \widehat{\lambda})(\alpha_L - \theta_L w_s - \theta_L w_p).$$

If the platform charges p_H, the L-types receive information rent $(\theta_H - \theta_L)(w_s - \widehat{w})$. Since $k < (\widehat{\lambda} - \lambda)(\theta_H - \theta_L)(w_s - \widehat{w})$, the firms would invest k and always join the platform. Then the platform's profit is

$$\Pi^H = \lambda(\alpha_H - \theta_H w_s - \theta_H w_p) + (1-\lambda)(\alpha_H - \theta_H w_s - \theta_L w_p).$$

If the platform charges p_0, the firms would invest k but the H-types would not join the platform. Then the platform's profit becomes

$$\Pi^0 = (1-\lambda)(\alpha_L - \theta_L w_s - \theta_L w_p) - k(1-\lambda)/(\widehat{\lambda} - \lambda).$$

Note that

$$\Pi^0 - \Pi^H = (1-\lambda)(\theta_H - \theta_L)(w_s - \widehat{w}) - \lambda(\alpha_H - \theta_H w_s - \theta_H w_p) - k(1-\lambda)/(\widehat{\lambda} - \lambda)$$

increases in w_p, while

$$\Pi^0 - \Pi^L = (\widehat{\lambda} - \lambda)(\alpha_L - \theta_L w_s - \theta_L w_p) - k(1-\lambda)/(\widehat{\lambda} - \lambda)$$

decreases in w_p. It can be verified that, when $w_s = w^k$, $\Pi^0 - \Pi^H \geq 0$ if and only if $w_p \geq (\alpha_H - \theta_H w_s)/\theta_H > 0$, and $\Pi^0 - \Pi^L \geq 0$ if $w_p = (\alpha_H - \theta_H w_s)/\theta_H$ and

$$\Big(1 - \frac{\widehat{\lambda} - \lambda}{1-\lambda}\Big)(\alpha_L - \theta_L w_s) \leq \Big(1 - \frac{\theta_L(\widehat{\lambda} - \lambda)}{\theta_H(1-\lambda)}\Big)(\alpha_H - \theta_H w_s),$$

which holds if θ_L is close to 0 and $\widehat{\lambda}$ is close to 1. Moreover, given $w_s \in (w^k, \frac{\alpha_H}{\theta_H})$, if there exists $w_p > 0$ under which $\Pi^0 - \Pi^H \geq 0$ and $\Pi^0 - \Pi^L \geq 0$, then for any $w_s' = w_s + \varepsilon$ with arbitrarily small $\varepsilon > 0$, $\Pi^0 - \Pi^H \geq 0$ and $\Pi^0 - \Pi^L \geq 0$ if platform liability is set at $w_p' = w_p - \varepsilon > 0$. Hence, there exists a unique threshold $\overline{w} \in [w^k, \frac{\alpha_H}{\theta_H}]$ such that, given

$w_s \in (w^k, \overline{w})$, only under a non-empty set of $w_p > 0$, the platform charges p_0 and the first-best outcome is achieved.[127] That is, if $w_s \in (w^k, \overline{w})$, platform liability is socially desired.

If $w_s = \overline{w}$, $\Pi^0 - \Pi^H \geq 0$ and $\Pi^0 - \Pi^L \geq 0$ only under $w_p = 0$, so it is optimal to set $w_p = 0$. If $w_s \in (\overline{w}, \frac{\alpha_H}{\theta_H})$, the platform never charges p_0. Since it is efficient for all the firms to invest k and the profit difference $\Pi^H - \Pi^L$ decreases in w_p, it is optimal to set $w_p = 0$, under which the platform charges p_H and the firms invest k.

Case 2.3: $w_s \in (\widehat{w}, w^k)$. Given $w_s < w^k$, we have $k > (\widehat{\lambda} - \lambda)(\theta_H - \theta_L)(w_s - \widehat{w})$, which implies $p_0 < p_H$. If the platform charges p_L, the firms would not invest k and the platform's profit is

$$\Pi^L = (1 - \widehat{\lambda})(\alpha_L - \theta_L w_s - \theta_L w_p).$$

If the platform charges p_H, the L-types receive information rent $(\theta_H - \theta_L)(w_s - \widehat{w})$. Since $k > (\widehat{\lambda} - \lambda)(\theta_H - \theta_L)(w_s - \widehat{w})$, the firms would not invest k but always join the platform. The platform's profit is

$$\Pi^H = \widehat{\lambda}(\alpha_H - \theta_H w_s - \theta_H w_p) + (1 - \widehat{\lambda})(\alpha_H - \theta_H w_s - \theta_L w_p).$$

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

$$\Pi^0 = \alpha_L - \theta_L w_s - k/(\widehat{\lambda} - \lambda) - [\lambda\theta_H + (1 - \lambda)\theta_L]w_p.$$

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

Summarizing the above analysis, we have

Proposition 7. *(Firm Moral Hazard.) Suppose that firm liability is $w_s \in [0, d]$ and the firms can take effort with costs k. The socially-optimal liability, w_p^k, is as follows:*

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

[127] Note that $\overline{w}$ may equal w^k or $\frac{\alpha_H}{\theta_H}$ under certain parameter values.

B3. False Positives (Type-I Errors)

Now we extend the baseline model by considering false positives. Suppose that the auditing effort of the platform may erroneously 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:

$$S(e) = v + \lambda(1-e)(\alpha_H - \theta_H d) + (1-\lambda)(1-\delta e)(\alpha_L - \theta_L d) - c(e). \tag{55}$$

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

$$-\lambda(\alpha_H - \theta_H d) - \delta(1-\lambda)(\alpha_L - \theta_L d) - c'(\widetilde{e}^{**}) = 0. \tag{56}$$

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

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

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

$$\Pi'(e^f) = S'(e^f) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s) - [\lambda\theta_H + (1-\lambda)\delta\theta_L](d-w) = 0. \tag{57}$$

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 8. *(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.*

B4. Litigation Costs

We extend the baseline model by considering litigation costs. When a user gets harmed by a firm and files a lawsuit, the litigation costs are z_p, z_s, z_b, respectively for the platform, the firm, and the user. Denote $z = z_p + z_s + z_b$. Assume that $z_b \leq w_s + w_p$ and $\alpha_L - \theta_L d - z > 0$.[128] 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

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

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

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

The two types of firms have the same rent when:

$$w_s + z_s = \widehat{w} = \frac{\alpha_H - \alpha_L}{\theta_H - \theta_L}. \tag{58}$$

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

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

The platform's profits can be written as

$$\Pi(e) = S(e) - (1-e)\lambda(\theta_H - \theta_L)(\widehat{w} - w_s - z_s) + [(1-e)\lambda\theta_H + (1-\lambda)\theta_L](d + z_b - w) - v.$$

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

$$\Pi'(e^z) = S'(e^z) + \lambda(\theta_H - \theta_L)(\widehat{w} - w_s - z_s) - \lambda\theta_H(d + z_b - w) = 0. \tag{59}$$

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

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

1. *If $\alpha_H - \theta_H(w + z_p + z_s) \geq r_H^z(w_s)$ then the platform does not audit, $e^z = 0 < \overline{e}^{**}$.*
2. *If $\alpha_H - \theta_H(w + z_p + z_s) < r_H^z(w_s)$ then $e^z > 0$.*
   (a) *If $\theta_H(d + z_b - w) > r_H^z(w_s)$ then $0 < e^z < \overline{e}^{**}$.*

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

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

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

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

$$\lambda(\alpha_H - \theta_H(w + z_s + z_p)) > (1 - \lambda)(\theta_H - \theta_L)(w_s + z_s - \widehat{w}), \tag{60}$$

which is less likely to hold when z_s or z_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* $w_s + z_s > \widehat{w}$. *Let* $r_L^z(w_s) = (\theta_H - \theta_L)(w_s + z_s - \widehat{w})$ *denote the L-type firm's information rents.*

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

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

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

Proposition 9. *(Litigation Costs) The socially-optimal platform liability for harm to users,* w_p^z, *is as follows:*

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

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

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

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

When $w_s + z_s \in (\widehat{w}, \widetilde{w}^z)$, with litigation costs, the platform has stronger incentives to deter the 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, $\underline{w}_p^z < \underline{w}_p$.