THE UNIVERSITY OF CHICAGO

AN ANALYSIS OF PRECAUTION-TAKING BEHAVIOR: VICTIMS, GATEKEEPERS, AND ONLINE COPYRIGHT ENFORCERS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE LAW SCHOOL
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF JURISPRUDENCE

BY
CHUNG-CHIA HUANG

CHICAGO, ILLINOIS
AUGUST 2023

For my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

As a teenager, my mom gave me a little black notebook that recorded in detail how people had helped our family during various occasions, from my parents' wedding to my birth to our elders' funerals. Its purpose was to teach us gratitude and the importance of reciprocating kindness when the time comes. She asked me to keep the notebook and pass it down to the next generation. Back then, I did not fully understand her sentiments and did not take it seriously. However, it all changed when I completed my dissertation and reached this acknowledgments section. I felt the need to record and express my heartfelt thanks to those who had supported me greatly throughout this journey.

First and foremost, I want to express my deep gratitude to my committee members: Professors Omri Ben-Shahar, Lee Anne Fennell, and Daniel Jacob Hemel. I vividly remember the excitement I felt when I received an email from Professor Ben-Shahar during my time at Mount Emei. At that moment, I believed I was ready to abandon my academic dreams and pursue a career as a defense lawyer after receiving rejections of my J.S.D. applications. However, emotions cannot be deceived, and their mentorship proved that this academic journey was worth pursuing. Professor Ben-Shahar always kept me focused on the bigger picture and the potential development of my entire project, without neglecting the importance of structure and readability. Professor Fennell provided gentle yet incisive critiques of my arguments, helping me refine and strengthen them. Professor Hemel consistently posed insightful questions that encouraged me to explore other possible legal frameworks and rules. I am also grateful that Professor Hemel continued advising me even after his move to NYU. They are not only knowledgeable and intelligent mentors but also incredibly supportive, and I feel truly fortunate to have worked with them during my J.S.D. studies.

Beyond my committee, I want to extend my appreciation to several teachers who have played a significant role in shaping this dissertation. Professor William Hubbard has been the de facto advisor for all J.S.D. students at the Law School. He actively participated in

our colloquium and provided thoughtful and critical feedback on our work. I have benefited immensely from his suggestions and comments on every draft. Professors Holger Spamann and Howell E. Jackson at Harvard Law School advised me on my projects on corporate criminal liability and anti-money-laundering laws, which became crucial foundations for the third chapter of this dissertation. It is important to note that acknowledging their support and guidance does not absolve me of any errors present in my work.

I am also thankful for the countless individuals who have made this dissertation possible. Working on an empirical project with Professor Tom Ginsburg was a valuable experience, and I learned a great deal from him. He always made himself available to students and offered unwavering support. Professor Yun-chien Chang opened for me the gate to the world of economic analysis of law in the United States and provided me with opportunities to collaborate with him and other leading scholars on various empirical projects. I would like to extend special thanks to several professors at my *alma mater*, National Taiwan University. Professor Huang-Yu Wang, my master's thesis advisor, wholeheartedly supported my decision to pursue my advanced studies in the United States. Professor Wang-Ruu Tseng introduced various funding opportunities. The encouragement from Professors Yang-Yi Chou, Heng-Da Hsu, Yu-Wei Hsieh, Ying-Hsin Tsai, and Neng-Chun Wang kept me going during challenging times. I would also like to acknowledge the administrative assistance from Associate Dean Justin Swinsick at the Law School, as well as Sunny Lin and Wendy Chen at the Taiwan Economic and Cultural Office in Chicago. Additionally, I am grateful for the financial support provided by the Law School, the Ministry of Education in Taiwan, and the Himalaya Foundation throughout my four-year studies.

Lastly, I am immensely grateful to my family. They have respected my choices and supported me unconditionally. I find it difficult to put into words the full extent of their contributions to my life. They truly are the best.

ABSTRACT

This dissertation investigates the optimality of precautions taken by victims, gatekeepers, and online copyright enforcers. The overarching theme is to identify the inefficiencies of private precaution-taking behavior, along with their underlying causes, and create legal regimes to address such inefficiencies. On one hand, precautions can help deter socially inefficient misconduct, which is already regulated by law enforcement. On the other hand, individuals may not fully internalize the costs associated with their precautions. Therefore, private precautionary decisions can be both socially suboptimal and sensitive to the strategies or focuses of law enforcement. Designing appropriate legal regimes should consider the social benefit of deterrence and the total social cost borne by precaution takers, wrongdoers, and non-perpetrator parties.

This dissertation contains three main chapters. Chapter 2 deals with the most essential case: victims, investigating how to regulate their precautionary measures when inflicting harm on both wrongdoers and non-perpetrator third parties. Chapter 3 extends the analysis and applies the insights from the victim-precaution literature to the context of gatekeeper liability, identifying the externalities of gatekeeping and its sensitivity to law enforcement strategies, and providing guidance on designing appropriate gatekeeper liability regimes. Chapter 4 shifts to a specific context of online copyright enforcement and proposes a regime to hold copyright holders accountable for the over-removal problem.

CHAPTER 1

INTRODUCTION

Precautionary measures can help deter socially inefficient misconduct. However, parties taking precautions may not always act in a socially desirable manner. Specifically, they often neglect the effects of their precautions on other precaution-takers, wrongdoers, or non-wrongdoer parties. Moreover, their precautions are influenced by law enforcement, which is intended to deter the same misconduct, making them sensitive to law enforcement. This dissertation, therefore, investigates how parties' precaution-taking behavior deviates from the social optimum and designs legal tools to address the inefficiencies.

In each chapter, I discuss different precaution-takers in various contexts. In Chapter 2, my focus is on essential precaution-takers, potential victims. Potential victims may sometimes take defensive precautions to mitigate their losses from victimization. However, they might also take aggressive precautions that can be harmful and, therefore, deterrent to wrongdoers. Such harmful precautions can misfire, causing harm to third parties. For instance, a pedestrian might be hurt by barbed wire intended to guard against burglars, or a car-theft alarm triggered by thunder might disturb someone who has no intent to steal the car. These examples illustrate the costs imposed on non-wrongdoer parties by potential victims' precautionary measures. Unfortunately, potential victims often do not consider these costs when legal interventions are absent. As a result, Chapter 2 investigates legal tools to hold victims accountable for the externalities of their harmful precautions.

To analyze this harmful precaution-taking behavior, I apply a framework of bi-dimensional precautions: precaution intensity and accuracy. The former concerns the scale of harm inflicted on both wrongdoers and non-wrongdoer parties, while the latter is about how precaution-takers can identify and exclude non-wrongdoer parties from being subject to their harmful precautions. This bi-dimensional framework serves as the foundation for the analysis in Chapter 2. Based on this framework, I analyze how potential victims behave when

they are subject to tort liability. I find that they are induced to take due care and invest optimally in precaution accuracy to ensure the optimal number of identified non-wrongdoer parties under both strict liability and the negligence rule. However, their chosen precaution intensity could vary with liability regimes due to differences in precaution costs. Moreover, it is ambiguous whether victims would choose more harmful precautions under either regime — it depends on the level of harm they face.

This ambiguity reveals an interesting interplay between the choice of liability regimes and law enforcement intensity. In short, the optimal law enforcement intensity and the optimal choice of liability regimes are interdependent. Law enforcement intensity affects the residual harm faced by victims and affects their precaution intensity, which, in turn, dictates the optimal choice of liability regime. Conversely, the choice of liability regimes determines the cost incurred by precaution-taking victims and affects the scale of the crowd-out effect on deterrence, thus impacting the optimal level of enforcement intensity. This interplay can help explain the current doctrinal puzzle of inconsistent dog-bite rules among states and the distinction between deadly and non-deadly traps. Additionally, I suggest decoupling liability and payment to harmed wrongdoers to fully address this suboptimal precaution-taking behavior. Finally, I comment on academic proposals to regulate precaution-taking behavior via public law enforcement, specifically sanctions on wrongdoers, arguing that such proposals may not work as expected in the context of harmful precautions.

In chapter 3, I compare gatekeepers and precaution-taking victims. Gatekeepers are legally required to prevent misconduct by others, making their goal similar to victims: avoiding misconduct. While they can dissolve their liability by meeting the legally required standard, they may also want to take some precautions to avoid serving wrongdoers that might risk them liability. Drawing from this comparison, insights from the victim-precaution literature can be applied to gatekeeper liability.

To begin with, the wrongdoers that gatekeepers are required to gatekeep are also subject

to law enforcement. Therefore, gatekeeping and law enforcement can be both substitutes and complements. In aggregate, more rigorous law enforcement deters more wrongdoers with fewer wrongdoers being left, so gatekeepers are less worried about liability. In this sense, law enforcement and gatekeeping are substitutes. However, when we dissect law enforcement into sanction and detection, and gatekeeping into interdicting and reporting, we find that when law enforcement increases the imposed sanction and reduces detection accordingly, interdicting gatekeepers are discouraged because the liability risk is lower due to fewer identified wrongdoers, but the increased sanction does not make interdicting more deterrent. In contrast, the effect on reporting gatekeepers is uncertain because their reporting becomes more deterrent, but at the same time, they are facing a lower liability risk. The interaction between law enforcement and gatekeeping implies that the choice of required gatekeeping act matters.

In addition, gatekeeping as precaution-taking can impose costs on others, including other gatekeepers, counterparties, and third parties. For instance, gatekeepers might want to increase their level of gatekeeping to scare wrongdoers away and avoid liability. However, doing so does not necessarily deter those wrongdoers but only diverts them to other gatekeepers, resulting in social waste from over-gatekeeping. Likewise, gatekeepers may impose costs on their clients. For example, big banks may refuse to serve some money-laundering suspects, who then seek financial services from smaller money-service providers. While the gatekeeper and the gatekept client are in a contractual relationship and are supposed to be able to bargain with each other, bargaining can be hindered by information asymmetry or imperfect market competition. As a result, when the gatekeeper does not fully internalize the cost imposed on the gatekept counterparty by gatekeeping, it is likely to over-gatekeep. For instance, online platforms often comply with takedown requests filed by copyright holders to avoid liability because they do not fully internalize the benefit of the content. Finally, gatekeeping itself can incur costs borne by third parties. For example, employers may reject

ex-offenders to avoid liability, but doing so incurs costs of recidivism imposed on communities. Likewise, banks rejecting undocumented customers may leave them underbanked and contribute to lower financial inclusion, which has been shown to cause poverty.

With this analysis, this chapter then provides more general but nuanced guidelines for designing gatekeeper liability, including the required gatekeeping act, the choice of liability regimes, and the amount of penalties. Based on these guidelines, I then provide policy recommendations for revising gatekeeper liability to address three social problems: boosting ex-offenders' employment to mitigate recidivism, increasing financial inclusion to combat poverty, and revisiting platform liability to prevent data pollution.

Chapter 4 focuses on the specific problem in the context of online copyright enforcement. Section 512 of the Digital Millennium Copyright Act stipulates a notice-and-takedown regime that allows rightsholders to request platforms to remove alleged infringing content. However, such a regime is often claimed to be abused by rightsholders, resulting in the problem of over-removal. This problem intertwines the victim (copyright holders) and gatekeepers (platforms). Several solutions have been proposed to address the problem, such as holding platforms liable, encouraging uploaders' responses, and holding rightsholders accountable. However, such proposals often focus too much on specific parties in the notice-and-takedown process and are thus unable to fully address the problem.

This chapter argues that taking down alleged infringing content can be understood as a form of harmful precaution, which can impose costs on non-infringing uploaders when their content is erroneously removed. Hence, the framework of harmful precaution — precaution intensity and accuracy — can be applied to this context. In short, removing alleged infringing content is socially desirable only if the investment in verification is optimal to exclude non-infringing content from the takedown requests (optimal accuracy) and if the expected benefit from stopping infringement outweighs the cost of error (precaution intensity). The prescription in Chapter 2 can mostly be applied. To begin with, rightsholders should be

held strictly liable to induce their investment in verification and exclude non-infringing content from the takedown requests. Also, being held strictly liable prevents them from filing suboptimal takedown requests that yield less deterrent benefit than the error costs.

Nevertheless, unlike the context in Chapter 2, where precaution-takers can themselves determine their precaution intensity, rightsholders cannot choose the level of precaution intensity as it depends on the value of the requested content, which may be unknown to rightsholders. As a result, rightsholders may make suboptimal decisions under imperfect information. To improve their decisions, I propose a committed maximum liability regime, where rightsholders can specify their committed maximum liability and have the platforms, which have better information about the content values, decide for them. Namely, platforms will only remove the requested content if its value is below the specified CML. Such a regime allows collaboration between rightsholders and platforms to share the information they have and can address the informational problem that previous proposals focusing on only a single player underappreciate.

CHAPTER 2

HARMFUL PRECAUTION

2.1 Introduction

In 2021, a person invited his neighbor for a tour of his home renovations. During the tour, he showed off the Ring doorbell installed on a shed in a driveway. The Ring doorbell not only captured images of the neighbor's house but collected audio data. The neighbor felt this to be an invasion of her privacy and filed a lawsuit. Despite the owner's claimed purpose to deter burglars, the court ruled that the doorbell violated U.K. data laws (Wakefield 2021). This story illustrates how a precautionary measure, while serving good-faith deterrent purposes, can harm non-perpetrator third parties, and introduces the question of regulating harmful precautions.

Harmful precautions, as in the story above, are not unusual, but omnipresent. Unlike "defensive" precautions, which serve to mitigate potential harm, such as wearing a bullet-proof vest, harmful precautions are aggressive and are intended to impose costs on perpetrators. Such costs, however, can sometimes fall on non-perpetrator third parties. In daily life, many of us have returned home from clothes shopping only to find unremoved security ink tags that require additional time to return to the store for their safe removal, or risk staining the clothes in an attempt to remove them ourselves. Collectively, we spend five hundred years, per day, deciphering reCaptcha texts to prove that we are not computer bots intent on launching cyberattacks on websites (Ahn et al. 2008; Meunier 2021). Certainly, we have all, at some time, been disturbed or startled by car anti-theft alarms mistakenly triggered by pedestrians or thunder. In the U.S., more than 5,400 postal employees were attacked by dogs in 2021 (United States Postal Service 2022-06-02). Indeed, in extreme cases, trespassers without criminal intent can be severely injured or killed by home-protection traps such as wires or spring guns.

Of course, such precautionary measures are meant to protect those implementing them from the misconduct of malicious perpetrators. For instance, ink security tags on clothes and anti-theft car alarms deter thieves. reCaptcha technology helps identify and block bot-attacks. Surveillance cameras, guard dogs, and traps help to protect property owners from burglars and intruders. While such precautionary measures are meant to target and impose costs on perpetrators, thus deterring misconduct, they are not neutral with respect to non-perpetrator third parties, as shown above. The costs falling on non-perpetrator third parties, even if small in each incidence, should not be ignored as such costs can be tremendous in aggregate given the comparatively large number of affected third parties. This chapter addresses the social cost related to "harmful precautions" and suggests how legal mechanisms might address the issue.

Precautionary measures, whether or not potentially harmful, might well deviate from the social optimum if precaution-takers do not internalize how their precautions affect others. The literature has already identified causes of socially suboptimal precautionary decisions. First, precaution-takers may over- or under-invest in precautions due to the externalities their measures impose on other precaution-takers. Clotfelter (1978) points out that when a precautionary measure is observable by potential perpetrators, it is easy to identify and target those who do not take similar measures. In this way, people who take observable precautions can divert perpetrators and avoid victimization. Such an individually rational decision incurs some cost but yields no social benefit, as misconduct is transferred rather than eliminated. Conversely, unobservable precautions deprive perpetrators of the opportunity to preemptively distinguish between those who do not take precautions and those who do and force them to rely on an overall probability of precaution-taking. Under these circumstances, the level of misconduct is lower for all precaution-takers, resulting in a social benefit of deterrence at individual precaution-takers' cost. Clotfelter's analysis is partially supported by empirical evidence. Ayres and Levitt (1998) find a harm-reduction effect from

unobservable Lojacks installed in automobiles that help track car thieves. Still, Koo and Png (1994) find the harm-displacement effect between London underground stations with and without surveillance cameras to be insignificant. They argue that for harm to be displaced by observable precautions requires the transferability of perpetrators' knowledge and a degree of similarity in the victims' profiles. Moreover, when the value at risk is private information, precaution-takers may not want to invest in observable precautions that signal the value. In this case, over-precaution due to harm-displacement effect may be replaced by under-precaution to avoid attracting perpetrators (Baumann and Friehe 2013). The harm-displacement effect can also occur when tortfeasors are subject to the negligence rule or enjoy the defense of contributory negligence; both induce the victim to take due care that might displace the harm (Givati and Kaplan 2020).

The first cause identified by Clotfelter illustrates a collective-action problem, where precaution-takers fail to collaborate on an efficient collective action. Shavell (1991) identifies another cause of suboptimal precaution-taking, even when precaution-takers successfully collaborate as a group. He indicates that they would still take excessive precautions if the harm suffered by an individual precaution-taker when victimized is higher than the actual social harm. Since precautions are taken to prevent the infliction of harm, higher levels of harm will lead to stronger incentives to take precautions. As the harm precaution-takers would suffer is not offset by the perpetrator's gain, precaution-takers will take more precautions than society would desire.

This chapter extends the analysis of precaution-taking behavior by identifying how an additional affected party contributes to suboptimal precaution-taking. The impact of precautions is not limited to those precaution-takers and perpetrators mentioned above, but also imposes costs on non-perpetrator third parties.[1]

1. Dam (1999) observes that precautionary measures could sometimes benefit counterparties by reducing transaction costs. This idea also touches upon the effect of precautions on parties other than precaution-takers and perpetrators.

When precaution-takers do not internalize costs, they are incentivized to take socially excessive precautions. This chapter aims to investigate what legal tools might address the social problem of over-precaution.

This chapter develops a formal model to analyze the problem. I consider precaution-taking behavior in two dimensions instead of one, as in the literature. That is, a precaution-taker chooses not only the level of precaution *intensity* but also the level of precaution *accuracy*. The former determines how harmful a precautionary measure is, while the latter denotes how accurately a precaution-taker can identify non-perpetrator third parties and prevent them from the harm inflicted by precautions. To see these two dimensions more clearly, consider cases of anti-theft car alarms and guard dogs. Car owners can choose the loudness (intensity) and the accuracy of the installed alarm. A less loud alarm would impose lower costs on neighbors (third parties) and perpetrators at some cost to the effectiveness of the deterrence; a more accurate alarm would be less likely to be triggered by non-perpetrators. Considering guard dogs, owners can choose between a mild-tempered Corgi that barks and a fierce Tibetan mastiff that bites; they can decide how much they will invest in training their dogs to distinguish between good-faith mail carriers and burglars.

As mentioned above, precaution-takers are incentivized to take socially excessive measures as they do not internalize the costs borne by third parties. Suboptimal precautionary decisions can be improved by either lowering the *intensity* of precautions or enhancing their *accuracy*. To align private precaution-taking with what is socially optimal, I consider two tort liability regimes — strict liability and the negligence rule — that hold precaution-takers liable for third-parties' losses. When held strictly liable, precaution-takers internalize all the costs of third parties. In comparison, precaution-takers subject to the negligence rule are only liable for the third parties' cost when failing to take due care. In this context, due care is defined as optimal accuracy in identifying non-perpetrator third parties. When a precaution-taker invests optimally in precaution accuracy, giving due consideration to pre-

caution intensity, she is exempted from liability.

As indicated in the literature, both regimes induce optimal care (precaution accuracy). However, they do not necessarily induce the same level of precaution intensity. The difference in precaution intensity is due to the differing marginal costs of precautions incurred under each of the regimes. To see this, consider closely the cost-benefit structure of precaution-taking. If held strictly liable, a precaution-taker who wants to enhance precaution intensity will incur the marginal cost of precaution intensity (C1), the marginal cost of required accuracy (C2), and the marginal liability due to the greater harm imposed on victimized non-perpetrator third parties by a slightly more intensive precaution (C3). In return, she gets the marginal deterrence benefit (B1) and the marginal benefit of reduced liability due to the higher required accuracy of more intensive precautions (B2). In comparison, those subject to the negligence rule incur only C1 and C2 and get only B1. Therefore, under the negligence rule a precaution-taker would choose a level of precaution intensity such that C1+C2=B1. A counterpart under strict liability would choose the point at which C1+C2+C3=B1+B2. By definition, C2=B2, so the chosen point lies where C1+C3=B1.

$$\begin{aligned} \text{C1+C2+C3} &= \text{B1+B2} && \text{strict liability} \\ \text{C1+C3} &= \text{B1} && \text{strict liability} \\ \text{C1+C2} &= \text{B1} && \text{negligence rule} \end{aligned}$$

If the marginal liability from increased precaution intensity (C3) does not coincide with the marginal cost of precaution accuracy (C2), then precautionary intensity would diverge under both regimes. This finding shares a similar flavor with the activity-level argument, but not necessarily in the same direction. That is, it is possible that the marginal cost of precaution accuracy will outweigh the marginal liability so that precaution-takers will take less, rather than more, intensive precautions under the negligence rule than under strict liability.

Remember, the precautionary decision can also be sensitive to law enforcement against

perpetrators, as observed in the literature (Clotfelter 1977). Therefore, law enforcement and liability affect each other through pivotal precaution-taking behavior. Assume a negative-slope distribution of perpetrators' gain from the misconduct and, therefore, a diminished deterrence effect. Further assume that law enforcement and precaution are substitutes. As enforcement becomes more severe (i.e., higher deterrence), fewer perpetrators could be further deterred by a marginal increase in precaution intensity, resulting in a lower marginal deterrence benefit of precautions for precaution-takers. With this assumption, it could reasonably expect that stronger enforcement against perpetrators leads to a lower deterrence benefit by precaution (B1). In brief, enforcement crowds out precaution. However, the scale of such a crowd-out effect depends on the liability regime. To begin with a unit increase in enforcement severity, precaution-takers are now faced with a lower marginal deterrence benefit from precaution (B1), which inevitably translates into a decrease in the marginal precaution cost. Since the marginal precaution costs do not necessarily coincide under both regimes, the decrease in precaution intensity could differ. As a result, we can expect the scale of the crowd-out effect of enforcement on precaution intensity to vary with liability regimes. With different levels of the crowd-out effect and the same function of enforcement cost, this chapter shows that the severity of the chosen enforcement mechanism may be related to the choice of liability regimes regulating precaution-taking behavior, potentially revealing a relationship between enforcement severity and liability regime.

Second, the desirability of liability regimes can be indeterminate and hinge on enforcement severity. As shown above, marginal precaution costs differ under different regimes. In some cases, the order of the marginal precaution costs under both regimes may flip with precaution intensity, which is potentially affected by enforcement severity. To see this, compare the marginal costs under two regimes: the marginal liability from increased intensity (C3) and the marginal cost of accuracy (C2). C3 should decrease as precautions become more intensive because the marginal liability from increased intensity is lower when more third

parties are identified due to the higher accuracy required by more intensive precaution. However, the trend of C2 depends on the function of the precaution accuracy cost. If accuracy comes at a lower cost, then the required improvement should be more significant. Likewise, expensive accuracy requires a lower level of improvement. Ultimately, it is uncertain whether the level of required improvement times the marginal cost is higher or lower when precaution becomes more intensive. If C2 increases as precautions grow more intensive, then C2 and C3 may intersect with each other at some point. The order relationship between C2 and C3, or broadly, the marginal costs of precaution under both regimes, would flip before and after the intersection. This flip implies that a precaution-taker may take more intensive precaution under strict liability than under the negligence rule when enforcement against perpetrators is strong, but otherwise chooses to take more intensive precaution under the negligence rule when enforcement is modest. When neither regime is optimal, such a flip also implies the desirability of a liability regime depends on the enforcement severity. This second finding again shows the relationship between enforcement and liability.

These findings allow this chapter to provide doctrinal, policy, and theoretical implications. Doctrinally, this chapter revisits the divergence of dog bite rules among states and the distinction between privileged (precaution-takers are not liable) and unprivileged (precaution-takers are liable) traps used to protect property. I argue that the reluctance to apply the negligence rule to regulate such precautions can be explained by the uncertainty of desirableness. Also, the tolerance toward a dog's first bite in some states and less harmful property-protection traps in common law could be rationalized by the tradeoff between aggregate deterrence and the error cost borne by non-perpetrator third parties. Therefore, I predict jurisdictions where people take less intensive precaution due to lower crime rates and stronger enforcement severity are more likely to apply the one-bite rule. However, the explainability of doctrines does not mean it is desirable or efficient. It can still be improved. In policy terms, I argue for decoupling the liability on precaution-takers' and their transfers

to perpetrators in cases of unprivileged traps. A decoupled liability for perpetrators' loss can maintain deterrence from private precautions while helping precaution-takers to internalize the cost. Lastly, this chapter engages in the scholarly proposals of regulating precaution-taking behavior in criminal enforcement. I argue that when private precaution is harmful, inducing optimal precaution by lowering punishment can backfire. A lower punishment may instead encourage potential victims to take precaution in defense of themselves when public enforcement retreats. In addition, whether lower sanctions can encourage perpetrators to target precaution-takers is a misery as they might prefer fines or humane imprisonment imposed by the states than bodily injury caused by harmful precautions, such as electrified fences or spring guns. Therefore, such proposals should be, at least, modified when precautions are harmful.

This chapter consists of six sections. This introductory section overviews the problem, the tool, and the main findings and is followed by the model in Section 2.2. In Section 2.3, I discuss how precaution-taking behavior changes when liability and enforcement are introduced. Then I enumerate three possible function forms to concretize the findings in Section 2.4. Based on the findings, I turn to the discussion in Section 2.5, where I discuss the qualifications and implications of the findings to doctrines, policies, and theories. Section 2.6 concludes this chapter.

2.2 The Model

To illustrate how precaution-takers' behavior deviates from the social optimum, I model their precautionary decisions, including precaution intensity and precaution accuracy.

2.2.1 Setup

The model contains three parties: the precaution-taker (PT, she), the perpetrator (P), and the non-perpetrator third party (NP, he). All three parties are risk-neutral. The only

decision-maker modeled here is the precaution-taker. Also, while grammatically singular, they could be understood as a "group" with multiple members.

To begin, the perpetrator gains private benefit from committing misconduct against the precaution-taker. P's benefit, $b \in [0, \bar{b}]$, conforms a probability distribution, $b \sim G(\cdot)$, where $G(0) = 0$, $G(\bar{b}) = 1$, $G'(b) = g(b) > 0$ and $G''(b) = g'(b) = m < 0 \quad \forall b \in [0, \bar{b}]$, and $G(z) = 1 \; \forall z \geq \bar{b}$. This distribution of P's benefit from the misconduct presumes that it is more likely that P enjoys a lower benefit than a higher benefit from the misconduct. (Or, there are more low-benefit Ps than high-benefit ones.) P will be subject to law enforcement and punishment. Also, P will face the harm imposed by precautions taken by PT. Both the sanction imposed by law enforcement and the harm imposed by private precautions constitute the cost of committing the misconduct. Assuming the expected sanction is s, and the precaution taken by PT imposes x harm on P, then P's cost of misconduct equals $s + x$. In that case, there is a $G(x + s)$ chance that P is deterred. Yet there is a $1 - G(x + s)$ chance that P is not deterred. When deterred, P would abandon the plan and would not be imposed any cost from the misconduct, including the expected sanction and the harm from precaution. Otherwise, undeterred P will suffer both the expected sanction and the precautionary harm.

In cases where P is not deterred, P will inflict h harm on PT. I assume that the misconduct is never efficient as the maximum benefit derived from the misconduct is always lower than the harm P causes to PT, i.e., $\bar{b} < h$. To avoid the harm imposed on her, PT will want to deter P by imposing cost on P. To do so, she needs to spend κx.

Such precautions can also harm NP, who suffers τx harm. Note that τ could be higher or lower than 1, meaning that NP could suffer greater or less harm than P does for two reasons. First, the realized harm could be different. For instance, P may be more prepared, more accustomed, or even more affected by the precautionary measure and therefore suffer less or more harm. For instance, burglars (P) may wear thicker gloves to defend against

barbed wire than a walking neighbor (NP) does. A thief, however, could be more harmed by a security ink tag that they cannot easily remove. Another source of disparity of harm comes from the relative proportion of P and NP. In reality, the size of P and NP may vary with geographic areas or crime types. When NPs outnumber Ps, NP, as a group, would, on average, suffer greater harm than P does. For simplicity, I assume no transfer of the error cost between NP and PT and no contributory defense available for PT. That is, absent legal intervention, all τx is borne by NP, and NP doesn't and cannot make any efforts to avoid τx. In reality, however, NP can to some extent avoid or transfer the cost to PT, depending on the relationship between PT and NP. For instance, an unremoved ink tag may cost both the customer (NP) and the shop owner (PT) because the cost incurred for removing tags can be perceived as a price by customers and thus be partially transferred to the shop owner in the form of decreased sales. NP can also prevent the harm by checking the purchased clothes before exiting the shop. Therefore, the model is more suitable for scenarios with high transaction costs.

Such a cost on NP is, however, avoidable when PT invests in precaution accuracy. Precaution accuracy is negatively defined as the error-rate, $e(y)$, when PT invests $y \geq 0$, where $e(0) = 1$, $e'(y) < 0$, and $e''(y) > 0$. Also, $0 < e(y) \leq 1 \quad \forall y \in \mathbb{R}^+$. With this assumption, when PT invests nothing in precaution accuracy, NP will certainly suffer harm ($e(0) = 1$). As NP invests more, the error rate decreases, and NP will be more likely to be identified and thus less likely to be harmed by such precautions. Note that for parsimony, PT never makes mistakes in targeting P in the model. That is, the precaution would certainly impose harm on P. While this assumption disallows the model to dynamically investigate the relationship between the type-II error and precautionary decisions, it can be statically incorporated into the precaution intensity. Namely, a precautionary measure with non-zero type-II error makes the expected harm from precautions lower and thus the precaution pricier (greater κ).

Finally, the state can determine the expected sanction, s, on P at some cost of $c(s)$.

Two assumptions should be stated here. First, I assume the sanction is monetary without incurring any cost other than the administrative cost. Second, the state's decision is not modeled. Instead, I focus on the decisions made by PT, who perceives the state's enforcement severity as given. Therefore, I will omit the enforcement cost $c(s)$ from the model and treat enforcement severity as given until Section 2.3.2.

Notations are summarized below for readers' convenience.

$$
\begin{aligned}
h &= \text{the harm imposed by P on PT.} \\
G(\cdot) &= \text{the cumulative distribution of P's gain from misconduct.} \\
x &= \text{P's harm inflicted by PT's precaution.} \\
\kappa x &= \text{PT's investment in precaution intensity.} \\
\tau x &= \text{NP's harm inflicted by PT's precaution when he is unidentified.} \\
y &= \text{PT's investment in precaution accuracy.} \\
e(y) &= \text{the error rate: the likelihood that NP is subject to precautions.} \\
s &= \text{law enforcement severity: expected sanctions on P.}
\end{aligned}
$$

2.2.2 The first-best outcome

In this section, I first investigate the first-best decision taken by PT as a benchmark. Note that since PT cannot alter the enforcement severity s, she will take it as given and make her precautionary decision accordingly. Hence, the first-best precautionary decision should be a function of s. The first-best precautionary decision, including precaution intensity (x^*) and

precaution accuracy (y^*), maximizes the social welfare function below:

$$\begin{aligned}\max_{x,y} W = &- \kappa x - y - [1 - G(x+s)]\, h \\ &+ \int_{x+s}^{\bar{b}} (b-x) g(b) db \\ &- e(y)\tau x\end{aligned} \tag{2.1}$$

The right-hand side of Equation 2.1 contains three lines. The first line denotes PT's payoff, which includes the total precaution cost, incurred for both intensity and accuracy, plus the expected harm inflicted by P when P is undeterred. The second line represents the net social benefit generated by P, who gains some utility from misconduct but suffers costs imposed by PT's precaution and law enforcement. Since the sanction is monetary and merely transfers the wealth from P to the state without incurring additional social cost, it itself is not included in the social welfare function. Note that, in the model setup, the gain from misconduct, while in dispute,[2] is recognized and included in the social welfare calculus.

The third line is NP's payoff. NP would suffer τx when PT fails to identify him, which occurs with probability $e(y)$. Again, I focus on scenarios where NP cannot (easily) bargain with and transfer the cost to PT.

To learn the first-best outcome, I begin with calculating the optimal expenditure in accuracy y^*, which solves the equation that the first-order condition of W with respect to y is zero. That is

$$-1 - e'(y^*)\tau x = 0 \tag{2.2}$$

From Equation 2.2, we learn that the optimal spending in accuracy lies where the marginal cost (1) equals the marginal benefit ($-e'(y)\tau x$), which is the reduced cost borne by NP due to higher precaution accuracy that identifies NP. By checking the second-order

2. Becker (1968) takes perpetrators' gain from misconduct into social welfare calculation. Stigler (1970) however, disagrees with this idea. For discussions, see Lewin and Trumbull (1990) and Curry and Doyle (2016).

condition, which is $-e''(y^*) < 0$, we ensure that the solver (y^*) is the welfare maximizer.

It should be noted that when PT spends more on precaution intensity, i.e., higher x, she should also invest more in accuracy. The intuition is simple: improving accuracy becomes more valuable when the stake, the error cost imposed on NP, is higher due to higher precaution intensity. Formally speaking, when x increases, the marginal error rate, $e'(y)$ should be lower. Owing to negative $e'(y)$, a lower $e'(y)$ implies a higher y. As a result, a higher x leads to a higher y. Accordingly, we can rewrite the first-best level of y^* as a function of x, i.e., $y^* = y^*(x)$, where $\frac{d}{dx}y^*(x) > 0$. This intuition can be formally shown by applying the implicit function theorem to Equation 2.2:

$$\frac{dy^*}{dx} = \frac{e'(y^*)\tau}{-e''(y^*)\tau x} > 0 \tag{2.3}$$

Therefore, we can ensure a positive relationship between the investment in precaution intensity, x, and the correspondent optimal expenditure in precaution accuracy, $y^*(x)$.

Likewise, we can also calculate the optimal level of precaution intensity, x^*. I take the first-order condition of W with respect to x and get

$$\begin{aligned} -\kappa - \frac{dy^*(x^*)}{dx} + g(x^*+s)(h-s) - [1-G(x^*+s)] - \frac{dy^*(x^*)}{dx}e'(y^*)\tau x - e(y^*(x^*))\tau = 0 \\ -\kappa + g(x^*+s)(h-s) - [1-G(x^*+s)] - e(y^*(x^*))\tau = 0 \end{aligned} \tag{2.4}$$

Equation 2.4 showcases that the optimal precaution intensity x^* solves the equation in which the marginal cost of precaution intensity (κ) is equal to the net marginal benefit, which consists of (1) net marginal deterrence benefit by precautions $(g(x+s)(h-s))$, (2) marginal cost borne by P $(-[1-G(x+s)])$, and (3) the marginal error cost borne by NP due to the increase in precaution intensity $(-e(y^*)\tau))$. The marginal cost of precaution accuracy is, by definition, canceled out by the marginal benefit from the reduction of the error cost borne by NP. To verify x^* is the welfare maximizer, we then check the second-order condition,

which is $g'(x+s)(h-s)+g(x+s)+e'(y^*)\frac{d}{dx}y^*(x)\tau$. The second-order condition is strictly decreasing but can be positive or negative. Hence, we can at most find one x^* such that the first-order condition is zero and the second-order condition is negative. If we cannot find such one, then the first-best precaution intensity is zero, implying no precautions should be taken. In this chapter, I focus on the case of positive optimal precaution intensity, $x^* > 0$.

Note that P is marginally indifferent between committing or abandoning the misconduct. Hence, P's reduced gain when P is deterred and therefore abandons misconduct is offset by the reduced cost P bears if undeterred. Therefore, the net gain from misconduct remains s.

2.2.3 Precaution-taker's decision

Unlike the social welfare function, which aggregates social cost and benefit from all parties, PT is only concerned with her payoff, which is the first line of the social welfare function in Equation 2.1. Hence, PT does not internalize the effect of her precaution on both P and NP. Instead of maximizing social welfare, she maximizes her payoff $-\kappa x - y - [1 - G(x+s)]\,h$.

PT, in this model, does not invest in precaution accuracy because she enjoys nothing therefrom. As a result, her investment in precaution accuracy, denoted by $y^{**}(x)$, is zero, regardless of her chosen precaution intensity. Accordingly, $e(y^{**}(x)) = e(0) = 1$. Therefore, absent legal intervention, her chosen level of investment in precaution intensity (x^{**}) solves the following equation.

$$-\kappa + g(x^{**}+s)h = 0 \tag{2.5}$$

Namely, her investment in precaution intensity lies where the marginal private deterrence benefit ($g(x+s)h$) equals the marginal cost (κ). To recap, P's precautionary decision is $(x^{**}, 0)$. The second-order condition, $g'(x^{**}+h) < 0$, ensures that x^{**} maximizies PT's payoff.

Proposition 2.1 (Social problem)**.** *Without legal intervention, a rational, risk-neutral precaution-taker's precaution intensity and precaution accuracy fail to meet their first-best levels.*

Proof. Let's plug P's decision $(x^{**}, 0)$ in Equation 2.4. We then get $-\kappa + g(x^{**} + s)(h - s) - [1 - G(x^{**} + s)] - e(0)\tau = 0 - g(x^{**} + s)s - [1 - G(x^{**} + s)] - \tau < 0$. The negative sign indicates the marginal social cost outweighs the marginal social benefit, implying over-precaution. □

Proposition 2.1 shows a social problem: private precaution-taking behavior without legal intervention is socially suboptimal. The negative sign indicates that PT invests socially excessively in precaution intensity. Such inefficient precaution-taking behavior results from three sources. First, PT overestimates the harm brought by P because PT does not internalize the gain from misconduct enjoyed by P. This result has been observed by Shavell (1991, p. 131). Second, PT does not internalize the cost imposed by her precaution on P. Lastly, PT does not internalize the cost borne by NP.

Such a socially suboptimal precautionary decision can be improved from two dimensions. First, it is required to discourage PT from taking over-precaution. That is, we need to reduce PT's precaution intensity from x^{**} to x^*. It is not sufficient to achieve the first-best outcome, however. In addition to precaution intensity, we also need to induce PT to invest optimally in precaution accuracy. As a result, the social problem illustrated in Proposition 2.1 requires legal interventions in both dimensions.

2.3 Legal Interventions

In the last section, I identify a social problem, which illustrates that P's precautionary decision is not socially optimal. Moreover, both dimensions of precautions require legal interventions to achieve their first-best levels. I first consider whether imposing liability on PT can solve the problem. Two liability regimes — strict liability and the negligence rule — are discussed. I find that while both liability regimes guarantee optimal precaution accuracy, they do not ensure that PT's precaution intensity is optimal. Then I turn to public law enforcement, which directly targets and deters P and thus indirectly affects PT's

precaution intensity. The findings are interesting: The desirability of liability regimes may vary with different enforcement severity. Also, optimal enforcement severity may hinge on the choice of liability regimes imposed on PT.

2.3.1 Liability on precaution-takers

Holding precaution-takers liable for the loss they inflict on non-perpetrator third parties helps them internalize the cost and induces their investment in precaution accuracy. To achieve this end, two liability regimes can be devised: strict liability and the negligence rule. The former asks PT to compensate NP whenever NP is unidentified and harmed, but the latter only asks for compensation when PT fails to exercise due care. Here, I define due care as optimal precaution accuracy. While both regimes expect to align PT's level of care with the first-best one (Shavell 1987c), they might lead PT to take different levels of precaution intensity. Below, I first investigate how PT behaves under strict liability and then turn to the negligence rule.

Strict liability

When held strictly liable for NP's losses, PT bears the total error cost incurred by NP. Since I do not model NP's contributory negligence, strict liability is equivalent to absolute liability as both regimes require PT to compensate for the harm inflicted by her precautions. Without considering NP's contributory negligence, PT faces the same payoff function and reaches the same decision, regardless of the choice of strict liability or absolute liability. Therefore, the analysis below would apply to a regime of absolute liability as well. Under such regimes, PT's payoff becomes:

$$-\kappa x - y(x) - [1 - G(x+s)]\, h - e(y(x))\tau x \tag{2.6}$$

In this case, her investment in precaution accuracy y_S solves

$$-1 - e'(y_S(x))\tau x = 0 \tag{2.7}$$

The equivalence of Equations 2.7 and 2.2, with monotonically decreasing $e'(y)$, implies that $y_S(x) = y^*(x)$ for all x. Therefore, strict liability ensures that PT invests optimally in precaution accuracy. Based on this result, we can further calculate PT's chosen level of precaution intensity by replacing $y(x)$ with $y^*(x)$. To abuse the notation a bit, I will replace $e(y^*(x))$ with $e^*(x)$ afterward.

PT's investment in precaution intensity under strict liability would be x_S, which solves the equation in which the first-order condition of her payoff with respect to x equals zero.

$$-\kappa + g(x_S + s)h - e^*(x_S)\tau = 0 \tag{2.8}$$

The precaution intensity taken by PT under strict liability lies where the marginal cost of precaution intensity plus the marginal liability equals to the marginal private deterrence benefit. Still, PT does not internalize P's gain from misconduct and cost inflicted by PT's precaution. As a result, PT will still take over-precaution under strict liability.

Proposition 2.2. *When a precaution-taker is held strictly liable for the non-perpetrator third party's loss from erroneous precautions, her precaution accuracy is optimal with regard to her precaution intensity, but her precaution intensity is socially excessive.*

Proof. To see whether PT invests optimally (with regard to the precaution intensity) in precaution accuracy, compare Equations 2.7 and 2.2. Let's again plug x_S into Equation 2.4. We get $-\kappa+g(x_S+s)(h-s)-[1-G(x_S+s)]-e^*(x_S)\tau = -g(x_S+s)s-[1-G(x_S+s)] < 0$. The negative result indicates the marginal social cost outweighs the marginal social benefit, implying over-precaution. □

Negligence rule

In contrast with strict liability, the negligence rule allows PT to claim that she has already taken due care in defense of NP's claim. In the model, due care is defined as optimal precaution accuracy, y^*. Under such a regime, PT needs not compensate for the harm done by her precautions if she invests optimally in precaution accuracy, which is set according to precaution intensity. As a result, PT would be treated as taking due care and refrained from compensation when her investment in precaution accuracy meets the first-best level based on her chosen precaution intensity. I assume the court can observe the harmfulness of precaution (i.e., precaution intensity) and commits no error in setting due care level. Hence, PT will comply with the due-care level and invest optimally in precaution accuracy, i.e., $y^*(x)$, to avoid liability. Accordingly, PT's payoff becomes

$$-\kappa x - y^*(x) - [1 - G(x+s)]\, h \tag{2.9}$$

The only decision for PT to make concerns the level of precaution intensity because the level of due care is optimally set according to precaution intensity. To learn the rational level of precaution intensity under the negligence rule, we again take the first-order derivative of PT's payoff with respect to x and solve the equation in which the derivative equals zero. We get

$$-\kappa - \frac{dy^*(x_N)}{dx} + g(x_N + s)h = 0 \tag{2.10}$$

From Equation 2.10, we learn that PT's precautionary decision under the negligence rule is different from the one under strict liability or no liability. In fact, PT can take over- or under-precaution under the negligence rule. In some cases, PT can take optimal precautions even if she does not internalize P's cost and gain associated with misconduct.

Lemma 2.1. $\frac{d}{dx}y^*(x) = -\tau x \frac{d}{dx}\left[e^*(x)\right]$

Proof. $\frac{d}{dx}\left[e^*(x)\right] = \frac{d}{dx}\left[e(y^*(x))\right] = e'(y^*(x))\frac{d}{dx}\left[y^*(x)\right] = \frac{-1}{\tau x}\frac{d}{dx}\left[y^*(x)\right] \implies \frac{d}{dx}\left[y^*(x)\right] =$

$-\tau x \frac{d}{dx}\left[e^*(x)\right]$ □

Lemma 2.1 can also be explained economically: the marginal cost of optimal precaution accuracy should equal to its marginal benefit from the reduction of loss.

Proposition 2.3. *Under the negligence rule, PT invests optimally in precaution accuracy. However, PT's precaution intensity, in comparison with strict liability and the first-best level, can be lower, equal, or higher, depending on* $\frac{d}{dx}\left[e^*(x_N)x_N\right]$. *In concrete, when* $\frac{d}{dx}\left[e^*(x_N)x_N\right] = 0$, *PT's precaution intensity would be the same under both strict liability and the negligence rule. If* $\frac{d}{dx}\left[e^*(x_N)x_N\right] > 0$, *the level of precaution intensity chosen by PT under the negligence rule would be higher than under strict liability, implying more significant over-precaution. If, otherwise,* $\frac{d}{dx}\left[e^*(x_N)x_N\right] < 0$, *PT's precaution intensity under the negligence rule is lower than under strict liability, implying the negligence rule is superior to strict liability. In a subset of such cases, it can also be equal to or lower than the first-best outcome.*

Proof. By definition, PT takes optimal precaution accuracy. For precaution intensity, I plug PT's decision under the negligence rule, x_N, into Equation 2.4 and get

$$\begin{aligned}
&-\kappa + g(x_N+s)(h-s) - \left[1-G(x_N+s)\right] - e(y^*(x_N))\tau \\
=&-\frac{d}{dx}\left[e^*(x_N)\right]\tau x_N - g(x_N+s)s - \left[1-G(x_N+s)\right] - e^*(x_N)\tau \\
=&-\tau\left[\frac{d}{dx}e^*(x_N)x_N + e^*(x_N)\right] - g(x_N+s)s - \left[1-G(x_N+s)\right] \\
=&-\tau\frac{d}{dx}\left[e^*(x_N)x_N\right] - g(x_N+s)s - \left[1-G(x_N+s)\right]
\end{aligned}$$

Interestingly, the result can be decomposed into two parts: the marginal effect of precaution intensity on NP's payoff under the negligence rule and the marginal effect on P, deterred or undeterred. The marginal effect on P remains unchanged between strict liability and the negligence rule. So, we only need to focus on $\frac{d}{dx}\left[e^*(x_N)x_N\right]$ to compare strict liability and the negligence rule. If $\frac{d}{dx}\left[e^*(x_N)x_N\right] = 0$, the net marginal social costs (i.e., the marginal

social cost plus the marginal social benefit) are the same, implying the same precaution intensity chosen by PT under both regimes. However, when $\frac{d}{dx}\left[e^*(x_N)x_N\right] < 0$, the net marginal social cost under the negligence is lower and closer to the first-best level (i.e., zero). When $\frac{d}{dx}\left[e^*(x_N)x_N\right] = \frac{-1}{\tau}\left(g(x_N+s)s + [1-G(x_N+s)]\right) < 0$, the net marginal social cost equals to zero, meaning PT's precaution intensity under the negligence rule achieves the first-best outcome. If, otherwise, $\frac{d}{dx}\left[e^*(x_N)x_N\right] > 0$, the net marginal social cost is higher than under strict liability rule, implying more significant over-precaution than under strict liability. □

This result is striking. While the negligence rule does not force PT to internalize all the costs associated with her precaution, it could, in some cases, induce PT to achieve the first-best outcome. The reason behind this is that the negligence rule helps curb over-precaution. When PT is held negligently liable, she is required to spend more on precaution accuracy when improving her precaution intensity. However, the additional expenditure is not rewarded as under strict liability, in which PT's marginal investment in precaution accuracy is offset by the marginal benefit of reduced liability. Therefore, PT bears a higher marginal cost of *precaution accuracy* than her counterpart does under strict liability. However, PT does not bear the marginal liability resulting from higher precaution intensity. If, however, the marginal cost of *precaution accuracy* outweighs the marginal liability from *precaution intensity*, which is exactly the marginal effect of precaution intensity on NP's welfare, then the marginal precaution cost for PT can be higher under the negligence rule than under strict liability. As mentioned in Section 2.2.3, the misalignment comes from three sources, which strict liability cannot completely resolve. Given the presence of over-precaution, a higher marginal cost that leads to a lower precaution intensity can make the negligence rule be superior to strict liability.

This result might sound both familiar and counter-intuitive. On the one hand, it sounds similar to the activity-level argument. On the other hand, it leads to the opposite result that

the activity-level argument predicts. Namely, tortfeasors subject to the negligence rule would engage in a suboptimally high activity level. Rather, NP under the negligence rule takes under-precaution. The difference between this result here and the activity-level argument is that the court can observe precaution intensity and apply the due care level accordingly but is assumed unable to do so for activity level. When the court can set the due care level that incorporates the activity level, the tortfeasor will be forced to internalize the additional cost of due care for a higher level of activity, which may yield a similar result as described here.

2.3.2 Law enforcement against perpetrators

Aside from the liability imposed on PT, PT's precautionary decision-making is also affected by law enforcement against perpetrators. P fears being sanctioned and are deterred from committing socially undesirable offenses. Thus, the likelihood of victimization is lower due to fewer perpetrators. With the assumption of diminished deterrence effect, PT would enjoy less private deterrence benefit from her precaution and thus invests less in the presence of law enforcement.[3] By deterring perpetrators and reducing the victim's precaution intensity, the introduction of law enforcement also saves on cost borne by P.

3. This statement may be qualified when law enforcement and precaution are complementary. For instance, law enforcement with a harsh sanction may lead the same detection level to be more deterrent and thus incentivize PT to make more effort in improving detection than when the sanction is weaker. See Ben-Shahar and Harel (1996).

Effects on precaution intensity

To see how the level of enforcement severity affects PT's behavior, I summarize PT's decision rules under different regimes.

$$\begin{aligned} \kappa &= g(x+s)h && \text{no liability} \\ \kappa + e^*(x)\tau &= g(x+s)h && \text{strict liability} \\ \kappa - x\left[e^*(x)\right]'\tau &= g(x+s)h && \text{negligence rule} \end{aligned}$$

As we can see above, regardless of legal regimes, the marginal benefit of deterrence (right-hand side) is the same negative-slope line. Moreover, as the severity of law enforcement (s) becomes larger, the marginal deterrence benefit is smaller, further driving the chosen precaution intensity down. This result can be illustrated both algebraically and geometrically. Algebraically, the function of perpetrators' marginal gain is decreasing because of negative $g'(\cdot)$, implying that a higher s lead to a lower $g(x+s)$. Lower marginal benefit leads to a lower precaution intensity. Geometrically, the line of marginal deterrence benefit moves leftward when s increases. Hence, its intersections with the marginal costs under each regime must also move leftward, resulting in a smaller precaution intensity in equilibrium.

The intuition behind formal equations is straightforward. As assumed in the model, P as a group has many potential perpetrators. There are more low-gain Ps than high-gain ones. Under such a circumstance, the same unit of precaution can thus deter more low-gain Ps than high-gain ones. When law enforcement is introduced, low-gain Ps are already deterred by legal sanctions, leaving the pool of Ps composed of fewer Ps with high gain. As a result, marginally, such a unit of precaution would deter fewer Ps. In sum, given the diminished effectiveness of deterrence, the introduction of law enforcement would make PT's precaution less rewarding.[4]

4. This may not hold when P's gain conforms to other distributions that do not have a diminished deterrence. For instance, when perpetrators' gain conforms to the uniform distribution, the marginal benefit

With cost unchanged, a lower marginal benefit would lead to a lower level of precaution intensity.

This crowd-out effect resulting from both the diminished deterrence and the fact that public enforcement and private precautions are substitutes should not be unrealistic. There are examples in various contexts where the introduction of law enforcement leads to lower private precautions (Bell and Parchomovsky 2012). Empirical evidence also shows that people do reduce their level of precaution when enforcement is present, such as police patrol, etc.(Vollaard and Koning 2009).

Proposition 2.4 (Crowd-out effect)**.** *When the probability density function of P's gain is strictly decreasing, more severe law enforcement would crowd out private precaution, and the scales of the crowd-out effect of enforcement on precaution vary with liability regimes.*

Proof. Assume the enforcement severity increases δ_s and the crowd-out effect is δ_x^i, where $i = \{NL, S, N\}$, denotes no liability, strict liability, and the negligence rule, respectively. When PT is not held liable for NP's losses, we then solve $\kappa = g(x - \delta_x^{NL} + s + \delta_s)$. It is obvious that $\delta_x^{NL} = \delta_s$.[5] Use δ_x^{NL} as a benchmark and plug it into the decision rule under strict liability. We find that $\kappa + e^*(x - \delta_x^{NL})\tau > \kappa + e^*(x)\tau = g(x + s)h$. The marginal cost outweighs the marginal benefit, so PT under strict liability would further reduce her precaution intensity, implying $\delta_x^S > \delta_x^{NL} = \delta_s$. Likewise, plug the benchmark crowd-out effect, δ_x^{NL}, into the decision rule under the negligence rule, we get $\kappa - (x - \delta_x^{NL}) \left[e^*(x - \delta_x^{NL})\right]' \tau$ which can be greater or less than $g(x + s)h$.

□

Proposition 2.4 not only shows the existence of the crowd-out effect but unearths different

---

of private precaution remains unchanged until aggregate deterrence, the sum of private precaution and law enforcement, deter all perpetrators.

5. It should be noted that the result $\delta_x^{NL} = \delta_s$ comes from the assumption of a constant marginal cost of precaution intensity. If the marginal cost of precaution intensity is increasing, then this result may no longer hold. Instead, $\delta_x^{NL} < \delta_s$. The comparison of the crowd-out effect across different liability regimes here should not be affected, however.

scales of the crowd-out effect of law enforcement on private precautions under different liability regimes. The result in Proposition 2.4 implies the possibility that the optimal enforcement severity could hinge on the choice of liability regimes. The net marginal benefit of increasing enforcement severity includes (1) savings on precaution costs incurred by PT, (2) savings on the social cost borne P, (3) marginal change of the cost borne by NP, and (4) net marginal deterrence benefit. Each of them is determined by the scale of the crowd-out effect and the starting points of both x and s. However, given different levels of the crowd-out effect, the marginal benefits under different regimes are unlikely to coincide. As a result, the fact that marginal benefits vary with liability regimes also implies that the optimal level of enforcement severity varies with the choice of liability regimes. An example will be provided in Section 2.4.3 to show this possibility.

Effect on parties' payoff

The level of law enforcement severity also affects each party's payoff. Moreover, the effect is not uniform across liability regimes. Below I will investigate the effect of increases in enforcement severity on each party's payoff under different regimes.

To begin, I first see how increasing enforcement severity affects PT's payoff. PT would never be worse off simply because of increasing enforcement severity. To see this, let's compare her decisions before and after law enforcement severity increases. By sticking to what she chose before enforcement severity increases, PT already benefits from lower expected harm from victimization and is therefore better off. If she further chooses to deviate from the status quo, she, as a rational actor, must benefit from such a deviation. Therefore, she could never be worse off and would unambiguously be better off when enforcement becomes more vigorous. Therefore, we can conclude that PT is always better off from stronger law enforcement.

Now I turn to perpetrators. P's payoff is represented by the net gain from committing

misconduct, which is negatively determined by aggregate deterrence. When aggregate deterrence is lower, P is less likely to be deterred and is more likely to commit misconduct that is beneficial to P. As aggregate deterrence gets higher, P is more likely to be deterred and is left with a smaller chance to gain from misconduct. The aggregate deterrence depends on the level of the crowd-out effect, which further hinges on the choice of liability regime, as Proposition 2.4 states. In this model, the crowd-out effect under the strict-liability regime could outweigh the increase in enforcement severity, leading to a decrease in aggregate deterrence. P under the strict-liability regime will be better off from increased enforcement severity. This echoes the phenomenon observed by Philipson and Posner (1996), which demonstrates potential increases in crime when law enforcement is so strong that crowds out private precaution. However, this is not necessarily the case for P when PT is held negligently liable.

Lastly, NP's situation can also be better or worse when law enforcement becomes more severe. When PT is not liable for NP's loss, NP would definitely suffer the cost imposed by PT's harmful precaution because PT does not invest in precaution accuracy. Under such a circumstance, reduced precaution intensity owing to stronger law enforcement would make NP better off. Under the strict liability regime, NP is indifferent as they are made whole. However, he could suffer greater costs when increasing law enforcement severity under the negligence rule. Again, the cost suffered by NP is multiplied by precaution intensity and the error rate. While stronger law enforcement reduces precaution intensity, it also reduces the accuracy when PT invests optimally in accuracy according to her chosen intensity. Specifically, NP bears the total error cost, $\tau x e^*(x)$. When $\frac{d}{dx}\left[xe^*(x)\right] < 0$, the total error cost increases as x decreases. As a result, when the negligence rule is chosen to regulate PT's effort in precaution accuracy, NP may be worse off as law enforcement becomes more severe.

Effect on the choice of liability regimes

As illustrated above, the optimal enforcement severity may vary with different liability regimes due to different scales of the crowd-out effect. Such a relationship may also hold in reverse. That is, under certain conditions, the choice of liability regimes may depend on enforcement severity.

Proposition 2.5. *When the total error cost borne by NP, $E(x) = \tau x e^*(x)$, is not monotonically increasing or decreasing in the range of positive $x \leq \bar{b} - s$, the choice of liability regimes may vary with enforcement severity.*

Proof. Let's first assume that $E(x)$ is not monotonically increasing or decreasing, implying that there exist x_1 and x_2, where $x_1 < x_2$, such that $E'(x_1)E'(x_2) < 0$. Furthermore, we can find an $\bar{x} \in [x_1, x_2]$ such that $E'(\bar{x}) = 0$. If we can find an $\bar{x}$ such that $E'(\bar{x}) = 0$, then we can find an $\underline{s}$ such that $\kappa + e^*(\bar{x})\tau = g(\bar{x} + \underline{s}) = \kappa - \bar{x}\left[e^*(\bar{x})\right]'\tau$. Accordingly, if the state slightly increases s from $\underline{s}$, then PT will take x^-, which is smaller than $\bar{x}$. If, otherwise, the state slightly decreases s from $\underline{s}$, PT will take $x^+ > \bar{x}$. Suppose $x_1 < x^- < \bar{x} < x^+ < x_2$. The order implies $E'(x^+)E'(x^-) < 0$, meaning the choice of efficient liability regimes is different under different levels of enforcement severity, and the order flips around $\bar{x}$ (see Proposition 2.3). Hence, when the conditions are met, the change in s may imply the optimal choice of liability regime. □

Proposition 2.5 demonstrates the possibility that, in the context of precaution-taking behavior, the choice of efficient liability regime may depend on the enforcement severity. That is, the negligence rule may be more desirable than strict liability when modest enforcement is introduced. But strict liability could outperform the negligence rule when enforcement severity becomes higher. An example will be offered in Section 2.4.1.

From the discussion above, precaution-taking behavior links both law enforcement against perpetrators and liability imposed on precaution-takers. Due to its pivotal position, the

choice of liability regimes and enforcement severity can sometimes be interrelated and interdependent.

2.4 Simulation

The analysis above is not sufficient to allow us to have a concrete comparison between strict liability and the negligence rule. As mentioned above, the scales of the crowd-out effect vary with liability regimes (Proposition 2.4). Also, whether the choice of liability regimes varies with different levels of enforcement severity hinges on certain conditions (Proposition 2.5). Both claims require specifications of the function type of precaution accuracy. As a result, I simulate the model with concrete function types to showcase potential results from the analysis. In this section, three types of functions will be discussed: radical, exponential, and fractional. All would conform to the basic assumptions in the setup; namely, $e(0) = 1, e'(y) < 0$, $e''(y) > 0$, and $0 < e(y) \leq 1 \ \forall y \geq 0$.

2.4.1 Radical error-rate function

In this case, I will specify the function of the error rate along with other parameters to showcase the potential results of the aforementioned analysis. In brief, the error rate is specified as $e(y) = 1 - \frac{\sqrt{y}}{2}$. As a result, we can solve the first-best error rate as a function of precaution intensity $e^*(x) = 1 - \frac{\tau x}{2}$, and the correspondent expenditure in precaution accuracy is $y^*(x) = \frac{(\tau x)^2}{4}$. Also, I assume that $h = 5, \bar{b} = 1, \tau = 1, \kappa = 1$. Accordingly, the lines of marginal cost under each regime are illustrated in Figure 2.1. As Figure 2.1 shows, the intersects between marginal costs and benefits under all regimes are smaller as law enforcement becomes more severe (moves leftward). Moreover, the relationship between law enforcement severity and precaution intensity can be demonstrated in Figure 2.2. Figure 2.2 showcases that x decreases as s increases, implying law enforcement crowds out private precautions. However, the scales of the crowd-out effect, which can be observed by the

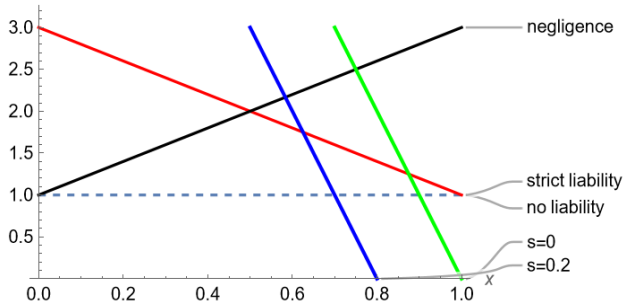

Figure 2.1: Marginal cost and marginal benefit under each regime

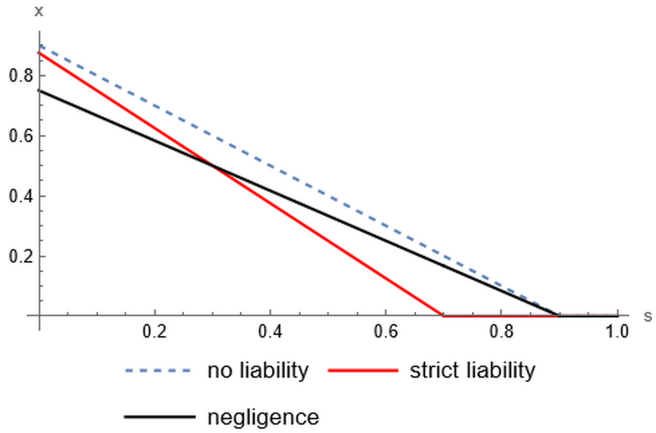

Figure 2.2: Relationship between primary enforcement severity and victim precaution intensity

slope, differ across different regimes. For instance, the slope of the dashed line is -1, implying aggregate deterrence remains unchanged under the no-liability regime because a unit decrease in x is offset by a unit increase in s. In comparison, the slope of the red line (strict liability) is steeper than -1, implying a unit increase in s would result in a more-than-one unit decrease in x. By contrast, the situation reverses under the negligence rule, where aggregate deterrence improves as s increases.

Also, we can specify parties' payoffs under different regimes. As Figures 2.3, 2.4, and 2.5 show, the parties' payoffs also vary with different levels of primary enforcement severity. As mentioned above, PT would never be worse off as primary enforcement becomes more severe. Also, P is better off under strict-liability regimes. NP, instead, is better off only in the no-liability regime but is worse off under the negligence rule when a modest level of law

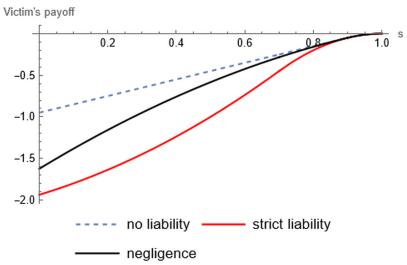

Figure 2.3: PT's payoff

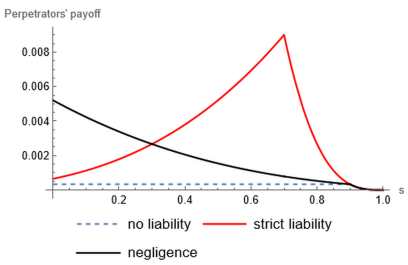

Figure 2.4: P's payoff

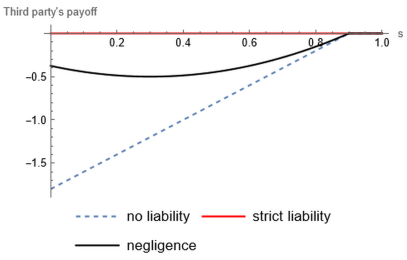

Figure 2.5: NP's payoff

enforcement is introduced.

Lastly, with a specified error-rate function and parameters, we can further calculate and compare social welfare under each regime. As Figures 2.6 and 2.7 illustrate, which regime is more desirable depends on the severity of primary enforcement. In reverse, the marginal benefit of increasing s (i.e., the slope of each line) also differs across different regimes, implying that optimal enforcement severity could be different as the liability regime changes.

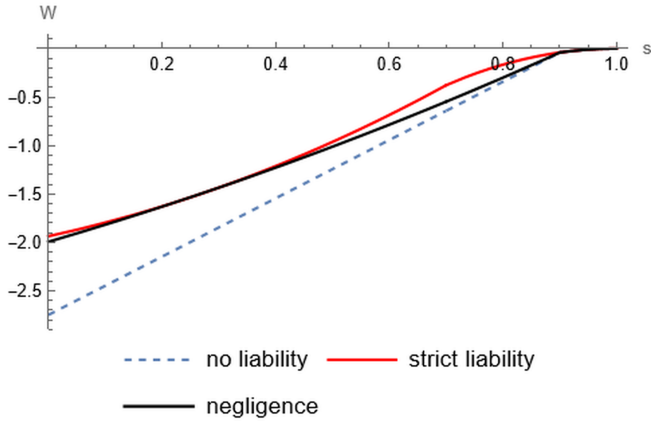

Figure 2.6: Social welfare with respect to s under different regimes of secondary enforcement

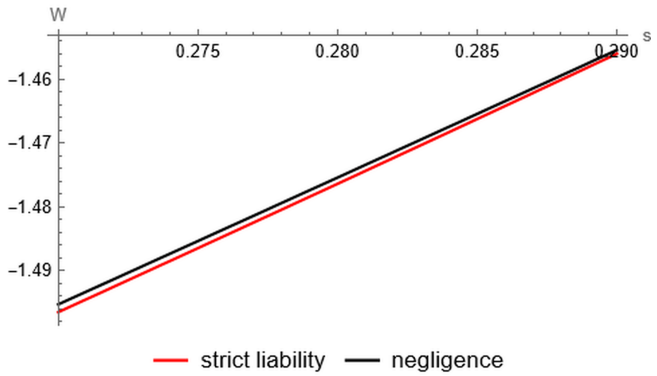

Figure 2.7: Social welfare w/r/t $s \in [0.27, 0.29]$ under strict liability and the negligence rule

2.4.2 Exponential error-rate function

Now I turn to the case in which the error rate is an exponential function of its investment, say, $e(y) = e^{-y}$. In this case, we can solve $e^*(x)$ as follows:

$$e'(y^*) = -e^{-y^*} = -e(y^*) = -e^*(x) = \frac{-1}{\tau x} \implies e^*(x) = \frac{1}{\tau x}$$

Hence, $e^*(x) = \frac{1}{\tau x}$. In addition, the total error cost $E(x) = \tau x e^*(x) = 1$. Since the total error cost is a constant, its first-order condition would be zero, implying $e^*(x) + x\left[e^*(x)\right]' = 0$. Accordingly, the marginal precaution cost borne by PT under both regimes would be equal. Therefore, based on Proposition 2.3, both strict liability and the negligence rule are equivalent regardless of the chosen precaution intensity x. That is, choosing one of the regimes would not have an efficiency effect. Instead, it would have a distributive effect since the error cost is allocated between PT and NP under different regimes. Aside from the choice of optimal regimes, the same marginal-cost lines also imply the same crowd-out effect when PT is held liable and that the crowd-out effect is greater than in the case where PT is not held liable for NP's losses.

2.4.3 Fractional error-rate function

Lastly, I turn to a possible fractional function of error rate. Here, I assume $e(y) = \frac{1}{1+y}$. By doing so, $e(0) = 1, e'(y) < 0$ and $e''(y) > 0$. Again, we can solve the optimal error rate as a function of precaution intensity $e^*(x)$.

$$e'(y^*) = \frac{-1}{(1+y^*)^2} = \frac{-1}{\tau x} \implies e^*(x) = e(y^*(x)) = \sqrt{-e'(y^*)} = \frac{1}{\sqrt{\tau x}}$$

As a result, the total error cost $E(x)$ becomes $\sqrt{\tau x}$, and $E'(x) = \frac{\sqrt{\tau}}{2\sqrt{x}} > 0$, implying strict liability always outperforms the negligence rule, regardless of the chosen level of precaution intensity.

2.4.4 Summary

As simulations show, different function forms imply different choices of liability regimes to regulate PT's effort in error reduction. To be clear, not all propositions given above are sensitive to function forms. Also, the aforementioned examples are not meant to exhaust all possible function forms and the relationships between the error rate and the investment. Rather, they are exemplary to provide a clearer sense of how function types affect the results derived from the model. With that in mind, how function types affect the results discussed above can be summarized in Table 2.1

	Radical error rate	Exponential error rate	Fractional error rate
Examples: $e(y)$=	$1 - \frac{\sqrt{y}}{2}$	e^{-y}	$\frac{1}{1+y}$
Total error cost E(x)	not monotonically increasing or decreasing	constant	monotonically increasing
Crowd-out effect (Proposition 2.4)	$\delta_x^{SL} > \delta_x^{NL} > \delta_x^N$	$\delta_x^{SL} = \delta_x^N > \delta_x^{NL}$	$\delta_x^{SL} > \delta_x^N > \delta_x^{NL}$
Optimal liability regime	depending on s (Proposition 2.5)	equivalent	strict liability

Table 2.1: Function forms compared

2.5 Discussion

Now I proceed to discuss the findings above and deliver the implications. All models have their limits, and the one in this chapter is no exception. For parsimony, I have to discard some real-world features. Some of them will be discussed in Section 2.5.1 With these qualifications in mind, I then move to the implications to see how these findings allow us to explain current doctrines and design a better legal tool.

I first investigate two tort law doctrines that regulate precautionary measures: dog bites and traps in Section 2.5.2. I argue that from the findings derived above, the rare use of the negligence rule in precautionary measures compared to other tortious activities and the tolerant attitude toward PT are, to some extent, explainable. Having said that, current doctrines could still be improvable. I propose in Section 2.5.3 the court should replace the current doctrines of unprivileged traps with a decoupled liability to avoid dilution of deterrence. Finally, I will revisit the line of literature that proposes to regulate precaution-taking behavior through criminal sentencing in Section 2.5.4. I argue that when precautions are harmful, such proposals may sometimes backfire and thus requires further scrutiny.

2.5.1 Qualifications

In the real world, both P and PT would be subject to both criminal and civil liability. In concrete, P is liable for PT's loss and is required to compensate PT. Meanwhile, when PT takes harmful precautions, PT can be held criminally liable for NP's injury inflicted by her precautions. However, the model only includes P's criminal liability and PT's civil liability. To wit, I do not include P's civil liability for PT's harm nor PT's criminal liability for NP's injury. In addition, the model's assumption of law enforcement can deviate from the practice. I assume that sanctions imposed by law enforcement are monetary and thus merely wealth transfers that do not incur additional social costs. Furthermore, law enforcement is assumed more accurate than private precautions. In reality, non-monetary sanctions are widely used in

the U.S., and innocent non-perpetrators might be wrongfully convicted. Finally, this chapter does not include potential effects discussed in the literature, such as the harm-displacement effect, on other precaution-takers.

First, PT, as a victim, is entitled to obtain recovery from P for the harm caused by P's misconduct. Tort damages awarded to PT significantly reduce the level of their expected harm. In this regard, the model may exaggerate the level of precaution intensity. Such an exaggeration, however, should be limited in some contexts where harms are irreparable by monetary awards (such as bodily injuries), the detection rate is low (burglary or theft on unattended premises), or even Ps are judgment-proof. In these contexts, private enforcement should be less frequent and less influential on PT's precautionary decisions.

Second, PT may be held criminally liable for the injury caused by their harmful precautions. In criminal cases, what PT can do is to claim self-defense. However, self-defense is not boundless and supposed to be used against P. To claim self-defense against NP, PT should at least reasonably believe that NP is creating an imminent threat. Otherwise, it is treated as "imperfect self-defense," which does not exempt her from criminal liability (see LaFave 2017, pp. 727–28). The threat of criminal liability should encourage PT to invest more in precaution accuracy to avoid criminal punishment. However, if the standard of reasonableness in criminal law is set below or equal to the standard of due care in tort law, then PT's precautionary decision would not be affected by the lower criminal law standard of reasonableness; PT would already invest optimally in precaution accuracy. Criminal liability would not induce more investment in precaution accuracy. Admittedly, if the criminal law standard of reasonableness is higher than the tort law standard of due care, then the actual investment in precaution accuracy will be higher than the model predicts. However, the criminal law standard should be less demanding considering the procedural protections and the idea of criminal law as *ultima ratio*.

Third, the model assumes that sanctions of law enforcement are monetary and thus

less socially costly than private precautions. However, in reality, most criminal sanctions are non-monetary, which deter P by imposing disutility on them as victim precautions do. Non-monetary sanctions make no difference from victim precautions in the sense that both measures are incurring net social costs when imposing *ex post*. Therefore, when sanctions are non-monetary, stronger law enforcement no longer guarantees lower social costs. However, this does not necessarily overturn the findings. In fact, we can tweak the model by incorporating the cost of non-monetary sanctions. Ultimately, it is the efficiency comparison between law enforcement and private precautions. Since sanctions are no longer less costly, the optimal enforcement severity should be lower than the model predicts.

Of course, law enforcement may commit errors as PT does. The concern of wrongful conviction is valid but should be less problematic given the procedural protection provided by constitutions. Even if, in practice, law enforcement is not as accurate as private precautions, we can incorporate the cost of wrongful enforcement into the enforcement cost, which again will drag down the optimal enforcement severity from the predicted value.

Lastly, this model does not consider the potential harm-displacement effect of PT's precautions on other PTs. To be clear, this model implicitly assumes the harm-reduction effect of precautions since they "deter" P from committing socially undesirable misconduct, rather than divert them to other PTs. Accordingly, the model would over-estimate the first-best level of precaution intensity. However, this overestimation would not severely nullify our analysis. On the contrary, it might reinforce the conclusion that PT takes an excessive level of precautions. If the actual first-best level of precaution intensity is lower, then it means PT's precautionary decisions are way more excessive from the social perspective. This would imply that the state should consider fortifying law enforcement to further discourage victims from taking such harmful but socially ineffective precautions.

2.5.2 Doctrinal applications: dog bites and traps

First and most relatedly, the model speaks to victim precautions. In tort law, two case types are most readily in terms of victim precautions: dog bites and traps. For instance, a homeowner can keep a dog or install a spring gun to protect herself and her home from illegal invaders. However, given their same purpose, no unified approach consistently regulates both measures purported to protect victims.

Currently, rules regulating dog bites diverge across states. Some states adopt the one-bite rule, meaning that dog owners are exempted from the liability of their dogs' first bite to learn the dogs' vicious propensity. In contrast, other states hold dog owners strictly liable for first and future bites. While both rules converge after the dog's second bite and after, they can differ if most dogs bite only once. Interestingly, the negligence rule is rarely used in the context of dog bites. Of course, dogs are usually kept as companion animals rather than guard dogs. However, the doctrine still applies to the context where dogs are used as a precautionary measure.

Another doctrine related to precautions is the tort liability for setting traps, including spring guns. The common law attempts to distinguish whether a trap is deadly or not and, based on the "deadliness," determines whether a trap is privileged (non-deadly) or unprivileged (deadly). Cases concerning deadly traps, such as spring guns (see e.g., *Phelps v Hamlett* (207 S.W. 425 [1918]) and *Grant v Hass* (31 Tex. Civ. App. 688 [1909])), are distinguished from non-deadly traps, such as barbed wire (see e.g., *Burrill v Alexander* (75 N.H. 554 [1910]) cf. *Quigley v Clough* (173 Mass. 429 [1899]) and *Mazey v Loveland* (133 Minn. 210 [1916])). If a trap is deadly and thus unprivileged, then plaintiffs (i.e., P and NP) are entitled to recovery. If, otherwise, the trap is privileged, defendants (i.e., PT) are not held liable for both P's and NP's losses.

The aforementioned doctrines share two common characteristics that could be puzzling. First, courts do not apply the negligence rule to such precautionary measures, given the

dominance of the negligence rule in tort law. The imposition of strict/absolute liability on precautionary measures seems to deviate from courts' inefficient tradition of limiting strict liability to "uncommon activities." (Shavell 2018). Second, they immunize some precaution-takers from liability. Specifically, a dog's first bite in some states and non-deadly traps in common law are both tolerated.

The findings above can, to some extent, explain courts' reluctance to apply the negligence rule. As we know, the negligence rule requires the court's knowledge of the optimal level of care. In the context of precaution-taking, for the negligence rule to be desirable requires more information to be optimally implemented. As shown in the Simulation in Part 2.4, the optimality of the negligence rule depends not only on the error-rate function types but also enforcement severity. Both are costly for the court to learn. The error-rate function could be specific to a certain precautionary measure and is too nuanced to apply. For instance, even if the court decides to apply the negligence rule to dog bites, can the court ensure that the radical error-rate function across all species of dogs? Needless to say, the court may also need to learn the enforcement severity for a specific defendant precaution-taker that makes the negligence rule to be optimal. Such details may, currently, significantly increase the cost of both implementation and compliance. Without knowing specific error-rate functions or the enforcement severity pertaining to the specific defendant precaution-taker, the court would not risk choosing the negligence rule, which may lead to over-deterring precaution (in the case of the radical error-rate function with low enforcement severity) or under-deterring precaution (in the case of the fractional error-rate function). Compared to strict/absolute liability, of which the inefficient result is determinate and easy to learn, the negligence rule brings more uncertainty, which the court may dislike.

The tolerance for some costs incurred by modestly harmful precautions can be explained by the tradeoff between lower aggregate deterrence due to the crowd-out effect and the total error cost. Based on the assumption of decreased deterrence benefit, enforcement crowds

out private precautions. The level of the crowd-out effect is, however, dependent on the underlying liability regimes (Proposition 2.4). Strict liability, while forcing PT to internalize NP's cost, has a greater crowd-out effect that undermines aggregate deterrence. Also, strict liability incurs administrative cost to enforce. If we assume a downward, convex-up optimal error-rate function, then it can be observed that when enforcement is severe, and precaution is less intensive, the crowd-out effect is significant, but the reduced error cost is limited. By contrast, when enforcement is weak, and PT takes intensive precautions, holding PT strictly liable can save a lot on the error cost with a limited crowd-out effect. Hence, the choice of liability regimes based on precaution intensity can be rationalized by the tradeoff between the crowd-out effect on aggregate deterrence and the savings from the reduced error cost. Based on the explanation provided by this model, it makes sense that low-intensity (non-deadly) precautions are privileged. States applying the one-bite rule should have a higher investment in enforcement and a lower level of private precautions.

2.5.3 Policy implication: decoupled liability

Proposition 2.2 claims, even though PT is held strictly liable for NP's loss and makes an optimal effort in precaution accuracy, her precaution intensity still remains excessively high. One source of the unsolved over-precaution is that PT does not internalize the harm she imposes on P. In practice, courts in some cases recognize that a felonious plaintiff (P) is entitled to recovery (see e.g., *Katko v Briney* (183 N.W.2d 657 [1970]) and *Allison v Fiscus* (100 N.E.2d 237 [1951])) but in other cases hold that PT is not liable (see e.g., *Scheuerman v Scharfenberg* (163 Ala. 337 [1909])). For me, the two conflicted lines of cases show a dilemma between deterrence and optimal precautions.

Clearly, holding PT liable for the harm inflicted by their precautions on P helps make their incentives more aligned. However, such alignment comes at the cost of reduced deterrence. If P is compensated and made whole for the loss caused by PT's precautions, such precautions

are ineffective. An ineffective precaution would never justify its cost. As a result, PT would not take any precautions anymore in equilibrium.

A similar effect could also be found when such harms imposed by PT's precaution should be considered "non-legal sanction" and be deducted from legal sanctions. Even when P is not entitled to recovery, deducting such non-legal sanctions from their legally imposed sanctions as proposed by Cooter and Porat (2001) can dilute the deterrence of private precaution. Likewise, the precaution intensity in equilibrium would reduce to zero as in the previous case. Hence, it seems not recommended to deduct harms done by PT's precautions from the sanction imposed on P. The disagreement between their proposal and this chapter is mainly attributed to their concern of "over-deterrence," which this chapter assumed to be non-existent. All misconduct is assumed inefficient, so over-deterrence is not a concern of this chapter. If, otherwise, some misconduct is efficient, then their proposal would be more appropriate.

Of course, precautions should not be eliminated but controlled as courts have been implicitly recognized. In particular, when private precautions are cheaper than public law enforcement in deterring inefficient misconduct, they could be socially desirable. As a result, holding PT liable for P's harm and eliminating the precaution-taking incentives would not be recommended. Irrefutably, PT can still take other defensive precautionary measures when harmful precautions are prohibited and abandoned. However, such situations are not the focus of this chapter.

Since liability does not work, Pigouvian taxes could be a potential candidate to solve this misalignment. Endorsed by both Clotfelter (1978) and Shavell (1991), such taxes reflect the social costs externalized by PT, and can be reflexively imposed on her and adjust her level of precaution intensity. However, such a scheme would be difficult, if not possible, to implement. To set an optimal tax, the policymaker should know the expected harm suffered by undeterred perpetrators. While the level of harm could be learned *ex ante* by observing the

type of precaution (e.g., dog's species or spring gun), the number of undeterred perpetrators is determined by both the distribution of P's gain as well as the underlying law enforcement severity. The distribution of P's gain may fluctuate with macroeconomic conditions, such as unemployment rate and wages (Gould, Weinberg, and Mustard 2002), and can therefore be obscure for policymakers. Moreover, law enforcement severity is subject to change over time, such as the priority of combating crime in the political agenda during the election and non-election years (Levitt 1997). All these reasons would prevent policymakers from setting an optimal tax that perfectly reflects the cost borne by perpetrators *ex ante*.

A decoupled liability regime can solve the information problem encountered by Pigouvian tax without deterrence dilution faced by the current case law of landowner's liability. First, decoupled liability allows the court to observe and impose liability *ex post*. Ideally, all undeterred perpetrators would enter the criminal justice system. Hence, the court is not required to learn the distribution of perpetrators' gain to know how many undeterred perpetrators there are. Instead, they can learn this through direct observation. While not all perpetrators are apprehended and prosecuted, an *ex-post* approximation based on real observations should be more precise than *ex-ante* estimation. Second, decoupling the payment from one party to the transfer to another party allows the state to flexibly regulate both parties' behavior without affecting one another (Polinsky and Che 1991). This regime has the advantage of solving the deterrence problem encountered by the current case law. On the payment side, PT paying for the harm would adjust her precaution intensity to align with the first-best level. From the transfer side, P is not granted such payment, so private precautions can still deter P.

2.5.4 Regulating precaution by criminal law

Aside from the discussions on tort doctrines and policy, this model also complements economic theories of public enforcement. Since the seminal piece of Becker (1968), the literature

on optimal public enforcement often suggests the harm as exogenous to the enforcement level. Their goal is to balance deterrence benefit and enforcement cost (Polinsky and Shavell 1979; Polinsky and Shavell 1984; Shavell 1987a). However, the cost of crime is not limited to the actual harm from the misconduct. As Anderson (1999; 2021) indicates, the aggregate cost of crime can way higher than the actual harm done by perpetrators. Such costs, unlike the harm, are endogenous to law enforcement. Namely, the level of law enforcement also affects the total aggregate cost of crime, particularly costs related to precautions.

Some scholars have already recognized this point and paid attention to optimal enforcement by considering the role of private precaution in public enforcement. For instance, Hylton (1996) argues that the optimal penalties should be set lower than the Beckerian optimal penalties to induce victim precautions. In the same vein, Ben-Shahar and Harel (1996) also note that lower penalties for criminal attempts are justified by their effect on curbing PT's over-precaution. They argue that by lowering penalties on criminal attempts, perpetrators are encouraged and directed to those victims with suboptimally high precaution, incentivizing such victims to reduce their precautions. Clements (2003) also agrees with the idea to use criminal sanctions to incentivize optimal precaution when precaution costs are different among PTs and unobservable. Given their distinct focuses on over- and under-precaution, they both recognize the role of law enforcement in intervening in private precautions.

While this chapter agrees with them that criminal sanctions should consider private precaution and assume the role of intervening through adjusting sanctions, it points in the opposite direction. Specifically, when PT takes harmful precautions, it is less likely that she will take insufficient precautions. On the contrary, they take excessive precautions because they fail to internalize the total social cost. As a result, lowering sanctions proposed by both Hylton and Ben-Shahar and Harel, and Clements would exacerbate the problem. On the one hand, PT is forced to take more precautions to protect herself when public enforcement

retreats. On the other hand, P might not be effectively encouraged to target PT taking more precautions because their harmful precautions, say spring guns or electrified fences, may impose greater harm than legal sanctions, such as fines or humane imprisonment.

2.6 Conclusion

Harmful precaution imposes costs on perpetrators and non-perpetrator third parties. Therefore, liability should be considered to impose on precaution-takers to induce their effort in precaution accuracy. When precaution-takers are held liable, their precautionary decision-making is influenced by both law enforcement against perpetrators and the liability imposed on them. On the one hand, law enforcement affects the private deterrence benefit by deterring perpetrators and reducing the need for private precaution. On the other hand, liability increases the precaution cost by incentivizing precaution-takers to invest in precaution accuracy and/or bear the loss borne by non-perpetrator third parties. The cost and benefit of precaution-taking are thus affected by liability and law enforcement, respectively. Such a pivotal position links liability and law enforcement, implying their interdependence on each other. In concrete, in the context of precaution-taking behavior, the optimal law enforcement severity may hinge on the choice of liability regimes. In reverse, the choice of liability regimes may depend on the enforcement severity. This interrelationship explains the choice of liability regimes for different treatment of precautionary measures, such as dog bites and traps. In addition, this chapter proposes a decoupled liability regime that aligns precaution-takers' incentives with the first-best outcome without sacrificing deterrence. This chapter also adds to the scholarly discussion of regulating precautions by adjusting criminal sanctions imposed on perpetrators.

CHAPTER 3

GATEKEEPERS AS PRECAUTION TAKERS

3.1 Introduction

In May 2023, the United States Supreme Court unanimously decided that Google is not liable for aiding terrorism via its users' use of YouTube in *Gonzalez v. Google LLC* (598 U.S. _____ [2023]). This decision, together with another concerning Twitter, resolves concerns that Big Tech would be subject to gatekeeper liability for online user conduct. In contrast, a half year before these cases were decided, the Department of Justice secured a guilty plea from Lafarge, a French cement conglomerate that "conspir[ed] to provide material support to designated foreign terrorist organizations," including ISIS and ANF." (U.S. Department of Justice 2022-10-18) This guilty plea led to a $780 million penalty. Looking further back, HSBC, a London-headquartered global bank, was fined $85 million for inadequate monitoring of money laundering and terrorist financing scenarios and poor risk assessment. (Withers 2021-12-17)

These three examples involve the same underlying misconduct — terrorist assisting — and subjects benefitting from others' misconduct. Specifically, Google benefits from users' content by gaining traffic, Lafarge secures competitive advantages through having subsidiaries do business with parties related to terrorists, and HSBC earns fees from customers and transactions potentially related to terrorist financing. Nonetheless, in these three cases, the outcomes differ: Google enjoys immunity, while Lafarge and HSBC face severe penalties. To avoid further liability, Lafarge undertakes to refrain from doing business with specific parties, and HSBC to improve its due diligence processes and report suspicious transactions. In contrast, Google faces no such obligations.

This disparity raises two questions. The first, which has been discussed extensively, is *why* and *how* third party gatekeepers are held liable for others' misconduct. The second

question, which stems from the first, asks what obligations designated gatekeepers must meet to be absolved of liability. The first question focuses primarily on the conventional wisdom around gatekeeper liability and the choice of optimal liability regimes; this chapter seeks to answer the nuances of the second question by drawing insights from the literature on victim precautions.

Gatekeepers are parties legally required to prevent or detect misconduct committed by others; failure to do so exposes them to legal liability. Examples of gatekeepers include accountants, banks (such as HSBC above, in terms of screening customers for compliance with anti-money laundering regulations), and employers held vicariously liable for employee misconduct. There are other *de facto* gatekeepers, such as journal editors who act as gatekeepers in reviewing which papers get published. However, this chapter focuses on gatekeepers who are *legally* responsible for the aforementioned tasks.

The short answer to the first question, regarding why gatekeepers are designated, is cheap deterrence. States designate gatekeepers to achieve cost-effective deterrence against misconduct. Contrary to Becker's (1968) suggestion, law enforcement, being budget-constrained, cannot achieve perfect deterrence by enhancing penalties alone. This is due to challenges like judgment-proof wrongdoers, the need for marginal deterrence, and the cost of non-monetary sanctions (see *infra* 3.2.1). To maintain deterrence, law enforcement must resort to higher and thus costlier detection. The cost of detection calls for the need for front-line gatekeepers better equipped to detect misconduct earlier and at a lower cost. By delegating partial responsibility to gatekeepers, society achieves the same level of deterrence at a lower cost.

Gatekeepers themselves are not wrongdoers, but are held responsible for misconduct committed by others (the gatekept) already subject to state law enforcement. Additional justifications are required to allocate partial responsibility for countering misconduct from the state to private gatekeepers, who face legal liability.[1]

1. The following introduction focuses on law-and-economics literature, for other discussions, such as moral justifications of gatekeeper liability, see Alzola (2017).

Kraakman's (1986) seminal work enumerates three conditions for a party to be an appropriate gatekeeper: efficacy, cost, and incentive. Parties possess the ability to identify, intervene, or stop misconduct effectively before law enforcement can do so. For instance, employers can screen or discipline employees to discourage them from committing misconduct before employees are pursued by law enforcement *ex post*. However, mere efficacy is insufficient; the party should also be able to perform these tasks at a *lower cost*. Gatekeepers incurring higher costs do not justify shifting responsibility from law enforcement. Lastly, private gatekeepers must already have gatekeeping incentives. Incentivized gatekeepers save enforcement resources spent on disciplining them. Although gatekeepers meeting all three requirements may not be perfect, Kraakman's conditions are helpful in selecting appropriate candidates for sharing enforcement responsibilities.

Kraakman's framework demonstrates how to identify ideal gatekeepers and establishes the bedrock for gatekeeper liability. However, in reality, not all candidates meet all the conditions or have incentives perfectly aligned with society. Gatekeepers' incentives vary with context, leading to context-specific discussions and designs of gatekeeper liability. This context-specific scholarly discussion encounters two shortcomings and thus can be improved upon. First, it fails to generate knowledge applicable and transferrable across different gatekeeper contexts, resulting in numerous papers focusing on auditors, lawyers, financial institutions, and employers without horizontal linkages to generate new ideas. Additionally, the choice of required acts is often path-dependent and rarely challenged, leaving the implications of such choices unexplored. Therefore, the compartmentalized scholarly research is underdeveloped in generating horizontal insights that answer the second question: what we expect designated gatekeepers to do.

This chapter aims to address the gaps in the current gatekeeper literature by comparing precaution-taking victims and gatekeepers, and add granular details to Kraakman's framework. By comparing victims and gatekeepers, this chapter views gatekeeping from

a precaution-taking perspective, drawing insights from the literature on victim precautions. Precaution-taking victims and gatekeepers share similarities, as both take precautions in response to potential harm caused by wrongdoers. Potential victims suffer direct harm when victimized by wrongdoers, while gatekeepers are legally liable and face sanctions when wrongdoers commit misconduct. To avoid or mitigate harm, both victims and gatekeepers can invest in precautions (or gatekeeping). By understanding gatekeeping as precaution-taking, the analysis of precaution costs and incentives can be applied to gatekeeper liability. Through this comparison, two observations arise: gatekeeping incentives are sensitive to enforcement strategies against gatekept first-order wrongdoers, and gatekeeping can impose costs on counterparties, other gatekeepers, and third parties.

First, gatekeepers' incentives can vary based on enforcement strategies against gatekept first-order wrongdoers. These incentives have significant implications for deterrence but are often underappreciated. Gatekeepers have different incentives when they are required to act differently. Some gatekeepers, such as auditors, are required to interdict wrongdoers, have different incentives compared to those required to report wrongdoers, such as banks. The scale and direction of changes in gatekeeping incentives also differ depending on the enforcement strategies. For example, if the law enforcement agency improves detection and lowers imposed sanctions without altering overall deterrence, interdicting gatekeepers would invest more in gatekeeping as their expected liability escalates due to more identified first-order wrongdoers. When more wrongdoers are identified, gatekeepers who fail to detect and interdict them will also be identified and punished. As a result, as the detection probability increases, interdicting gatekeepers would be incentivized to invest more in gatekeeping.

If, on the other hand, law enforcement imposes higher sanctions and a lower detection probability, gatekeeping incentives would change. With deterrence unchanged, fewer wrongdoers will be detected when law enforcement increases the severity of sanctions with reduced detection. The smaller number of detected wrongdoers disincentivizes gatekeepers from in-

terdicting wrongdoers. Gatekeepers are less willing to interdict wrongdoers because doing so yields a lower net benefit. Gatekeepers enjoy a lower level of deterrence benefit because wrongdoers would not be deterred by the additional sanction when interdicted, absent the punishment of inchoate misconduct. Also, gatekeepers face a lower cost for not gatekeeping because they are less likely to be detected due to a lower detection probability. Hence, interdicting such misconduct when law enforcement invests in sanctions rather than detection would be less attractive to gatekeepers.

However, the effect of such enforcement strategies on reporting gatekeepers' incentives is ambiguous. Assume that law enforcement emphasizes sanctions rather than detection, i.e., high-sanction-low-detection. In this case, reporting gatekeepers will, on the one hand, face a lower liability risk as there will be fewer detected wrongdoers. On the other hand, greater sanctions will make gatekeepers' reporting more deterrent to wrongdoers. As a result of more robust deterrence, the marginal deterrence benefit reflected by the associated reduced liability from a higher level of gatekeeping justifies the increased cost of gatekeeping. Therefore, there will be two forces simultaneously affecting the gatekeepers' reporting decision: one motivating more diligence, while the other disincentivizes gatekeeping. Ultimately, the result will depend on which effect trumps the other. From the analysis above, this chapter uncovers two points. First, the acts required of gatekeepers matter and vary with enforcement strategy. Second, more specifically, this chapter identifies a potential perverse effect of gatekeeper liability: when gatekeepers are liable for interdicting certain misconduct, increasing sanctions without punishing inchoate interdicted misconduct can crowd out gatekeeping incentives, leading to lower aggregate deterrence.

Aside from the dynamic relationship between enforcement strategies and gatekeeping incentives, comparing precaution-taking victims and gatekeepers also helps in identifying previously under-appreciated gatekeeping costs. Gatekeeping not only incurs costs borne by gatekeepers themselves, but can impose costs on other parties. As already noted in the

literature, gatekeeping can impose costs on counterparties (the gatekept clients), ranging from mere hassle to market unraveling. In addition, learning from the literature on victim precaution, gatekeeping can impose costs on other gatekeepers and unrelated third parties. These costs are rarely mentioned and discussed in the gatekeeper literature.

To begin with, gatekeeping can impose costs on other gatekeepers. When there are multiple competing gatekeepers, potential wrongdoers can shop around, leading to redundant costs from replicative gatekeeping processes. Gatekeepers may be incentivized to take observable gatekeeping measures to divert potential wrongdoers to other gatekeepers. Take banks as anti-money laundering gatekeepers, for example. Consider a case where two gatekeepers adopt a socially optimal level of gatekeeping. Unfortunately, this efficient state is unstable, as either gatekeeper would want to enhance their gatekeeping level to divert wrongdoers to the other gatekeeper. The first-moving gatekeeper incurs additional gatekeeping costs but benefits from there being fewer wrongdoers who choose a less scrupulous gatekeeper.[2] Anticipating more wrongdoers in the customer pool, the other gatekeeper may match its gatekeeping level to the one taken by the first gatekeeper. Ultimately, both gatekeepers would engage in the same level of gatekeeping, which is socially excessive. This phenomenon can be observed in the real world, where banks opt to leave Caribbean countries to de-risk instead of staying to profit from a larger market share (Mohammed 2022-11-11).

Gatekeeping can also impose costs on third parties. Using the example of de-risking in the anti-money laundering context, banks' de-risking can be understood as an extreme form of precaution-taking behavior that prevents banks from being punished for serving as money launderers. Unfortunately, individuals with risky profiles are already under- or unbanked. Precautionary de-risking further deprives them of financial services, reducing their financial inclusion. Evidence shows that lower financial inclusion is causally related to poverty, which can have spillover effects on society. These costs of poverty to society

2. Here I assume that gatekeeping imposes higher costs on wrongdoers than on innocent customers. Therefore, enhancing the level of gatekeeping will divert more wrongdoers than counterparties.

are not internalized by the gatekeepers (banks) or the rejected customers (counterparties). Therefore, bargaining cannot solve this problem. In such cases, increasing deterrence through heightened gatekeeping can be socially costly. Similar examples can be found in the contexts of corporate employers and online platforms.

This chapter makes three contributions. First, it links gatekeeper liability with victim precaution and allows for insights to be borrowed between these areas. Second, it distinguishes between two types of required acts of gatekeepers—interdicting and reporting—and identifies how different enforcement strategies against first-order wrongdoers affect gatekeeping incentives and aggregate deterrence. Finally, based on this comparison, the chapter provides novel considerations for designing gatekeeper liability regimes in various contexts.

This chapter presents a more detailed liability design for gatekeeper candidates selected from Kraakman's framework. When a candidate is selected, I suggest first determining the required gatekeeping behavior, considering the enforcement strategies and available resources. For instance, when misconduct is more likely to be deterred by escalating sanctions, reporting rather than interdicting should be prioritized. Additionally, a liability regime for the gatekeeper can be designed. The choice between act-based liability, negligence, or strict liability should consider the market structure of the gatekeeper (monopolistic or competitive) and the cost-benefit analysis of duplicate gatekeeping processes. Strict liability is preferred to reduce the cost of duplicate gatekeeping, while act-based liability or the negligence rule can address the problem of harm displacement among gatekeepers. Lastly, it is important to consider the third-party costs incurred by the gatekeeping process and adjust the gatekeeper liability regime accordingly. If gatekeeping imposes externalities on third parties, the amount of liability should be adjusted, or the liability regime should shift towards a negligence-based rule to ensure the adoption of the socially optimal level of gatekeeping.

This chapter is divided into four sections. Section 3.2 provides a review of the literature on gatekeeper liability. Section 3.3 compares gatekeepers and precaution-taking victims,

drawing insights from the literature on victim precaution. Section 3.4 addresses contested questions regarding gatekeeper liability, including the required gatekeeping act, the choice of liability regime, and the optimal level of liability. Section 3.5 applies the analysis to three specific gatekeepers — corporate employers, banks, and online platforms — and proposes reforms to address social issues potentially associated with gatekeeper liability.

3.2 Three Questions about Gatekeeper Liability

3.2.1 Why do we have gatekeeper liability?

Gatekeepers are held liable for the misconduct of other parties when they have the ability to control, interdict, or withdraw assistance to such misconduct. This misconduct often violates the law and is subject to law enforcement. Given the presence of law enforcement, one may question the rationale behind holding gatekeepers accountable for the wrongdoing of others. Law-and-economics scholars justify gatekeeper liability by arguing that gatekeepers can enhance deterrence at a lower cost (see e.g., Kraakman 1986). According to Becker (1968), rational offenders engage in misconduct when the gains outweigh the costs associated with such behavior. Other factors being equal, increasing the expected cost would make misconduct less attractive and deter potential offenders (p. 177). Hence, deterrence relies on the expected cost to potential offenders, which is a combination of the imposed sanction and the detection probability. Since rational, risk-neutral offenders only consider the expected cost and benefit, the state can enhance deterrence by increasing either the sanction or the detection probability. Additionally, the state can maintain deterrence by adjusting one component to compensate for changes in the other. This Beckerian perspective on deterrence through law enforcement suggests that a cost-minimizing state should maximize the costless monetary sanction to save on detection costs (p. 183). However, in the real world, excessively high fines do not guarantee strong deterrence due to the judgment-proof problem (see e.g.,

Shavell 1986). The effectiveness of monetary sanctions in deterrence is limited by offenders' limited financial resources, necessitating more costly non-monetary sanctions (Shavell 1987c). The higher cost of non-monetary sanctions, combined with the need for marginal deterrence, reduces the attractiveness of severe sanctions (Stigler 1970; Friedman and Sjostrom 1993). Therefore, increasing the detection probability becomes a relevant consideration for law enforcement agencies. However, this raises enforcement costs for the state since deterrence can no longer be achieved cheaply as in theory. Instead, the state should explore alternative measures to reduce enforcement costs while pursuing deterrence.

Gatekeepers are introduced to mitigate the problem of costly deterrence. They can help reduce enforcement costs by detecting, preventing, and deterring misconduct. Compared to law enforcement agencies, gatekeepers are in a position which enables them to observe and identify misconduct at an earlier stage (Gadinis and Mangels 2016, p. 875).[3] Moreover, gatekeepers are engaged in business relationships in which they can withdraw services from potential wrongdoers to prevent harm and provide deterrence (Kraakman 1986, p. 63). Furthermore, gatekeepers may already be incentivized to do so when their reputation is at stake (p. 61).

Take auditors — a common gatekeeper in the initial public offering (IPO) context — for example. Auditors are required to verify the financial statements prepared by issuers. They have access to such information before an IPO and possess the expertise to detect potential misrepresentations within those statements. As a result, they are more capable of detecting and preventing securities fraud *ex ante* than are law enforcement agencies *ex post*. Moreover, they operate in the context of a business relationship wherein they provide a crucial service to issuers seeking access to the capital market. This positions them to deter ill-intentioned issuers by threatening to withdraw their services. Furthermore, auditors have the incentive to do so. When an auditor firm fails to detect fraud in their client's financial statements,

3. Paradoxically, the ability to gatekeep create an information asymmetry between gatekeepers and the state, allowing them to evade enforcement actions by exploiting the information gap. See Manns (2006)

investors may penalize their future clients by charging a higher risk premium for their stocks. Consequently, potential clients may be less inclined to hire that auditor. Conversely, a high-performing auditor can leverage its reputation to benefit its clients, enabling them to secure capital at a lower cost. This lower cost translates into higher service fees collected by the auditor. Hence, gatekeepers are naturally well-suited to enforce the law.

3.2.2 How do we design gatekeeper liability regime

In the last section, Kraakman's framework outlines the conditions necessary for effective gatekeeper candidates, but does not provide guidance on how to design gatekeeper liability regimes. Specifically, Kraakman aids in filtering potential candidates for gatekeeping misconduct, but says little about how to hold different gatekeepers liable. Should we hold them strictly liable or apply the negligence rule? Which liability regime best suits each gatekeeper in their specific context? He does not delve into the how question.

Some literature addresses these issues, focusing on specific contexts and problems. For instance, after the Enron scandal, scholars extensively discussed how to enhance gatekeeper liability to prevent similar incidents. In this specific context, they focused on the issue of conflicted interests between the gatekeeper as an enforcer and as a service provider to the entity being monitored. Coffee (2002), for instance, emphasizes that the gatekeeper should be held responsible for the scandal because it failed to intervene against the illegal practices. He further explores why gatekeeper liability did not effectively function in the Enron case. Coffee (2001) argues that the failure of gatekeeper liability resulted from intensified conflicts of interest between the gatekeeper as a law enforcer and as a service provider to the party being monitored. The business relationship that enables the gatekeeper to closely monitor the client and prevent misconduct also creates a conflict of interest. The once advantageous close relationship between the gatekeeper as a trusted law enforcer and a for-profit service provider now becomes problematic.

In response to conflicts of interest, Coffee (2004) advocates for a "stricter" strict liability regime, where gatekeepers who breach their duties would face penalties based on their annual revenue. By imposing penalties tied to annual revenue, gatekeepers' incentive to comply aligns with their financial interest in serving clients, thereby mitigating conflicts of interest. Similarly, Partnoy (2004) argues that strict liability is preferable to address the challenges faced by gatekeepers, but for different reasons. He posits that gatekeepers are investing too much in meeting the requirements of due-diligence-based defenses. Shifting to strict liability would enable gatekeepers to efficiently monitor their clients (Partnoy 2001). Radical proposals also exist, such as Choi's suggestion of utilizing market mechanisms where gatekeepers can signal their credibility by selecting the level of enforcement severity (Choi 1997-1998). By opting for greater enforcement severity, a gatekeeper can gain more from the increased credibility but face harsher threats, aligning their financial benefit with the cost of non-compliance.

The discussion also raises other considerations regarding specific gatekeepers. Specifically, imposing strict liability on capital market gatekeepers for securities misconduct can give rise to a free-rider problem among collaborative gatekeepers (Tuch 2010). When multiple gatekeepers collaborate on a project, such as initial public offerings (IPOs), they share joint liability for the issuer's misconduct. However, they have the incentive to withhold their efforts and benefit from reduced liability resulting from the efforts of other gatekeepers (see also Darrow 2014). In response to proposals of strict liability, Hamdani (2003) reminds us that if gatekeepers cannot perfectly distinguish bad clients from good, the increased price charged for their services due to heightened liability would drive away good customers, leading to the adverse-selection problem and potentially destabilizing the market.

3.2.3 The gaps in the literature: what must gatekeepers do?

The focus on the problems faced by gatekeepers in the capital market, motivated by the Enron scandal, may overshadow other gatekeepers in other contexts that also require attention. For instance, there are proposals to hold online shopping platforms liable for defective products sold (Martin 2021). Also, retailers are expected to enforce laws against their suppliers (Rogers 2010). Furthermore, leading firms in certain industries, such as big tech firms or petroleum companies, are implicitly tasked with enforcing the law against their clients and thus becoming *de facto* gatekeepers (Loo 2020). This trend extends beyond financially resourceful companies capable of preventing misconduct by their counterparties. In the realm of data protection, there are arguments for imposing gatekeeper liability on individual users for the failures or leaks of the parties with whom they share information (Hu 2021).

These recent developments and the exploration of other gatekeeper contexts highlight the inadequacy of previous discussions. The challenges faced by capital market gatekeepers may differ from those encountered by other gatekeepers. For example, banks providing basic deposit services, compared to investment banks (see Tuch 2012), may not face significant conflicts of interest, as the revenue from such services is standardized among customers and minimal in a low-interest era. Likewise, other contexts may not involve multiple collaborative gatekeepers. Given these different conditions, the aforementioned recommendations may not be applicable to other gatekeeper contexts. Consequently, there is a need to develop a more comprehensive yet nuanced framework that can guide the design of gatekeeper liability.

Additionally, the emergence of various gatekeeper candidates unveils a previously overlooked aspect of gatekeeper liability: what actions must gatekeepers take? The diversity of gatekeepers serves as a reminder that they may not be required to respond to misconduct in the same manner. Some gatekeepers may be tasked with interdicting wrongdoers by avoiding relationships with or disciplining them, such as auditors. In contrast, others may be mandated to report them, like banks. Corporate employers may interdict a risky applicant

ex ante and are required to report the employee's misconduct *ex post*. Imposing gatekeeper liability is not the end of the analysis but rather a beginning with detailed gatekeeper liability designs once their gatekeeper status is confirmed.

This chapter aims to provide a detailed framework that can offer guidance for tailoring gatekeeper liability regimes to different gatekeepers. The design of liability regimes will focus on three dimensions: the required gatekeeping act, the choice of liability regimes, and the determination of penalties. By comparing gatekeepers with precaution-taking victims, this chapter explores the relationship between law enforcement against primary wrongdoers and gatekeeping incentives. It also examines how the gatekeeper's market structure influences liability design. Through this analysis, the chapter intends to outline relevant factors for designing and customizing gatekeeper liability regimes (see Section 3.3) and provide applications to different gatekeeper contexts (see Section 3.4).

3.3 Gatekeepers as Precaution Takers

In this section, I construct arguments based on the analogy of victim precaution. Gatekeepers are similar to victims in that both suffer disutility from wrongdoers' misconduct. While victims endure direct harm imposed by another's misconduct, gatekeepers encounter indirect harm through the sanctions imposed for gatekeeping failures. Consequently, both have incentives to take precautions to minimize the adverse impacts done by wrongdoers. To this end, gatekeepers must perform certain required tasks, such as interdicting or reporting misconduct, to avoid liability. On top of that, gatekeepers might want to reduce the chance of serving wrongdoers to minimize the risk. In this regard, their calculation of risk is akin to that of potential victims.

The incentive of potential victims to take precautions is imperfect in terms of social welfare. Potential victims take precautions in response to misconduct, which is sensitive to the intensity of law enforcement. Potential victims also take precautions for their own

private deterrence benefit and neglect the cost incurred by non-wrongdoer third parties. Consequently, their precaution-taking is often socially suboptimal in the absence of legal intervention. Gatekeepers, as precaution-takers, also exhibit these characteristics. Therefore, their gatekeeping incentives are potentially sensitive to the underlying law enforcement strategies and could be socially suboptimal.

In what follows, I will briefly review how potential victims behave in response to misconduct in Section 3.3.1 and show the resemblance and difference between gatekeepers and potential victims in Section 3.3.2. After clarifying their common characteristics, I will draw insights from existing victim-precaution literature and show the relationship between law enforcement and gatekeeping in Section 3.3.3 and the total cost of gatekeeping in Section 3.3.4.

3.3.1 How do victims take precaution?

Potential victims respond to harm by taking precautions, regardless of whether the harm they wish to avoid is caused by nature or humans. Such preventative measures can reduce the probability of its occurrence or mitigate the severity of the harm. For instance, people living near a river can build levees to reduce the risk of flood damage. Also, they may also elevate their homes to minimize potential flood-induced harm when the levees are breached. Potential victims similarly take precautions against harm inflicted by others by deterring wrongdoers or mitigating the resulting harm. This section investigates potential victims' precautionary behavior, emphasizing its relationship to law enforcement and misalignments between private and social incentives.

Precautions against misconduct

When addressing harm done by others, potential victims can focus on mitigating harmful outcomes, as described above. For example, an individual can choose to reduce the possibil-

ity of theft by installing locks or reduce exposure to harm by storing valuables in a bank safe deposit box. Potential victims can also take precautions against wrongdoers, altering their decision-making calculations. In other words, precaution-taking victims can deter wrongdoers by making crime less attractive. Applying the Beckerian model of criminals' rational calculation, a criminal decides to commit misconduct because the expected payoff outweighs the expected sanction, which is the multiple of the detection probability and the imposed sanction (Becker 1968). As a result, potential victims might change potential wrongdoers' chosen behavior by (1) reducing the gains from misconduct, (2) enhancing the probability of detection, and (3) imposing additional sanctions (see e.g., Cook 1986).

Everyday observations illustrate these precautionary directions taken by potential victims. Shops install security ink tags on merchandise to reduce the value of the goods stolen to thieves. They also install surveillance cameras in the stores and sensors at the exits to identify potential thieves. Some shop owners, when they find someone has stolen from them, publicly post photos of the suspects, inflicting reputational harm. These measures reduce the wrongdoer's gain from the misconduct, enhance the probability of detection, and impose sanctions on wrongdoers.

Relationship between law enforcement and private precautions

The harms potential victims wish to avoid is subject not only to their precautions but is also regulated by the actions of law enforcement agencies. Potential victims' precautionary decision-making do not operate in a vacuum but must consider the extent and effectiveness of law enforcement when determining the nature and extent of their precautions.

In the abstract, it is unsurprising that law enforcement and private precautions serve as substitutes for each other. In concrete terms, when law enforcement, whether private or public, is more effective, potential wrongdoers are deterred to a greater extent, lessening the need for private precautions. This, in turn, indirectly disincentivizes potential victims from

undertaking costly but less beneficial precautions.

This interplay between law enforcement and private precautions has been explored and explained in formal models of law enforcement. For instance, Professor Philipson and Judge Posner develope a model showing that this quasi-substitute relationship between law enforcement and private precaution would lead to a natural beyond-zero crime rate (Philipson and Posner 1996). According to the model, as law enforcement intensifies its efforts, private precaution retreats, so adjustments in private precautions, induced by more rigorous law enforcement, could result in a bounce back in the crime rate (p. 421). Noticing the effect of law enforcement on victim precaution, Hylton (1996) argues that optimal law enforcement should consider victims' precautionary efforts and be adjusted accordingly to induce victims' optimal precautions. This theory of a relationship between law enforcement intensity and victim precaution finds some supporting empirical evidence (see e.g., Vollaard and Koning 2009).

Nevertheless, the relationship can be more subtle when we regard law enforcement and private precaution as multi-dimensional. Namely, law enforcement and private precaution can be dissected into detection and sanction. While the aggregate investment in private precaution and law enforcement can substitute for one another, as shown above, each dimension can also complement the other (Ben-Shahar and Harel 1995). To elaborate, private investment in detection can be a substitute for its counterpart from law enforcement, but also complement the investment in sanctions. Likewise, private (public) investment in sanctions complements public (private) investment in detection.

Consider the following numerical example. Assume there are 100 wrongdoers and that both law enforcement and private detection have a 50% chance of detecting them. Further assume that these methods of detection work independently. Initially, wrongdoers would face a 50% chance of detection if victims make no efforts to detect them. Once victims invest in detection, the chance of detection becomes 75%, meaning an additional 25 wrongdoers

would be detected and face sanctions. If law enforcement raises its investment and improves the detection rate from 50% to 60%, the difference between investing and not investing in private detection shrinks from 25% to 20%. Given the cost is unchanged, private detection becomes less attractive to potential victims. Hence, they would be less likely to invest in detection. The improvement of law enforcement's detection lowers the level of private investment in detection. This is the substitute effect between direct law enforcement and private precautions.

When law enforcement enhances the sanction imposed on detected wrongdoers, the results are reversed. Assuming the imposed sanction is one year, then the expected sanctions faced by wrongdoers with and without private detection are nine and six months, respectively. That is, the investment in private detection brings additional deterrence benefits of three-month longer sanctions. Once the imposed sanction is lengthened from one to two years, the additional deterrence benefit increases as the gap between expected sanctions doubles. Therefore, the increased sanctions imposed by law enforcement encourage potential victims to invest more in private detection. From these scenarios, it can be observed that law enforcement and private precaution can complement each other, depending on the dimension. As a result, the decision-making of private precaution should be sensitive to the intensity and strategic focus of law enforcement.

Misaligned incentives between victims and society

Aside from the sensitivity of private precaution to law enforcement, it should be noted that private precaution itself is not necessarily socially optimal. On the contrary, absent legal intervention, it often deviates from the social optimum due to various misalignments of incentives. Potential victims seldom internalize the cost and benefit of other peer victims, third parties, and even wrongdoers. Hence, their precaution-taking calculus fails to internalize total social costs, leading to over-investment in precautions.

There is a collective-action problem that results in suboptimal precaution-taking. Take riverside levees, for example. A higher levee on one riverbank may lead to flooding on the other bank (see Song et al. 2018-03-30; Smith 2019-05-07). In such a case, the harm is not mitigated, only displaced. From a social perspective, which takes note of the degree rather than the distribution of harm, the cost of additional levee height may be unjustified as it does not mitigate, but only redirects, the floods. Worse, residents on the other bank may also choose to raise their levees in response, thus guarding against redirected floods. Ultimately, both sides may engage in a competition to raise their levees, leading merely to the displacement of flood waters.

Similarly, if potential victims take precautions that reduce the net gain from misconduct, wrongdoers will target other potential victims who have not taken such measures. Thus, the harm done is not eliminated, only shifted. From a social perspective, such private investments are socially inefficient because those so investing incur an additional precautionary cost without reducing social harm. This misalignment between private and social benefit results from the assumption that wrongdoers can observe the precautionary measures. Such observable precautions lead Clotfelter (1978) to conclude that potential victims would over-invest in observable precautions because they fail to internalize the social cost borne by other victims to which the misconduct is directed. However, Clotfelter's conclusion may not hold when wrongdoers' knowledge cannot be easily transferred among victim groups (Koo and Png 1994, p. 88) or the perverse signaling effect attracts wrongdoers by informing them of the value at risk via the amount of investment in observable precautions (Baumann and Friehe 2013). Accordingly, the harm-displacement effect of observable precautions and the private incentives to take them could be less severe than Clotfelter expects.

Unobservable precautionary measures could result in under-investment. When wrongdoers are aware that specific precautionary measures have been taken, but cannot directly observe who adopts them, they can only assume that every potential victim has some prob-

ability of taking them. This probabilistic estimation, by its nature, inevitably applies to potential victims who do and do not adopt such measures. Unable to distinguish those adopting from those not adopting such precautionary measures, wrongdoers are forced to reduce their level of misconduct based on the increased prior. Such unobservable precautions hence have a deterrent instead of a harm-displacement effect. The evidence of the use of a device such as a LoJack in automobiles also supports this argument (Ayres and Levitt 1998).

Unfortunately, such a deterrent benefit is not fully captured by those who adopt it because wrongdoers cannot tell who does. A potential victim who does not adopt such an unobservable precautionary measure is likely to benefit from others' precaution-taking. The effect is twofold: precaution-taking victims do not fully enjoy the benefits of their actions, while their inactive neighbors free-ride on their efforts. The free-rider problem discourages potential victims from making optimal efforts to take such precautions. In consequence, victims could under-invest in unobservable precautions (Clotfelter 1978). Nevertheless, while the argument holds in theory, it does not consider the possibility that victims may have private information about the value at risk and want to avoid attracting wrongdoers by signaling it through taking observable precautions. When that is the case, the choice between observable and unobservable precautions becomes endogenous to the property value (Baumann, Denter, and Friehe 2019).

Misaligned incentives to take precautionary measures can also come from the different components of private benefit and social welfare (Shavell 1991). Precaution-taking victims are motivated to mitigate or avoid the harm inflicted by wrongdoers. Therefore, their incentives to take precautions are positively related to the value at risk. However, such harm may not be equivalent to social harm. Take car theft, for example. A car owner may experience a total loss of her car. Society, nevertheless, only regards the car as transferred and still calculates the benefit of the usage of the car by people other than the victim-owner (see *infra* footnote 2 in Chapter 2). In brief, victims inflate the level of harm and are more

motivated to take precautions than society expects. The different calculations contribute to a misalignment between privately rational and socially optimal precaution-taking.

Beyond the difference between the gain in private and social welfare calculus, another source of misaligned precautionary incentives can be attributed to a subset of precautions. Potential victims sometimes take precautions that intentionally inflict harm on wrongdoers to deter them. Such harmful precautions can unintentionally impact non-wrongdoers. Non-wrongdoers can be counterparties that have transactional relationships with precaution-taking victims or unrelated third parties (see Chapter 2). For example, barbed wires could cut the fingers of burglars (wrongdoers), food-delivery persons (non-wrongdoer counterparties), or tourist passersby (third parties), and likewise, spring guns may injure intruders, suppliers, or neighbors. Both home-protection measures illustrate how precautions can be harmful. Unlike criminal gains, such precaution-inflicted harm should unequivocally be included in social welfare. Hence, it should be considered a social cost. Absent legal interventions, potential victims rarely internalize such costs borne by wrongdoers and non-wrongdoer third parties, except for counterparties who can bargain with them. For those wrongdoers and non-wrongdoer third parties, case law distinguishes privileged and unprivileged traps in the context of property-protection traps (see generally Dobbs, Hayden, and Bublick 2016, pp. 143–47). The former immunizes defendant-precaution-takers from compensating injured wrongdoers and third parties. Hence, potential victims could over-invest in privileged traps than the socially desirable level, as they do not internalize the social cost borne by wrongdoers and non-wrongdoer third parties.

3.3.2 Gatekeepers and precaution-taking victims compared

In the previous section, I summarized how victims behave by analyzing the relationship between their decisions and law enforcement, as well as the sources of misalignment between private and socially optimal precaution-taking decisions. Now, I turn to gatekeepers and

compare them with precaution-taking victims to show the similarity in their behavior. The comparison lays the bedrock for importing insights from the analysis above.

As potential victims suffer harm inflicted by wrongdoers, gatekeepers also suffer "harm" from wrongdoers' misconduct. However, unlike harm suffered by victims, which is direct and encompasses bodily injury or property loss, the harm suffered by gatekeepers results from the penalty imposed by law enforcement for their failure to honor their legal obligation. Moreover, the scale and conditions of harm suffered by gatekeepers for the gatekept's misconduct are prescribed by law and determined at the discretion of law enforcement and the courts.

To illustrate, imagine an employee in a chain store fighting with, and injuring, a customer. The injured customer is a victim who suffers bodily injury. The employer of that tortfeasor-employee is also legally liable for the customer's loss. However, the liability is financial and conditional. The employer does not directly suffer the same bodily injury as the customer but bears the financial liability of compensating the injured customer. The compensation is conditional and subject to what is prescribed in law and how much the tortfeasor-employee pays the employer. This financial liability often tracks the actual level of harm suffered by victims, but sometimes does not. Other gatekeepers may incur liability imposed by law enforcement agencies that do not determine the amount based on the real harm suffered by potential victims. For instance, banks are responsible for monitoring customers and their transactions to avoid facilitating money laundering and terrorist financing (see e.g. 12 U.S.C.A. §§ 1952–53, USA PATRIOT Act Sec. 326). While the direct harm a victim may suffer can be fear, property loss, bodily injury, or death caused by criminal or terrorist organizations, banks are merely fined. Moreover, the amount is often determined based on the conduct of banks as gatekeepers rather than the potential harm the failure can create (Financial Industry Regulatory Authority 2022-09-01, pp. 83–85).

Given the disparity in the scale, type, and condition of harm, gatekeepers share similar

precaution-taking behavior with potential victims. While potential victims take precautions in response to expected harm, gatekeepers also target the harm imposed by law enforcement. That is, they take precautions to avoid liability. As a result, what they are incentivized to do is directly linked to the conditions for imposing liability. When the liability is designed to be harm-based or result-based, as potential victims anticipate, gatekeepers may take multi-dimensional precautions to avoid the harmful results done by wrongdoers. First, they can invest in identifying and detecting potential misconduct or the gatekept who commits it. By enhancing the investment in detection, the gatekept faces higher expected sanctions and is more likely to be deterred by legal sanctions. With the ability to tell which gatekept is wrongful, gatekeepers can further decide to reduce the risk of harm posed by wrongful gatekept. With reduced exposure, harm could be mitigated, and the associated liability can be lower. Also, gatekeepers can impose sanctions on wrongful gatekept. Such sanctions can range from the simple withdrawal of engagement to real sanctions imposed by gatekeepers (Hill 2022-08-25).

Again, take the employer's liability, for example. A corporate employer who wants to avoid the liability of an employee's criminal act may first screen the job applicants and assign supervisors to monitor employees. When the employer is aware of potential misconduct risk before or during the employee's tenure, it may transfer the employee to a position with less customer exposure (Holzer and Stoll 2004, p. 40). In addition, the employer may discipline the employee by reducing wages, terminating their contract, and disclosing misconduct during background checks to potential future employers. All these mechanisms help the employer to reduce the likelihood of employee misconduct and, thus, their liability for the misconduct. When the law holds gatekeepers liable for the harmful result, as employers are in this context, gatekeepers will deploy all measures and allocate the resource in each dimension for their benefit.

In some cases, rather than holding gatekeepers liable for the result, the law holds them

liable for certain (in)actions. With specified requirements, the law may ask gatekeepers to act in a certain way, including inaction. For instance, banks are required to file reports on suspicious transactions. When gatekeepers meet the specified requirements, they are not held liable, regardless of the result. That is, even though the reported transactions are ultimately proved to be money laundering and inflict some harm, banks are not liable for facilitating such transactions. As liability is imposed conditional to the requirement of specific conduct, gatekeepers bear no liability for acts outside the scope of specified requirements. Hence, the law provides no incentive for those acts outside the scope.[4] As a result, banks required to file reports detailing suspicious transactions have no additional incentive to, say, share the information with fellow banks or track the transaction to verify its merit, even if such measures help reduce actual harm. Instead, they only take measures that can reduce their cost, including liability and compliance costs.

Of course, the analysis above does not claim that gatekeepers and potential victims are comparable in every aspect. Instead, the claim is modest and built on a focused comparison: potential victims and gatekeepers would both take precautions in response to human misconduct, with the qualification that the latter is subject to legal requirements and enforcement. There are, however, differences between their behavior if we look beyond precaution-taking behavior against wrongdoers. A significant difference is the source of harm. Since gatekeepers suffer "harm" only if liability is imposed and enforced, gatekeepers can target processes that impose and enforce liability. Potential gatekeepers may lobby legislation to remove their status as gatekeepers *ex ante*. They can also impede enforcement *ex post* by concealing their failures, challenging the disposition, or, worse, bribing relevant officials and fabricating evidence.

That said, comparing victims' and gatekeepers' precaution-taking behavior against wrong-

4. However, it is possible that, aside from liability, there are additional compliance cost associated with the liability. In such cases, liability may affect those acts that do not directly reduce liability but decrease future compliance costs.

doers can still bring insights from the victim-precaution literature to the literature on gatekeeper liability. No matter how gatekeepers direct legislative decisions or affect enforcement effectiveness, they will naturally have the incentive to take some precautions against the gatekept as long as their liability is associated with the gatekept's misconduct. Even if they choose the aforementioned ex-post measures, they will face non-zero harm (indirectly) from the gatekept's misconduct whenever the liability is not entirely eliminated. The non-zero harm should motivate them to take precautions, as potential victims do. As long as gatekeepers are incentivized to take precautions against wrongdoers, their behavior should be qualitatively comparable to precaution-taking victims. This result builds the foundation for borrowing insights from the literature on victim precaution in the following sections.

3.3.3 The relationship between law enforcement and gatekeeping

As mentioned above, gatekeeper liability is utilized to improve law enforcement by reducing the enforcement cost. This rhetoric implicitly suggests that gatekeeper liability is used to substitute or replace direct enforcement against wrongdoers when the latter is weak and ineffective. By treating gatekeeping as a substitute for law enforcement as substitutes, the discussion is inevitably tilted to the cost-effectiveness comparison between gatekeepers and law enforcement, such as direct regulation (see e.g., Wan 2008), as if the introduction of gatekeeper liability is an either-or question.

In the real world, direct enforcement does not retreat but coexists with gatekeeper liability, with both sometimes being robust, however. Therefore, the question should not be whether we should use gatekeeper liability or direct enforcement but how to allocate social resources between gatekeepers and direct enforcement. Since both direct enforcement and gatekeepers coexist, they would affect each other. As the literature argues that gatekeeping can substitute direct enforcement, the latter could also affect the former. In this Section, I will discuss the reverse effect of law enforcement on gatekeeping incentives. The discus-

sion will be segmented into two subsections. The first subsection investigates the effect of overall enforcement intensity on gatekeeping incentives. The second subsection unearths the potential relationship between enforcement focuses, such as detection and sanctions, and the dimensions of gatekeeping. Note that I focus on direct enforcement against gatekept wrongdoers and leave the enforcement against gatekeepers untouched here. Intuitively, the enforcement against gatekeepers can be understood as a determinant of gatekeeper liability that incentivizes gatekeeping efforts.

Effect of enforcement intensity on gatekeeping

Direct enforcement can share a negative relationship with gatekeeping incentives. The negative relationship comes from two sources. First, direct enforcement reduces the amount of misconduct and the associated gatekeeper liability. In addition, direct enforcement decreases the value of gatekeeping services. Both reasons depreciate the benefit of gatekeeping. As a result, as direct enforcement becomes more rigorous, we should expect that gatekeepers invest less in gatekeeping, on average.

As mentioned above, the harm faced by gatekeepers comes from their liability, which is conditional on the wrongful gatekept's misconduct. When wrongful gatekept are deterred, there are fewer wrongdoers. Fewer wrongdoers mean that gatekeepers' expected liability is lower because there are fewer instances where gatekeepers are held liable. Since the motivation to gatekeep is fueled by the expected liability, the lower the latter is, the less incentivized gatekeepers are to take precautions. In a model developed by Ewert and Wagehofer (2019), they find the same result that a strategic gatekeeper (auditor) could reduce their efforts in gatekeeping when law enforcement increases the intensity. The underlying rationale is that increasing enforcement intensity reduces the likelihood of misconduct and, therefore, liability faced by gatekeepers.

Rigorous enforcement also decreases the value of gatekeeping when gatekeepers monetize

their gatekeeping service. One of the values gatekeepers provide is information. Gatekeepers provide information regarding which counterparties are more trustworthy. Take auditors for example. They are hired to show the credibility of financial statements prepared by companies. With higher credibility, investors need not deduct the risk premium from the stock price. Hence, the company hiring a reputable auditor can enjoy a higher stock price, which translates into a lower cost of capital. With that, the companies can raise capital more efficiently.

The value provided by auditors to the companies is, in fact, dependent on the proportion of bad actors on the market. When many bad actors prepare incorrect or fraudulent financial statements, the risk reflected in the stock price by investors is higher. By discounting the stock price, investors hedge their risk of being defrauded. The existence of gatekeeping auditors assures investors of the credibility of financial statements by lending their reputational assets to their client companies. However, when bad actors are deterred by rigorous enforcement, and the risk of financial fraud is lower, the stock price discount suffered by the good companies could be minimal. The lower demand drives down the profit from such a service, disincentivizing gatekeeping efforts.

Both legal and market mechanisms above demonstrate the possibility that stronger law enforcement reduces the need for gatekeeping, namely lower liability and profit. In such cases, gatekeepers could behave like potential victims who withdraw from taking precautions in the face of rigorous law enforcement against wrongdoers.

Effect of enforcement strategies on gatekeeping

As analyzed by Ben-Shahar and Harel (1995), law enforcement strategies can affect victims' allocation of precautionary resources among different dimensions. Likewise, gatekeeping efforts can be directed by different law enforcement focuses. In this subsection, I will apply the Beckerian framework of enforcement. Similarly, I will segment gatekeeping into two main

acts required by law: interdicting and reporting.

In the analysis below, I will keep enforcement intensity equal. I do so to keep the analysis simple and avoid potential compounding effects from changing deterrence, which overlaps with the previous subsection. In other words, the expected sanction faced by wrongdoers remains the same, so the deterrent effect from enforcement intensity is unchanged and can be excluded. For that purpose, when detection doubles, the imposed sanction halves, and vice versa.

Increased sanctions I first consider situations where law enforcement focuses on imposed sanctions by, say, doubling penalties or the lengths of imprisonment to save the cost of detection. When law enforcement increases the imposed sanction, gatekeepers will shift their gatekeeping focus accordingly. In short, gatekeepers would invest more in reporting but reduce their investment when required to interdict wrongdoers.

Gatekeepers' efforts to report wrongdoers will increase because the marginal deterrence benefit is enhanced due to the increased sanctions. As in the case of potential victims, the benefit of gatekeepers' detecting and reporting is backed by sanctions imposed by law enforcement. As a result, when the sanctions increase, the deterrence benefit added by gatekeepers' reporting also increases. The increased marginal benefit with the unchanged marginal cost implies a higher level of effort. Hence, gatekeepers will be encouraged to invest more in detecting and reporting.

In comparison, the increase in sanctions imposed by law enforcement does not have the same effect on gatekeepers' interdiction. When interdicted, wrongdoers cannot complete their misconduct and will not be subject to sanctions imposed by law enforcement. For those wrongdoers who are already interdicted, the change in the scale of sanctions does not matter. Therefore, gatekeepers cannot harvest additional deterrence benefits from interdicting more wrongdoers as they do from reporting more wrongdoers.

The analysis above considers how increasing sanctions affects gatekeepers' efforts to in-

terdict and detect wrongdoers. However, the corresponding adjustment in law enforcement's detection may exacerbate or qualify the effects. As assumed above, law enforcement divests from detection when increasing imposed sanctions. The reduced detection can have an ambiguous effect on gatekeepers' reporting behavior but unambiguously dilute gatekeepers' interdicting efforts.

The lower investment by law enforcement in detection leaves more wrongdoers undetected. The greater number of undetected wrongdoers, on the one hand, increases the marginal benefit of private detection from gatekeepers if both law enforcement and gatekeepers work independently. Private detection plays a role in identifying those wrongdoers undetected by law enforcement. When the number of undetected wrongdoers by law enforcement increases, the value of private detection also increases.

However, more undetected wrongdoers also lead to a lower detection rate of gatekeeper liability, on the other hand. Here, I assume that the finding of wrongdoers also leads to potential gatekeeper liability, as it can easily be traced back to the gatekeeper associated with the misconduct and the wrongdoer. Hence, a lower detection rate also implies a lower expected gatekeeper liability, which disincentivizes gatekeepers from making the same amount of effort. Whether gatekeepers would invest more in reporting depends on which effect — the increased marginal deterrence benefit from detection and the decreased marginal deterrence benefit of lower expected liability — dominates. Ultimately, the comparison hinges on the marginal deterrence effect on wrongdoers or gatekeepers, which further depends on the relative scale of sanctions imposed on both parties. When the sanctions imposed on wrongdoers are significantly higher than those on gatekeepers, the increase in deterrence from private detection should outweigh the dilution of gatekeeping incentive from a lower expected gatekeeper liability.

The case is more straightforward for gatekeepers' interdicting incentives, as increased sanctions should not affect interdiction. However, the effect of lower liability due to lower

detection still applies. In aggregate, gatekeepers' interdicting incentives are not encouraged by aggravated sanctions but diluted by a lower expected liability. Therefore, we should expect a lower gatekeeping incentive for interdicting wrongdoers.

To see the arguments above more clearly, let me provide a simple model similar to the in 3.3.1. Assume that the chance of a wrongdoer being detected by law enforcement and the gatekeeper is p and q, respectively. Once detected, sanctions s would be imposed on the wrongdoers, and liability l will be imposed on the gatekeeper for failing to interdict or report. Further assume that the benchmark case is that p_0 and s_0 and the gatekeeper's effort is denoted by q_0. Now, law enforcement increases its sanction from s_0 to s_1 with associated changes in p from p_0 to p_1. By assumption, $p_0 s_0 = p_1 s_1$ because the enforcement intensity and the associated deterrent effect from law enforcement remain unchanged. In that setting, the effect of that change in enforcement focus on gatekeepers' interdicting decisions is zero if those interdicted will not be punished when they are interdicted. In contrast, the effect on the gatekeeper's reporting can be deconstructed into two parts. On the one hand, the marginal deterrence against wrongdoers from the gatekeeper's reporting increases from $(1 - p_0)s_0$ to $(1-p_1)s_1$, which is equivalent to s_0 to s_1. On the other hand, the expected liability faced by the gatekeeper reduces from $p_0 l$ to $p_1 l$. Ultimately, whether the gatekeeper will increase the investment depends on whether the marginal deterrence benefit to the wrongdoer from s_0 to s_1 outweighs the reduced expected liability from $p_0 l$ to $p_1 l$. If the wrongdoer is very sensitive to the sanction, then the condition is more likely to be met. If, otherwise, the wrongdoer is insensitive to the sanction, or the liability imposed on the gatekeeper (l) is significantly high, then the direction of the gatekeeper's behavior change can be the opposite.

Enhanced detection When law enforcement instead shifts its focus and resources to enhancing detection, the effect on gatekeeping behavior will also change. Intuitively, the effect should be the opposite of what is described above, as we can consider the case of enhanced detection as the case of "decreased" sanctions. Namely, we can reuse the formal

example above with the enforcement policy shifts from $p_1 s_1$ to $p_0 s_0$.

Motivated by increased expected liability, gatekeepers should increase their investment in interdicting wrongdoers. Again, I assume the detection of wrongdoers would also lead to the detection of their gatekeepers' failure. As a result, enhanced detection of wrongdoers will also increase the expected liability of gatekeepers, given other things equal. Unlike increased sanctions, enhanced detection encourages gatekeepers to interdict more because doing so can save them from liability. By interdicting wrongdoers, they can offset the number of more detected wrongdoers and the associated liability. To see this, let's assume that the detection rate doubles with imposed sanctions halved. Now, other things being equal, the chance of gatekeeper liability also doubles. Hence, the marginal benefit of interdicting an additional wrongdoer also doubles. With the marginal cost unchanged, the increased marginal benefit implies a higher level of interdiction.

The effect on gatekeepers' reporting warrants a close inspection. Similarly, the increased chance of liability increases the marginal benefit of detecting and reporting. Investing in reporting now saves more by avoiding higher liability, $p_0 l$ compared to $p_1 l$. However, at the same time, the marginal deterrence benefit of gatekeepers' detecting also decreases, given a higher level of detection from law enforcement. That is, the marginal deterrence provided by the gatekeeper's reporting decreases from s_1 to s_0. The same ambiguous effect can again be found here as it is the mirrored case of the case above. The reasoning remains the same: it ultimately depends on the effect of changing gatekeeper liability and the effect of deterrence against wrongdoers. Whichever effect dominates, the gatekeeper will choose the direction. One thing certain is that the effect is less significant than the change in the gatekeeper's interdicting behavior when law enforcement enhances detection.

Summary: sensitive gatekeeping incentives

The analysis above unearths a more nuanced view of how direct enforcement affects gatekeeping incentives. Rigorous direct enforcement deters more wrongdoers, which dilutes gatekeeping incentives due to both reduced liability and reduced value provided by gatekeeping. This result showcases that direct enforcement and gatekeeping can be substitutes, as shown in the context of victim precaution.

However, this substitutive effect can be qualified or exacerbated by the complementary effect of specific gatekeeping and direct enforcement dimensions. In other words, the focus of enforcement resources can have different effects on different dimensions of gatekeeping. The analysis above unearths that when direct enforcement increases its sanctions on wrongdoers, gatekeepers could invest more in reporting but would certainly divest from interdicting. This finding results from the change in the marginal deterrence benefit from gatekeeping and the change in marginal liability due to the shift of enforcement focus. The findings are further summarized in Table 3.1 below.

		Required Gatekeeping Conduct	
		Interdicting	Reporting
Direct Enforcement Focus	Increased sanctions	–	+/–
	Enhanced detection	++	+/–
Direct Enforcement Intensity		–	

Table 3.1: The effect of direct enforcement on gatekeeping

The summarized findings can be promising or problematic. From an optimistic standpoint, socially inadequate gatekeeping, as most literature suggests, can be improved and incentivized by redirecting enforcement focus at zero cost. However, redirecting the focus could be a double-edged sword when law enforcement is unaware of the potential effect on gatekeeping incentives. In particular, it can exacerbate socially excessive gatekeeping, as indicated in the next section, or discourage already socially inadequate gatekeeping. In sum-

mary, this analysis serves as a reminder to law enforcement of the potential impact of its enforcement strategies or policies on gatekeepers.

3.3.4 Misaligned gatekeeping incentives

Gatekeepers' incentives to engage in gatekeeping can be misaligned with society's interests. In order for private gatekeeping incentives to align with social ones, decision-makers, namely gatekeepers, should consider the same factors as society. Unfortunately, there are several stakeholders whose interests are not taken into account in gatekeepers' calculations. As a result, gatekeepers' incentives often deviate from the social optimum.

Below, I will enumerate three sources of misalignment between private and social gatekeeping incentives. They result from failing to internalize the costs borne by the gatekept counterparties, other gatekeepers, and society.

Costs borne by counterparties

As precautions can impose costs on innocent counterparties, gatekeeping can incur costs that are borne by counterparties. These costs include expenses related to the gatekeeping process and the cost of errors, particularly false positives. For example, in the context of initial public offerings (IPOs), firms seek to raise capital from the capital market by issuing stocks. To do so, they must hire underwriters and auditors, both of whom are legally liable when issuing firms engage in securities misconduct. In this scenario, underwriters and auditors act as gatekeepers, while the issuing firms are counterparties. During the IPO process, gatekeepers may conduct various verification procedures, such as reviewing documents, conducting on-site due diligence, and preparing prospectuses. All of these tasks require cooperation from the counterparties. Despite these verification efforts, gatekeepers may be unable to ensure full compliance with applicable laws on the part of the counterparty. Errors may occur, resulting in innocent gatekept parties being charged a higher price due to uncertainty or,

worse, being mistakenly rejected by the gatekeeper.

Adverse selection Gatekeepers and counterparties are typically already in a contractual relationship, which should enable them to share these costs and reach an efficient outcome, as argued by Coase (1960) Ideally, when both the gatekeeper and the gatekept negotiate, they can allocate the costs and risks accordingly. If the cost of gatekeeping exceeds the reduced liability, both parties would agree to lower the intensity of gatekeeping by sharing the liability risk. In other words, the gatekept party can pay more to the gatekeeper to subsidize the increased liability in exchange for reduced gatekeeping. Such Coasian bargaining prevents inefficient over-gatekeeping. However, this type of bargaining may be hindered by imperfect information and market structure. Moreover, legal liability can exacerbate the problem.

When there is a lack of perfect information, the adverse selection problem can hinder optimal bargaining. Hamdani (2003) demonstrates that imposing liability on gatekeepers may increase the price, driving away good gatekept parties and disrupting the market. His argument is primarily based on the assumption that gatekeepers may have imperfect information to distinguish bad actors from good ones, and wrongdoers may benefit more from their actions compared to law-abiding clients, along with the costly nature of ex-post cost-shifting (pp. 74–75). Hamdani's findings support the idea that imposing liability on imperfect gatekeepers may have undesirable outcomes even with Coasian bargaining.

One could argue that Hamdani's assumptions may deviate from reality, as we expect liability to be imposed on capable gatekeepers who can identify bad actors. However, when considering the cost of gatekeeping rather than error judgment, the adverse selection problem may still persist and prevent optimal bargaining. To illustrate this, let's consider a simple example where a gatekeeper can choose between low and high levels of scrutiny. Unsurprisingly, high scrutiny costs more than low scrutiny for both the gatekeeper and the gatekept party. Let's further assume that high scrutiny is inefficient as it incurs more cost than the harm that can be prevented by interdicting wrongdoers. Without bargaining, the gatekeeper

may not internalize the cost borne by the gatekept party, leading to an over-investment in gatekeeping. What if bargaining is allowed and occurs? Ideally, the law-abiding party may offer a higher price to compensate for the increased liability in exchange for reduced scrutiny. The bad actor would move in the same direction but to a greater extent. Apart from the cost of gatekeeping, the bad actor can benefit from a higher expected illicit gain resulting from reduced scrutiny. In fact, it could be in the bad actor's interest to bargain for no scrutiny. If this is the case, it would be undesirable as the good actor is subjected to low scrutiny, which provides no benefit, while the bad actor escapes gatekeeping and inflicts social harm. The outcome could be worse than disallowing bargaining and subjecting every gatekept party to scrutiny.

To provide a more concrete argument, let's introduce some numerical values. Let's assume that the auditor has two levels of scrutiny: low and high, with different costs and accuracy. We'll also assume that there are two types of firms: good and bad, with an equal probability. Each good firm gains $500 from going public, while each bad firm inflicts social harm of $800 and gains $1,000 through fraudulently acquiring funds from investors. When the auditor adopts a low level of scrutiny, it costs the gatekeeper $50 and each firm $60, with a 50% chance of identifying bad actors. On the other hand, high scrutiny costs the auditor $110 and the firm $110, with a 75% chance of identifying bad firms.

First, let's consider a universal standard of scrutiny without individualized arrangements. The socially optimal standard would be low scrutiny. However, it is in the gatekeeper's interest to adopt a high level of scrutiny because the reduced liability justifies the additional cost borne by the gatekeeper.

If bargaining is allowed, the gatekept party can bargain with the gatekeeper for a tailored level of scrutiny. In exchange, the gatekept party should compensate for the additional cost borne by the gatekeeper, which is the associated expected liability minus the savings on the gatekeeping cost. The gatekeeper can provide a price list, as shown in Table 3. By comparing

	No Scrutiny	Low Scutiny	High Scrutiny
Gatekeeper's Payoff			
Gatekeeping (A1)	0	-100	-220
Expected Liability(A2)	-800	-400	-200
Total Payoff (A3)	-800	-500	-420
Good Firm's Payoff			
Gatekeeping (B1)	0	-60	-110
Going Public (B2)	500	500	500
Total Payoff (B3)	500	440	390
Bad Firm's Payoff			
Gatekeeping (C1)	0	-60	-110
Going Public (C2)	1000	500	250
Total Payoff (C3)	1000	440	140
Social Welfare (A3+B3-C1)	-300	-120	-140

Table 3.2: Payoffs of the gatekeeper (for **two** firms) and both firms (no bargaining)

the payoffs of the firms and the offered price, we can see that the good firm would choose low scrutiny,[5]

while the bad firm chooses no scrutiny. However, this arrangement leads to a social welfare of –$410, which is worse than universally low or high scrutiny.

From this numerical example, we observe that bargaining does not improve social welfare. In fact, social welfare deteriorates in this example due to the adverse selection problem. Good firms, who do not require scrutiny, opt for scrutiny to avoid additional payment for the gatekeeper's expected liability. On the other hand, bad firms, who we expect to be subject to higher scrutiny, escape it through bargaining.

To address this issue of corruptive bargaining by bad actors, some proposals suggest enhancing gatekeeper liability, such as Coffee (2004). However, doing so could exacerbate the adverse selection problem. Let's assume that the liability is 20% higher than the actual harm, as shown in Table 3.4. With higher expected liability, the gatekeeper and the good firm are induced to engage in socially excessive gatekeeping, while the bad firm can still avoid scrutiny by paying for no scrutiny. Other factors such as judicial hindsight bias or

5. The gatekeeper can design a mechanism to separate good firms from bad ones by exempting those firms optting in for high scrutiny from paying the expected liability. Even if the gatekeeper does so, the bad firm would not choose high scrutiny. As a result, the price for high scrutiny would be lower, and the good firms are induced to enroll in high scrutiny. If so, the problem is further worse. The total gatekeeping cost become $220, and social welfare decreases to -$520.

errors can further worsen the distortion.

	No Scrutiny	Low Scutiny	High Scrutiny
Gatekeeper's Price (per firm)			
Gatekeeping	0	-50	-110
Expected Liability	-400	-200	-100
Price (A)	400	250	210
Good Firm's Payoff			
Gatekeeping cost (B1)	0	-60	-110
Going Public (B2)	500	500	500
Total Surplus (B2+B1-A)	100	190	180
Bad Firm's Payoff			
Gatekeeping cost (C1)	0	-60	-110
Going Public (C2)	1000	500	250
Total Surplus (C2+C1-A)	600	190	-70

Table 3.3: Gatekeeper's price and firms' surpluses

	No Scrutiny	Low Scutiny	High Scrutiny
Gatekeeper's Price (per firm)			
Gatekeeping	0	-50	-110
Expected Liability	-480	-240	-120
Price (A)	480	290	230
Good Firm's Payoff			
Gatekeeping cost (B1)	0	-60	-110
Going Public (B2)	500	500	500
Total Surplus (B2+B1-A)	20	150	160
Bad Firm's Payoff			
Gatekeeping cost (C1)	0	-60	-110
Going Public (C2)	1000	500	250
Total Surplus (C2+C1-A)	520	150	-90

Table 3.4: Gatekeeper's price and firms' surpluses (enhanced liabiilty)

In conclusion, bargaining might not necessarily solve the problem of socially excessive gatekeeping. Due to the adverse selection problem, the cost of gatekeeping may inefficiently fall on law-abiding gatekept parties, making scrutiny meaningless. Meanwhile, those who society regards as needing scrutiny may escape it through bargaining.

Market structure Apart from adverse selection, bargaining may face challenges due to the market structure. Market structure can prevent bargaining through price caps, such as capping the auditor fee at $240. In other cases, legal rules can hinder bargaining, such as banks being obligated not to discriminate based on race, ethnicity, national origin, or

religion in some transactions(see generally Levitin 2018, pp. 487–88). These factors limit the flexibility of individualized arrangements.

Even if bargaining is allowed, it can be costly to establish tailored contracts. For example, a gatekeeping service provider serving a large number of gatekept clients, like an airline company verifying travel documents (see McDonnell 2002-04-06) for hundreds of passengers per week,[6] would find it challenging to individually bargain with each passenger regarding the level of scrutiny and associated payment for liability.

Fortunately, market competition can solve the problem of costly bargaining. When multiple gatekeepers exist, clients can choose among them based on their specific needs. For example, during the Covid-19 pandemic, each country stipulated different entry rules that frequently changed. Passengers began purchasing tickets from national airlines of the destination due to their superior understanding of the rules, reducing the risk of erroneous denial.

However, the market may not always be competitive. Take one of gatekeepers of corporate misconduct, corporate executives, for example. While their incentives to monitor are qualified by incentives contracts, it is found that corporations have hard times attracting competent lawyers without such contracts (Morse, Wang, and Wu 2016). This example shows that market forces can work against desirable gatekeeping. Additionally, monopolistic gatekeepers and prohibitive bargaining costs can limit the choices of gatekept clients, forcing them to accept the arrangements preferred by the gatekeeper. For instance, when a single carrier monopolizes a flight to a destination, passengers have no option but to accept the gatekeeping intensity and associated price without the possibility of bargaining for adjusted scrutiny.

Recent news can illustrate the suboptimality resulting from a lack of market competition.

6. For example, the route from Taipei (TPE) to Chicago O'Hare (ORD) operated by Eva Air has carry 3,787 passengers in April 2023. Approximately, near 1,000 per week. Data available at Civil Aeronautics Administration (2023).

In a previous news article regarding Google, the father's Google account was banned because he sent his son's nude photo to a doctor for medical advice (Hill 2022-08-25). The photo was identified as child-abusive content and illegal, resulting in the father losing access to his photos, emails, and Google Fi phone number. The father seemed unable to negotiate with Google for a thorough and accurate review, and alternatives to Google may not have been readily available.

Costs incurred when other gatekeepers exist

In real life, it is common to see multiple homogeneous gatekeepers competing to provide the same service. For instance, there are Big-Four auditor firms offering auditing services, and multiple law firms assist in issuing securities. When multiple homogeneous gatekeepers exist, gatekept parties can avoid the problems associated with bargaining by seeking services from other gatekeepers. However, the market of multiple gatekeepers can introduce other costs, namely redundant gatekeeping and harm displacement.

Before diving into the costs related to multiple homogeneous gatekeepers, I should pause here and distinguish it from the similar term "multiple gatekeepers" used by Tuch (2010). In Tuch's article, he focuses on multiple "collaborative" gatekeepers. Namely, multiple gatekeepers work collaboratively to monitor and stop the same client's misconduct, and each is responsible for partial monitoring. For instance, in the IPO context, lawyers are responsible for legal due diligence, and auditors focus on financial statements. Both gatekeepers monitor the same client (issuer) and ensure their client complies with the requirement, but each limits their focus on different issues. In sum, multiple gatekeepers in Tuch's paper focus on the same project of ensuring the client's compliance with different focuses and divisions of labor. As they work on the same project, gatekeepers may want to shirk and free-ride other gatekeepers' efforts on some overlapped areas, resulting in the free-rider problem (pp. 1624–26).

In comparison, the multiple homogeneous gatekeepers are competitive rather than collaborative. Multiple competitive gatekeepers are responsible for the same type of monitorship and compete with each other for law-abiding clients. As a result, there is no qualitative difference in what they do. For example, all auditors are responsible for verifying financial statements prepared by companies. Also, they compete with each other for clients. Once the service-providing gatekeeper engages in the relationship with the client, it is the only gatekeeper responsible for monitoring the gatekept client. For each client at a time, there will be only one, rather than many, gatekeepers responsible for monitoring. In other words, the plurality of gatekeepers only exists in the competition stage. This single gatekeeper would be identified and held liable when the gatekept's wrongdoing is found. In this case, the gatekeeper is solely responsible, and there is no other gatekeeper to rely on. Therefore, the free-rider problem described in Tuch's paper would not exist among multiple competitive gatekeepers.

Rather than free-riding on other gatekeepers, the problem in the context of multiple competitive gatekeepers is redundant gatekeeping costs and the displacement of wrongdoers. When one gatekeeper identifies and interdicts a wrongdoer, the wrongdoer can seek the services of another gatekeeper for a second attempt. Without knowledge of the previous denial, the second gatekeeper may engage with the wrongdoer and incur the costs of gatekeeping again. This duplication of processes is socially less beneficial, as it is less likely to generate new information that overrules the existing information obtained by previous gatekeepers. Moreover, when gatekeepers could disagree, leading wrongdoers to evade scrutiny if they continue seeking different gatekeepers until one allows them through.

The possibility that an interdicted wrongdoer can seek a second attempt from other gatekeepers echoes the finding in victim precaution. Wrongdoers are not deterred. Instead, they are displaced. They can either observe a precautionary measure ex ante or experience it by testing the water *ex post* before being committed to a specific victim. If they judge that the

gain does not justify the cost, they would not necessarily give up but find another victim. Similarly, wrongdoers may not simply be deterred by a single gatekeeper's denial. Moreover, gatekeepers may share the same rationale with victims to take observable gatekeeping processes that scare away potential wrongdoers.

When gatekeepers invest resources to increase the intensity of gatekeeping to deter wrongdoers, they reduce their expected liability. However, this does not reduce social harm; instead, it displaces social harm to other gatekeepers who do not match the same level of intensity. Ultimately, the remaining gatekeepers may end up handling all the displaced wrongdoers or be forced to match the efforts of the over-precautious gatekeepers, resulting in an inefficient pooling equilibrium.

To see this, let's revisit the numerical example with a slight change. I will revise the assumption that bad firms can enjoy $1,400 when successfully going public, which would also make them accept high-scrutiny gatekeeping. The prices and payoffs are as follows. Since the level of harm remains at $800, the socially optimal level of scrutiny is still low. Let's assume two auditors are competing for clients. Initially, both auditors adopt low scrutiny. Now, Auditor A may be incentivized to move to high scrutiny. By doing so, Auditor A ensures that its service is less attractive to bad firms compared to Auditor B. As a result, Auditor A can offer a lower price due to a lower expected liability,[7] making it more appealing to good firms. In equilibrium, the price would be $110, and the effective price for good firms is $220, which is cheaper than Auditor B, who sticks to low scrutiny. With this move, all good firms contract with Auditor A, leaving bad firms with Auditor B. However, Auditor B's price is not sustainable because the expected liability becomes $400 since the pool of clients is now full of bad firms. Auditor B may be forced to adjust the price to reflect the expected liability or raise its scrutiny level to match Auditor A. If both auditors adopt high scrutiny,

7. The lower expected liability is due to the lack of attractiveness to bad firms, so the prior of bad firms is lower than 0.5. If the prior is lower than 0.45, then high scrutiny is sufficiently attractive to good firms for them to move.

an inefficient equilibrium emerges (see Table 3.2) Moreover, once they enter this equilibrium, there is no incentive for either auditor to move, as the first mover would be forced to take on all the bad firms.[8] This phenomenon mirrors how harm is displaced through victim precaution. The underlying rationale is that higher scrutiny imposes different costs on law-abiding gatekept parties and wrongdoers. As a result, increasing the level of scrutiny can divert more wrongdoers than law-abiding clients, making it a profitable and rational move.

	No Scrutiny	**Low Scutiny**	**High Scrutiny**
Gatekeeper's Price (per firm)			
Gatekeeping	0	-50	-110
Expected Liability	-400	-200	-100
Price (A)	400	250	210
Good Firm's Payoff			
Gatekeeping cost (B1)	0	-60	-110
Going Public (B2)	500	500	500
Total Surplus (B2+B1-A)	100	190	180
Bad Firm's Payoff			
Gatekeeping cost (C1)	0	-60	-110
Going Public (C2)	1400	700	350
Total Surplus (C2+C1-A)	1000	390	30

Table 3.5: Gatekeeper's price and firms' surpluses (higher bad firms' payoff)

The harm-displacement effect among gatekeepers can be more severe than among victims. First, gatekeepers are more homogeneous than victims, as they possess professional expertise in providing similar services. Moreover, the service they provide—be it endorsement or access to specific benefits—is qualitatively homogeneous to the gatekept parties, with only quantitative differences apparent. For example, auditors provide certification services for financial statements prepared by issuers, and the certification requirement remains consistent regardless of the firm. However, factors like the level of scrutiny, price, and associated reputational assets may vary. This qualitatively homogeneous service simplifies the transfer of knowledge among wrongdoers, resulting in more severe harm-displacement effects (Koo and Png 1994). It's worth noting that gatekeepers may specialize in specific areas, leading

8. Of course, the first mover can prepare itself to move by setting an optimal price. Nevertheless, it can only hold in theory. I doubt that in practice there will be gatekeeper who devoted to solely serving bad firms.

to market segmentation. In such cases, the extent of harm displacement depends on market competitiveness. If multiple gatekeepers are competing, harm displacement is likely to occur. Conversely, if there is only one monopolistic gatekeeper, harm may not be displaced, but the problem of uninternalized counterparties' costs can still arise.

Gatekeepers, unlike victims, do not worry about the perverse signaling effect. Victims may hesitate to take observable precautions because high precaution intensity might signal a high value at risk, thus attracting wrongdoers (Baumann and Friehe 2013). However, this effect does not occur among gatekeepers. The value at risk for gatekeepers is the expected liability, largely depending on the harm inflicted and probably known by the wrongdoer. In most cases, a higher expected liability does not lead to higher illicit gains for the wrongdoer using the gatekeeper's service. Therefore, the perverse signaling effect does not deter gatekeepers. Gatekeepers might invest in gatekeeping to maintain their reputational assets, and in this scenario, higher gatekeeping might signal higher reputational assets that wrongdoers can exploit. However, even if the signaling effect exists, gatekeepers are unlikely to refrain from such actions as they can also benefit from the signal when dealing with law-abiding firms. This benefit does not exist for victims, who cannot attract and benefit from non-wrongdoers by signaling the value at risk. Consequently, the perverse signaling effect that qualifies harm displacement is less likely among gatekeepers.

In summary, when multiple homogeneous gatekeepers compete to provide services to clients, they incur redundant gatekeeping costs due to duplicative scrutiny and increase the likelihood of wrongdoers inflicting social harm. When heightened gatekeeping imposes greater cost on wrongdoers than on law-abiding clients, gatekeepers may increase their gatekeeping intensity, resulting in socially excessive gatekeeping. However, these problems can be mitigated through altering liability regimes, which will be discussed in more detail in Section 3.4.

Cost on third parties

Gatekeeping can impose costs on third parties, costs which neither gatekeepers nor counterparties internalize. These third-party costs are comparable to car alarms disturbing neighbors or traps harming pedestrians in victim precaution. But the costs imposed by gatekeeping on third parties or society at large might be more indirect and less apparent.

The gatekeeping process itself can be costly. Revisiting the example of a father's Google account being banned illustrates this. To better identify wrongdoers, Google and other service providers may collect more data, for instance, a child's photo or the father's browsing history. Ideally, collecting more data can improve accuracy, it may, however, invade the user's privacy. This privacy invasion is a cost borne by the user who can weigh the costs and benefits of consenting to such data collection.

What is concerning is that rational consent to data collection by users can create a spillover effect on other dissenting (non)users. On a micro level, the data shared by a user may concern other individuals who do not wish to share their data. For instance, sharing the child's photo may not be in his best interest, even if the parents are his legal guardians. On a macro level, allowing data collection can indirectly contribute to gathering information about others who also do not desire such collection. This phenomenon, referred to as "data pollution" by Ben-Shahar (2019), clearly illustrates this cost of gatekeeping — specifically, data collection — on non-contracting third parties.

Other gatekeeping practices, particularly when abused, can yield similar effects. Anti-money-laundering regulations serve as an example. Financial institutions are legally required to collect customers' information to combat money laundering or terrorist financing activities. However, this process can be abused to enable authoritarian regimes to access information about human rights activists (see France 2023, p. 4), opposition parties, or dissidents (see Mehra 2023-01-06). The fear of abusive access to information can restrict financing to these entities by intimidating donors and supporters. While these regulations are beneficial

for combating crime, heightened processes may inadvertently lead to adverse consequences, potentially harming civil society and democracy (see France 2023, p. 3).

In the gatekeeping process, the decision to interdict or report can impose costs on external parties. When contracting with gatekeepers has positive externalities, interdicting can result in externalized costs. For instance, an employer may reject a job applicant with a criminal record to mitigate potential liability (see e.g., Heydon and Naylor 2018; Sugie, Zatz, and Augustine 2020). By rejecting ex-felons, employers reduce their liability risk since hiring such individuals increases the likelihood of employee misconduct. However, this precautionary response can jeopardize community safety as unemployed ex-felons are more likely to recidivate and pose harm to communities (see Tripodi, Kim, and Bender 2010; contra Visher, Winterfield, and Coggeshall 2005). When employers cannot precisely determine who has a criminal record due to Ban-the-Box (BTB) laws, they resort to other statistically related but imperfect attributes (see Agan and Starr 2018, p. 194). Relying on such attributes puts employers at risk of over-rejecting certain groups and imposing costs on those groups.

A similar scenario can be observed in banking services. Banks offer essential services like money deposits, transfers, and withdrawals, but they are also required to comply with anti-money laundering regulations, which mandate the collection and verification of customer information. As a result, certain individuals without qualified documentation — like immigrants, self-employed individuals, and residents in high-risk geographic areas — may be denied banking services, a practice known as de-risking. De-risking deprives these underbanked or unbanked individuals of access to essential financial activities. Empirical evidence supports the notion that lower financial inclusion, i.e., a higher number of unbanked or underbanked individuals, can contribute to poverty (Tran and Le 2021). Poverty incurs costs to society beyond the contracting parties. Although those denied services may not aspire to be impoverished, they may not fully internalize the total social cost, particularly in a welfarist system where government-funded social safety nets help share their burden.

Other gatekeeping decisions, such as reporting, can also impose costs on third parties. For example, banks are obligated to report suspicious transactions to financial intelligence units, such as FinCEN in the U.S. By filing these reports, banks gain immunity from liability associated with such criminal transactions. However, the sheer volume of reports can overwhelm the agencies responsible for analyzing the data. When the data exceeds the agency's processing capacity, it can lead to ineffective enforcement due to paralysis. Consequently, reporting as a gatekeeping decision can impose costs on the receiving agency. Fortunately, in most contexts, the receiver is the state acting as a social planner and can address this issue through various mechanisms, such as charging filing fees (see Takáts 2011, p. 57).

Summary

In conclusion, this section has outlined additional sources of what Kraakman (1986, pp. 77–78) would categorize as "tertiary costs" that arise from imposing gatekeeper liability. Gatekeeping, as a precautionary behavior, is not always neutral to other parties. It can impose costs on gatekept counterparties or be used to divert, rather than deter, wrongdoers. Additionally, gatekeeping itself, including the processes and decisions made by gatekeepers, can have social costs that neither the gatekeepers nor the gatekept consider. Importantly, the existence and extent of these costs cannot be analyzed in abstract terms. They require careful examination through concrete analysis of the obligations and the market of designated gatekeepers. This analysis paves the way for a list of more concrete factors for policy prescription, which will be developed in the next part.

3.4 Designing Gatekeeper Liability Regimes

The comparison between gatekeepers and victims prepares us for importing insights from the victim precaution literature, shedding new light on the discussion of gatekeeper liability. Victim precautions are sensitive to law enforcement intensity and focus. Also, precaution

can impose costs on other people. Gatekeeping as a form of precaution-taking can share the same characteristics as victim precaution. Appreciating these characteristics, I argue, helps us to design better, tailored gatekeeper liability regimes.

In the subsequent sections, I will utilize the preceding analysis to explore the factors involved in designing gatekeeper liability regimes. Specifically, I will focus on three dimensions: required conduct, liability regimes, and penalty amounts.

The first dimension pertains to the kind of conduct we require from gatekeepers. Essentially, do we expect gatekeepers to interdict or report wrongdoers? The relationship between the choice of interdiction and reporting, in relation to direct enforcement, seems relatively under-explored. The second dimension involves the choice among strict liability, the negligence rule, and the act-based regime. I contend that the choice of liability regime should pay close attention to the market structure of the services provided by gatekeepers and the associated gatekeeping costs. The last dimension briefly discusses proposals to increase penalties to address the issue of corrupt gatekeepers. While such proposals can reduce corruption, they might inadvertently intensify the issue of excessive gatekeeping toward law-abiding customers.

3.4.1 Required gatekeeping conduct

When a party is designated as a gatekeeper, it's necessary to define what is required. Some gatekeepers, such as auditors, must withdraw their services or interdict the wrongdoer to avoid liability. Others, like banks dealing with suspicious money-laundering transactions, are simply required to monitor and report wrongdoings. A gatekeeper may perform both functions in different contexts. For instance, banks are required not to engage in business relationships with undocumented persons (interdiction) and report certain transactions that are suspicious (reporting).

As mentioned in 3.3.3, gatekeeping incentives under different enforcement focuses vary

with the required conduct. To recap, a gatekeeper required to interdict would likely invest more in gatekeeping when the direct enforcement focus shifts from sanction to detection. In contrast, when sanctions are elevated, gatekeeping can be crowded out because interdiction does not bring additional deterrent benefits to the gatekeeper. Instead, the elevated sanctions, coupled with the same or lower detection rate, can dilute the expected liability faced by interdicting gatekeepers.

The reactions of gatekeepers required to report when direct enforcement changes focus are different from those required to interdict. Gatekeepers responsible for reporting could invest more or less under the same circumstances. On the one hand, elevated sanctions make their reporting more deterrent, as the expected sanctions on the wrongdoers increase. On the other hand, lower detection from direct enforcement would dilute the expected liability faced by gatekeepers, leading them to divest from gatekeeping. Therefore, the overall effect on reporting incentives is uncertain.

Of course, this analysis assumes that the intensity of direct enforcement remains constant, ensuring the same level of deterrence from direct enforcement to wrongdoers. However, this is often not the case in reality. In practice, imposed sanctions are determined by a legislatively stipulated range of sanctions coupled with judges' discretion, somewhat restrained by sentencing guidelines. In contrast, detection can be more volatile, fluctuating with prosecutorial focuses (see e.g., Saadatmand, Toma, and Choquette 2012, p. 287), election cycles (Levitt 1997), or international pressures (Reuters Newsroom 2022-09-21). Therefore, compared to sanctions, detection is more fluid.

To provide a more realistic picture, let's assume that imposed sanctions are relatively fixed in the short term for the reasons mentioned above, and detection is the only variable. When direct enforcement invests more resources in and improves detection, enforcement becomes more rigorous. More wrongdoers are deterred by direct enforcement, resulting in a lower risk of liability for gatekeepers. At first glance, such a lower liability risk could mo-

tivate gatekeepers to divest from gatekeeping. However, this disincentive has a disparate effect on gatekeepers responsible for interdicting and reporting. Those responsible for interdicting would have a stronger gatekeeping incentive than those responsible for reporting, owing to a lower marginal benefit of reporting by gatekeepers. In sum, the decision on the required conduct for designated gatekeepers should consider the volatility of direct enforcement. When direct enforcement is variable, requiring gatekeepers to report could weaken the effectiveness of direct enforcement's redirected focus. Specifically, with reporting gatekeepers in place, a shift of focus in direct enforcement towards misconduct could be less fruitful than anticipated.

Besides the differing effects of direct enforcement strategies, the costs of gatekeeping can differ between the two types of required conduct. As discussed, interdicting imposes direct error costs on law-abiding counterparties. Counterparties may not always be able to negotiate with the gatekeeper to minimize these error costs efficiently due to market structure or the problem of adverse selection. Even if they can negotiate a scrutiny level that benefits both parties, they might fail to internalize the true social cost of errors, such as poverty resulting from lower financial inclusion and recidivism from unemployed ex-inmates.

Compared to interdicting, reporting seems less problematic as it does not directly reject service to law-abiding customers and thus avoids incurring the error cost. However, reporting is not cost-free. Apart from the financial cost associated with filing and its preparatory process, reporting also incurs a cost of processing information by the receiver. This cost is also a third-party cost considered by both the gatekeeper and the gatekept but can be mitigated by charging a reporting fee. Therefore, the problem of externalized cost appears less of a concern for reporting gatekeepers.

The above analysis distinguishes between interdicting and reporting. However, this dichotomy is not absolute. When the compliance cost and the expected liability for failing to report are both prohibitive, reporting gatekeepers may opt to withdraw their service to

counterparties, effectively becoming interdicting gatekeepers. Despite reporting already immunizing gatekeepers from liability, complying with reporting requirements can be costly. For instance, gatekeepers may be required to follow specific formats to prepare reports before filing. As a result, gatekeepers are compelled to weigh between the costly filing and the expected liability from not filing. Alternatively, they can simply avoid such a costly tradeoff by rejecting certain customers who are likely to trigger the reporting requirement. When the total expenditure, including the compliance cost and the liability of non-compliance, exceeds the revenue, gatekeepers would find it beneficial to interdict such customers from the outset. Under such circumstances, reporting is replaced by interdicting. However, the reverse does not hold. Interdicting gatekeepers never report as reporting does not save them from liability or compliance costs.

Despite these complexities, the distinction between interdicting and reporting gatekeepers remains crucial for policymakers. They respond differently to changes in enforcement policy and impose different types of gatekeeping costs on different parties. As a result, policymakers should consider these factors before delegating specific gatekeeping responsibilities to designated gatekeepers. Moreover, policymakers should be aware of the possibility that drives reporting gatekeepers to interdict certain customers and become de facto interdicting gatekeepers, especially in situations of costly reporting with draconian liability.

3.4.2 The choice of liability regimes

The choice of liability regimes is a hotly debated topic in gatekeeper literature, often raised to address specific issues faced by gatekeepers. For instance, Coffee (2004) and Partnoy (2004) advocate for strict liability in their respective articles out of different concerns. In contrast, Hamdani (2003) and Tuch (2010) favor the negligence rule. Hamdani (2003) supports the negligence rule because he believes that strict liability may destabilize the market when gatekeepers cannot flawlessly distinguish between unscrupulous and honest individuals. Tuch

(2010) supports the negligence rule as well, asserting that it ensures multiple cooperative gatekeepers will fulfill their responsibilities without capitalizing on the efforts of others.

However, these debates over liability regimes often focus on specific problems and fail to offer a comprehensive view of how the choice of regime may impact others. They also do not provide guidance for gatekeepers outside the specific context. Therefore, this section broadens the discussion by considering how liability regime choices can influence gatekeeping behavior and costs.

Before analyzing how gatekeeping behavior and costs differ across liability regimes, it is essential to define the liability regimes under discussion. Here, I adhere to Shavell (2004, pp. 474–479)'s categorization of liability regimes and focus on three: the act-based regime, harm-based negligence rule, and harm-based strict liability. Gatekeepers can be held liable either when they fail to meet the gatekeeping requirement (act-based) or when the failure results in harm (harm-based). The act-based liability is straightforward: gatekeepers are held liable for failing to conduct required acts, irrespective of whether the failure results in harm. In the harm-based liability category, there are two types, distinguished by the required level of care. If gatekeepers are held liable only when they fail to meet the necessary level of care (negligent), it is called the negligence rule. However, under strict liability, gatekeepers are held accountable whenever harmful results occur, regardless of their level of care.

The advantages and disadvantages of each liability regime have been extensively discussed in literature, including aspects like courts' information, administrative costs, and so on (pp. 474–479). In this section, I will focus solely on how these regimes impact gatekeeper responses to different direct enforcement strategies and the costs imposed on other parties. This evaluation should not be considered definitive but viewed as complementary to existing findings, providing policymakers with additional perspectives when deciding which liability regime best suits a specific context.

Act-based liability

Act-based liability demands that a gatekeeper performs specific required actions, often setting a minimum standard. For example, banks are required to collect basic customer information (customer identification program) and verify the information provided by the customers (customer due diligence) (see generally Stessens 2009, pp. 146–178). Noncompliance with such requirements can lead to penalties, irrespective of whether the customer actually uses the bank for money laundering. In this context, knowing their customers can be viewed as a legal obligation backed by act-based liability.

Gatekeeping backed by act-based liability is insensitive to direct enforcement. As long as gatekeepers meet the standard, they are free from liability. Since they are compliant, they need not adjust their behavior in response to changes in enforcement focus. Further, they do not internalize additional deterrence benefit from their more intensive gatekeeping. For example, gatekeepers with reporting obligations under act-based liability are not incentivized to report more when detection from direct enforcement weakens. Hence, regardless of changes in direct enforcement, they adhere to the required standard. In sum, as long as gatekeepers are compliant with the requirements backed by act-based liability, they would not respond to the change in the expected liability resulting from the alterations in enforcement intensity and focus.

Act-based liability also disincentivizes parties from bargaining for doing more than the required standard. In the setting in Table 3.2, both law-abiding parties and wrongdoers have incentives to bargain for their preferred standard to obtain optimal cost-liability combinations. Hence, law-abiding counterparties may ask for a higher standard to lower the payment for expected liabilities. However, under the act-based liability regime, law-abiding parties need not do so because there is no liability for gatekeepers who meet the standard, and there is no additional benefit from improving beyond the standard. The only exception is that the court or law enforcement agencies may commit errors in ascertaining the liability.

Hence, both gatekeepers and law-abiding counterparties have no incentives to move above the standard.

Contrarily, parties can still have incentives to bargain for a lower but non-compliant standard if they are willing to bear the cost of liability. This finding has practical implications. First, act-based liability alone does not necessarily prevent corrupt gatekeeping — it needs to be accompanied by high penalties. Second, the benefits of a lower standard may not cover the expected liability, leading parties, particularly law-abiding ones, to adhere to the required standard. If the standard is set too high, the overall gatekeeping cost borne by both parties would be excessive.

Act-based liability has both advantages and disadvantages in the context of multiple competing gatekeepers. On the one hand, it discourages gatekeepers from diverting wrongdoers to other gatekeepers because they would be delighted to take such customers without liability by simply meeting the standard. On the other hand, it disallows gatekeepers to rely on other gatekeepers' findings if reliance is not allowed and qualified for the required standard. Therefore, the second and further gatekeepers are forced to go through their own gatekeeping process and incur a redundant gatekeeping cost. Ultimately, which cost is more relevant depends on the service provided by gatekeepers. In other words, how many gatekeepers can the gatekept simultaneously be subject to? A person can have, say, five bank accounts at the same time and probably one to three employers, but the person can only board one plane at a time. When the gatekept can be subject to multiple gatekeepers at a time, then information sharing and reliance on the shared information from the previous gatekeeping process can be optimal for saving redundant gatekeeping costs. Otherwise, redundant costs would be less of a concern if the gatekept would only be subject to one gatekeeper at a time. There is no information to be shared and relied upon — every gatekeeping process matters as it represents different judgments of different facts at different times. In this case, it is more likely that those previously rejected gatekept — often wrongdoers — would replicate

the gatekeeping process and incur the cost. If the proportion of wrongdoers is low, or the history of rejection can be inferred (such as a long period of unemployment), the problem is less significant.

Act-based liability has the advantage of addressing the problem of third-party costs incurred by gatekeeping. When gatekeeping incurs costs borne by third parties who are not in the contract, such costs are externalized by gatekeepers. If the state has information about such costs, it can design proper act-based liability by requiring specific acts with a certain level of intensity. By complying with the standard, gatekeepers can reach socially optimal gatekeeping.

In summary, gatekeepers under the act-based liability regime are insensitive to changes in direct enforcement. The regime also discourages parties from bargaining around the standard, implying a higher likelihood of adherence to the required standards. Lastly, it deters harm displacement among competing gatekeepers but may lead to redundant gatekeeping costs.

Harm-based negligence rule

The harm-based negligence rule holds gatekeepers liable when they fail to meet the due level of care in gatekeeping wrongdoers whose wrongdoing ultimately results in harm. Since both the negligence rule and the act-based liability regime both set the required level of due care, which often is in the form of certain acts, they function similarly, except that liability is not imposed until harm results under the negligence rule.

Much like act-based liability, gatekeepers subject to the negligence rule have no incentive to take care beyond the due level. Ideally, they would not be held liable as long as they meet the due care level. Therefore, gatekeepers should incur the same cost of due care, regardless of the number of deterred or detected wrongdoers. Nevertheless, the claim presumes zero litigation cost and the perfect information possessed by the court. In practice, even though

gatekeepers take due care, they can still incur litigation costs to defend their case against enforcement agencies or, worse, be imposed with liability because of the court's error from hindsight bias (see e.g., LaBine and LaBine 1996) or lack of information. Under such circumstances, gatekeepers may be motivated to adjust their due care level to save expected litigation costs by interdicting or reporting suspicious customers. However, the extent of such adjustment should be limited to the scale of litigation cost and the expected liability resulting from court's error. As a result, gatekeepers under the negligence rule could still exert care above the due care level when detection from direct enforcement improved. Likewise, when detection is so low that the marginal expected liability is lower than the marginal cost of due care, gatekeepers may also take care below the due care level. This differs from the act-based liability regime, where detecting wrongdoers does not imply gatekeepers' non-compliance so long as the standard does not require zero wrongdoers.

Under the negligence rule, both the gatekeeper and the gatekept may have incentives to bargain around the due care level. Despite being free from liability, gatekeepers may fear costly litigation. In response, they would either enhance the care level above the due care level to avoid more potential wrongdoers that cannot be interdicted or reported under the due care level. Alternatively, gatekeepers may also price the risk of litigation cost in their service price. Hence, both gatekeepers and the gatekept may want an enhanced care level to the extent of litigation costs and expected liability from courts' errors. Hence, it may be beneficial to the gatekeepers and the law-abiding clients to exercise more care and reduce the expected litigation cost. Of course, the gatekeeper and the wrongdoer may want and are free to bargain for a lower level of care when they both agree to share the expected liability.

Under the negligence rule, gatekeepers should not benefit from a lower liability by diverting wrongdoers to other gatekeepers. However, as mentioned above, they benefit from a lower expected litigation cost and lower liability risk due to courts' errors. It can be problematic when courts rely on industry standards to set the due care level. When gatekeepers

deviate from the due care level by diverting wrongdoers to save litigation costs, the care level in equilibrium would be higher than the socially optimal one. In such a case, industry standards can lead gatekeepers to exercise an excessive level of care. In fact, this is expected to happen in the area of online platforms (see Riis and Schwemer 2018, p 13).

Finally, similar to act-based liability, the negligence rule may not be able to address the problem of redundant costs when due care is defined to exclude reliance on other gatekeepers' findings. However, if due care is set properly, it can induce gatekeepers to take socially desirable care that considers externalities imposed on third parties.

Harm-based strict liability

Under strict liability, gatekeepers are held accountable for the misconduct of the gatekept, irrespective of their care level. When held strictly liable for harm, gatekeepers are encouraged to invest in gatekeeping to balance the cost of gatekeeping with expected liability. As such, when liability is set equal to social harm, it is anticipated that gatekeepers will adopt a socially optimal level of care (Shavell 1987b, p. 9).

Gatekeepers, when held strictly liable, are extremely sensitive to potential liability (Cooter 1984, p. 1539). Any minor change in expected liability prompts them to adjust their level of gatekeeping. Thus, when direct enforcement becomes more rigorous and deters more wrongdoers, the risk of liability, in general, decreases. Responding to this change, gatekeepers would reduce the level of gatekeeping to save costs, as indicated in 3.3.3. On the other hand, when direct enforcement retreats and wrongdoers become more active, gatekeepers might choose to improve their level of gatekeeping. The increased cost of gatekeeping, previously considered inefficient, is now justified by the benefit of avoiding additional liability from undeterred wrongdoers.

Gatekeepers under strict liability are also sensitive to changes in direct enforcement focus. As reasoned earlier, the detection of wrongdoers always uncovers the gatekeeper's failure

to block or report them. When direct enforcement improves its detection, the associated gatekeeper is also more likely to be detected and held liable. Consequently, gatekeepers would want to interdict more counterparties they perceive as potential wrongdoers.

However, they may not necessarily want to report more wrongdoers. Although reporting gatekeepers also face greater expected liability as interdicting gatekeepers do, the benefit they can derive from rigorous reporting decreases. Reporting not only immunizes gatekeepers from liability but deters potential wrongdoers. As detection from direct enforcement improves, the marginal deterrence benefit from gatekeepers' reporting diminishes. Hence, while incurring more costs, more rigorous reporting may not be justified by its benefit as before. Therefore, the effect of improved detection on reporting is unclear. Furthermore, when the change in detection also accompanies a change in direct enforcement intensity, the outcome is even more uncertain. While the increased enforcement intensity may suppress reporting on one hand, the improved detection could either increase or decrease reporting on the other hand. Ultimately, it could be that gatekeepers are investing more or less in reporting.

Given gatekeepers' sensitive reactions to enforcement changes, they can easily negotiate with counterparties to prevent inefficient gatekeeping costs borne by counterparties under strict liability. Assuming that liability is set equal to harm, both the gatekeeper and the gatekept can share the liability risk and agree on a level of gatekeeping beneficial to both parties. In the absence of the adverse selection problem, both law-abiding and illegal counterparties are pooled together and share the same liability risk. Under such a circumstance, law-abiding counterparties would not agree on a level of gatekeeping that incurs a cost higher than the reduction in liability. Therefore, the level of gatekeeping is optimal for both parties.

While strict liability allows the gatekeeper and the gatekept to negotiate and, to some extent, avoids excessively costly gatekeeping, it encourages harm displacement among competing gatekeepers. By slightly improving their gatekeeping level, they can divert sophisticated wrongdoers and then discount their price to reflect the lower expected liability. With

such a low price, gatekeepers who increase their gatekeeping level can attract law-abiding customers and create a separating equilibrium. Other gatekeepers, in response, may follow to match the increased gatekeeping level or remain with wrongdoers.

Both outcomes would be inefficient. In the former case, the mutual gatekeeping level would deviate from the socially optimal level, resulting in an inefficient equilibrium characterized by the collective action problem. In the latter case, those who require stringent scrutiny (wrongdoers) are subject to a lower gatekeeping standard, while those who are supposed to be law-abiding experience an unnecessary process solely to signal themselves.

Fortunately, such harm displacement can, to some extent, be mitigated by bargaining. Gatekeepers may contract with counterparties and offer a discount when they share the findings of previous gatekeepers. With this financial incentive, counterparties can also "buy" their record from previous gatekeepers. When previous gatekeepers' findings are shared at a modest cost,[9] other gatekeepers may easily free-ride on such findings from costlier and more stringent gatekeeping. When other gatekeepers can cheaply or easily free-ride on their effort, gatekeepers cannot guarantee that they fully attract law-abiding customers and capture entire benefits from their businesses. This kind of free-rider problem that worries Tuch in the context of multiple collaborative gatekeepers can oppositely mitigate the undesirable harm-displacement effect. In addition to harm displacement, such bargaining can also avoid redundant costs incurred by replicative gatekeeping processes.

Again, to what extent bargaining can mitigate the harm-displacement effect and redundant costs hinges on the relationship between the gatekeeper and the gatekept. If the relationship is exclusive at a time, law-abiding gatekeepers would have little need to share the information with the unnecessary "next" gatekeeper, and wrongdoers would rather want to hide than share their history of rejection. Hence, competition still exists among gatekeepers, and harm displacement can be found, while information sharing plays a minor role in

9. In fact, such findings can be free to shared. For example, simply showing the debit card from another bank may suffice to prove that the customer passed the customer due diligence from another bank.

mitigating the problem.

Finally, strict liability cannot directly solve the problem of third-party costs unless the amount of liability is adjusted. If the gatekeeper is held liable for the entire harm from the wrongdoing, it would invest in gatekeeping according to the level of harm. In a more technical way, it would invest to the point where the marginal gatekeeping cost is equal to the marginal benefit from harm reduction. Nevertheless, gatekeepers may not fully internalize all the costs associated with gatekeeping. Some costs are borne by parties who are not even in the contract with the gatekeepers and thus unable to bargain with them. Therefore, with externalized costs, gatekeepers may take on too much gatekeeping. To fix the over-gatekeeping problem under strict liability, the state should adjust the amount of penalties imposed on liable gatekeepers, which will be discussed in the next section.

Comparison

The choice of liability regime hinges on many factors, such as information and litigation cost. In the context of gatekeeper liability, other factors may come into play, such as the free-rider problem among multiple collaborative gatekeepers and the risk of corrupt gatekeepers. The analysis above further points out several points worth consideration in the framework of precaution-taking behavior.

If we want stable gatekeeping efforts despite changes in direct enforcement, we should prefer act-based liability or the negligence rule. Moreover, both rules avoid potential harm displacement among competing gatekeepers. In addition, when the required standard is set appropriately, the level of gatekeeping can effectively be optimal, given the existence of third-party costs. However, both regimes prevent parties' bargaining, which can exacerbate the problem of counterparty costs and redundant costs.

In comparison, strict liability allows bargaining that solves counterparty and redundant costs. However, gatekeepers are more sensitive to changes in direct enforcement intensity and

focus, making gatekeeping more volatile. In cases where sanctions imposed on wrongdoers by direct enforcement are enhanced, a crowd-out effect may be produced. Gatekeepers are more incentivized to divert wrongdoers to other competing gatekeepers, which generates no social benefit but imposes costs. Additionally, gatekeepers would not reach the socially optimal level of gatekeeping when they fail to internalize third-party costs.

3.4.3 Amount of penalties imposed on liable gatekeepers

In the previous section, I discuss how gatekeepers behave under different liability regimes. Liability regimes are concerned with under what conditions gatekeepers are held liable. There is another dimension of their liability when they are ultimately found liable. That is, what is the magnitude of the penalties that must be imposed on gatekeepers when they are liable? This question is the focus of this section.

Before diving into the main discussion, I should clarify that the discussion is most relevant to the strict liability regime. As mentioned above, under strict liability, gatekeepers are very sensitive to the change in their liability. Therefore, the penalty is relevant to their gatekeeping decisions. However, this is not to say that the penalty amount plays no role in affecting gatekeeping behavior under the act-based liability regime or the negligence rule. On the contrary, the penalty amount could still affect their behavior to the extent of judicial errors. When courts commit errors in verifying whether gatekeepers fulfill the act-based requirements or the due care level, the greater the penalty amount, the more distortion judicial errors can create in their gatekeeping behavior. Gatekeepers may find it beneficial to increase their gatekeeping levels when due care is upward-biased. As the penalty for gatekeepers who are negligent or do not comply with required acts increases, gatekeepers are more willing to invest more in gatekeeping to hedge this liability risk. Hence, the penalty amount's effect still exists but is not as sensitive as it is under strict liability.

Focusing on strict liability, I first investigate the basic case as a benchmark, where lia-

bility is set equal to social harm from the wrongdoing that gatekeepers fail to interdict or report. Ideally, equalizing harm to liability induces gatekeepers' optimal level of gatekeeping. However, as mentioned above, gatekeepers may not fully internalize the gatekeeping cost borne by third parties. Under such a circumstance, the private marginal cost is lower than the social marginal cost of gatekeeping, leading private gatekeeping incentives to be socially excessive.

Another problem arises when multiple competing gatekeepers exist, such that wrongdoers can have multiple attempts to carry out their plans. If, on average, a wrongdoer has, say, two attempts, such a penalty amount could be excessive because the marginal deterrence benefit is lower than expected when all wrongdoers only have one attempt. To see the underlying intuition, let's compare two scenarios: the wrongdoer has only one attempt in the first scenario but has two attempts in the second one. When the wrongdoer has only one attempt, the gatekeeper's failure to interdict or report the wrongdoer leads to direct social harm. Hence, the gatekeeper should spend until the marginal gatekeeping cost equals the marginal reduction of social harm. In comparison, in the second scenario, the marginal deterrence benefit is lower. To elaborate, let's consider two possibilities: the gatekeeper is the first or the second one. If the gatekeeper is the first one, then her successful detection yields a lower social benefit as the wrongdoer can have a second attempt from the second gatekeeper. If, otherwise, the gatekeeper is the second one, successful detection prevents social harm, but the failure is less harmful because the gatekeeper only lets go of a smaller subset of wrongdoers compared to the single gatekeeper. Hence, for both gatekeepers, the marginal benefit of successful detection is lower, implying a lower socially desirable level of gatekeeping.[10] Of course, the analysis here presumes no information sharing and reliance on other gatekeepers.

10. Assume that each gatekeeper can spends x so that social harm, h, can be prevented with a chance $p(x)$, where $p'(x) > 0$, $p''(x) < 0$. A gatekeeper's objective function, as well as social welfare function, is to maximize $-x - [1 - p(x)]h$, and its socially optimal gatekeeping level would be where $1 = p'(x)h$, meaning that the marginal gatekeeping cost equals to the marginal reduced harm. When both gatekeepers work

If sharing and relying is possible, then holding the first gatekeepers fully liable should be optimal. From the analysis above, setting a penalty amount equal to the wrongdoing's social harm does not seem ideal when multiple gatekeepers exist. Indeed, it points to the idea that we should lower the penalty amount below social harm from un-interdicted or unreported wrongdoing.

This recommendation can trigger another conventional problem that perplexes scholars. When gatekeepers are prone to be conflicted or even corrupted because of their financial interest in the gatekept, they may acquiesce to the gatekept's wrongdoing. A lower penalty may comfort, if not encourage, them to do so. Also, with a lower penalty, the gatekeeper and the gatekept would be more willing to bargain down the scrutiny, as the threat of non-compliance is weakened. The dilemma between the problem of conflicted or corrupted gatekeepers and the problems of externalized and redundant gatekeeping costs implies that Coffee's proposal of "stricter" strict liability could have a downside. In response to the gatekeepers' corruption unearthed by the Enron scandal, Coffee proposes a stricter strict liability, calling for enhanced penalties calculated based on a gatekeeper's annual revenue.[11]

By coupling the penalty with the revenue, the gatekeeper's financial interest is linked to the legal consequences of incompliance. Hence, the gatekeeper is expected to achieve the required gatekeeping.

Coffee's proposal can rightly address the problem when gatekeepers are reluctant to gatekeep their clients because of financial ties. However, his proposal could exacerbate other issues identified above. Namely, gatekeepers under such a stricter strict liability regime

---

independently, their joint objective function is $-2x-[1-p(x)]^2h$, and the socially optimal joint gatekeeping level is $1=[1-p(x)]p'(x)h<p'(x)h$. The marginal benefit of reduced harm is lower because of the existence of another gatekeeper.

11. In fact, a perverse effect can occur with this proposal. When the gatekeeper engages in a relationship with law-abiding clients, it is also increasing its penalty amount when it fail to interdict "one" wrongdoer. This would increase the cost of serving additional law-abiding customer and be translated into a surcharge on law-abiding customers. Moreover, this surcharge cannot be avoided by heightened scrutiny as it has nothing to do with the law-abiding customer's riskiness.

would incur more costs borne by third parties and redundant costs as well. Furthermore, the enhanced penalties make gatekeeping decisions more sensitive to the changes in direct enforcement. That is, the enhanced penalties would exacerbate the crowd-out effect or make gatekeeping more volatile. Finally, the enhanced penalties may provide competing gatekeepers with a stronger incentive to divert wrongdoers.

The discussion above is not meant to provide a determinative, one-size-fits-all prescription for setting penalty amounts for gatekeepers. On the contrary, its purpose is to remind us that we must consider several factors after holding gatekeepers liable. In short, we should first consider whether there are third parties bearing the cost of the gatekeeping process. Second, we should look at the market structure of the service provided by gatekeepers. Are there multiple competing gatekeepers? To what extent can a wrongdoer exercise multiple attempts by seeking assistance from those competing gatekeepers? If so, can a gatekeeper rely on other previous gatekeepers' wrongdoing? Third, we should know more about the essence of the transaction between the gatekeeper and the gatekept to evaluate the risk of conflicted and corrupted gatekeeping incentives. Lastly, we should ensure that changes in the law of direct enforcement should not be too volatile, making gatekeeping incentives frequently deviate from the expected level.

3.5 Applications: Addressing Societal Issues through Gatekeeper Liability

This section applies the analysis of gatekeeper liability to three societal issues. The first pertains to the enhancement of the employment rate among ex-offenders, aiming to mitigate their collateral consequences. The second revolves around boosting financial inclusion hampered by stringent anti-money laundering laws. The third discusses online platforms' liability for users' misconduct.

In each context, pertinent gatekeepers exist. For ex-offender employment, employers

could potentially bear liability for employees' criminal misconduct. The fact that banks act as anti-money laundering gatekeepers significantly influenced people's financial inclusion. In the digital realm, online platforms stand at the crossroads of liability for their users' misdeeds. However, these issues may not always directly invoke gatekeeping behavior in scholarly discourse. This section seeks to contribute to the literature by providing fresh perspectives and demonstrating how effective gatekeeper liability design can address these problems.

3.5.1 Ex-offender employment and recidivism

Ex-inmates often struggle to find employment after their release. Lack of job opportunities constitutes a severe collateral consequence of criminal convictions (Nikolaides 2022-08-31). Unlike other collateral consequences prescribed by law (see National Inventory of Collateral Consequences of Conviction, n.d.), which are subject to judicial review, unemployment is purely an employer's decision. Despite laws prohibiting employers from collecting information like criminal records (see e.g., National Conference of State Legislatures 2021-06-29), statutes cannot force them to hire ex-offenders. Why do employers hesitate to hire ex-offenders, even at the expense of information collection? One reason is that employers are liable for their employees' misconduct,[12] both criminally and civilly.

Corporate employers: gatekeepers of employee misconduct

In the United States, corporate employers face criminal liability for their employees' misconduct, not merely civil liability. This corporate criminal liability was first recognized by the

12. Admittedly, there are other reasons for employers to reject job applicants with criminal records. For instance, they might be afraid of victimized by the hired employee's criminal misconduct, such as embezzlement or violent attacks. Or, employers may be concerned about the reputation risk when the employee's criminal history is known to customers. Here, I do not exclude the possibility of other reasons, nor do I argue that those reasons are dwarfed by liability concerns. Instead, the claim I make is a modest one: the liability could play a role in refusing to hire an ex-offender. As long as there are some ex-offenders being refused out of the reason of liability, the analysis below is to that extent applicable.

U.S. Supreme Court in *New York Cent. & H.R.R. Co. v. U.S.* (212 U.S. 481 [1909]) and expanded rapidly alongside the widespread adoption of the strict liability regime, Congress's drive to criminalize regulatory offenses, and the issuance of Sentencing Guidelines for Organizations (see Khanna 1996, p. 1479).

The U.S. corporate criminal liability theory revolves around the principle of *respondeat superior* (see generally LaFave 2017, pp. 556–58), which holds corporations accountable for their employees' actions. However, corporations aren't held liable for all employee misconduct. The imputation of the wrongdoing necessitates two prerequisites: the employee commits a crime (1) within the scope of employment (2) intending to benefit the corporation (see generally Strader 2017, pp. 21–22). Economically speaking, a rationale behind this theory is to induce corporations to internalize the social cost generated by their agents' misconduct (see Sykes 1984, pp. 1231–80). The notion is that when corporations internalize these costs, the social benefit aligns with the corporation's cost-saving efforts in monitoring and deterring misconduct (Fischel and Sykes 1996, pp. 324–25).

Nonetheless, internal disciplinary actions alone might not be sufficient to deter misconduct (Polinsky and Shavell 1993, p. 248). When an external sanction is required to ensure deterrence, the parallel existence of punishments against criminal employees and the corporation surfaces. However, the corporation may be reluctant to report an employee's criminal conduct due to liability under the strict liability regime (Arlen 1994, pp. 833–67). This unintended consequence of strict liability suggests a duty-based liability regime might better facilitate optimal corporate reporting (Arlen 2014, p. 172).

Over-gatekeeping by employers

The refusal to hire ex-offenders can be viewed as a form of precautionary interdiction. To minimize liability, corporate employers interdict risky employees, thus lowering liability risk. However, gatekeepers may be overzealous, as the information upon which they rely can be

inaccurate. Commercial databases may contain errors or outdated information (see National Consumer Law Center 2023, pp. 17–23). Even if the criminal history information is correct, it doesn't necessarily reflect an applicant's propensity to reoffend. For instance, the Michigan experience of expungement shows that the ex-offenders, while not having changes in recidivist risk, enjoy a wage raise and higher likelihood of employment (Prescott and Starr 2020, p. 29), indicating employers' decisions do not closely track the underlying risks. This discrepancy between an employee's crime risk and gatekeeper behavior suggests the possibility of over-gatekeeping.

Excessive gatekeeping burdens those ex-offenders, who bear the cost of gatekeeping cost but cannot mitigate the cost through bargaining. Low-risk ex-offenders who are rejected cannot offer more to get hired, probably because of insufficient assets. Employers might also deliberately apply excessive gatekeeping to circumvent the adverse selection problem. Employers refuse to bargain and thus apply heightened scrutiny to offer a more competitive salary to attract more capable law-abiding applicants. Otherwise, they would set a salary discounted by expected liability, which then is unattractive to applicants without criminal records.

Additionally, even when ex-offenders and employers successfully negotiate, the agreed-upon gatekeeping level (i.e., the employment outcome) may not be socially optimal. Employers might still over-gatekeep because they don't internalize the social benefit of hiring an ex-offender. Evidence supports the positive relationship between employment and recidivism: A more robust labor market seems to lead to fewer property crimes (see Prescott and Pyle 2019). Communities enjoy a safer environment when ex-offenders get employed. As a result, whenever the employer rejects hiring an ex-offender, it implicitly imposes a social cost on the community. The social benefit of employment, or the social cost of unemployment, is neither fully internalized by ex-offenders nor employers. Hence, employers can still overly refuse to hire ex-offenders, even if they can bargain with each other.

Lastly, over-gatekeeping could impact applicants without criminal records if employers are denied the information. A field experiment comparing employer hiring behavior before and after the passage of the Ban-the-Box (BTB) law finds that racial disparity increased when employers were no longer allowed to directly gather information from job applications (Agan and Starr 2018). Those without criminal records but who share certain stereotypical characteristics are more likely to be rejected, marking an additional cost borne by law-abiding individuals.

How can re-designing employer liability help?

The discussion above identifies the problems in the context of corporate criminal liability. Specifically, corporate employers, acting as employees' gatekeepers, can commit an excessive level of gatekeeping due to difficulties in bargaining and third-party costs borne by communities. Redesigning gatekeeper liability can address this problem, potentially improving employment outcomes among ex-offenders.

First, adopting a multi-faceted, fault-based liability regime could curb over-gatekeeping. Arlen (2014) proposes this fault-based liability approach to counter the perverse effect of strict liability on reporting incentives. Such a regime, similar to the negligence rule, can prevent employers from over-gatekeeping because those who meet the standards will not be held liable. More precisely, we can exclude hiring decisions from the component of employers' fault, focusing instead on monitoring and reporting. Under this regime, an employer would not be liable for hiring an ex-offender.[13] Instead, the employer would be liable for failing to monitor and report the employee's misconduct. When employers are subject to specific requirements of monitoring and reporting, whether act-based or negligence-based, they will meet the standards and remain free from liability. In this case, hiring ex-offenders would not be burdensome.

13. A similar direction would be limiting employers' potential liability for hiring ex-offenders (see Garcia 2013, pp. 943–44).

Of course, one caveat is that the monitoring standard should not be adjusted according to employees' criminal records. In other words, the level of due care in monitoring an employee with a criminal history should not be set higher than that for monitoring an employee without a criminal record. Also, reporting should be inexpensive. Suppose the due care level equals the risk, or reporting is costly. In that case, hiring ex-offenders can still be expensive because the employer would incur a higher compliance cost due to an elevated due care level or more frequent reporting. Hence, to significantly increase and improve job opportunities for ex-offenders, we should ensure that compliance cost is not prohibitive, dissuading employers from hiring them in the first place.

A lower penalty amount can be considered when specifying the due care level or when the required acts by employers are costly for judicial or legislative sectors. Deducting the benefit of hiring an ex-offender from the employer's liability allows employers to internalize that social benefit and make socially optimal gatekeeping decisions. This ex-post liability reduction can align better with the current trend of BTB laws than an ex-ante Pigouvian tax. An ex-ante payment calculated based on the crime risk of the applicant could inadvertently signal the applicant's criminal history, which the Ban-the-Box law aims to prevent. Arguably, when employers receive subsidies, they are more likely to hire ex-offenders, making the Ban-the-Box law seem less necessary. However, setting the optimal amount of subsidy requires the state's superior information about the risk level of ex-offenders. If the state does not have such information, but employers do, then employers would be encouraged to hire low-risk ex-offenders rather than *average* ex-offenders. The selection may leave high-risk ex-offenders unemployed, imposing greater recidivism risk on communities. Furthermore, employers have various reasons to reject ex-offenders, which the Ban-the-Box law wants to address. In this case, signaling criminal history could counteract the purpose of the Ban-the-Box law. Instead, reducing liability to mitigate the disincentives of hiring ex-offenders may work better.

3.5.2 Financial inclusion and poverty

Banks are irreplaceable institutions at the heart of modern capitalism (Baer 2020-11-17). They provide essential services, such as transforming liquidity, maturity, and credit, without which capitalism cannot function (see generally Armour et al. 2016, pp. 277–78). However, such vital services in this modern era are not accessible to everyone. Now, banks are beginning to withdraw their service from certain geographic areas or groups of customers, leaving them unbanked or underbanked. This practice is called "de-risking." The reason behind de-risking is the increasingly stringent compliance programs required by anti-money laundering (AML) laws.

Banks as AML gatekeepers

As previously stated, banks stand in a position that provides services essential to modern capitalist society. They facilitate lawful businesses that benefit the economy by serving law-abiding customers. However, banks can be exploited by both law-abiding customers and criminals. With the help of banking services, criminals can quickly transfer their illicit gains and obscure their sources to evade law enforcement's tracing and confiscation. This process of transferring, hiding, and reusing the illicit gains from crime is money laundering (Financial Crimes Enforcement Network 2023). By money laundering, criminals can keep their criminal proceeds and further reinvest in criminal activities, thus imposing a greater social risk.

Due to their position in the flow of tainted money, banks are delegated with the mission of combating crime (see Stessens 2009, pp. 143–182). Since the passage of the Bank Secrecy Act of 1970 and subsequent related acts and international agreements, banks are required to conduct customer due diligence (CDD), keep a record of all transactions, and file suspicious activities reports (SARs) and currency transaction reports (CTRs) for transactions exceeding the stipulated amount. Banks are penalized for inadequate CDD programs and failure to

report money-laundering transactions. Hence, the liability regime of CDD can be seen as an act-based liability regime, as no harmful result is required for imposing penalties on banks. Furthermore, banks can shield themselves from the liability of facilitating money-laundering transactions by reporting. Reporting can be seen as an obligation backed by strict liability.

The social cost of banks' gatekeeping

Banks incur tremendous costs to comply with AML regulations, and these costs are on the rise (LexisNexis 2023). Such escalating costs not only disrupt banks' operations (Gill and Taylor 2004, p. 588) but also force them to withdraw services from riskier customers(Jojarth 2013, p. 17). However, the real cost extends beyond the financial. The withdrawal of service from certain customers can impose a cost on financial inclusion (Saperstein, Sant, and Ng 2015, p. 5; Financial Actions Taks Force 2017, p. 10). Worse, compared to the astronomical cost, the effectiveness seems limited (Levi 2002, p. 190). Despite the development of AML regulations and enforcement, laundered money remains steady at 5% of the global GDP (see Alkaabi et al. 2010, p. 3). In the following discussion, I will dissect the social cost of bank gatekeeping and explore how changing their liability can address the problem.

I argue that there are three types of inefficiencies in AML compliance: redundant gatekeeping processes, harm displacement, and decreased financial inclusion. First, banks can incur redundant costs of gatekeeping. Imagine a person who wants to open two bank accounts in one week. The person would go to the first bank and go through the CDD process, then visit the second bank for another similar process. In this case, the second bank incurs a redundant gatekeeping cost through a repetitive process. Given a short period and the requirements stipulated by law, the second process should not generate much additional information. However, the second bank would still undergo the process for compliance reasons. Without generating more information, the second CDD process can waste social resources.

Second, faced with increasing compliance costs and draconian penalties, banks are now

opting to "de-risk." De-risking means that banks withdraw services from certain customers due to their high risk. De-risking can take several forms, from terminating certain services like correspondent remittance, closing existing bank accounts, to denying the opening of new accounts. On the surface, de-risking can be socially beneficial because it prevents channels that money launderers or terrorists can exploit. However, this may not be the case in reality. In fact, money launderers and terrorists are not deterred by de-risking practices. On the contrary, they are diverted to smaller banks or financial institutions with inadequate capacity to identify the risk (Durner and Shetret 2005-11, p. 19). As a result, de-risking does not eliminate the risk. Instead, it reallocates the risk. Without the benefit of risk elimination, de-risking can be socially costlier than the de-risking banks anticipate.

Besides harm displacement, de-risking also imposes real costs by depriving people of access to financial services, leading to lower financial inclusion. Evidence shows that lower financial inclusion can contribute to poverty (Tran and Le 2021). In other words, when people are denied financial services, they are likely to become poorer. The cost of poverty, while borne mainly by the customer who is denied service, is not fully internalized by both parties. First, poverty can lead to social problems, such as crime (Webster and Kingston 2014, pp. 10–11) and social instability. These social costs are definitely not internalized by the bank and the customer. Second, in a welfarist regime, poor people often enjoy subsidies from the state, which are financed through taxes. As a result, poverty in a welfarist state directly imposes a cost on the public. Even if the cost of poverty is fully borne by the customer, the decision can still be suboptimal due to behavioral reasons.[14] The customer may have a hyperbolic discount rate or be subject to optimistic bias that leads to an underestimation of the poverty cost. With these challenges, the gatekeeping could be suboptimally high even if parties can bargain.

All the inefficiencies above can be attributed to one reason: the misalignment between

14. Consumers often suffer from misperceptions that prevent them from making neoclassically rational decisions. (see Bar-Gill 2008).

banks' incentive to gatekeep and the socially optimal goal to deter money laundering and terrorist financing efficiently. Banks only want to avoid liability, not the real harm of money laundering or terrorist financing. They would do whatever they can if it reduces their liability at a privately reasonable cost. As a result, they conduct redundant CDD processes and divert risky customers to other financial institutions by de-risking without being concerned about the potential social cost of poverty.

Modifying liability regime

Before evaluating how liability regimes can address this problem, we need to understand what motivates banks to adopt the inefficient gatekeeping measures described above. Banks conduct redundant and replicative Customer Due Diligence (CDD) processes because they cannot rely on information from other banks. The act-based requirements compel banks to go through these processes. Additionally, act-based filing requirements, combined with draconian penalties (Noonan and Smith 2023-01-18),[15] incentivize banks to over-file to hedge against liability risks from imperfect enforcement.

To solve this problem, I propose two key changes to the liability regime. In essence, banks would be held strictly liable for failing to report money laundering and terrorist-financing transactions, but with a cap on customer due diligence efforts. Under the proposed regime, the two-step requirement would no longer exist. Specifically, banks would not be fined for failures to conduct customer due diligence and failures to file suspicious transactions separately. Instead, they would only be punished for the latter. Despite the removal of liability for the first step, banks would still have incentives to invest in collecting customer data, as they would need to determine which transactions should be reported.

Under this consolidated strict liability regime, redundant and counterparty costs could be significantly reduced. Firstly, banks would be relieved from the redundant CDD process.

15. For instance, Wells Fargo was fined $7 million due to insufficient compliance. (see United States Securities and Exchange Commission 2022-05-20).

They could rely on other banks' results, provided they consider those banks credible. By relying on previous gatekeeping findings, banks could reduce redundant gatekeeping costs. Secondly, strict liability would also enable customers to negotiate with banks, increasing financial inclusion to some extent. Although unbanked customers do not fully internalize the cost of poverty caused by lower financial inclusion, they do partially absorb some of the cost. As a result, transitioning to a strict liability regime that allows negotiation could lead to progress in financial inclusion compared to the act-based liability regime, where negotiation is not permitted.

However, a shift to strict liability could also incentivize the diversion of wrongdoers, as discussed earlier. Therefore, I suggest limiting the scope of information collected by banks. With this limitation, banks would not compete in gatekeeping intensity to divert wrongdoers or risky customers to other banks, preventing an inefficient equilibrium. Still, banks could optimize their algorithms or risk analysis tools freely, given the same amount of information. Such algorithms and analyses, along with their final filing decisions, would be invisible and confidential to customers, resulting in a lower risk of harm displacement. Overall, this regime of strict liability with an information scope limit should encourage competition in better information processing, not in information collection, which can be more easily observed and displacing harm.

The second aspect of the proposal involves reducing the penalty amount. Restricting the scope of collected information with liability for failures to file suspicious transactions could prompt banks to withdraw their service from risky customers. Faced with draconian penalties, banks often err on the side of caution, choosing to file conservatively and incur higher filing costs when they cannot ascertain the quality of the transactions. In line with the proposal of Takáts (2011, p. 57), I agree that reducing the penalty amount could mitigate the problem of defensive filing and the associated inflated filing cost. With a lower penalty amount, banks would file fewer reports, incurring lower filing costs. Consequently, they

would be less likely to withdraw their services when the associated risk is justified by the expected revenue.

Lastly, the third-party cost of poverty cannot be properly addressed through strict liability alone. Unbanked individuals may not wish to endure the hassle of undergoing an additional process to access banking services. If their intolerance results in poverty, the cost is externalized to society. While act or negligence-based requirements enable the state to set standards to accommodate unbanked people, these regimes could introduce the problems we aim to solve. Moreover, if we truly want to accommodate these individuals with a lowered standard, we sacrifice the information from those who are banked. In this case, a subsidy for unbanked or underbanked individuals would be superior to the current act-based or negligence-based regime. By subsidizing unbanked individuals and supporting them to undergo the universally required customer due diligence program, we ensure that banks have the same quality of information about them for evaluating their money-laundering and terrorist-financing risks.

In summary, to address the problems associated with banks as anti-money-laundering gatekeepers, a consolidated strict liability regime for filing failures, coupled with an information scope cap, can facilitate negotiation and sharing without promoting harm displacement. Furthermore, a reduced penalty amount can mitigate the drawbacks of de-risking. Lastly, subsidizing unbanked or under-banked individuals can help alleviate the externality of poverty.

3.5.3 Platform liability and data pollution

Online platforms are ubiquitous in this digital era. They host content provided by users and facilitate communication between them. For instance, Twitter and Facebook allow users to post their ideas online and share them with friends or even strangers around the globe. Amazon and eBay provide marketplaces connecting sellers and buyers. YouTube creates

a platform for creators to profit from uploading videos and attracting viewers' attention. Simultaneously, such online platforms can foster criminal activity. For example, Twitter and Facebook may be filled with defamatory, terrorist-supporting (Cohen-Almagor 2017), or hate speech. Amazon and eBay can enable transactions of defective, counterfeited, or contraband goods (Dinwoodie 2014; Janger and Twerski 2020). YouTube can host content that infringes other creators' copyright (see e.g., Elkin-Koren 2014) or even harms teenagers.

Due to their position regarding user misconduct, they are often called upon to take responsibility and proactively enforce the law. They are required not only to passively comply with the notice-and-takedown regime but also to be active players in proactive screening and removal in copyright areas (Frosio 2017-2018). Outside of copyright, people are challenging the long-standing immunity granted by Section 230 of the Communications Decency Act, resulting in two Supreme Court decisions in Section 3.1.

How do platforms gatekeep

Considering the current trend, this section projects the cost of gatekeeping for online platforms when they are held liable and assesses which liability regime best suits them as gatekeepers. First, online platforms can remove illegal content uploaded by users, such as posts, pictures, or videos. Their gatekeeping approach is simple; like other gatekeepers, they must collect information to distinguish between law-abiding and illegal activities and content. For instance, for YouTube to identify copyright-infringing content, it must first know the existence and the nature of copyrighted works. Similarly, to identify posts concerning terrorism, child pornography, and sexual abuse, Twitter needs to understand both the content of the posts and the definitions of these terms to evaluate whether such posts are illegal and should be removed.

In addition to the content information, platforms can collect further data. For example, eBay could gather more information about the seller to determine whether a product infringes

on a trademark. eBay can examine the transaction history, price, or even the location and occupation of the seller to evaluate risk. Similarly, Google could learn about users' private lives to identify potential wrongdoers. For instance, Google might have acted differently in the example where it banned the father's account, had it known the baby in the photo was the sender's son and that the conversation aimed to solicit medical advice.

Platforms' gatekeeping cost

Platforms, when designated as gatekeepers, can incur several gatekeeping costs. They might fail to internalize the cost borne by counterparties, resulting in inefficient over-enforcement. Their gatekeeping actions can also impose a societal cost.

First, it is unsurprising that platforms can over-enforce when designated as gatekeepers. We already witness over-removal in the context of online platforms liable for copyright infringement (Bar-Ziv and Elkin-Koren 2018; Gabison and Buiten 2020; Erickson and Kretschmer 2020, p. 108), and Google's self-imposed, non-legally required over-gatekeeping. This over-gatekeeping can be attributed to market structures where platforms often enjoy significant market power due to their network effect, potentially facing less competition than other gatekeepers.

Moreover, for platforms to identify and remove illegal content, they must collect information that, in aggregate, could cause harm. We already see how the vast data collected and shared by Facebook can harm democracy (Granville 2018-03-19). If platforms are designated as gatekeepers and are held liable for more precise enforcement, they may extract more information from users. This data generation can exacerbate the so-called "data pollution" problem (Ben-Shahar 2019), imposing a cost on individuals other than the information provider who does not internalize this cost. For example, adult websites could ask visitors to share access to their Facebook/Instagram friend list to confirm the visitor's age. Law-abiding visitors may find it rational to avoid the hassle of clicking "I am over 18" or

scanning their photo ID.[16] However, by releasing the friend list, such websites may gather additional information to tailor their recommended videos to the users. This information can be used for customized experiences targeting the user's friends without their consent. Even worse, such information, combined with existing data, could be reverse-engineered to predict visitors' political affiliations (Markey and Markey 2010). If misused, this data can make disinformation more effective, potentially harming democracy (Shen 2019-11-16).

Liability regimes for gatekeepng platforms

Based on the analysis of liability regimes and the diagnosis of platforms' gatekeeping costs, I advocate for strict liability for two reasons: bargaining between parties and the uncertainty surrounding the level of due care.

First, while act-based and negligence-based regimes can ideally mitigate third-party costs (see Hamdani 2002, pp. 936–38), they might prevent parties from direct bargaining, thus exacerbating the problem of over-gatekeeping costs borne by counterparties. In this context, a strict liability regime can at least provide platforms and users with opportunities to bargain and avoid inefficiencies of over-gatekeeping. Namely, users can negotiate with platforms that remove their content or deactivate their access if their (lawful) benefit outweighs the liability faced by platforms. For instance, under strict liability, the father could pay Google for a thorough review or immunize Google from liability to reclaim his account. However, Google might not be willing to do this if it is held negligently liable for misconduct with a higher-than-harm penalty.

Second, the widespread use of machine-learning algorithms in gatekeeping can prevent both the court and the platform from determining the appropriate level of due care. Machine-learning algorithms can be a "black box" that makes the result-generating process unexplainable. Under such circumstances, courts may not know how to define due care. More tech-

16. For instance, Utah recent passed the law requiring porn websites to verify visitors' age, probably by checking their IDs (Metz 2023-05-03).

nically, can the court determine how to set a penalty parameter used in machine-learning algorithms to optimize the tradeoff between biases and variances (see generally James et al. 2013, pp. 33–37)? Alternatively, platforms need information about what content is illegal and how harmful it is to improve their algorithms. This information can be lacking under the negligence rule because platforms are not always liable for every piece of illegal content.

Admittedly, a strict liability regime that facilitates bargaining cannot address the problem of data pollution externalities. For this, we need to apply a Pigouvian tax regime, as Ben-Shahar (2019, pp. 138-43) suggests, to online platforms that actively collect data, regardless of whether it's for gatekeeping or other purposes. Pigouvian taxes are better suited for the adjustment of penalty amounts discussed in Section 3.3.3. Reducing the penalty amount is expected to lower the level of gatekeeping behavior. However, platforms collect data not only to mitigate liability but also for profit reasons. Therefore, while a liability reduction can alleviate over-collection resulting from liability, it cannot fully solve the problem of data pollution. Sometimes, the liability is far lower than the harm caused by data collection. In this case, the liability reduction cannot entirely address data pollution. After all, if a Pigouvian tax is employed, it can work more effectively than a reduced liability and should replace it due to administrative cost considerations.

3.6 Conclusion

This chapter compares gatekeepers and precaution-taking victims. Gatekeepers can be perceived as indirect "victims" because they, too, suffer harm in the form of legal liability when those they supervise commit misconduct. Consequently, their behavior can parallel that of victims. Both gatekeepers and victims adjust their level of precaution in response to the potential for misconduct in the real world, a factor influenced by the intensity of law enforcement. Furthermore, the precautionary measures undertaken by both gatekeepers and victims can affect non-wrongdoers, such as their counterparts (in the gatekeeper context)

or third parties. These two findings — gatekeeping incentives in response to law enforcement and the burden of gatekeeping costs on non-wrongdoers — constitute the pillars of this chapter.

By comparing these entities, this chapter draws insights from the literature on victim precaution and adds nuance to the framework developed by Kraakman. More specifically, gatekeeping incentives fluctuate with the intensity and focus of direct enforcement against wrongdoers. When direct enforcement improves detection, gatekeepers' efforts to interdict are encouraged because their failures are more likely to be detected as more wrongdoers would be discovered. However, while reporting gatekeepers are also more likely to be detected, the marginal deterrence benefit diminishes as direct enforcement enhances detection, leading them to reduce reporting. Consequently, the effect becomes ambiguous. In contrast, when direct enforcement heightens the imposed sanction to conserve investigation costs, gatekeepers' incentive to interdict weakens, and the effect on their incentive to report remains uncertain for the same reasons. Efforts to interdict misconduct yield no deterrent effect when wrongdoers, if interdicted, will not be punished for their inchoate misconduct. Additionally, fewer wrongdoers are detected due to a decreased detection rate. In conclusion, interdicting gatekeepers invest less when direct enforcement raises sanctions. The trajectory of reporting efforts, however, remains unclear.

Apart from the effect of law enforcement strategies on gatekeeping incentives, the comparison also highlights the often under-appreciated costs of gatekeeping. Gatekeeping can be expensive, and these costs are not solely borne by gatekeepers. When multiple gatekeepers compete, they can incur redundant gatekeeping costs, which is socially inefficient. Moreover, gatekeepers may seek to divert wrongdoers to avoid liability or reduce compliance costs, where the cost is not justified by the social benefit. Gatekeeping can also impose costs on unrelated third parties in various contexts. For instance, within the realm of corporate criminal liability, corporate employers may take precautions by refusing to hire ex-offenders.

This practice could lead to an increase in recidivism, negatively impacting communities. For banks acting as anti-money-laundering gatekeepers, they may turn away under-banked individuals, undermining financial inclusion and potentially leading them to poverty. When online platforms are held liable for their users' misconduct, they are likely to collect more personal data to identify illegal content. Such data collection can pose risks of data pollution.

Understanding the incentives and costs of gatekeeping allows this chapter to add detailed information for the design of gatekeeper liability. This chapter discusses how to construct gatekeeper liability with considerations of required acts, liability regimes, and penalty amounts. With a deeper understanding of gatekeeping costs, this analysis can offer guidelines for the design of gatekeeper liability. Using the three examples, this chapter provides policy recommendations to reform liability regimes to address issues of ex-offenders' unemployment due to corporate criminal liability, decreased financial inclusion because of AML laws, and data pollution exacerbated by platform liability.

CHAPTER 4

HOLDING RIGHTSHOLDERS ACCOUNTABLE: TACKLING THE OVER-REMOVAL PROBLEM ON ONLINE PLATFORMS

4.1 Introduction

In 1998, the U.S. Congress passed the Digital Millennium Copyright Act (DMCA). The Act establishes the notice-and-takedown (NTD) regime, allowing copyright holders (rightsholders) to request that platforms (platforms) remove infringing content uploaded by users (uploaders). However, this NTD regime is often abused, resulting in the problem of over-removal in which legal content is wrongfully removed, accidentally or intentionally, by actual or (non-)rightsholders.

According to YouTube, approximately five percent of takedown requests filed in the first half of 2022 via their public web form were found to be wrongfully submitted (YouTube 2023a, p. 6). Acting on some of the takedown requests resulted in public anger or litigation. On July 10, 2022, Lofi Girl, a famous YouTube livestream channel with over ten million subscribers (Astley 2022-07-21), was taken down due to a complaint filed under the DMCA (Perrett 2022-07-11). AThe response was swift, with loyal fans criticizing YouTube's erroneous decision in more than 3,500 tweets (Liu 2022-07-12). Responding to public anger, YouTube confirmed that the "takedown requests were abusive," restored the account, and terminated the account from which the requests were filed (Perrett 2022-07-11). This incident was not the first faced by Lofi Girl, nor for YouTube, nor the broader online platform industry. Lofi Girl faced a similar takedown in 2020, which also proved wrongful.

Some erroneous takedowns led to litigation. In 2007, Ms. Lenz uploaded a video of her child dancing with Prince's song, "Let's Go Crazy" playing in the background in *Lenz v. Universal Music Corp.* (2013 WL 271673 at *1). The video was removed at the request of the rightsholder, Universal Music. The video was only reinstated after Lenz contested the

takedown. Lenz then filed a lawsuit against Universal Music, though ultimately she was not fully compensated for the erroneous removal (*Lenz v. Univesal Music Corp.* 815 F.3d 1145, 1157 [2016]). The issue is not confined to copyright infringement or media platform such as YouTube. For example, recently Amazon discovered that some of its users were abusively filing takedown requests targeting legitimate rightsholders for illegitimate reasons (Jahner 2023-03-30).

The examples above showcase how the NTD regime can be abused. Rightsholders often file too many takedown notices without verifying their requests, leading to the erroneous removal of legal content. Unfortunately, instead of guarding their own and uploaders' interests, platforms often comply with such inaccurate notices out of fear of liability. Worse yet, uploaders rarely fight back against takedown requests, and those fail to reverse erroneous removals. While scholars have identified this as a problem and proposed various solutions since the passage of the Act, the development of technology and the industry have both encountered rapid changes. These changes, particularly the automatic enforcement and monetization function, inspire the provision of novel solutions in this chapter. Specifically, automation, particularly the algorithms behind automation, allows platforms to have a better idea of potential revenue brought by a piece of content, and the monetization function offers the possibility for both rightsholders and uploaders to bargain, avoiding inefficient takedowns.

The literature has already identified the problem of over-removal and proposed several solutions. Scholars have provided evidence that the quality of takedown notices is unsatisfactory, with between 10% and 40% of takedown notices filed by rightsholders being problematic. Most empirical evidence focuses on the number of erroneous takedown requests. However, emphasizing the number, rather than the cost, can be misleading. In fact, there can be uncertainty with neither party able to determine conclusively whether or not a piece of uploaded content is infringing. Therefore, erroneous takedown requests are, to some ex-

tent, inevitable. What matters is not the amount of legal content removed but whether costs incurred from removal are justified by the expected reduction in harm from potential infringement. From the perspective of social welfare, our goal should not be to unconditionally reduce the amount of removed legal content, but rather to avoid the removal of content for which the benefit derived from avoiding infringement harm is outweighed by the cost of erroneous removal. As such, the problem of "over-removal" should be reframed as "inefficient over-removal."

In light of this understanding, I argue that we ought to encourage parties to make socially optimal takedown decisions. Optimal removal requires two conditions. First, parties should invest in verification to ensure that legal content is identified and excluded from the takedown requests. Second, given conditions of uncertainty, the takedown decision itself should be efficient, meaning that the benefit from preventing infringement will outweigh the cost of erroneous removal. Unfortunately, both directions of decision-making necessitate information: the social value of content should be devised for the former, and the private harm inflicted by infringement is necessary for the latter.

Unsurprisingly, the current good-faith requirement does not incentivize parties to make optimal decisions in both directions. The good-faith requirement fails to make rightsholders internalize the full cost of errors and to induce optimal verification, leading to the filing of too many takedown requests targeting legal content. Platforms over-comply because they do not fully internalize the harm removal imposes on uploaders, and possibly overestimate the harm done to rightsholders due to the repercussions of imperfect information. Finally, uploaders may not want to challenge takedown requests, either by filing counter-notice or litigation, because they have no idea whether the requests meet the legal standard of good faith action.

Through the lens of bi-dimensional takedown decisions, previous scholarly proposals to address the problem seem inadequate. Most narrowly and individually focus on each player

in the takedown process instead of collaboration between them to make optimal takedown decisions. Specifically, they either adjust platform liability to induce platforms to make efforts to scrutinize takedown requests, empower uploaders to challenge or sue wrongful rightsholders making erroneous takedown requests, or penalize rightsholders for unscrupulous requests. However, these proposals cannot effectively address the problem of parties' imperfect information and neglect the potential upsides of pushing parties to collaborate.

In this chapter, I propose holding rightsholders strictly liable for costs associated with erroneous removal. Moreover, I suggest rightsholders disclose their committed maximum liability (CML) with their takedown requests. This regime has three advantages. First, being held strictly liable, rightsholders are induced to invest in verification and reduce the possibility of legal content being flagged for takedown. In addition, strict liability ensures that takedown decisions are more efficiently made than under the negligence rule or good faith requirement. Strict liability also encourages uploaders to file litigation against rightsholders who file erroneous takedown requests. Second, the committed maximum liability allows for the utilization of information regarding content value, avoiding inefficient takedowns. By specifying CML, platforms will only remove content if the value is lower than the specified CML. On the one hand, if erroneously removed, all low-value content can secure compensation from the rightsholder. In contrast, high-value content will remain online because the value lost will not be fully compensated. This mechanism prevents high-value content from being removed and, therefore, inefficient takedown decisions. On top of the proposed regime, this paper also evaluates and considers other liability regimes, such as the negligence rule, charging a modest filing fee, and joint liability between both rightsholders and platforms.

This chapter makes three contributions to a new understanding of the over-removal problem. First, epistemologically, it challenges the common presumption of "over-removal." Previous empirical evidence tends to validate the over-removal problem by showing the number of incorrect takedown requests. Nevertheless, under uncertainty, it is costly, if not impossible, to

avoid type-I errors. Hence, what should be a greater concern is the cost of erroneous removal and benefit of preventing harm infringement. Second, based on epistemological reframing, a conceptual framework is provided to approach the problem. It is argued that socially inefficient over-removal should be improved in two dimensions: verifying (non-)infringing content and taking down content, corresponding to the level of care and the level of activity, respectively. This bi-dimensional decision-making introduces both problems of misaligned incentives and information asymmetry. Third, I provide concrete policy implications, I substantiate potential legal reforms to better hold rightsholders accountable and reduce socially inefficient removal.

This chapter consists of four sections. Section 4.2 provides a background introduction that reviews the current notice-and-takedown regime stipulated in Section 512 of the DMCA and the recent developments in practice. In addition, I also compare the notice-and-takedown regime with other legal regimes adopted outside the U.S. to evaluate whether other legal regimes are superior. Section 4.3 turns to the long-lasting concern of the abuse of the notice-and-takedown regime: over-removal. Reviewing the literature reveals how the problem emerges and why existing proposals may not satisfactorily address the issue. Section 4.4 lays out the argument for holding rightsholders accountable and the directions of reforms. Section 4.5 substantiates the concrete reform proposals.

4.2 An Overview of Legal Regimes in Online Copyright Enforcement

In this section, specific attention is paid to the current NTD regime in the United States, namely Section 512 of the DMCA, and its practical implication. After going through the legal details, the practices currently adopted by platforms, including automated enforcement and monetization function, are reviewed. Finally, there is a comparison of the NTD regime with other well-known regimes, such as the notice-and-notice (NN) and the notice-and-staydown

(NSD) regimes. This section then paves the way for further analysis in the following sections.

4.2.1 An overview of the notice-and-takedown regime of DMCA

Section 512 of DMCA stipulates that rightsholders can file takedown requests to have hosting platforms remove allegedly infringing content. Upon receipt of such requests, the platform removes the content should notify the uploader. The uploader can choose to file a counter-notice, explaining why the content is not infringing. Such a counter-notice will be forwarded by the platform to the requesting rightsholder, who may decide to file a lawsuit. Failing to do so will result in the content being restored.

Congress drafted Section 512 of the DMCA in the era of nascent Internet industries. At that time, copyright holders were worried that copyright on their work would be infringed and disseminated online; online service providers (OSPs), in turn, were concerned about the uncertainty of liability (United States Copyright Office 2023, p. 8). To strike a balance and encourage cooperation between content and Internet industries, Congress crafted section 512 with a dual purpose: provide tools for copyright owners to address online infringement, and to build a friendly environment in which OSPs could thrive without fear of liability for their users' infringement (see S. Rep. No. 105-190, at 40 [1998]). In summary, Section 512 provides safe harbors to qualifying OSPs with the condition that they cooperate with rightsholders' enforcement actions

To qualify for safe harbor protection, an OSP must first engage in at least one of the following activities stipulated in Sections 512(a) – (d). These activities are (a) mere conduit: an OSP automatically transmitting materials with third parties' direction; (b) "system caching": temporarily store material for third parties' transmission via the Internet; (c) "information residing on systems or networks at direction of users", such as Dropbox; and (d) "information location tools", such as Google search (p. 23).

OSPs engaging in qualifying activities other than being mere conduits must also maintain

a notice-and-takedown process, in which they must, upon receipt of proper notice from a copyright holder or its authorized agent, expeditiously remove or disable access to the alleged infringing material or be subject to secondary liability for their users' infringement.

Notice is properly effective only when meeting the requirements specified in section 512(c)(3)(A), including (i) signature, identification of (ii) the alleged infringed copyrighted work and (iii) the alleged infringing material, (iv) contact information to notice filer, (v) a statement of good-faith belief in infringement, and (vi) a statement of accuracy. OSPs are only required to respond to a proper notice that complies with aforementioned requirements. Receiving an improper notice does not constitute an OSP's actual knowledge of infringement and does not constitute a basis for establishing liability.

An OSP, when acting in good faith to remove content at the request of the rightsholder, is also immune from liability if it notifies the subscriber (i.e., users) of the removal and forwards the subscriber's counter-notification to original requesting rightsholder (512(g)(1) and (2)). When conveying the counter-notification, an OSP should also inform the requesting rightsholder that it will restore the content in ten business days. Receiving the counter-notification forwarded by the OSP, the requesting rightsholder can file a lawsuit against the subscriber, but if it fails to do so the alleged content will be restored. If the requesting rightsholder does file a lawsuit, the content will remain unavailable on the platform. The counter-notice procedure involves a set of formal requirements similar to the takedown procedure, including (A) signature, (B) identification of the removed material, (C) a statement of good faith belief under penalty of perjury, and (D) contact information with consent to federal jurisdiction. (512(g)(3)).

In summary, the NTD process specified in the current DMCA has five steps:

1. The rightsholder files notification to remove alleged infringing content;

2. The platform removes the content and notifies the uploader of the content;

3. The uploader files a counter-notice;

4. The platform notifies the requesting rightsholder of the counter-notice; and

5. The requesting rightsholder files a lawsuit, otherwise the content is restored.

4.2.2 NTD in practice

In the last section, I presented a brief introduction to the current NTD regime. However, this describes how the law expects players to act. In practice, players may behave differently from what the law expects in response to technological developments, the market, and the political environment. In the following paragraphs, I will highlight three main practices embedded in the notice-and-takedown regime. Each of these practices can play a significant role in our policy evaluation in the upcoming sections.

Automated enforcement

The first practice worth mentioning is the spread of automated enforcement. Automated enforcement encompasses attempts to automate the process of enforcing copyright, including identifying infringing content, the filing of requests by rightsholders, and the review or even proactive filtering adopted by online platforms. In what follows, these uses of automation in the notice-and-takedown process are discussed.

Rightsholders, in particular rights enforcement organizations (REOs), have begun relying on automated systems to identify potentially infringing content and file takedown requests. Evidence shows that the majority of takedown requests are issued by high-volume rightsholders utilizing automated website tools (Erickson and Kretschmer 2020, p. 110). Aside from developing their own in-house automation systems, rightsholders can also resort to third-party REOs who use automated systems to file takedown requests after discovering potentially infringing content (Urban, Karaganis, and Schofield 2017b, p. 374). As the scope of rights possessed by rightsholders increases, and there is large scale infringement, rightsholders are more likely to adopt automated systems (p. 379), probably for economic reasons.

Unsurprisingly, when automated system are as yet not well-developed, requests filed under such a system will often be problematic. Records show that nearly 10% of takedown notices received between 2011 and 2015 were erroneous (Seng 2015, p. 3). High volumes of flawed takedown requests invite platforms to invest in automated systems. While platforms' automated review systems save on human labor, they can suffer from much the same flaws as rightsholders' automated systems. Consider takedown requests from 2011 to 2015, for example. Among erroneous takedown requests, Twitter rejected 59% of them were rejected by Twitter, while the rejection rate was only 2.5% under Google's automated review. Of course, the essence of the services provided by both platforms may, to some extent, explain the disparate rejection rate. However, some of the disparity can be attributed to the review process, which is automated for Google but remains manual for Twitter (p. 3). Of course, the use of automated systems helps detect potential infringement at a lower cost than manual procedures, but such systems are not a free lunch — they comes at some cost to accuracy (Depoorter and Walker 2013, p. 326). Admittedly, the precision of automated systems can improve as the technology develops and algorithms learn more from the growing data. However, the disparity still showcases that automated systems can engage in overkill and exacerbate the over-removal problem.

Platforms not only rely on automated systems to passively review takedown requests filed by rightsholders, but also use them proactively to filter out potentially infringing content. For instance, YouTube, followed by Vimeo, developed proactive filtering systems, Content ID and Copyright Match. These systems filter out content uploaded by users matching rightsholders' files in their database (Bridy 2016, p. 191). On its face, people may wonder why platforms want to invest considerable resources in developing filtering systems that are beyond what is legally required (Urban, Karaganis, and Schofield 2017b, p. 399). However, it is less surprising when we trace the origin of such investments. These platforms began developing such systems after facing high-stake lawsuits: YouTube launched Content ID after

being sued by Viacom, and Vimeo its project after receiving a summary judgment ruling in litigation with Capitol Records and other plaintiffs (Bridy 2016, p. 195).

As proactive filtering is not required by the current notice-and-takedown regime, such voluntary practices which go beyond what is required are often dubbed "DMCA+." Aside from fear of legal liability, political pressure motives platforms to adopt such voluntary measures. Platforms may strategically adopt voluntary measures to reassure governmental or regulatory bodies to avoid the imposition of more stringent legislation or regulation (p. 186). That said, proactive filtering itself is not risk-free. Adopting proactive filtering can furnish platforms with the fact of infringement, thus exposing them to the legal risk of losing immunity. Therefore, scholars have called for explicit affirmation of such proactive measures to encourage their adoption (see e.g., Buiten, Streel, and Peitz 2020, p. 163).

Proactive filtering prevents infringement from harming rightsholders in the first place, but is not neutral to other stakeholders. Voluntary measures motivated by legal liability and political pressure inevitably favor rightsholders, in particular those who manage massive copyrighted works, and disadvantage uploaders and other stakeholders (Bridy and Keller 2017, p. 17). Even for rightsholders, automated enforcement can have disparate impacts. For instance, while large rightsholders rely on automated systems to enforce their rights, middle-class artists, who lack the means or access to such systems, may find tackling online infringement challenging (Cusey and Carrington 2016, p. 2).

Reliant on automation, proactive filtering may suffer the same inaccuracies seen in rightsholders' filing and platforms' review. Not only are the errors committed by automated systems more difficult to identify and avoid, but the decision-making processes of such systems are often ill-suited to legal issues (Urban, Karaganis, and Schofield 2017b, p. 386). Such inaccurate judgments lead to erroneous removal more severe than under the current legal regime and impose greater harms on uploaders (Bridy 2016, p. 187).

Besides middle-size rightsholders and uploaders, DMCA+ practices also affect other plat-

forms and the industry as a whole. Implementing DMCA+ practices often come with high initial costs, and DMCA+ practices have therefore not yet been widely adopted across all platforms (Urban, Karaganis, and Schofield 2017b, p. 399). On the contrary, only a relatively small number of platforms have implemented such practices in their daily operations (p. 383). However, a larger number of platforms that have not yet opted into proactive filtering worry that legal standards may be strengthened to match expectations set by DMCA+ practices (p. 398). Although the legal standard remains, as yet, unchanged, the adoption of such practices may invite further enforcement actions by rightsholders in the form of a greater number of takedown requests (p. 399).

Such matters concern both existent platforms and new market entrants as the high initial cost of implementing such practices becomes a new entry barrier (p. 399). Take Content ID for example. As the system becomes more mature, the database will contain more content provided by both uploaders and rightsholders, improving the precision of the content matching system. Such advantages may dwarf the competitive edge of new entrants, and have anti-competition implications for the market (Bridy 2016, p. 208).

In sum, the current use of automated enforcement and DMCA+ practices warrant further discussion and evaluation. To ensure such practices are socially optimal, cost-benefit analysis should at least consider the interest of those previously under-included groups of stakeholders, such as uploaders (p. 208). While the main focus of this paper is not to evaluate whether or not such practices are laudable, the discussion above is insightful: practices that benefit copyright enforcement might come at the expense of other stakeholders.

Monetization of content

Another trend worth mentioning is the monetization of content by rightsholders. With the help of Content ID developed by YouTube, rightsholders who are notified of uploaded content that matches their copyrighted content can choose to file a takedown request or monetize

the content YouTube.. Given the existing notice-and-takedown regime, most rightsholders seem to not resort to filing takedown requests, but rather choose monetization, claiming the revenue generated by advertising against the matched content (Bridy 2016, p. 196).

Furthermore, the selection of monetized content is not random. Research finds that rightsholders often elect to monetize popular content with high-production values, and request the removal of low-volume content (Erickson and Kretschmer 2018, p. 87). One study by Erickson and Kretschmer (p. 87) finds that "[h]igh-quality and popular parodies might remain live on the platform." They argue that their choice is conscious and rational. Rightsholders make such monetization-or-takedown decisions by weighing the potential negative effects of substitution and the benefit of monetization (p. 87).

The monetization function directly challenges the notice-and-takedown regime. Imagine an extreme case in which all alleged infringing content is monetized rather than removed. In such a situation, the notice-and-takedown regime has been superseded. To be sure, this is not now the case. There is a rational separation between content that is monetized and that is removed. It is such rational decision making that renders the sample of content or to-be-removed, or not, unrepresentative. Content that is flagged for removal, compared to that which is monetized, is more likely to be low-traffic, less viewed, and thus the real cost of removal will be lower than perceived. With this possibility in mind, the concern of over-removal, at least on platforms with monetization functions such as YouTube, should be less problematic.

There are two implications of analysis above. First, as a result of monetization the content creates more value than harm, and the opportunity to seize that value persuades the rightsholder to accept monetize rather than seek to remove the content. Such content, if removed, is socially undesirable. However, without the choice of monetization, rightsholders would have no choice but removal. The existence of the monetization function unearths the possibility of bargaining among rightsholders, platforms, and uploaders, so that parties can

share in the revenue and avoid inefficient takedowns. Monetization provides a novel solution to the problem of content removal, facilitating bargaining and avoiding inefficient takedowns.

Second, while monetization seems a promising means of addressing inefficient takedowns, it does not fully eliminate the problem. Instead, it may simply discourage content creation by allowing rightholders to claim the profit from uploaders' efforts. Under this system, profit derived from legal, non-infringing content, can be erroneously monetized and flow to the rightsholder, diluting the uploader's incentive to create new content. Moreover, the monetization process itself can be technically burdensome to adopt for small platforms. For instance, YouTube currently allows only Content ID claims to monetize the content identified by their Content ID system (YouTube 2023b), which is expensive and unaffordable to other platforms. Likewise, such a monetization function may not be widely available for rightsholders whose content is not yet in the Content ID system. Hence, while such a function seems promising, it may not, at this point, be a solution to the over-removal problem.

4.2.3 Other regimes compared

Before moving to the analysis of the over-removal problem in Part II, two other existing legal regimes, in other jurisdictions, should be introduced and compared. These regimes are the notice-and-notice regime in Canada, and the notice-and-staydown regime in Europe. After introducing both regimes, a brief comment is offered on why neither regime is particularly well-suited for solving the over-removal problem.

Notice and notice

Unlike the notice-and-takedown regime stipulated in the DMCA, platforms (formally defined as providers of "Information Location Tools") are not required to take down allegedly infringing content according to the Canadian Copyright Act. Instead, platforms are only obliged to forward the notice filed by the rightsholder to the uploader (Copyright Act (R.S.C. 1985,

c. C-42) 41.26(1)(a)). Platforms are also obligated to keep records to help identify infringers for future litigation (41.26(1)(b)). To take down the alleged infringing content, the rightsholder must secure a court order (Kuczerawy 2020, p. 537). Even though injunction relief is possible, the scope is limited by the effect on platforms (41.27(4.1)(b)). The Chilean regime follows the same approach, in which takedown is only possible with a court order (see generally Center for Democarcy and Technology 2023, p. 3). While each jurisdiction varies in the details in its notice-and-notice regime (Kuczerawy 2020, p. 537), the core characteristic here is that platforms are not required to remove allegedly infringing content.

The notice-and-notice approach shifts the evaluation of, and responsibility for, takedowns from platforms to courts. Such a regime, compared to the notice-and-takedown regime, avoids the over-removal problem and is more likely to protect free speech and preserve other values. Nevertheless, this system can also impose a great burden on the judiciary. Take YouTube for example, there are, on average, more than 37,000 takedown requests (excluding from Content ID) every day. Even if 10% of takedown requests appear on the court's docket, this would constitute an untenable burden on the court. Moreover, court proceedings are time-consuming, and securing an injunction against an uploaded may take weeks, if not months, and such delays may render a judgment moot.

Notice and staydown

More radical than the notice-and-takedown regime is notice-and-staydown. The notice-and-takedown regime is sometimes criticized as being too passive and unsuited to incentivizing platforms to identify known illegal content (see e.g., Mann and Belzley 2005, pp. 270–71). That is, under the notice-and-takedown regime, when the rightsholder identifies infringing content and files a takedown request, the platform is obliged only to take down the identified content. When the same content is again uploaded, the platform need not act unless the rightsholder again file a takedown request.

The notice-and-staydown regime requires the platform to monitor the content on their platform and to ensure that the same infringing content is not uploaded again (Kuczerawy 2020, p. 538). Such a regime was introduced by the new Directive in the Digital Single Market in Europe (see Directive 2019/790/EU and Directive 96/9/EC). Under Article 17 of the Directive, platforms only enjoy safe harbor protection if they remove infringing content and "ma[k]e best efforts to prevent [its] future uploads." (see Directive 2019/790/EU and Directive 96/9/EC) This notice-and-staydown regime effectively imposes an obligation to monitor future uploads. The implementation of the Directive depends on national legislations' incorporation as well as courts' interpretations (p. 538).

An issue that might trigger concern is the scope for monitoring future infringing content. In short, what constitutes "future uploads" of alleged infringing content? Regarding this issue, Kuczerawy believes that the monitoring obligation can lead to the prevention of "similar" infringements (p. 538). The broader scope of prevention could effectively construct a positive obligation of general screening and monitoring. Nevertheless, German courts and scholars believe that such an obligation to prevent similar infringing content from being uploaded again is a specific rather than a general monitoring obligation. Therefore, the interpretation is compatible with the e-Commerce Directive (Nordemann 2011, p. 42).

Scholars have criticized the notice-and-staydown regime on cost-benefit grounds. Bridy and Keller (2017) argue that the cost of introducing screening, and the erroneous removals associated with that screening, can be too dear to society, and is not justified by the potential benefits accruing to rightsholders (p. 18). Indeed, the screening requirement, together with the expansion to similar infringing content, can work as effectively as the DMCA+ practices illustrated above. In other words, the screening requirement mandated by the notice-and-staydown regime can again invite problems encountered under the DMCA+, and the over-removal problem grow even more severe. Moreover, the staydown requirement can prevent rightsholders from monetizing content or bargaining with the uploader, therefore losing the

opportunity for an efficient "stay up" option. Again, rightsholders consciously choose to monetize some high-volume content, which may be related to who uploads it rather than what is uploaded. Therefore, a content-based staydown requirement may prevent efficient bargaining over the similar potentially infringing content between popular uploaders and rightsholders. Finally, the technologies required for staydown requirement can similarly play a role as an entry barrier that has an anti-competition impact on the market.

Remarks

The primary distinction among the aforementioned regimes is the platform's response to notices filed by rightsholders. The degree of the response, from light-touch to severe, ranges from forwarding a notice to removing alleged infringing content and preventing the content from again being uploaded.

Concerning the cost of the over-removal problem, it is evident that the notice-and-notice regime has minor concerns related to over-removal. However, it might not be the most cost-effective method as it fails to prevent immediate harm from infringement. Conversely, the notice-and-staydown regime exacerbates over-removal concerns faced by the notice-and-takedown regime. It not only prevents similar content from being uploaded, but deprives parties of the opportunity to bargain.

There seems no strictly better shift from one regime to another. Nevertheless, the discussion and evaluation above remain meaningful. They highlight the limitations in effectiveness when designing platform responses and encourage researchers to explore other possibilities. In fact, the varying responses from platforms may not help tackle the current problem. The choice among these regimes inevitably focuses on the desirable reaction of platforms. However, platforms' reactions may not be the only source of the problem, nor even the primary one. Hence, this paper should pause here and refrain from focusing on the details of each regime. On the contrary, a more fruitful way is to zoom out and take a broad perspective to

consider the incentives of all relevant parties, not only the platforms. In the next section, the problem of over-removal is clarified and dissected with a brief review of existing proposals to prepare for my own reform proposals.

4.3 The Problem of Over-Removal

This section demonstrates why over-removal is a social problem in Section 4.3.1, and then investigates the origin of the problem from an economic perspective in Section 4.3.2. Section 4.3.3 offers a review and evaluation of previous solutions in the literature. Finally, the problems that remain unsolved by previous proposals are laid out.

4.3.1 Social costs of over-removal

Over-removal has been a concern for the NTD regime since the passage of the DMCA. While it is labeled "over-removal," the form is not limited to removing alleged infringing content, but can also include other dispositions that make the alleged infringing content less accessible, such as down ranking the links to infringing content (Urban, Karaganis, and Schofield 2017b, p. 396). While down ranking does not completely remove the infringing content from the platform and is sometimes criticized as useless for popular websites (p. 396), it can be understood as a measure making certain content less visible to users. To some extent, the impact imposed by down ranking content is not qualitatively, but quantitatively, different. Hence, the following discussions of harm brought by over-removal can apply to the case of down ranking.

In comments on Section 512 prepared by Bridy and Keller (2017), three potential harms of removal are enumerated, ranging from harm to uploaders, the market, and the Internet ecosystem. First, and most directly, removing content inflicts harm on uploaders. The harm can be economic (p. 3). When a clip is removed from YouTube, the uploader cannot profit from the views and associated advertisement revenue. Similarly, on a platform such

as Amazon, when a seller is groundlessly accused of infringing trademarks and erroneously removed from the website, the seller cannot sell further goods online. The harm can also be non-monetary. For instance, in the *Lenz* case, the plaintiff sought damages for chilling her free speech (see *Lenz v. Universal Music Corp* 2013 WL 271673, at *8 [N.D.Cal.,2013]).

Second, removal can harm consumers. Directly, removing legal content deprives consumers of the opportunity to derive utility therefrom. A more subtle and remote concern relates to market distortion resulting not from removal *per se* but from the required technologies for removal. Bridy and Keller worry that the costly technology that supports removal regimes can deter future investment from new entrants and harm consumer welfare (Bridy and Keller 2017, p. 3).

Third, when removal is erroneous, legal content and speech are harmed (p. 1). Worse, when abused, content may be removed not for copyright issues but for other reasons. For instance, people might ask to remove certain content that complies fully with copyright law for other reasons. For instance, Bar-Ziv and Elkin-Koren (2018) find that during May and October 2013, in Israel, most takedown requests filed to Google were not copyright-related. Given the imperfect detection of erroneous takedown requests, people can disguise themselves as rightsholders and ask to remove the content they dislike. Such abuse may undermine free speech online.

Finally, a rarely discussed harm from over-removal is the dilution of deterrence effect on copyright enforcement. The conventional wisdom of law enforcement already indicates that erroneous enforcement can discount the deterrence effect (Png 1986; Polinsky and Shavell 2007, p. 427). Other things being equal, less deterrent copyright enforcement could induce more copyright infringement. Moreover, as the over-removal problem is specific to the Internet, we can expect the dilution of deterrence is more severe for online copyright infringement than offline infringement.

4.3.2 Over-removal as a social problem

Abundant empirical evidence shows that legal content is removed at rightsholders' request. For instance, in an archival study, Urban, Karaganis, and Schofield (2017 a) sample takedown notices related to Google Web Search from the Lumen13 database (p. 2). They find that 30% of notices are potentially problematic. Also, by surveying platform operators, they share stories of abuse of the NTD regime (p. 2). Out of the United States, researchers rely on the mystery shopper test to find that platforms in the United Kingdom overly remove websites without careful examination of the notices (Ahlert, Marsden, and Yung 2004, p. 26). In Israel, scholars analyze takedown notices sent to Google Search. Surprisingly, while they limit the sample size to those alleging copyright infringement, two-thirds of the notices are, in fact, not related to copyright issues (Bar-Ziv and Elkin-Koren 2018, p. 344).

While such evidence proves that some legal content is removed, it is insufficient to show that such removals are socially undesirable. In fact, no one, even the court, can ensure that no errors are committed when judging which content to remove. What really matters is not the amount of legal content that is removed, but whether the cost incurred for removing legal content is justified by the benefit of stopping the harm done by infringement. As a result, the percentage of legal content to the total requested content does not show that takedown decisions are socially inefficient.

To see this argument more clearly, consider a numerical example. Suppose that a piece of content, if infringing the rightsholder's right, inflicts social harm of $600. In contrast, if it is legal, the value for staying online is $300. The cost of erroneous removal of legal content is thus $300. In this case, removal can still be desirable if two-thirds of the takedown requests targeting legal content because the expected error cost, $200 ($300 \times \frac{2}{3}$), does not exceed the enforcement benefit $200 ($600 \times \frac{1}{3}$). Therefore, if the ratio of legal content to the total requested content is below 66.6%, it is monetarily acceptable.

However, acceptability is not the same as efficiency. Even if the takedown request itself

is justified by its benefit, it can still be improved. Parties can invest in verification that identifies legal content and exclude such from removal. Under the current regime, rightsholders can invest in verifying content to see whether it is legal or infringing instead of wholesale takedown. For instance, suppose the rightsholder can invest $90 in verification to filter 50% of legal content. With such an investment, the rightsholder can further reduce the error cost to $100, which is lower than the cost of verification.

From this illustration, we see that efficient takedowns necessitate two prerequisites: optimal verification and optimal takedowns. The former deals with acquiring more information to filter legal content and avoid the error cost associated with its removal. The latter considers whether a takedown decision is socially desirable, given its optimal verification. When, after verification, the expected cost of error still outweighs the expected benefit from stopping infringement, then the takedown should not be allowed. Conversely, socially problematic over-removal refers to two situations: (1) inadequate verification, or (2) costly takedowns. The optimality of both dimensions hinges on several factors. In concrete, optimal verification depends on the cost-effectiveness of verification and the social value of legal content. Optimal takedown, in turn, depends on the expected error cost of removal based on optimal verification and the expected benefit from stopping infringement. Without information pertaining to the aforementioned factors, we cannot contend that the current practice is, in fact, socially undesirable.

4.3.3 Dissecting over-removal: incentive misalignment and information asymmetry

While lacking direct empirical evidence, we can still analytically investigate whether parties behave optimally under the current regime. In fact, from the evidence on hand, the analytic reasoning is not baseless. Hence, in this section, I will use a simple numerical example to illustrate why under the current regime over-removal is highly likely.

Suppose there are two rightsholders, R1 and R2, each with equal probability. When the content infringes R1's right, it inflicts $100 harm. If it infringes R2's right, the harm is $20. The social value of the content uploaded by U1 is $100, of which 20% is enjoyed by the platform. Instead, the content of U2 creates $25 social value, with $5 enjoyed by the platform. Both U1 and U2 are equiprobable. The chance for the content to infringe one's right is 60%. Namely, there is a 40% chance that the content is legal. To figure out whether the content is legal, parties can choose among different levels of verification with differed prices: low scrutiny helps identify 10% of legal content at $1; intermediate scrutiny identifies 25% at $3; and high scrutiny identifies 40% at $8. The type of rightsholder, i.e., R1 or R2, is only known to the rightsholders. Likewise, the type of uploaders is only known to themselves and the platform. An analysis with formal modeling can be found in Appendix A.

In the subsequent paragraphs, I will analyze each party's decisions step by step, following the framework of Fiala and Husovec (forthcoming). Removal of content is a result of mutual decision-making of three parties: a rightsholder who sends the takedown request, the platform that complies with the requests and removes the content, and the uploader that chooses not to file a counter-notice. A question worth asking is why each party behaves as described above. To answer this question economically, we need to examine each party's incentive to send, comply, and tolerate. Aside from the analysis via the numerical illustration, empirical evidence will be cited to support the assumptions and arguments.

Over-requests by rightsholders

As the first step of the takedown process, rightsholders already file excessive requests. Under the current regime, rightsholders must file takedown requests in good faith. Here, the good-faith requirement is defined as rightsholders investing in a low degree of scrutiny to filter out some obviously legal content. Nevertheless, given the good-faith requirement, they still file excessive takedown requests because they do not fully internalize the cost of erroneous

removal and an inadequate level of required verification. In short, rightsholders' over-filing can be attributed to both suboptimal verification and suboptimal takedowns.

To see this, consider how rightsholders behave when filing takedown requests. Rightsholders file takedown requests when the benefit of stopping infringement outweighs the cost. R1 and R2 must invest \$1 to meet the good-faith requirement. In contrast, the benefit for R1 is \$60 (100×0.6) and \$12 (20×6) for R2. With low-scrutiny verification, 10% of legal content is identified and excluded from the takedown requests. As a result, 90% of the legal content is still requested. The expected cost of error is thus \$22.5 ($\frac{100+25}{2} \times 0.4 \times 0.9$).

Obviously, the verification mandated under the good-faith requirement is inadequate for U1's content. Regarding U1's content, it is still efficient to increase scrutiny to the high level, as the additional cost (\$7) is justified by the reduced social cost of erroneous removal (\$12). Suboptimal verification is not the case for U2's content, as the improvement is inefficient for low-value content. However, when rightsholders have no information pertaining to the type of content and use expected social value instead, intermediate scrutiny should be adopted because the additional cost, \$2 ($3 - 1$) is justified by the reduced social cost of error, \$3.75 ($22.5 - 62.5 \times 0.4 \times (1 - 0.25)$). Hence, the takedown requests already consists of *too much* non-infringing content.

Furthermore, with inadequate verification, R2 should not file the takedown requests, as the cost of error outweighs the enforcement benefit. In concrete, filing a takedown request can incur a \$22.5 error cost, but only yield expected benefit of \$12 (20×0.6). Therefore, it is not efficient to file such takedown requests. Unfortunately, R2 needs not bear the cost of error when he meets the good-faith standard. The good-faith standard cannot induce full internalization of social cost by rightsholders.

Given that rightsholders comply with the good-faith requirement, there is still in their takedown requests 37.5% of the requested content is legal ($\frac{0.4 \times 0.9}{0.4 \times 0.9 + 0.6}$) because the good-faith requirement itself is imperfect. As a result, it is inevitable that some legal content is

included in the takedown requests filed by rightsholders. With that in mind, it should not be surprising that various sources of evidence show that notices sent by rightsholders are far from perfect. For instance, the Urban, Karaganis, and Schofield (2017 a) study shows that around 30% of the takedown request in the sample drawn from the Lumen13 database is problematic (p. 2). Seng (2015) surveys takedown requests filed to Google and finds that 8.3% of them fail to meet the formality specified in the DMCA and 1.3% with substantial errors. A more recent study conducted again by Urban, Karaganis, and Schofield (2017 b) find that questionable requests account for 28.4% and 36.8% of the sample requests regarding Google Search and Google Image Search. In short, imperfect takedown requests may not necessarily be the product of imperfect compliance with the good-faith standard, it can well be the product of the imperfect standard.

The numerical exercise mirrors the real-world scenario where rightsholders focus on enforcement rather than uploaders and platforms' interests. They often pay less attention to facts favorable to uploaders, such as fair use (p. 389), unless required by law. Instead, from their perspective, they rely on counter-notices filed by uploaders or the platforms' intervention, rather than their own investment to avoid errors (p. 385). As a result, even though some rightsholders invest in checking their takedown requests, such as human cross-checking, the level of investment seems inadequate (pp. 385–86). In short, rightsholders have little incentive to take care of uploaders' and platforms' interests. They overly focus on the enforcement benefit brought by the takedown requests.

Moreover, the good-faith standard cannot stop rightsholders filing takedown requests out of copyright concerns. Assume that the harm for R1 is not copyright infringement, but defamatory statements that lead to reputational harm. R1 may still be willing to bear the liability and file takedown requests against U1 and U2 because the reduced harm is no less than the liability. Anecdotal evidence shows that platforms often complain about rightsholders gaming the takedown system for non-copyright purposes, such as gaining a

competitive advantage, silencing critics, etc (Urban, Karaganis, and Schofield 2017b, p. 389). Hence, liability itself cannot fully deter bad-faith takedowns.

To summarize, three factors contribute to rightsholders' excessive filing of takedown requests. First, the good-faith requirement is suboptimal for high-value content. Rightsholders should invest more in verifying high-value content and exclude it from removal. Second, the good-faith requirement is unable to restrain inefficient filing. Once rightsholders filter out some legal content in good faith, they are entitled to remove all content regardless of their reduced harm from infringement and the value of the content to be removed. Third, liability itself cannot prevent intentional abuse. Parties, even not rightsholders, may be willing to bear liability to request removing some harmful but not copyright-infringing content.

Over-compliance by platforms

Facing excessive requests filed by rightsholders, platforms could voluntarily invest in verification to sort out some non-infringing content from the requested content. Nevertheless, due to the under-internalization of the content's social value, the platform's incentive to verify takedown requests is inadequate. Moreover, they may over-comply with rightsholders' requests out of an over-estimation of their infringement harm.

Platforms' over-compliance can be observed via the numerical example. Assume that the platform knows the value of the content. While the platform can voluntarily invest in verification, it may not necessarily find it worth to do so. As the platform only enjoys 20% of the social value of the content, their benefit from verification is lower than the social benefit. The costs of verification and the accompanied benefits to the platform and society for U1's and U2's content are as the Table 4.1 below:

As we can see, the platform will not invest optimally in verification because it does not fully internalize the cost borne by the uploaders. Even if the platform and the uploaders can negotiate and share the cost of verification so that optimal verification is achieved, it

		Verification Cost	Error Cost	Total Social Cost	Platform's Error Cost	Platform's Total Cost
U1	None	\$0.0	\$40.0	\$40.0	\$8.0	**\$8.0**
	Low	\$1.0	\$36.0	\$37.0	\$7.2	\$8.2
	Intermediate	\$3.0	\$30.0	\$33.0	\$6.0	\$9.0
	High	\$8.0	\$24.0	**\$32.0**	\$4.8	\$12.8
U2	None	\$0.0	\$10.0	\$10.0	\$2.0	**\$2.0**
	Low	\$1.0	\$9.0	**\$10.0**	\$1.8	\$2.8
	Intermediate	\$3.0	\$7.5	\$10.5	\$1.5	\$4.5
	High	\$8.0	\$6.0	\$14.0	\$1.2	\$9.2

Table 4.1: Platform's verification decision for high- and low-value content

may still comply with the takedown requests filed by the low-harm right shoulder, R2, due to imperfect information.

Assume that the platform adopts optimal verification tailored for both types of content. The cost of complying with the takedown requests is equal to the total social cost. Namely, the error cost of compliance is \$16 for the high-value content and \$9 for the low-value content. In comparison, non-compliance exposes the platform to liability, which is rightsholders' infringement harm. Lacking the knowledge of the rightsholders' types, the platform will calculate the expected liability based on the proportion. In this case, the expected liability is \$36, which is higher than the error cost to the platform or even the error cost to both the platform and the uploaders. Therefore, the platform would choose to comply with the takedown requests when it is uncertain and thus over-estimates the expected liability.

Evidence also shows that platforms do not effectively review such requests. Hamdani (2002) already points out that platforms do not necessarily internalize the full benefit of the content. Ahlert, Marsden, and Yung (2004) upload J.S. Mill's *On Liberty* and follow up with takedown notices to platforms in the U.K. and the U.S., they find that platforms in the U.K. quickly remove the uploaded content without reviewing it. Similarly, Perel and Elkin-Koren (2017) upload content and file takedown requests to test the Israeli platforms' use of algorithms in the takedown regime. They find that platforms respond to takedown requests

without asking for further information (Perel and Elkin-Koren 2017, p. 208). Furthermore, Urban, Karaganis, and Schofield (2017 a) show that Google complies with 58.8% of takedown requests and removes the links. However, among the requests received by Google, more than 70% are problematic. The evidence shows that platforms, rather than providing inadequate enforcement, as Urban, Karaganis, and Schofield (2017a) may worry, are indeed over-compliant with takedown requests filed by rightsholders.

The same result is confirmed from the interviews of platform operators (Urban, Karaganis, and Schofield 2017b). What platforms really care about is liability and immunity (p. 379). In addition, one thing that this numerical example does not include but significantly affects platforms' behavior is the statutory damages to copyright infringement. The statutory damages are also one driver that makes them conservative (2017b, p. 388; Carroll 2016, p. 177). In fact, platforms are struggling between liability and users. They do engage in verification and spend a considerable amount of resources in edge cases (Urban, Karaganis, and Schofield 2017b, pp. 390–91). However, when there is doubt, they will still ultimately tilt to compliance, even if the uncertainty does not amount to legal liability (p. 388).

Under-response by uploaders

Uploaders are the final gate of the takedown process. Ideally, uploaders care the most about their content being online. Nevertheless, they, too, fail to behave optimally under the current regime. While the numerical example does not consider the uploaders' behavior, we can analyze it without concrete numbers.

Under the current regime, filing counter-notice does not guarantee the content is restored immediately. Instead, it may invite the rightsholder to file litigation to maintain the removal status. Such a design implies that filing counter-notice exposes the uploader to the risk of litigation, which incurs costs and potential liability when courts err. Even though the court correctly identifies the content legal, the uploader may not successfully restore its content

and secure compensation when the requesting rightsholder meets the good-faith standard. Hence, the benefit of filing counter-notice may be limited and not justify the associated costs.

Evidence shows that uploaders rarely challenge takedown requests by filing counter-notices. Seng investigates the database of takedown notices and finds that there are only two counter-notices compared to 54 million takedown notices in 2012 (Seng 2014, p. 439). In their comments, Bridy and Keller (2017) review transparency reports, which show that the rate of counter-notices is less than 1% of the total takedown requests. More recently, according to YouTube, only 0.5% of claims (Content ID and others) are challenged (YouTube 2021-12-06).

Fiala and Husovec (forthcoming) have summarized possible reasons. They argue that the cost of filing counter-notices, fear of liability, and the fear of retaliation from rightsholders can be the reason for uploaders' under-response. Also, Kuczerawy (2020) mentions that uploaders, without concrete legal knowledge, may be intimidated by the potential penalties of perjury (p. 535).

4.3.4 Comments on previous proposals

Scholars have proposed several solutions to address the over-removal problem. As the analytic framework of each party's misaligned incentives, the solutions can similarly be categorized by their targets. In the following paragraphs, I will review and segment different proposed solutions according to their targets.

Adjusting platform liability

In response to the platforms' over-compliance, several proposals have been made to address the problem. The first line of the proposals is to adjust the liability regime for platforms. While scholars regard platforms are the least-cost avoidant (Mann and Belzley 2005, p. 240) and should thus be held liable for online infringement (see e.g., Lichtman and Posner 2006),

it remains in dispute which liability regime is optimal and suitable for the over-compliance problem.

Hamdani (2002) first proposes a negligence-based regime or strict liability with reduced liability to address the over-compliance due to the under-internalization of social benefit. He argues that platforms do not fully internalize the benefit enjoyed by the users and society. Hence, holding them strictly liable for the harm from infringement would lead to over-deterrence (p. 905). In the same vein, Buiten, Streel, and Peitz (2020) recognize that either full exemption or strict liability would not induce optimal efforts. They suggest a baseline negligence liability with tailored co-regulatory measures varying with types of infringement, platforms, and costs. Moreover, they impose a positive obligation to review takedown requests on platforms (p. 164).

A more radical approach proposed by Arsham (2013) creates a brand-new opt-in safe harbor. Under such a regime, platforms pay a compulsory royalty fee, determined by professional royalty judges, for registered copyrighted works in exchange for immunity (p. 792). Arsham argues that this royalty regime has several advantages. It is clear and cheap; it facilitates information-sharing, promotes technological advances, and is compatible with the monetization function (p. 798). However, he also admits that the proposed regime cannot prevent double-charging. That is, platforms may be required to pay for fairly and legally used content (p. 801).

These proposals partially address the cause of platforms' over-compliance. However, they essentially assume that platforms can cheaply distinguish infringing and non-infringing content. This presumption does not always hold in the real world. Copyright or other right infringement can be very fact specific. Platforms may not know whether the request filer really holds the right. Alternatively, platforms may lack the knowledge of parties' internal contractual relationships, which allows the use of content in some forms but not another (e.g., the music can be played during the wedding ceremony but not broadcasted

online.) Therefore, given platforms are well-positioned to takedown the content swiftly without inflicting more harm, they may, to some extent, not be the most effective verifier compared to the rightsholders.

Considering the comparative advantages in verifying infringing content, the idea of imposing a positive review obligation on platforms warrants a close investigation. When platforms are obliged to review takedown requests, rightsholders would rely on their review and withdraw the efforts to verify the content. Such a proposal, thus, only shifts the responsibility and the cost without really enhancing social benefit, unless platforms can verify content more cost-effectively. If rightsholders can verify content more efficiently, then the obligation may dilute rightsholders' verification efforts.

Hamdani's proposed regime addresses the root cause of suboptimal compliance from the perspective of the type-II error. Namely, when in doubt, platforms subject to strict liability would always err on removing the content. Hamdani's proposals reduce the cost of the type-II error to balance the incentives between removing and keeping content online. Unlike Hamdani's premise, the numerical example assumes no type-I errors. That is, all infringing content is readily identified, but the non-infringing content is not. As the numerical example shows, over-compliance can still occur without type-II errors. It can solely result from the type-I errors caused by suboptimal verification. A possible tweak to implement his proposal of the negligence rule is that when the expected error cost under optimal verification outweighs the benefit from stopping infringement, platforms are regarded as taking due care and can defy the takedown requests without liability. In that case, platforms may still comply with all takedown requests because the expected benefit is \$37.5, exceeding both the error cost of the high-value content (\$16) and the low-value content (\$9). Such over-compliance results from the uncertainty about the rightsholders' type. If platforms can distinguish the low-harm rightsholders from the high-harm ones, then platforms will defy their requests for taking down high-value content. Reduced strict liability can only mitigate the problem

without completely solving it.

Finally, all these regimes, particularly Arsham's royal-fee proposal, impose costs on platforms. Such costs could flow to users of platforms, disincentivizing their content creation. In addition, burden on platforms may have perverse effects on competition among platforms, as phenomena described in the DMCA+ practices. Alternatively, competition among platforms can encourage harm-displacement among platforms. By taking observable precautions, such as algorithmic identification technologies or perjury disclaimer, infringers may be diverted to small platforms that have no such ability to identify infringement and hold infringers liable. Under such circumstances, infringement is not deterred but displaced. The harm-displacement effect may be more salient among platforms than among right enforcement organizations for two reasons. First, platforms providing homogeneous services are more replaceable to users. Second, in the copyright area, the value to the infringer from infringement is content-specific. The extent is limited that a copyright infringer, out of the concern of enforcement, shifts to the content owned by other less aggressive rightsholders.

In summary, imposing additional burden on platforms can be less cost-effective and cannot fully solve the problem caused by their imperfect information. Moreover, it can have ripple effects on the market of platforms, harming their competition and encouraging their socially undesirable harm displacement.

Encouraging uploaders' reaction

Regarding the under-response problem encountered by uploaders, scholars also propose reforms to assist uploaders in challenging the decisions made by platforms. Notably, Fiala and Husovec replicate the real world in the lab and introduce an alternative dispute resolution (ADR) mechanism (Fiala and Husovec, forthcoming, p. 20).

In the experiment, they created an ADR mechanism in which uploaders can challenge platforms' removal disposition. Furthermore, uploaders can secure damages if the ADR

board regards their challenge as meritorious (Fiala and Husovec, forthcoming, p. 10). In their experiment, they find that once the ADR mechanism is introduced, platforms are more cautious of their disposition (p. 24). Namely, over-compliance is mitigated. Also, uploaders are more willing to resort to the ADR mechanism to challenge the platform's disposition (p. 26). To incorporate the use of the ADR mechanism, they provide both carrot and stick options. The first option is to make such an ADR process an option to encourage platforms' adoption, with an emphasis on risk-free decision-making. Alternatively, it can be embedded in regulations that force platforms to use (p. 28).

Their proposal is based on several assumptions that are incongruent with the current regime. First, they assume that filing a counter-notice will not only restore the harm inflicted on uploaders, but grant them compensation. This cannot be done by merely filing counter-notice under the current regime; uploaders who file counter-notice at best have their content re-uploaded. Second, their game assumes that, when uploaders win in the ADR process, platforms are liable for the harm. Nevertheless, platforms only face a modest cost of forwarding counter-notices to requesting rightsholders. In reality, such forwarding can be done automatically, incurring negligible cost, rather than triggering liability.

The mechanism they design should be better characterized as encouraging litigation rather than filing counter-notice. Hence, they do not directly solve the problem of under-filing of counter-notices. However, even if reading their proposal as encouraging uploaders' responses by filing lawsuits against requesting rightsholders, the proposal may not work as expected because of two reasons. First, filing a lawsuit does not necessarily mean the content is restored immediately. Litigation takes time. Even if the court correctly identifies the content as non-infringing, the value of restoring the content may be limited. Second, under the current regime, rightsholders are not liable when they file the takedown requests in good faith. Therefore, even though uploaders do suffer harm from the removal, they may not be compensated. Worse, litigation can expose them to the risk of liability when the

court commits a mistake identifying the content as infringing. In this case, uploaders are not compensated but subject to liability for infringing. All these downsides may hinder uploaders wishing to file litigation against rightsholders who file erroneous takedown requests, and such hindrances cannot be easily cleared merely by subsidizing the cost of litigation. Aside from the effectiveness of their proposal, making platforms liable for erroneous removal can similarly encounter problems in the last paragraphs.

Holding rightsholders accountable

Finally, there are proposed solutions targeting rightsholders that file malicious takedown requests. The main idea here is to increase the sanction on malicious notices to make it more difficult for rightsholders to file groundless takedown requests (Karaganis and Urban 2015, p. 30). In detail, the liability for malicious notice filing should be increased (p. 30). Moreover, the subjective standard should be adjusted from bad faith to reckless (Urban, Karaganis, and Schofield 2017a, p. 128). That is, rightsholders who recklessly file takedown requests, even in good faith, can be held liable. Furthermore, for bad faith rightsholders, they should face the penalty under perjury (p. 128). Finally, it is suggested that the put-back period should be shortened (p. 128). Currently, the put-back period lasts for fourteen days. In other words, after uploaders challenge the takedown notice, the content would remain removed for fourteen days if the rightsholder does not file the lawsuit.

This chapter agrees with the direction to hold rightsholders accountable. Nevertheless, the detailed reform proposed above may not be sufficiently effective to solve the over-removal problem. Moving the subjective standard from good faith to reckless only mitigates the problem. Both the good-faith standard and recklessness specify the level of care. As a result, while they can ensure the effort to distinguish non-infringing from infringing content, they cannot guarantee that the takedown decisions are efficient. Second, such mandated care levels do not track the social value of non-infringing content and fail to induce optimal verification

tailored to content of different values. Therefore, they often fall short of inducing socially adequate verification. Penalizing bad faith rightsholders may only address partial erroneous removal. As shown in the numerical example, even if all rightsholders file takedown requests in good faith, there can be still a fair amount of non-infringing content being removed. Finally, shortening the put-back period mitigates the social cost of erroneous removal but seems to have limited impact on rightsholders' filing behavior because the shortened period of takedown also reduces their liability for bad-faith or reckless removal. If they comply with the legal standard, regardless of good-faith, reckless, or even negligence, they will not care about how many days the content will be restored, if the stopped harm outweighs the filing cost.

Overhauling the copyright law

There are also proposals that are relevant to the over-removal problem. For instance, one of the main reasons for platforms to over-comply and uploaders to under-respond is the intimidating statutory damages. As a result, adjusting the statutory damage provision should be able to encourage platforms and uploaders to challenge the groundless takedown requests filed by rightsholders (Depoorter and Walker 2013, p. 328; Urban, Karaganis, and Schofield 2017a, p. 129). In addition, Depoorter and Walker (2013) argue that the root cause is the unclear boundaries of copyright. Hence, they propose to register the copyright with a procedural review plus periodic renewals (p. 327). With the clearer boundary of copyright, the issue of false positives and the associated over-assertion of rights can be addressed (p. 358). Admittedly, such reforms rebalance the liability faced by platforms when deciding whether to comply with the takedown requests. However, such proposed reforms can similarly affect infringements outside the online arena. Whether they are appropriate requires a thorough investigation, which this chapter cannot afford.

4.4 Holding Rightsholders Accountable

The last section reviews existing proposals and evaluates their effectiveness. The argument is offered that the proposals are often narrowly focused without inducing collaboration among parties. This section will first characterize what optimal takedowns look like via numerical example. By setting the first-best case as the benchmark, then pointing out directions for improvement, I compare how allocating responsibilities between platforms and rightsholders can better achieve the first-best outcome. Based on the comparison, the argument is offered that we should hold rightsholders accountable.

4.4.1 Optimal removals

As illustrated above, optimal removals consist of two dimensions: optimal verifications and optimal takedown decisions. For the verification to be optimal, the verification investment should be aligned with the social value of the content. Namely, the higher social value of the content, the more investment should be made for verification. More formally, parties should invest in verification until the point where the marginal verification cost equals the marginal savings on error costs. As a result, verification efforts are content based.

Accordingly, optimal verification in this numerical example should be as Table 4.1. For the high-value content, verifiers should adopt high-scrutiny because it minimizes the total social cost, including both the verification cost and the cost of erroneous removal. Instead, the socially optimal verification level for the low-value content is low scrutiny.

Based on the results of optimal verification, socially optimal takedowns require that the benefit from stopping infringement should outweigh the social cost of errors. Also note that if, in the end, takedowns are inefficient, then verification should be avoided to save the cost of verification. Hence, the socially optimal takedown decisions are as follows:

As a result, R2 should never takedown U1's content. If R2 can know the uploader's type in advance, he should also not invest in verification, as the information plays no role in his

	R1's Expected Harm= $60	R2's Expected Harm=$12
U1's Error Cost=$24	Remove	Stay online
U2's Error Cost=$9	Remove	Remove

Table 4.2: Takedown decision under optimal verification

takedown decision. As each cell occurs with 25% probability, the expected social cost of the first-best outcome is $16.

	R1's Expected Harm= $60	R2's Expected Harm=$12
U1's Error Cost=$24	$8 + $24	$12
U2's Error Cost=$9	$1 + $9	$1 + $9

Table 4.3: Social cost of the first-best outcome

4.4.2 Allocating liability between rightsholders and platforms

The first-best outcome requires perfect information regarding both the types of uploaders (the social value of content) and the types of rightsholders (the harm of infringement). However, in the real world, neither party has such perfect information: rightsholders know their types, and platforms know the types of uploaders. As a result, allocating responsibility between rightsholders and platforms is comparing the efficiency of optimal verification (yet suboptimal takedowns) with that of optimal takedowns with suboptimal verification.

Individual liability

To see the comparison more concretely, consider their behavior when they are held strictly liable for the error cost. When platforms are held strictly liable, the total social cost becomes $21 because they comply with the requests filed by the low-harm rightsholder, R2.

In comparison, when rightsholders are held liable, they would adopt the intermediate level of scrutiny. The expected social cost is $16.875.

From the results above, we can see that holding either rightsholders or platforms liable

	R1's Expected Harm= $60	R2's Expected Harm=$12
	Expected liability for non-compliance=$36	
U1's Error Cost=$24	$8 + $24	$8 + $24
U2's Error Cost=$9	$1 + $9	$1 + $9

Table 4.4: Platforms' decision under strict liability

	Verification Cost	Liability for U1	Liability for U2	Expected Liability	Platform's Total Cost
None	$0.00	$40.00	$10.00	$25.00	$25.00
Low	$1.00	$36.00	$9.00	$22.50	$23.50
Intermediate	$3.00	$30.00	$7.50	$18.75	**$21.75**
High	$8.00	$24.00	$6.00	$15.00	$23.00

Table 4.5: Rightsholders' decision under strict liability

for the error cost is not optimal. Both are subject to imperfect information and make suboptimal decisions. Rightsholders fail to invest optimally in verification. Also, their takedown decisions are impacted by both suboptimal verification and imperfect information about uploaders' type. In contrast, platforms also fail to make optimal takedown decisions because they lack the information about rightsholders' type. One thing should be noted is that the comparison between holding platforms and rightsholders individually liable is not indicative. On the contrary, it is sensitive to the numbers plugged in. Therefore, the upshot of this numerical illustration is that holding either party individually liable cannot be socially optimal.

Joint liability

Since both parties have their own information to make optimal verification and takedown decisions, it is worthwhile considering holding both parties jointly liable to induce their use of private information. In concrete, four joint liability regimes are discussed in the subsequent paragraphs.

Consider first holding both parties strictly liable for the error cost. Under such a regime, when non-infringing content is removed, both parties are liable. To analyze how parties be-

have under such a regime, we can apply the backward induction. If the platform is fully liable for the error cost, then, as a second-mover, the platform would verify the content according to the content's value. In this case, socially optimal verification is guaranteed. However, when liability completely falls on the platform, rightsholders would have no incentive to correct their decision on takedowns. Hence, they will request takedowns for all content, regardless of its value, resulting in excessive takedown requests. If we adjust the share of the liability between platforms and rightsholders, say each party bears 50% of liability, then optimal verification is no longer guaranteed, neither is optimal for takedown decisions

	Verification Cost	Liability for U1	Total Cost	Liability for U2	Total Cost
None	\$0.00	\$20.00	\$20.00	\$5.00	**\$5.00**
Low	\$1.00	\$18.00	\$19.00	\$4.50	\$5.50
Intermediate	\$3.00	\$15.00	**\$18.00**	\$3.75	\$6.75
High	\$8.00	\$12.00	\$20.00	\$3.00	\$11.00

Table 4.6: Platforms' verification when being held jointly liable (50%)

		R1's Expected Harm= \$60	R2's Expected Harm=\$12
U1's Error Cost=\$18	Expected Liability	Remove	Remove
U2's Error Cost=\$5	\$11.5	Remove	Remove

Table 4.7: Rightsholders' takedown when being held jointly liable (50%)

To fix this problem, a possible solution is to introduce fines penalizing one parties to hold both parties strictly liable for the full cost of errors. Under such a circumstances, both optimal verification and takedown are secured. However, one caveat is that over-verification can be possible when parties bargain. When both rightsholder and platforms bargain, they may find it in their common interest to invest suboptimally higher in verification, as the marginal benefit from reducing liability is now doubled.

Another regime is that holding platforms negligently liable for their verification while imposing strict liability on rightsholders. Under such a regime, platforms would invest optimally in verification. However, rightsholders' takedown decision can still be suboptimal due to imperfect information.

	Verification Cost	Liability for U1	Total Cost	Liability for U2	Total Cost
None	\$0	\$80	\$80	\$20	\$20
Low	\$1	\$72	\$73	\$18	\$19
Intermediate	\$3	\$60	\$63	\$15	**\$18**
High	\$8	\$48	**\$56**	\$12	\$20

Table 4.8: Platforms' verification when being held jointly liable (enhanced liability/penalty)

		R1 (\$60)	R2 (\$12)
U1's Content (\$24)	Expected Liability	Remove	Stay online
U2's Content (\$9)	\$16.5	Remove	Stay online

Table 4.9: Rightsholders' takedown decision when the platform are held negligently liable

We can also reverse the regime to hold rightsholders negligently liable and platforms strictly liable. However, such a regime cannot outperform the aforementioned regimes because rightsholders have inferior information regarding the content value and cannot invest optimally in verification. Furthermore, when rightsholders are held liable, they are not fully internalized the total error costs, resulting in excessive takedowns. Platforms, facing strict liability, would also comply with the excessive takedowns because they have no idea about the types of rightsholders.

The last alternative is to hold both platforms and rightsholders negligently liable. However, aside from the reasons mentioned above — lack of information to perform optimal verification by rightsholders and incomplete information for optimal takedowns by platforms — such a regime can incur redundant verification process, as it requires both parties to conduct verification. Replicative verification processes yield less information but incur the same cost and could be socially undesirable. Therefore, such a regime could not be superior to the regime above.

4.4.3 Holding rightsholders accountable

As we can observe from the analysis above, merely holding platforms liable cannot achieve the optimal result. Hence, it is, at least, necessary to hold rightsholders liable to begin with.

Aside from the incentive reasons considered above, there are several arguments that may support holding rightsholders liable.

First, rightsholders are the first step of the over-removal problem. Correcting rightsholders' incentives can curb their excessive notices, thus saving the resource used by platforms to process their requests. A frivolous takedown request is time-consuming for the platform to verify. Worse, if the platform does not identify such a request, it can also take the uploader's time to collect evidence and respond. All the time and effort spent on processing such a request can be avoided if the rightsholder can readily retain from filing such a request. Indeed, if takedown requests are socially optimal, platforms' over-compliance and the uploaders' under-response should not be a problem. In contrast, it can still be socially desirable if platforms fully comply with the optimal takedown requests.

Irrefutably, it may be argued that platforms, as resourceful big Tech companies, have better knowledge (Gabison and Buiten 2020, p. 252) and technology in identifying which requests are groundless and take correspondent actions (Mann and Belzley 2005, p. 240). Nevertheless, as shown above, platforms themselves do not have perfect incentives and information. On the contrary, they are too conservative to overrule the takedown requests due to under-internalization of full social cost of error and imperfect information. Even if the legal rules are designed to align their incentives with socially desirable levels, they may not have sufficient information to make optimal takedown decisions.

Arguably, whether platforms can effectively distinguish infringing and non-infringing content is a premise worth challenging. Rightsholders also have superior knowledge of facts that are legally relevant. Take trademark-infringing sellers for example. A platform can best identify suspicious trademark-infringing products from their price and brand. Nevertheless, a suspiciously low price does not necessarily mean the product violates trademark law. The seller can sell the product in urgent need of money. Or the seller may have an internal contract with the trademark owner. Similarly, in the copyright context, a movie producer

may contract with a famous YouTuber to promote their content by authorizing movie clips. Or, in an extreme case, where the ownership of copyright is in dispute, platforms may not know who really owns the right and qualify for sending takedown requests. Imposing enhanced responsibility on platforms inevitably asks them to make judgment calls on issues about which they cannot cheaply acquire information. In contrast, rightsholders have such knowledge and can readily utilize it before filing a takedown request. The possibility at least supports the claim that holding rightsholders accountable for the cost incurred by erroneous removal.

The need for holding rightsholders accountable has been demonstrated. However, the problem is that merely holding them accountable is insufficient to reach the first-best outcome. A step further is to ask how to induce platforms' collaboration to improve the outcome. After all, optimal removal requires information of both the harm from infringement and the content value, which is, presumably, unknown to rightsholders. A readily impulse is to impose liability on platforms, as discussed in the part of joint liability. However, aside from the problems of misaligned incentives and imperfect information, there are other factors to consider for imposing liability on platforms.

The main concern is the spillover effect. As reviewed previously, regardless of voluntary or mandated, whenever platforms take on more responsibility, such as the DMCA+ or the staydown requirement, they inevitably impact the market. Enhanced requirements often rely on advanced technology, which can be costly and deter new entrants. The effect is similar to the harm-displacement effect found in the literature on victim precaution (see Clotfelter 1978). Wrongdoers can swiftly switch from one platform to another if the platforms are providing similar services. Hence, when YouTube develops Content ID, which deters potential infringers, Vimeo should do likewise. Otherwise, it may be flooded with infringers and risk tremendous liability. Such an effect can be less significant among rightsholders. Rightsholders themselves rarely compete with one another for enforcement. Often, the copyright of a

work is exclusively owned by one or a few entities. Also, from the infringers' perspective, each creative work seems irreplaceable. Hence, the harm-displacement effect is not as severe as in the case of platforms, which can be easily transferrable (see Koo and Png 1994).

In summary, holding rightsholders accountable is necessary but not sufficient. As a result, a question to be explored is how to induce platforms' collaboration to improve rightsholders' decision-making without inviting the potential adverse effects on platforms' market. In the next section, I follow this direction to substantiate my proposed reform.

4.5 Proposed Reforms

Based on the analysis above, three steps are proposed to address specific issues. To begin, it is suggested that rightsholders be held strictly liable to ensure that all takedowns are, in terms of expectation, justified by their enforcement benefit. In addition to strict liability, I then propose a disclosure mechanism in which rightsholders should disclose their committed maximum liability (CML) for platforms' reference. Platforms, based on the CML, then decide whether to remove the requested content. If the value of the requested content outweighs CML, then the content is kept. Such a mechanism prevents ex-post costly takedowns. Aside from the strict liability regime with disclosure of CML, the idea of charging filing fees to curb excessive takedown requests is also considered. I argue that such a fee could be used to address the social benefit that is not internalized by both platforms and uploaders.

4.5.1 Holding rightsholders strictly liable

The most modest reform is holding rightsholders who send erroneous removal requests liable for the loss incurred by uploaders and platforms. The liability regime is already built in Section 512 of the DMCA, except for the subjective standard of rightsholders. As a result, the revision proposed here is only pertaining to the subjective standard.

Compared to the current good-faith requirement or the proposed recklessness standard,

strict liability induce both optimal verification and takedowns, if parties have perfect information. The reason behind is that the good-faith and recklessness standards, or even the negligence rule only address the care level but not activity level. Hence, even if parties' due care, as defined verification investment here, is optimal, their activity level, namely takedowns, can be suboptimal because they do not fully internalize the cost they impose.

To see this, compare rightsholders' decisions under strict liability and the negligence rule.

	Verification Cost	Liability for U1	Liability for U2	Expected Liability	Platform's Total Cost
None	$0.00	$40.00	$10.00	$25.00	$25.00
Low	$1.00	$36.00	$9.00	$22.50	$23.50
Intermediate	$3.00	$30.00	$7.50	$18.75	**$21.75**
High	$8.00	$24.00	$6.00	$15.00	$23.00

Table 4.10: Rightsholders verification and takedown decisions under strict liability

When rightsholders are held strictly liable, R2 will not file takedown requests, as the expected cost of error exceeds his enforcement benefit, $12. However, under the negligence rule, R2 will still file such requests because taking due care ($3), defined as intermediate scrutiny, costs less than the enforcement benefit ($12). However, such requests are socially inefficient, because the enforcement benefit is not justified by the error cost. Consequently, while the negligence rule induces optimal verification *ex ante*, they cannot prevent socially costly takedowns.

Other concerns also come into play when deciding which regime is better in the context of filing takedown requests. One crucial factor is the required information for courts to determine due care. Theoretically, due care should be set according to the content's value. However, it is unlikely that the court will impose a variable due care standard fluctuating with the value of content in question. Such case-by-case judgment may have a disadvantage in forming legal standards. Moreover, the negligence rule also invites a problem of *how* to define the rightsholder's care level. Should it be the rightsholder's expenditure, the process of verification, or the ultimate accuracy of notices? Neither dimension seems to be

perfect. In concrete, when using expenditure as the level of care, it may perversely incentivize rightsholders to adopt costly rather than cost-effective measures to meet the standard, such as a complete human review rather than a combination of human and machinery judgment. Likewise, using the process of verification may face problems when rightsholders utilize machine-learning algorithms that are not transparent and explainable. In fact, rightsholders who use machine-learning algorithms to verify non-infringing content requires courts' decisions to optimize its model, which cannot be achieved when the court applies the negligence rule because the legality of content is not always litigated. Finally, while overall accuracy can serve as a benchmark of due care, it is unobservable within the case. The court cannot infer the overall accuracy in one piece of removed content.

The negligence rule can also discourage uploaders from filing litigation. As observed in the literature, uploaders already under-respond to erroneous removals. When the negligence rule is applied, uploaders may not want to file a costly lawsuit to sue the rightsholder, whose care level cannot be readily observable. Hence, the deterrence effect from uploaders' response is lower than in strict liability regime, where uploaders can secure compensation without proving the rightsholder is negligent.

In summary, while the negligence rule can also induce optimal verification, it fails to induce optimal takedown decisions. Furthermore, it imposes burden on courts to set due care and observe actual level of care. It also discourages uploaders from filing lawsuits. In contrast, strict liability can avoid such obstacles and help improve rightsholders' algorithms that are used for verification. Of course, strict liability incurs more litigation costs. However, the negligence rule also incurs costs of determining due care and observing the actual care level. Put them aside, strict liability should be more desirable than the negligence rule to hold rightsholders accountable.

4.5.2 Disclosure of committed maximum liability

Holding rightsholders strictly liable cannot solve the problem of imperfect information. To ensure that rightsholders remove content that is less valuable than the enforcement benefit, we require the platform's knowledge of the content value. As a result, I introduce a disclosure mechanism in which rightsholders can specify their committed maximum liability (CML) with their takedown requests. When receiving such requests with CML, then the platform will determine whether to remove the content. The platform would only remove the content whose value is lower than CML, preventing valuable content from being removed.

Being asked to specify CML, the rightsholder will set his CML according to his harm from infringement. Intuitively, higher harm allows higher CML. In concrete, R1 will set his CML to \$216.67. When setting CML to \$216.67, R1 will invest in high scrutiny of verification, and the expected maximum liability is \$52 ($216.67 \times 0.4 \times 0.6$). That is, even if the value of content is \$216.67, R1 is confident to remove that content because the benefit from stopping infringement is no less than the expected liability (\$52) plus the cost of verification (\$8). In contrast, R2's CML is only \$30.55 ($\frac{12-1}{0.4\times0.9}$. Accordingly, R2's verification level is low scrutiny, incurring \$1. R2's CML ensures that his takedown requests are justifiable. The enforcement benefit, \$12, exceeds the expected liability plus the verification cost, \$11.998 ($1 + 30.55 \times 0.4 \times 0.9$).

With such specified CML, the platform would remove all content requested by R1, but only remove U2's content for R2's request. The takedown decisions are optimal. Uploaders whose content is removed per rightsholders' requests are certainly compensated.

However, this is still not the first-best outcome, as the verification efforts fail to adjust according to the content value. As a result, rightsholders always over-verify the low-value content and the unremoved content. In the first-best scenario, R1 should be able to apply lower scrutiny for U2's content, but he cannot do so under this regime. Likewise, R2 can save his verification cost when he knows the requested content is uploaded by U2. However, this

regime is more feasible than the first-best scenario, as rightsholders often invest in verification before knowing the value of suspiciously infringing content. For instance, Warner Brothers probably invest in developing verification algorithms because of their content's high stake. It is less likely that they adjust down their verification efforts for low-value content after learning its value. Likewise, rightsholders may not readily increase their verification efforts when learning the content is more valuable. Therefore, pre-specifying CML and have the platform decide may be more realistic.

We should further confirm that parties behave as expected under such a regime. That is, rightsholders genuinely report their CML, and the platforms remove content according to the CML specified by rightsholders. I first investigate whether rightsholders' incentive to report genuinely is compatible with the design. To see this, let's first check whether rightsholders want to exaggerate their CML. Rightsholders may not want to do so as doing so may decrease their net payoff because the marginal liability could outweigh the enforcement benefit for content of which the value is higher than CML. Hence, exaggerating CML will lead the expected liability to exceed the expected harm of infringement, making rightsholders suffer a net expected loss. Rightsholders would not want to lower his CML either because doing so would risk him more infringement harm from not removing some content. For instance, if R2 lowers his CML from \$30.55 to, say, \$20, so that U2's content is not removed, he will suffer \$12 harm rather than the expected liability of the expected liability under low scrutiny plus the verification cost, \$10 ($1 + 25 \times 0.4 \times 0.9$). Likewise, rightsholders would committed to the optimal verification level according to their CML because otherwise their total cost will increase. Hence, we can see that under such a regime, rightsholders always disclose genuine CML with corresponding optimal verification.

Platforms also comply with the disclosed CML. For content of which the value is higher than CML, keeping such content online is expected to yield more benefit than the harm of infringement. Therefore, in terms of expectation, platforms and the uploaders should be

able to afford the liability even if it turns out to be infringing. Of course, the statutory damage and the courts' error may distort platforms' decision. To fix this problem, we might consider immune platforms from liability when conforming to rightsholders' disclosed CML. Since they are shielded from the liability, they no longer have incentive to remove high-value content. However, it is questionable whether platforms may exaggerate the value of the requested content to keep it online. They are unlikely to do so because they are fully compensated, so they should be indifferent between keeping it online and removing it. The only caveat is the litigation cost and errors, when litigation is costly and the court could commit errors, the expected compensation is lower than the actual loss, making platforms less willing to remove the content. Nevertheless, such impact is not special to this context.

4.5.3 Charging filing fee

Holding rightsholders strictly liable allows them to internalize the cost of error borne by the platforms and the uploader. Nevertheless, as in all liability regimes, strict liability has its full deterrence effect when the stakeholders sue. Reasons preventing uploaders from filing counter-notice can likewise discourage them from filing lawsuits against rightsholders, diluting the deterrence effect. Therefore, we might want to consider an ex-ante mechanism to curb over-filing without being worried about the under-response from uploaders.[1]

A possible candidate is to charge a filing fee. A filing fee can incentivize rightsholders to invest in verification to sort out non-infringing content to avoid paying for removing content that is not harmful. Such a filing fee can also prevent low-harm rightsholders from filing takedown requests. However, compared to liability, a filing fee should be tailored to the content value to induce optimal verification and optimal takedowns.

To see this, let's assume that platforms are allowed to charge a modest filing fee, $5. With such charging fee, we might expect that rightsholders may incentivize to verify the content to

1. I thank Professor Omri Ben-Shahar for raising this possibility.

save expenditure. However, as the table shows, such a modest filing fee is still insufficient to induce their optimal verification. Also, mere filing fee cannot guarantee optimal takedowns.

	Verification Cost	Filing Fee	Total Cost
None	\$0.0	\$5.0	**\$5.0**
Low	\$1.0	\$4.8	\$5.8
Intermediate	\$3.0	\$4.5	\$7.5
High	\$8.0	\$4.2	\$12.2

Table 4.11: Verification under a filing fee \$5

Let's consider raising the filing fee to \$25, equal to the value of U2's content. In this case, rightsholders are induced to opt in low scrutiny, optimal for U2's content. However, it prevents R2 from filing takedown requests as the filing cost outweighs the expected enforcement benefit.

	Verification Cost	Filing Fee	Total Cost
None	\$0.0	\$25.0	\$25.0
Low	\$1.0	\$24.0	**\$25.0**
Intermediate	\$3.0	\$22.5	\$25.5
High	\$8.0	\$21.0	\$29.0

Table 4.12: Verification under a filing fee \$25

We can further raise the filing fee to \$100, equal to the value of U1's content. In this case, rightsholders' verification level is optimal with regard to the high-value content. The takedown decision is, again, distorted. Even R1 will not file takedown requests as the expected enforcement benefit from stopping infringement is only \$60. The reason is that a filing fee is also applied to the infringing content, making enforcement more costly.

	Verification Cost	Filing Fee	Total Cost
None	\$0	\$100	\$100
Low	\$1	\$96	\$97
Intermediate	\$3	\$90	\$93
High	\$8	\$84	**\$92**

Table 4.13: Verification under a filing fee \$100

From the numerical examples, we can see that while charging a filing fee can induce rightsholders' optimal verification, it may distort the takedown decisions, resulting *too few*

takedowns. It is because rightsholders are now required to pay for removing infringing content, which they are not liable for under liability regimes.

That said, a filing fee can have an advantage over liability regimes. As mentioned in Section 4.3.1, over-removal, or more broadly, removing non-infringing content, imposes a cost on consumers and harms freedom of speech. Such costs are not internalized by platforms and uploaders. Indeed, they do not pertain to specific content. Instead, their impact on social welfare is observed in aggregation. Since neither platforms nor uploaders internalize the costs, they are not addressed in the previously proposed liability regime. Platforms would not consider them when deciding whether to remove the requested content. Uploaders and platforms can neither seek damages to address such costs. Even if strict liability functions perfectly, rightsholders' takedown decisions can still be socially excessive without considering such costs.

Charging a filing fee can be a suitable complement to the liability regime to address this problem. By charging a modest fee, we make rightsholders pay for the minor uninternalized social harm and incentivize them to consider it before making takedown decisions. A universal modest charging fee should not burden them as the one illustrated above because the amount is set far smaller than the one to change their verification decisions. It is also more acceptable to charge a filing fee to remove infringing content because such content may have its value from the perspective of freedom of speech.

4.6 Conclusion

The notice-and-takedown regime aims to strike a balance between the Internet and content industries. However, it is subject to abuse and fails to hold rightsholders requesting takedowns accountable, resulting in the problem of over-removal. Over-removal can be fixed in two directions: optimal verification to identify non-infringing content and optimal takedowns to ensure the prevented harm outweighs the error cost. However, both directions

require private information about the content value and the level of harm. Therefore, to address the over-removal problem, we should facilitate collaboration between parties possessing such information.

This chapter suggests holding rightsholders strictly liable for the error cost to induce their optimal verification efforts and ensure their takedowns are socially desirable. Rightsholders may not have perfect information about the content value and, therefore, may over- or under-invest in verification. To address this problem, I propose that rightsholders specify their committed maximum liability for the platform's reference. The platform with information about the content value can help distinguish high-value from low-value content and only remove the latter for requesting rightsholders. Inefficient takedowns are prevented. Furthermore, a modest charging fee can be introduced, not to induce optimal verification but to address the social harm that is not internalized by platforms and uploaders and cannot be properly dealt with in the proposed liability regime.

CHAPTER 5

CONCLUDING REMARKS

Precaution-taking behavior is a double-edged sword. On one hand, it deters misconduct and serves as a substitute for law enforcement. On the other hand, it incurs costs borne by other parties. Given this understanding, it becomes essential to analyze precaution-taking behavior in terms of intensity and accuracy. We aim for it to be sufficiently accurate to avoid the cost of errors, while remaining intensive enough to deter wrongdoers. These two dimensions reveal two interesting findings. First, precautions are sensitive to law enforcement. Additionally, they can be socially costly and require investments in verification. This dissertation investigates various contexts where precaution-taking behavior exists but is socially suboptimal, such as victim precaution, gatekeeping, and online copyright enforcement. After identifying their inefficiencies, this dissertation provides policy recommendations to address them.

Chapter 2 argues that the choice liability regime and optimal law enforcement intensity are interdependent in the context of harmful precautions. Chapter 3 unearths that the design of gatekeeper liability should consider the background law enforcement and the externalities imposed by gatekeeping. Chapter 4 addresses the problem of over-removal under the current NTD regime by holding copyright holders strictly liable for the loss borne by non-infringing uploaders with specified committed maximum liability that allows platforms to utilize the information of the value of the requested content.

APPENDIX A

A GENERAL MODEL FOR CHAPTER 4

A.1 Setup

In this game, there are two players: rightsholders (R) and platforms (P).

A.1.1 Players

1. **Rightsholders**. There are two types of rightsholders, R_1 and R_2. When the uploaded content infringes R_1's and R_2's right, it inflicts H_1 and H_2 harm, respectively. $H_1 > H_2$

2. **Platform**. There is only one platform, which can benefit from the uploaded content when the content stays online. In concrete, it shares $\alpha \in (0, 1]$ of the value of the uploaded content.

A.1.2 States

Nature decides the following states:

1. The type of the rightsholder, where $\pi(R = R_1) = q$.

2. The legality of the content, Prob(infringement)=p.

3. The value of the uploaded content, $v \sim \mathcal{U}(0, \bar{v})$.

A.1.3 Sequence of actions

1. Nature draws the type of rightsholders, the legality of the content, and the value of the content.

2. Rightsholders decide how much to invest in verification and file the takedown requests.

3. The platform decides how much to invest in verification and whether to comply with the takedown requests filed by rightsholders.

A.1.4 Assumptions

To simplify the analysis, several assumptions are made as follows:

1. The verification process costs each verifier kx^2, where $k > 0, x \in [0, 1]$ and helps identify x of non-infringing content. In other words, when spending kx^2, the verifier can ensure that only $1-x$ of non-infringing content is included in the takedown request.

2. The value of infringing content is not calculated in social welfare.

3. R's type is only known to R, and v is only known to P and the uploaders. By contrast, both p and q are common knowledge.

A.2 First-Best Outcome

The first-best outcome minimizes the total social cost, which consists of (1) the harm from unremoved infringing content, (2) the error cost of removing legal content, and (3) the cost of verification.

To begin with, let's check the socially optimal amount invested in verification, which balances the verification cost and the savings on error costs.

$$kx^2 + (1-p)v(1-x) \tag{A.1}$$

The socially optimal threshold of verification x^* solves the first-order condition of the equation above:

$$2kx^* - (1-p)v = 0 \implies x^* = \frac{(1-p)v}{2k}$$

By taking derivative of x^* wrt v, we can see that $\frac{dx^*}{dv} = \frac{1-p}{2k} > 0$, implying that the higher value the content has, the higher level verification should be.

Furthermore, let's investigate the threshold for society to remove the content. It is socially desirable to remove the content when the reduced harm outweighs the error cost plus the verification cost. Otherwise, takedown should not occur and no verification is needed. Namely, the condition for socially desirable removal is

$$pH - k(x^*)^2 - (1-p) \times v \times (1-x^*) \leq 0 \tag{A.2}$$

Let's assume that the decision of verification is always socially optimal and plug $x^* = \frac{(1-p)v}{2k}$ into Equation A.2. We get the condition for socially optimal removal:

$$\begin{aligned} & p \times H - k(x^*)^2 - 2kx^*(1-x^*) > 0 \\ & k(x^*)^2 - 2kx^* + pH > 0 \\ \implies & x^* \leq \frac{2k - \sqrt{4k^2 - 4kpH}}{2k} = 1 - \frac{\sqrt{k(k-pH)}}{k} \\ \implies & v \leq \frac{2k - \sqrt{4k^2 - 4kpH}}{(1-p)} \end{aligned} \tag{A.3}$$

For simplicity, let's denote $\tilde{v}(H) = \frac{2k - \sqrt{4k^2-4kpH}}{1-p}$. We find that $\frac{\partial \tilde{v}(H)}{\partial H} = \frac{4kp(4k^2-4kpH)^{-0.5}}{2(1-p)} > 0$ and $\frac{\partial^2 \tilde{v}(H)}{\partial H^2} = \frac{8(kp)^2(4k^2-4kpH)^{-1.5}}{2(1-p)} > 0$. Therefore, when the rightsholder suffers more harm, he can justifiably remove more content.

From both Equations A.2 and A.3, we can see that socially optimal verification depends on the value of the content, and the optimal removal has a cutoff that is a function of both the content value and the harm inflicted by rightsholders. As a result, the first-order outcome can only be achieved if the decisionmaker has perfect information and considers all three

types of social costs. In this case, the social cost is

$$\begin{aligned} SC^{FB} =& q \int_0^{\max(0,\tilde{v}(H_1))} \left(k(\frac{(1-p)v}{2k})^2 + (1-p)(1-\frac{(1-p)v}{2k})v \right) \frac{dv}{\bar{v}} \\ &+q \int_{\max(0,\tilde{v}(H_1))}^{\bar{v}} (pH_1) \frac{dv}{\bar{v}} \\ &+(1-q) \int_0^{\max(0,\tilde{v}(H_2))} \left(k(\frac{(1-p)v}{2k})^2 + (1-p)(1-\frac{(1-p)v}{2k})v) \right) \frac{dv}{\bar{v}} \\ &+(1-q) \int_{\max(0,\tilde{v}(H_2))}^{\bar{v}} (pH_2) \frac{dv}{\bar{v}} \end{aligned} \tag{A.4}$$

Each line of the right-hand side of Equation A.4 represents (1) the threshold for R_1 to verify contents and file takedown requests, (2) the threshold for R_1 to tolerate and withhold her takedown request, (3) the threshold for R_2 to verify contents and file takedown requests, and (4) the threshold for R_2 to tolerate and withhold her takedown request. As we can see, since $H_1 > H_2$, R_1 should be allowed to file more takedown requests than R_2 does because the harm inflicted by infringing her right is more severe than by infringing R_2's.

From the first-best outcome, we have following observations:

Observation A.1. *The first-best outcome allows us to observe*

1. *The optimal investment in verification depends on the content's value and the proportion of infringing/legal content.*

2. *Given the optimal investment in verification, type-I error is inevitable due to the efficiency reason.*

3. *The threshold of optimal takedown disposition varies with the inflicted harm of the rightsholders. High-value rightsholders enjoy a higher threshold for filing takedown requests.*

A.3 The Social Problem: Over-Removal

In this section, I discuss the social problem of over-removal in the current DMCA notice-and-takedown regime. Under the current regime, rightsholders must file takedown requests in *good faith*. After requests are filed, the platform then decides whether to comply. Here, let me assume that the good-faith standard requires rightsholders to invest in verification with a certain level, x^{GF}.

A.3.1 Rightsholders' over-filing

When the good-faith requirement is defined as the level of verification, x^{GF}, three scenarios are possible.

First, it is possible that the good-faith standard is lower than the first-best level of verification given the value of content, i.e., $x^{GF} < x^* = \frac{(1-p)v}{2k}$. Since rightsholders can avoid liability by meeting the good-faith standard, they need not consider the liability or the error cost borne by both the platform and the uploaders. Immune from liability, they are more willing to file takedown requests after they meet the good-faith standard. In concrete, rightsholders will file more takedown requests when the reduced harm outweighs the cost of verification, i.e.,

$$pH - k(x^{GF})^2 > pH - k(x^*)^2 > pH - k(x^*)^2 - 2kx^*(1-x^*)$$

Hence, when the required good-faith standard is lower than the optimal level for the content, rightsholders' investment in verification is socially inadequate ($x^{GF} < x^*$), and their takedown decision is socially excessive. Such decisions lead to two inefficiencies: less legal content is verified and excluded from being taken down, and more takedown requests for unverified content are filed.

Second, $x^{GF} > x^*$ and $k(x^*)^2 + 2kx^*(1-x^*) < k(x^{GF})^2$. For some low-value content,

the good-faith standard may be too costly to comply. In this case, rightsholders would rather bear the liability for non-compliance than comply with the good-faith requirement. When rightsholders choose to bear liability, they behave as if they are held strictly liable. Hence, they internalize the cost borne by the platform and the uploaders and are incentivized to make socially optimal takedown decisions.

If, otherwise, the condition is not met, namely $k(x^*)^2+2kx^*(1-x^*) > k(x^{GF})^2 > k(x^*)^2$, then rightsholders will still choose to comply with the good-faith standard. In this case, two efficiencies can be observed. First, the verification cost is suboptimally high for some content ($k(x^{GF})^2 > k(x^*)^2$). Second, given the higher-than-optimal verification threshold, the takedown decision is still excessive because rightsholders will file takedown requests because $pH - k(x^{GF})^2 > pH - k(x^*)^2 - 2kx^*(1-x^*)$.

To summarize, under the current regime that requires rightsholders to meet the good-faith standard, rightsholders would never file fewer requests than in the first-best case. Therefore, over-filing occurs.

A.3.2 Platform's over-compliance

One may wonder whether the platform may cure the problem by reviewing rightsholders' takedown requests. Admittedly, the platform can invest in verification and refuse to take down some content that is identified legal. However, the platform may over-comply for two reasons: First, the platform does not fully internalize the benefit of the content staying online (i.e., $\alpha \leq 1$). In addition, the platform has imperfect information for complying with takedown decisions.

To see this, let's focus on the case where the good-faith requirement is socially inadequate so that the platform can invest in the same verification to gain more information. The platform can invest in verification for its own benefit. It faces the cost minimization problem

as follows:

$$C^P = kx^2 - (1-p)\alpha v x \tag{A.5}$$

The rational level of verification solves the first-order condition of Equation A.5:

$$2kx^P - (1-p)\alpha v \implies x^P = \frac{\alpha(1-p)v}{2k} < \frac{(1-p)v}{2k}$$

We can see that the platform's investment in verification is lower than the socially optimal level because it only internalizes a partial cost of error removal. Based on such verification investment, its takedown decision is made by weighing potential liability for non-compliance against the benefit therefrom. That is, it would comply with the takedown requests if

$$p\left[qH_1 + (1-q)H_2\right] > k(x^P)^2 + (1-p) \times \alpha v(1-x^P) \tag{A.6}$$

Therefore, it will apply a cutoff strategy to comply with the takedown requests. Namely, it will take down the content of which the value is below the threshold:

$$\begin{aligned}
& p\left[qH_1 + (1-q)H_2\right] - k(x^P)^2 - 2kx^P(1-x^P) > 0 \\
& k(x^P)^2 - 2kx^P + p\left[qH_1 + (1-q)H_2\right] > 0 \\
\implies & x^P \leq \frac{2k - \sqrt{4k^2 - 4kp\left[qH_1 + (1-q)H_2\right]}}{2k} = 1 - \frac{\sqrt{k(k - p\left[qH_1 + (1-q)H_2\right])}}{k} \\
\implies & v^P \leq \frac{2k - \sqrt{4k^2 - 4kp\left[qH_1 + (1-q)H_2\right]}}{(1-p)\alpha}
\end{aligned} \tag{A.7}$$

When the platform only internalizes a modest amount of the content value, namely, α is sufficiently low, it is possible that $\frac{2k-\sqrt{4k^2-4kp[qH_1+(1-q)H_2]}}{(1-p)\alpha} > \tilde{v}(H_1) > \tilde{v}(H_2)$. That is, over-compliance can occur. Even if the platform captures the full benefit of the content, either due to bargaining or liability, it can still over-comply with R_2's requests because $\tilde{v}(qH_1 + (1-q)H_2) > \tilde{v}(H_2)$. However, whether it would over- or under-comply with R_1's

takedown requests, depend on the level of H_1 and $\bar{v}$. If H_1 is significantly larger than $\bar{v}$, then it is still possible that $\tilde{v}(qH_1 + (1-q)H_2) > q\tilde{v}(H_1) + (1-q)\tilde{v}(H_2)$. Otherwise, the platform could under-remove because normally $\tilde{v}(qH_1 + (1-q)H_2) < q\tilde{v}(H_1) + (1-q)\tilde{v}(H_2)$ due to the positive second derivative of $\tilde{v}(H)$ wrt H.

The above case analyzes the case where the platform voluntarily invests in verification. It is possible that $x^P < x^{GF}$ so the platform's investment does not provide additional information and is unnecessary. If that is the case, the takedown decision for the platform becomes evaluating the following

$$\begin{aligned} p\left[qH_1 + (1-q)H_2\right] &> (1-p) \times \alpha v(1 - x^{GF}) \\ v &< \frac{p\left[qH_1 + (1-q)H_2\right]}{\alpha(1-p)(1-x^{GF})} \end{aligned} \tag{A.8}$$

The threshold for the platform to remove the requested content could is higher than the first-best removal threshold (see Equation A.2 line 2) for both R_1 and R_2 because the platform no longer incurs the verification cost.

A.3.3 Summary

Under the current regime, which requires the rightsholders to file takedown requests in good faith, rightsholders could still over-file because they are not held fully liable for the total cost borne by the platform and the uploaders. Even if it does, when H_1 is significantly high, over-compliance can still occur. Ultimately, whether the platform, when fully internalizing the value of the requested content, would over- or under-comply depends on the level of harm of rightsholders. This indicates that holding platforms liable or having them review takedown requests filed by rightsholders, while saving verification cost in this model, may not guarantee efficiency. Hence, I will focus on holding rightsholders accountable and elaborate on the proposed regime.

A.4 Proposed Regime

In this section, I propose holding rightsholders strictly liable for the loss borne by the platform and the uploaders to induce their optimal takedown decisions. When rightsholders are strictly liable, social welfare improves compared to the current good-faith requirement. However, the first-best outcome is still not met. Rightsholders can be subject to suboptimal verification and filing due to imperfect information regarding the content value. An improvement can be made by having the rightsholder specify his maximum liability with the takedown requests for the platform to distinguish whether the content should be removed.

A.4.1 Holding rightsholders strictly liable

Let's first consider holding rightsholders strictly liable. Under strict liability, both $R1$ and $R2$ will first determine whether to take down the content if they invest in verification optimally. Here, they aim to minimize the sum of (1) the cost of verification, (2) the liability from removing non-infringing content, and (3) the harm from unremoved infringing content. Since they have no information about the content value, they can only rely on the expected content value, $\frac{\bar{v}}{2}$. The optimal verification is thus

$$x^{SL} = \frac{(1-p)\bar{v}}{4k}$$

Strict liability vs. the first-best outcome

Holding rightsholders strictly liable does not achieve the first-best outcome due to imperfect information. Since an un-informed rightsholder can only rely on the expected content value, he will either remove all content if $\frac{\bar{v}}{2} < \tilde{v}(H)$ or withdraw his takedown requests for all content otherwise. Let's compare the total social cost under strict liability with the first-best outcome. For simplicity, I use a more generic infringement harm, H.

Proposition A.1. *Holding rightsholders strictly liable for the error cost borne by the platform and the uploaders does not meet the first-best outcome.*

Removing all content under strict liability When all content is removed, the social cost exceeds the first-best case because the verification level is suboptimal for the content of which the value $v < \tilde{v}(H)$. Also, the rightsholder removes the content of which the value exceeds $\tilde{v}(H)$, resulting in inefficient takedowns.

$$\begin{aligned} SC_r^{SL} = & \int_0^{\bar{v}} \left[k(x^{SL})^2 + (1-p)v(1-x^{SL})\right] \frac{dv}{\bar{v}} \\ = & \int_0^{\max(0,\tilde{v}(H))} \left[k(x^{SL})^2 + (1-p)v(1-x^{SL})\right] \frac{dv}{\bar{v}} \\ & + \int_{\max(0,\tilde{v}(H))}^{\bar{v}} \left[k(x^{SL})^2 + (1-p)v(1-x^{SL})\right] \frac{dv}{\bar{v}} \\ > & \int_0^{\max(0,\tilde{v}(H))} \left[k(\frac{(1-p)v}{2k})^2 + (1-p)v(1-\frac{(1-p)v}{2k})\right] \frac{dv}{\bar{v}} \\ & + \int_{\max(0,\tilde{v}(H))}^{\bar{v}} (pH) \frac{dv}{\bar{v}} \\ = & SC^{FB} \end{aligned} \tag{A.9}$$

All content stays under strict liability When all content is kept online, the social cost is still greater than the first-best case because the rightsholder fails to remove some content

$(v < \tilde{v}(H))$ that expects to inflict more social harm than the error cost.

$$\begin{aligned} SC_s^{SL} = & \int_0^{\bar{v}} (pH)\,\frac{dv}{\bar{v}} \\ = & \int_0^{\max(0,\tilde{v}(H))} (pH)\,\frac{dv}{\bar{v}} \\ & + \int_{\max(0,\tilde{v}(H))}^{\bar{v}} (pH)\,\frac{dv}{\bar{v}} \\ > & \int_0^{\max(0,\tilde{v}(H))} \left[k(\frac{(1-p)v}{2k})^2 + (1-p)v(1-\frac{(1-p)v}{2k})\right]\frac{dv}{\bar{v}} \\ & + \int_{\max(0,\tilde{v}(H))}^{\bar{v}} (pH)\,\frac{dv}{\bar{v}} \\ = & SC^{FB} \end{aligned} \tag{A.10}$$

Strict liability vs. the good-faith standard

Here, we can compare the strict liability regime with the good-faith standard. To begin with, we can bifurcate the case into $x^{SL} > x^{GF}$ and $x^{SL} < x^{GF}$.

Proposition A.2. *Holding rightsholders strictly liable for the error cost borne by the platform and the uploaders is better than the good-faith standard, regardless of the verification level required by the good-faith standard.*

Case 1 $pH - k(x^{GF})^2 > pH - k(x^{SL})^2 - \int_0^{\bar{v}}(1-p)v(1-x^{SL})\frac{dv}{\bar{v}}$. In this case, a rightsholder who takes down all content $(\frac{\bar{v}}{2} < \tilde{v}(H))$ would choose to do so under the good-faith requirement because $pH - k(x^{GF}) > pH - k(x^{SL})^2 - \int_0^{\bar{v}}(1-p)v(1-x^{SL})\frac{dv}{\bar{v}} > 0$. We can

compare the social costs under both regimes.

$$\begin{aligned} SC_r^{SL} &= \int_0^{\bar{v}} \left[k(x^{SL})^2 + (1-p)v(1-x^{SL})\right] \frac{dv}{\bar{v}} \\ &= k(x^{SL})^2 + (1-p)(1-x^{SL}) \int_0^{\bar{v}} \frac{vdv}{\bar{v}} \\ &= k(x^{SL})^2 + (1-p)(1-x^{SL}) \frac{\bar{v}}{2} \\ &< k(x^{GF})^2 + (1-p)(1-x^{GF}) \frac{\bar{v}}{2} \\ &= k(x^{GF})^2 + (1-p)(1-x^{GF}) \int_0^{\bar{v}} \frac{vdv}{\bar{v}} \\ &= \int_0^{\bar{v}} \left[k(x^{GF})^2 + (1-p)v(1-x^{GF})\right] \frac{dv}{\bar{v}} \\ &= SC_r^{GF} \end{aligned} \tag{A.11}$$

We can see that the verification level is optimal under the strict liability regime. As a result, the total cost of error and verification should be lower than the case under the good-faith requirement.

It is also possible that a rightsholder removes all content under the good-faith requirement but withdraws his takedown requests when being held strictly liable. That is, $pH - k(x^{GF})^2 > 0 > pH - k(x^{SL})^2 - \int_0^{\bar{v}} (1-p)v(1-x^{SL})\frac{dv}{\bar{v}}$.

$$\begin{aligned} SC_s^{SL} &= pH \\ &< k(x^{SL})^2 - \int_0^{\bar{v}} (1-p)v(1-x^{SL}) \frac{dv}{\bar{v}} \\ &= k(x^{SL})^2 + (1-p)(1-x^{SL}) \frac{\bar{v}}{2} \\ &< k(x^{GF})^2 + (1-p)(1-x^{GF}) \frac{\bar{v}}{2} \\ &= k(x^{GF})^2 + (1-p)(1-x^{GF}) \int_0^{\bar{v}} \frac{vdv}{\bar{v}} \\ &= SC_r^{GF} \end{aligned} \tag{A.12}$$

If otherwise, a rightsholder refrains from removing all content under both regimes, then

$SC_s^{SL} = SC_s^{GF}$.

Case 2 $pH - k(x^{GF})^2 < pH - k(x^{SL})^2 - \int_0^{\bar{v}}(1-p)v(1-x^{SL})\frac{dv}{\bar{v}}$. In this case, a rightsholder would find it more cost-saving to bear the liability rather than meet the good-faith standard. Hence, he will actually subject himself to strict liability. Therefore, the good-faith standard fails. Even if a rightsholder does comply with the good-faith standard, the social cost of compliance such a standard would exceed the case of strict liability.

To see this, let's first that $pH - k(x^{SL})^2 - \int_0^{\bar{v}}(1-p)v(1-x^{SL})\frac{dv}{\bar{v}} > pH - k(x^{GF})^2 > 0$.

$$\begin{aligned} SC_r^{SL} &= k(x^{SL})^2 + \int_0^{\bar{v}}(1-p)v(1-x^{SL})\frac{dv}{\bar{v}} \\ &< k(x^{GF})^2 \\ &< k(x^{GF})^2 + \int_0^{\bar{v}}(1-p)v(1-x^{GF}) \\ &= SC_r^{GF} \end{aligned} \tag{A.13}$$

If, otherwise, $pH - k(x^{SL})^2 - \int_0^{\bar{v}}(1-p)v(1-x^{SL})\frac{dv}{\bar{v}} > 0 > pH - k(x^{GF})$

$$\begin{aligned} SC_r^{SL} &= \int_0^{\bar{v}}\left[k(x^{SL})^2 + (1-p)v(1-x^{SL})\right.\frac{dv}{\bar{v}} \\ &< \int_0^{\bar{v}}(pH)\frac{dv}{\bar{v}} \\ &= SC_s^{GF} \end{aligned} \tag{A.14}$$

In summary, no matter how the good-faith standard is set, social welfare under the strict liability regime would be better than that of the good-faith standard.

A.4.2 Disclosure of committed maximum liability (CML)

As analyzed above, mere strict liability does not meet the first-best outcome because it fails to align the verification level with the content value and only avails the rightsholder of a

binary choice: remove or keep all content. Such a binary choice inevitably results in over- or under-removal. To address this problem, I suggest rightsholders disclose their committed maximum liability (CML). By specifying CML, the platform only removes content of which the value is below the CML and keep other content online. This regime would be more efficient than strict liability.

To see this, let's see how rightsholders would determine their CML. The CML allows them to remove the content of which the enforcement benefit (stopped infringement) outweighs the expected liability plus the verification cost. Denote the CML as V. Formally, a rightsholder would first determine his verification investment to minimize his liability and verification cost:

$$kx^2 + \int_0^V [(1-p)(1-x)v] \frac{dv}{\bar{v}} \tag{A.15}$$

The correspondent verification level, $x^\star$ solves the first-order condition:

$$2kx^\star - \frac{(V)^2}{2\bar{v}}(1-p) = 0 \implies x^\star = \frac{(1-p)(V)^2}{4k\bar{v}} = \frac{(1-p)\frac{V}{\bar{v}}V}{4k} < x^{SL}$$

His cost, according to the specified CML, is thus

$$\begin{aligned} C^\star &= k(x^\star)^2 + \int_0^V [(1-p)(1-x^\star)v] \frac{dv}{\bar{v}} + \int_V^{\bar{v}} pH \frac{dv}{\bar{v}} \\ &= k(x^\star)^2 + (1-p)(1-x^\star)\frac{V^2}{2\bar{v}} + \frac{(\bar{v}-V)pH}{\bar{v}} \end{aligned} \tag{A.16}$$

Assume a rightsholder rationally chooses his CML, $v^\star$, which minimizes his total cost, including the verification cost, the expected liability, and the unstopped harm from infringing content. Such $v^\star$ should solve the first-order condition of $C^\star$ with respect to V:

$$\begin{aligned} & 2k(x^\star)\frac{dx^\star}{dV} + \frac{v^\star}{\bar{v}}(1-p)(1-x^\star) - (1-p)\frac{(v^\star)^2}{2\bar{v}}\frac{dx^\star}{dV} - \frac{pH}{\bar{v}} = 0 \\ \implies & \frac{1}{\bar{v}}\left[(1-p)(1-x^\star)v^\star - pH\right] + \left[2kx^\star - \frac{(1-p)(v^\star)^2}{2\bar{v}}\right]\frac{dx^\star}{dV} = 0 \end{aligned} \tag{A.17}$$

The second-order derivative is positive, implying that $v^\star$ is the minimizer.

Proposition A.3. *Allowing the rightsholders to specify their committed maximum liability is always better than the strict liability regime.*

Assume that the rightsholder will remove all content under the strict liability regime. Let's compare the social cost under this regime with that under strict liability.

$$\begin{aligned} SC^\star &= k(x^\star)^2 + \int_0^{v^\star} \left[(1-p)(1-x^\star)v\right] \frac{dv}{\bar{v}} + \int_{v^\star}^{\bar{v}} pH \frac{dv}{\bar{v}} \\ &< k(x^\star)^2 + \int_0^{\bar{v}} \left[(1-p)(1-x^\star)v\right] \frac{dv}{\bar{v}} + \underbrace{\int_{v^\star}^{\bar{v}} \left[pH - (1-p)(1-x^\star)v\right] \frac{dv}{\bar{v}}}_{<0, \text{ otherwise improveable}} \\ &< k(x^{SL})^2 + \int_0^{v^\star} \left[(1-p)(1-x^{SL})v\right] \frac{dv}{\bar{v}} + \int_{v^\star}^{\bar{v}} \left[(1-p)(1-x^{SL})v\right] \frac{dv}{\bar{v}} \\ &< SC_r^{SL} \end{aligned} \tag{A.18}$$

The second line of the right-hand side of Equation A.18 is because $pH - (1-p)(1-x^\star)v < 0$ (see Equation A.17. Otherwise, the rightsholder should elevate his $v^\star$. The third line is because $v^\star$ minimizes $C^\star$ and $\int_{v^\star}^{\bar{v}} \left[(1-p)(1-x^{SL})v\right] \frac{dv}{\bar{v}} > 0$.

Likewise, the superiority of the CML regime also holds when the rightsholder chooses not to remove any content at all under the strict liability regime. To see this, we can arbitrarily choose another CML, $V = 0$. Since $v^\star$ is the cost minimizer, $SC^\star$ should not exceed SC when $V = 0$.

$$\begin{aligned} SC^\star &= k(x^\star)^2 + \int_0^{v^\star} \left[(1-p)(1-x^\star)v\right] \frac{dv}{\bar{v}} + \int_{v^\star}^{\bar{v}} pH \frac{dv}{\bar{v}} \\ &\leq k(x^\star)^2 + \int_0^{0} \left[(1-p)(1-x^\star)v\right] \frac{dv}{\bar{v}} + \int_{0}^{\bar{v}} pH \frac{dv}{\bar{v}} \\ &= pH = SC_s^{SL} \end{aligned} \tag{A.19}$$

The intuition behind is simple: when being held strictly liable, the rightsholder bears the total social cost. Offering him a choice to specify his CML should not make him worse, as he can always replicate the case under strict liability by specifying his CML $v^\star = \bar{v}$ to remove

all content or $v^\star = 0$ to allow all content to stay online.

REFERENCES

Agan, Amanda, and Sonja B. Starr. "Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment." *Quarterly Journal of Economics* 133, no. 1 (2018): 191–235. 10.1093/qje/qjx028.

Ahlert, Christian, Chris Marsden, and Chester Yung. *How 'Liberty' Disappeared from Cyberspace: The Mystery Shopper Tests Internet Content Self-Regulation.* Programme in Comparative Media Law & Policy. 2004.

Ahn, Luis von, Benjamin Maurer, Colin McMillen, David Abraham, and Manuel Blum. "reCAPTCHA: Human-Based Character Recognition via Web Security Measures." *Science* 321, no. 5895 (2008): 1465–1468. 10.1126/science.1160379.

Alkaabi, Ali Obaid, George Mohay, Adrian McCullagh, and Nicholas Chantler. "A Comparative Analysis of the Extent of Money Laundering in Australia, UAE, UK and the USA." *SSRN Electronic Journal*, 2010. 10.2139/ssrn.1539843.

Alzola, Miguel. "Beware of the Watchdog: Rethinking the Normative Justification of Gatekeeper Liability." *Journal of Business Ethics* 140, no. 4 (2017): 705–721. 10.1007/s10551-017-3460-3.

Anderson, David A. "The Aggregate Burden of Crime." *Journal of Law and Economics* 42, no. 2 (1999): 611–642. http://links.jstor.org/sici?sici=0022-2186%28199910%2942%3A2%3C611%3ATABOC%3E2.0.CO%3B2-3.

Anderson, David A. "The Aggregate Cost of Crime in the United States." *The Journal of Law and Economics* 64, no. 4 (2021): 857–885. 10.1086/715713.

Arlen, Jennifer. "Corporate Criminal Liability: Theory and Evidence." In *Research Handbook on the Economics of Criminal Law*, edited by Alon Harel, 144–203. Edward Elgar, 2014.

Arlen, Jennifer. "The Potentially Perverse Effects of Corporate Criminal Liability." *Journal of Legal Studies* 23, no. 2 (1994): 833–867.

Armour, John, Dan Awrey, Paul Davies, Luca Enriques, Jeffrey N. Gordon, Colin Mayer, and Jennifer Payne. *Principles of Financial Regulation.* First edition. Oxford: Oxford University Press, 2016.

Arsham, Bryan E. "Monetizing Infringement: A New Legal Regime for Hosts of User-Generated Content." *Georgetown Law Journal* 101, no. 3 (2013): 775–806.

Astley, Sab. "An Anonymous Music Streaming YouTube Channel Has Built a Loyal Community of over 11 Million Fans, But Complex Copyright Rules Could Put its Future in Jeopardy." *Insider*, 2022-07-21. Accessed June 30, 2023. https://www.insider.com/lofi-girl-founder-calls-on-youtube-to-change-copyright-rules-2022-7.

Ayres, Ian, and Steven D. Levitt. "Measuring Positive Externalities from Unobservable Victim Precaution: An Empirical Analysis of Lojack." *Quarterly Journal of Economics* 113, no. 1 (1998): 43–77.

Baer, Justin. "History: Banks Are at the Heart of Capitalism." *Financial Times*, 2020-11-17.

Bar-Gill, Oren. "The Behavioral Economics of Consumer Contracts." *Minnesota Law Review* 93 (2008): 749–780.

Bar-Ziv, Sharon, and Niva Elkin-Koren. "Behind the Scenes of Online Copyright Enforcement: Empirical Evidence on Notice & Takedown." *Connecticut Law Review* 50, no. 2 (2018): 339–386.

Baumann, Florian, Philipp Denter, and Tim Friehe. "Hide or Show? Observability of Private Precautions Against Crime When Property Value is Private Information." *American Law and Economics Review* 21, no. 1 (2019): 209–245. 10.1093/aler/ahy009.

Baumann, Florian, and Tim Friehe. "Private Protection against Crime when Property Value is Private Information." *International Review of Law and Economics* 35 (2013): 73–79. 10.1016/j.irle.2013.03.002.

Becker, Gary S. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76, no. 2 (1968): 169–217.

Bell, Abraham, and Gideon Parchomovsky. "The Case for Imperfect Enforcement of Property Rights." *University of Pennsylvania Law Review* 160, no. 7 (2012): 1927–1954.

Ben-Shahar, Omri. "Data Pollution." *Journal of Legal Analysis* 11 (2019): 104–159. 10.1093/jla/laz005.

Ben-Shahar, Omri, and Alon Harel. "Blaming the Victim: Optimal Incentives for Private Precautions against Crime." *Journal of Law, Economics, and Organization* 11, no. 2 (1995): 434–455. 10.1093/oxfordjournals.jleo.a036879.

Ben-Shahar, Omri, and Alon Harel. "The Economics of the Law of Criminal Attempts: A Victim-Centered Perspective." *University of Pennsylvania Law Review* 145, no. 2 (1996): 299–351. 10.2307/3312659.

Bridy, Annemarie. "Copyright's digital deputies: DMCA-plus enforcement by Internet intermediaries." In *Research Handbook on Electronic Commerce Law*, edited by John A. Rothchild, 185–208. Research Handbooks in Information Law. Cheltenham, UK: Edward Elgar Publishing, 2016.

Bridy, Annemarie, and Daphne Keller. *Section 512 Study*. February 21, 2017.

Buiten, Miriam C., Alexandre de Streel, and Martin Peitz. "Rethinking Liability Rules for Online Hosting Platforms." *International Journal of Law and Information Technology* 28 (2020): 139–166. 10.2139/ssrn.3350693.

Carroll, Michael W. "Safe Harbors from Intermediary Liability and Social Media." In *Research Handbook on Electronic Commerce Law*, edited by John A. Rothchild, 168–184. Research Handbooks in Information Law. Cheltenham, UK: Edward Elgar Publishing, 2016.

Center for Democarcy and Technology. Chile's Notice-and-Takedown System for Copyright Protection: An Alternative Approach. Accessed June 30, 2023. https://cdt.org/wp-content/uploads/pdfs/Chile-notice-takedown.pdf.

Choi, Stephen. "Market Lessons for Gatekeepers." *Northwestern University Law Review* 92, no. 3 (1997-1998): 916–966.

Civil Aeronautics Administration. Passenger Load Factor of Scheduled International and Cross-Strait Flights by Lines in Taiwan Area. Accessed June 30, 2023. https://www.caa.gov.tw/Article.aspx?a=2998&lang=2.

Clements, Matthew T. "Precautionary Incentives for Privately Informed Victims." *International Review of Law and Economics* 23, no. 3 (2003): 237–251. 10.1016/j.irle.2003.09.006.

Clotfelter, Charles T. "Private Security and the Public Safety." *Journal of Urban Economics* 5 (1978): 388–402.

Clotfelter, Charles T. "Public Services, Private Substitutes, and the Demand for Protection against Crime." *American Economic Review* 67, no. 5 (1977): 867–877.

Coase, Ronald H. "The Problem of Social Cost." *Journal of Law and Economics* 3 (1960): 1–44.

Coffee, John C., Jr. "Gatekeeper Failure and Reform: The Challenge of Fashioning Relevant Reforms." *Boston University Law Review* 84, no. 2 (2004): 301–364.

Coffee, John C., Jr. "The Acquiescent Gatekeeper: Reputational Intermediaries, Auditor Independence and the Governance of Accounting," 2001. http://papers.ssrn.com/paper.taf?abstractid=270944. 10.2139/ssrn.270944.

Coffee, John C., Jr. "Understanding Enron: "It's About the Gatekeepers, Stupid"." *Business Lawyer* 57, no. 4 (2002): 1403–1420.

Cohen-Almagor, Raphael. "The Role of Internet Intermediaries in Tackling Terrorism Online." *Fordham Law Review* 86, no. 2 (2017): 425–454.

Cook, Philip J. "The Demand and Supply of Criminal Opportunities." *Crime and Justice* 7 (1986): 1–27.

Cooter, Robert D. "Prices and Sanctions." *Columbia Law Review* 84, no. 6 (1984): 1523–1560.

Cooter, Robert D., and Ariel Porat. "Should Courts Deduct Nonlegal Sanctions from Damages?" *Journal of Legal Studies* 30, no. 2 (2001): 401–422. 10.1086/322058.

Curry, Philip A., and Matthew Doyle. "Intergrating Market Alternatives into the Economic Theory of Optimal Deterrence." *Economic Inquiry* 54, no. 4 (2016): 1873–1883. 10.1111/ecin.12344.

Cusey, Rebecca, and Terrica Carrington. *Comments on Behalf of the Arts and Entertainment Advocacy Clinic at George Mason Univesity School of Law*. 2016. Accessed June 30, 2023. https://sls.gmu.edu/cpip/wp-content/uploads/sites/31/2016/04/Section-512-Study-Comments-on-Behalf-of-the-Arts-Entertainment-Advocacy-Clinic.pdf.

Dam, Kenneth W. "Self-Help in the Digital Jungle." *Journal of Legal Studies* 28 (1999): 393–412. 10.2139/ssrn.157448.

Darrow, Jonathan J. "Pharmaceutical Gatekeepers." *Indiana Law Review* 47, no. 2 (2014): 363–420.

Depoorter, Ben, and Robert Kirk Walker. "Copyright False Positives." *Notre Dame Law Review* 89, no. 1 (2013): 319–360.

Dinwoodie, Graeme B. "Secondary Liabilty for Online Trademark Infringement: The International Landscape." *Columbia Journal of Law and the Arts* 37, no. 4 (2014): 463–502.

Dobbs, Dan B., Paul T. Hayden, and Ellen M. Bublick. *Hornbook on Torts*. Second edition. Hornbook Series. St. Paul, MN: West Academic Publishing, 2016.

Durner, Tracey, and Liat Shetret. *Understanding Bank De-Risking and its Effects on Financial Inclusion: An exploratory study*. 2005-11.

Elkin-Koren, Niva. "After Twenty Years: Revisiting Copyright Liability of Online Intermediaries." In *The Evolution and Equilibrium of Copyright in the Digital Age*, edited by Susy Frankel and Daniel Gervais, 29–51. Cambridge University Press, 2014. 10.1017/CBO9781107477179.005.

Erickson, Kristofer, and Martin Kretschmer. "Empirical Approaches to Intermediary Liability." In *The Oxford Handbook of Online Intermediary Liability*, edited by Giancarlo F. Frosio, 104–121. Oxford handbooks. Oxford: Oxford University Press, 2020. 10.1093/oxfordhb/9780198837138.001.0001.

Erickson, Kristofer, and Martin Kretschmer. "This Video Is Unavailable." *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 9, no. 1 (2018): 75–89.

Ewert, Ralf, and Alfred Wagenhofer. "Effects of Increasing Enforcement on Financial Reporting Quality and Audit Quality." *Journal of Accounting Research* 57, no. 1 (2019): 121–168. 10.1111/1475-679X.12251.

Fiala, Lenka, and Martin Husovec. "Using Experimental Evidence to Design Optimal Notice and Takedown Process." *International Review of Law and Economics*, forthcoming. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3218286. 10.2139/ssrn.3218286.

Financial Actions Taks Force. FATF Guidance: Anti-Money Laundering and Terrorist Financing Measures and Financial Inclusion. https://www.fatf-gafi.org/media/fatf/content/images/Updated-2017-FATF-2013-Guidance.pdf.

Financial Crimes Enforcement Network. *What is money laundering?* Accessed June 30, 2023. https://www.fincen.gov/what-money-laundering.

Financial Industry Regulatory Authority. *Sanction Guidelines*. 2022-09-01.

Fischel, Daniel R., and Alan O. Sykes. "Corporate Crime." *Journal of Legal Studies* 25, no. 2 (1996): 315–349.

France, Guilherme. The impact of anti-money laundering and counter terrorist financing regulations on civi space and human rights. Accessed March 30, 2023. https://www.u4.no/publications/the-impact-of-anti-money-laundering-and-counter-terrorist-financing-regulations-on-civic-space-and-human-rights.

Friedman, David, and William Sjostrom. "Hanged for a Sheep: The Economics of Marginal Deterrence." *Journal of Legal Studies* 22, no. 2 (1993): 345–366. 10.1086/468168.

Frosio, Giancarlo F. "Reforming Intermediary Liability in the Platform Economy: A European Digital Single Market Strategy." *Northwestern University Law Review* 112 (2017-2018): 18–46.

Gabison, Garry A., and Miriam C. Buiten. "Platform Liability in Copyright Enforcement." *Columbia Science and Technology Law Review* 21, no. 2 (2020): 237–281.

Gadinis, Stavros, and Colby Mangels. "Collaborative Gatekeepers." *Washington and Lee Law Review Online* 73, no. 2 (2016): 797–914.

Garcia, Adriel. "The Kobayashi Maru of Ex-Offender Employment: Rewriting the Rules and Thinking outside Current Ban the Box Legislation." *Temple Law Review* 85, no. 4 (2013): 921–950.

Gill, Martin, and Geoff Taylor. "Preventing Money Laundering or Obstructing Business? Financial Companies' Perspectives on 'Know Your Customer' Procedures." *British Journal of Criminology* 44, no. 4 (2004): 582–594. https://www.jstor.org/stable/23639265.

Givati, Yehonatan, and Yotam Kaplan. "Harm Displacement and Tort Doctrine." *Journal of Legal Studies* 49, no. 1 (2020): 73–101. 10.1086/707601.

Gould, Eric D., Bruce A. Weinberg, and David B. Mustard. "Crime Rates and Local Labor Market Opportunities in the United States: 1979–1997." *Review of Economics and Statistics* 84, no. 1 (2002): 45–61. 10.1162/003465302317331919.

Granville, Kevin. "Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens." *New York Times*, 2018-03-19. Accessed June 30, 2023. https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html.

Hamdani, Assaf. "Gatekeeper Liability." *Southern California Law Review* 77 (2003): 53–122.

Hamdani, Assaf. "Who's Liable for Cyberwrongs." *Cornell Law Review* 87, no. 4 (2002): 901–957.

Heydon, Georgina, and Bronwyn Naylor. "Criminal Record Checking and Employment: The Importance of Policy and Proximity." *Australian & New Zealand Journal of Criminology* 51, no. 3 (2018): 372–394. 10.1177/0004865817723410.

Hill, Kashmir. "A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal." *New York Times*, 2022-08-25. https://www.nytimes.com/2022/08/21/technology/google-surveillance-toddler-photo.html.

Holzer, Harry J., and Michael A. Stoll. "How Willing Are Employers to Hire Ex-Offenders?" *Focus* 23, no. 2 (2004): 40–43.

Hu, Ying. "Individuals as Gatekeepers against Data Misuse." *Michigan Technology Law Review* 28, no. 1 (2021): 115–152.

Hylton, Keith N. "Optimal Law Enforcement and Victim Precaution." *RAND Journal of Economics* 27, no. 1 (1996): 197–206.

Jahner, Kyle. "Amazon Targets Phony IP Takedown Bids in 'Novel' Trio of Suits." *Bloomberg Law*, 2023-03-30.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications In R*. Springer texts in statistics, vol. 103. New York: Springer, 2013.

Janger, Edward J., and Aaron D. Twerski. "The Heavy Hand of Amazon: A Seller Not a Neutral Platform." *Brooklyn Journal of Corporate, Financial & Commercial Law* 14, no. 2 (2020): 259–274.

Jojarth, Christine. "Money Laundering: Motives, Methods, Impact, and Countermeasures." In *Transnational Organized Crime*, edited by Heinrich Böll Stiftung, Regine Schönenberg, and Etannibi E. O. Alemika, 17–34. Political science. Bielefeld: Transcript, 2013. https://www.jstor.org/stable/j.ctv1fxh0d.5.

Karaganis, Joe, and Jennifer M. Urban. "The Rise of the Robo Notice." *Communications of the ACM* 58, no. 9 (2015): 28–30. 10.1145/2804244.

Khanna, V. S. "Corporate Criminal Liability: What Purpose Does it Serve?" *Harvard Law Review* 109, no. 7 (1996): 1477–1534.

Koo, Hui-Wen, and I. P. L. Png. "Private Security: Deterrent or Diversion?" *International Review of Law and Economics* 14, no. 1 (1994): 87–101. 10.1016/0144-8188(94)90038-8.

Kraakman, Reinier H. "Gatekeepers: The Anatomy of a Third-Party Enforcement Strategy." *Journal of Law, Economics, and Organization* 2, no. 1 (1986): 53–104.

Kuczerawy, Aleksandra. "From 'Notice and Takedown' to 'Notice and Stay Down': Risks and Safeguards for Freedom of Expression." In *The Oxford Handbook of Online Intermediary Liability*, edited by Giancarlo F. Frosio, 524–543. Oxford handbooks. Oxford: Oxford University Press, 2020.

LaBine, Susan J., and Gary LaBine. "Determinations of Negligence and the Hindsight Bias." *Law and human behavior* 20, no. 5 (1996): 501–516. https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1007/bf01499038&casa_token=wqnr2h3t-woaaaaa:5fxp91eggovrpdrh0wtlbtqyoq3thxg19yzfmszklzerpiqzlufjvjmq5l1ypxoju3woierh6uxc3dde. 10.1007/BF01499038.

LaFave, Wayne R. *Criminal Law*. Sixth edition. Hornbook Series. St. Paul, MN: West Academic Publishing, 2017.

Levi, Michael. "Money Laundering and Its Regulation." *The Annals of the American Academy of Political and Social Science* 582 (2002): 181–194. https://www.jstor.org/stable/1049742?seq=1&cid=pdf-reference#references_tab_contents.

Levitin, Adam Jeremiah. *Consumer Finance: Markets and Regulation*. Aspen casebook series. New York: Wolters Kluwer, 2018.

Levitt, Steven D. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime." *American Economic Review* 87, no. 3 (1997): 270–290.

Lewin, Jeff L., and William N. Trumbull. "The Social Value of Crime?" *International Review of Law and Economics* 10 (1990): 271–284.

LexisNexis. *Explore the Global Cost of Financial Crime Compliance: Global Summary*. Accessed June 30, 2023. https://risk.lexisnexis.com/insights-resources/research/true-cost-of-financial-crime-compliance-study-global-report.

Lichtman, Doug, and Eric A. Posner. "Holding Internet Service Providers Accountable." *Supreme Court Economic Review* 14 (2006): 221–260.

Liu, Pheobe. "Lofi Girl Returns: YouTube Apologizes For Removing Popular Music Stream Due To 'Abusive' Copyright Notice." *Forbes*, 2022-07-12. Accessed June 30, 2023. https://www.forbes.com/sites/phoebeliu/2022/07/12/lofi-girl-returns-youtube-apologizes-for-removing-popular-music-stream-due-to-abusive-copyright-notice/?sh=5cc8722680f7.

Loo, Rory van. "The New Gatekeepers." *Virginia Law Review* 106, no. 2 (2020): 467–522.

Mann, Ronald J., and Seth R. Belzley. "The Promise of Internet Intermediary Liability." *William and Mary Law Review* 47, no. 1 (2005): 239–308.

Manns, Jeffrey. "Private Monitoring of Gatekeepers: The Case of Immigration Enforcement." *University of Illinois Law Review* 2006, no. 5 (2006): 887–974.

Markey, Patrick M., and Charlotte N. Markey. "Changes in Pornography-Seeking Behaviors Following Political Elections: An Examination of the Challenge Hypothesis." *Evolution and Human Behavior* 31, no. 6 (2010): 442–446. https://www.sciencedirect.com/science/article/pii/S1090513810000711. 10.1016/j.evolhumbehav.2010.06.004.

Martin, Austin. "A Gatekeeper Approach to Product Liability for Amazon." *George Washington Law Review* 89, no. 3 (2021): 766–800.

McDonnell, Patrick J. "Airline Faces Fines for Ferrying Illegals." *Los Angeles Times*, 2002-04-06. Accessed June 30, 2023. https://www.latimes.com/archives/la-xpm-2002-apr-06-me-ins-story.html.

Mehra, Ajay K. "The ED Wasn't Created to Target the Opposition and Dissenters." *The Wire*, 2023-01-06. Accessed June 30, 2023. https://thewire.in/politics/enforcement-directorate-target-opposition-dissenters.

Metz, Sam. Utah Law Requiring Porn Sites Verify User Ages Takes Effect. https://apnews.com/article/porn-age-verification-utah-8f8f4960ad1ec4afc5d59fd7d34c3b9d.

Meunier, Thibault. Humanity Wastes about 500 Years per Day on CAPTCHAs. It's Time to End this Madness. Accessed January 21, 2023. https://blog.cloudflare.com/introducing-cryptographic-attestation-of-personhood/.

Mohammed, Kenneth. "Banks Are Leaving the Caribbean. It's Unfair and Will Backfire on the West." *The Guardian*, 2022-11-11. Accessed June 29, 2023. https://www.theguardian.com/global-development/commentisfree/2022/nov/11/banks-leaving-caribbean-unfair-backfire-on-west.

Morse, Adair, Wei Wang, and Serena Wu. "Executive Lawyers." *Journal of Law and Economics* 59, no. 4 (2016): 847–888.

National Conference of State Legislatures. Ban the Box. Accessed June 30, 2023. https://www.ncsl.org/civil-and-criminal-justice/ban-the-box.

National Consumer Law Center. Broken Records Redux: How Errors by Criminal Background Check Companies Continue to Harm Consumers Seeking Jobs and Housing. Accessed June 30, 2023. https://www.nclc.org/wp-content/uploads/2022/09/report-broken-records-redux.pdf.

National Inventory of Collateral Consequences of Conviction. Collateral Consequences Inventory. https://niccc.nationalreentryresourcecenter.org/consequences.

Nikolaides, Kira. "Collateral Consequences of Conviction: Barriers to Employment." *Berkeley Journal of Criminal Law Blog*, 2022-08-31. Accessed June 30, 2023. https://www.bjcl.org/blog/collateral-consequences-of-conviction-barriers-to-employment.

Noonan, Laura, and Alan Smith. "Global Anti-money Laundering Fines Surge 50%: New Data Fuels Doubts over Effectiveness of Crackdown on Financial Crime since 2008 Crisis." *Financial Times*, 2023-01-18. Accessed June 30, 2023. https://www.ft.com/content/7a4821e6-96f1-475c-ae55-6401e402061f.

Nordemann, Jan Bernd. "Liability for Copyright Infringements on the Internet: Host Providers (Content Providers) -The German Approach." *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 2 (2011): 37–49. Accessed June 30, 2023. https://www.jipitec.eu/issues/jipitec-2-1-2011/2962/JIPITEC_Nordemann.pdf.

Partnoy, Frank. "Barbarians at the Gatekeepers: A Proposal for a Modified Strict Liability Regime." *Washington University Law Quarterly* 79, no. 2 (2001): 491–548.

Partnoy, Frank. "Strict Liability for Gatekeepers: A Reply to Professor Coffee." *Boston University Law Review* 84, no. 2 (2004): 365–376.

Perel, Maayan, and Niva Elkin-Koren. "Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement." *Florida Law Review* 69, no. 1 (2017): 181–222.

Perrett, Connor. "YouTube Removed Popular 'Lofi Girl' Streams after It Received False and 'Abusive' Copyright Claims." *Insider*, 2022-07-11. Accessed June 30, 2023. https://www.businessinsider.com/youtube-lofi-girl-suspended-beats-to-study-copyright2022-7.

Philipson, Tomas J., and Richard A. Posner. "The Economic Epidemiology of Crime." *Journal of Law and Economics* 39 (1996): 405–433.

Png, I. P. L. "Optimal Subsidies and Damages in the Presence of Judicial Error." *International Review of Law and Economics* 6, no. 1 (1986): 101–105. 10.1016/0144-8188(86)90042-6.

Polinsky, A. Mitchell, and Yeon-Koo Che. "Decoupling Liability: Optimal Incentives for Care and Litigation." *RAND Journal of Economics* 22, no. 4 (1991): 562–570. https://www.jstor.org/stable/2600989.

Polinsky, A. Mitchell, and Steven Shavell. "Chapter 6 The Theory of Public Enforcement of Law." In *Handbook of Law and Economics. Volume 1*, edited by A. Mitchell Polinsky and Steven Shavell, 403–454. Handbook of Law and Economics series, 1574-0730, vol. 1. Amsterdam: Elsevier; London: 2007. 10.1016/S1574-0730(07)01006-7.

Polinsky, A. Mitchell, and Steven Shavell. "Should Employees Be Subject to Fines and Imprisonment Given the Existence of Corporate Liability?" *International Review of Law and Economics* 13 (1993): 239–257.

Polinsky, A. Mitchell, and Steven Shavell. "The Optimal Tradeoff between the Probability and Magnitude of Fines." *American Economic Review* 69, no. 5 (1979): 880–891.

Polinsky, A. Mitchell, and Steven Shavell. "The Optimal Use of Fines and Imprisonment." *Journal of Public Economics* 24 (1984): 89–99.

Prescott, J. J., and Benjamin Pyle. "Identifying the Impact of Labor Market Opportunities on Criminal Behavior." *International Review of Law and Economics* 59 (2019): 65–81.

Prescott, J. J., and Sonja B. Starr. "The Power of a Clean Slate." *Regulation* 43, no. 2 (2020): 28–34.

Reuters Newsroom. Cambodian Police Raid Alleged Cybercrime Trafficking Compounds. Accessed June 30, 2023. https://www.reuters.com/world/asia-pacific/cambodian-police-raid-alleged-cybercrime-trafficking-compounds-2022-09-21/.

Riis, Thomas, and Sebastian Felix Schwemer. "Leaving the European Safe Harbor, Sailing Towards Algorithmic Content Regulation." *Journal of Internet Law* 22, no. 7 (2018): 1–21. 10.2139/ssrn.3300159.

Rogers, Brishen. "Toward Third-Party Liability for Wage Theft." *Berkeley Journal of Employment and Labor Law* 31, no. 1 (2010): 1–64.

Saadatmand, Yassaman, Michael Toma, and Jeremy Choquette. "The War On Drugs And Crime Rates." *Journal of Business & Economics Research* 10, no. 5 (2012): 285–290. https://clutejournals.com/index.php/jber/article/view/6980. 10.19030/jber.v10i5.6980.

Saperstein, Lanier, Geoffrey Sant, and Michelle Ng. "The Failure of Anti-Money Laundering Regulation: Where is the Cost-Benefit Analysis?" *Notre Dame Law Review* 91, no. 1 (2015): 1–10.

Seng, Daniel. "Who Watches the Watchmen": An Empirical Analysis of the Reasons for Rejecting Copyright Takedown Notices. https://ssrn.com/abstract=3687861.

Seng, Daniel. "The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices." *Virginia Journal of Law & Technology* 18, no. 3 (2014): 369–473.

Shavell, Steven. "A Model of Optimal Incapacitation." *American Economic Review* 77, no. 2 (1987): 107–110.

Shavell, Steven. *Economic Analysis of Accident Law*. Cambridge Mass. Harvard University Press, 1987.

Shavell, Steven. *Foundations of economic analysis of law*. Cambridge, Mass. Belknap Press of Harvard University Press; London: 2004.

Shavell, Steven. "Individual Precautions to Prevent Theft: Private versus Socially Optimal Behavior." *International Review of Law and Economics* 11 (1991): 123–132.

Shavell, Steven. "The Judgment Proof Problem." *International Review of Law and Economics* 6, no. 1 (1986): 45–58.

Shavell, Steven. "The Mistaken Restriction of Strict Liability to Uncommon Activities." *Journal of Legal Analysis* 10 (2018): 1–45. 10.1093/jla/lay004.

Shavell, Steven. "The Optimal Use of Nonmonetary Sanctions as a Deterrent." *American Economic Review* 77, no. 4 (1987): 584–592.

Shen, Po-Yang. "Shen, Po-Yang: Zhongguo Zhengfu Ruhe Liyong Jiaxiaoxi Yingxiang Taiwan Xuanju." *Up Media*, 2019-11-16. https://www.upmedia.mg/news_info.php?Type=2&SerialNo=75337.

Smith, Mitch. "Midwestern Floods Pit Communities Against One Another as Levees Rise Ever Higher." *New York Times*, 2019-05-07. Accessed March 7, 2023. https://www.nytimes.com/2019/05/07/us/flood-midwest-levees.html.

Song, Lisa, Al Shaw, Patrick Michels, and Alex Heeb. "New Model Shows Towns on the Wrong Side of an Illinois Levee District Are Treading WaterWater." *ProPublica*, 2018-03-30. Accessed March 6, 2023. https://www.propublica.org/article/new-model-shows-towns-on-the-wrong-side-of-an-illinois-levee-district-are-treading-water.

Stessens, Guy. *Money Laundering: A New International Law Enforcement Model*. Cambridge Studies in International and Comparative Law. Cambridge: Cambridge University Press, 2009.

Stigler, George J. "The Optimum Enforcement of Laws." *Journal of Political Economy* 78, no. 3 (1970): 526–536. https://www.jstor.org/stable/1829647.

Strader, J. Kelly. *Understanding White Collar Crime*. Fourth edition. Understanding series. Durham North Carolina: Carolina Academic Press, 2017.

Sugie, Naomi F., Noah D. Zatz, and Dallas Augustine. "Employer Aversion to Criminal Records: An Experimental Study of Mechanisms." *Criminology* 58, no. 1 (2020): 5–34. 10.1111/1745-9125.12228.

Sykes, Alan O. "The Economics of Vicarious Liability." *Yale Law Journal* 93, no. 7 (1984): 1231–1280.

Takáts, Előd. "A Theory of "Crying Wolf": The Economics of Money Laundering Enforcement." *Journal of Law, Economics, and Organization* 27, no. 1 (2011): 32–78.

Tran, Huong Thi Thanh, and Ha Thi Thu Le. "The Impact of Financial Inclusion on Poverty Reduction." *Asian Journal of Law and Economics* 12, no. 1 (2021): 95–119.

Tripodi, Stephen J., Johnny S. Kim, and Kimberly Bender. "Is Employment Associated with Reduced Recidivism?: The Complex Relationship between Employment and Crime." *International Journal of Offender Therapy and Comparative Criminology* 54, no. 5 (2010): 706–720. 10.1177/0306624X09342980.

Tuch, Andrew F. "Conflicted Gatekeepers: The Volcker Rule and Goldman Sachs." *Virginia Law & Business Review* 7, no. 2 (2012): 365–420.

Tuch, Andrew F. "Multiple Gatekeepers." *Virginia Law Review* 96, no. 7 (2010): 1583–1672. https://ssrn.com/abstract=1577405.

U.S. Department of Justice. Lafarge Pleads Guilty to Conspiring to Provide Material Support to Foreign Terrorist Organizations. Accessed June 29, 2023. https://www.justice.gov/opa/pr/lafarge-pleads-guilty-conspiring-provide-material-support-foreign-terrorist-organizations.

United States Copyright Office. Sectio 512 of Title 17: A Report of the Register of Copyrights. Accessed June 30, 2023. https://www.copyright.gov/policy/section512/section-512-full-report.pdf.

United States Postal Service. U.S. Postal Service Releases Dog Attack National Rankings. Accessed August 21, 2022. https://about.usps.com/newsroom/national-releases/2022/0602-usps-releases-dog-attack-national-rankings.htm.

United States Securities and Exchange Commission. *SEC Charges Wells Fargo Advisors With Anti-Money Laundering Related Violations*. 2022-05-20. Accessed June 30, 2023. https://www.sec.gov/news/press-release/2022-85?utm_medium=email&utm_source=govdelivery.

Urban, Jennifer M., Joe Karaganis, and Brianna L. Schofield. *Notice and Takedown in Everyday Practice*. March 2017.

Urban, Jennifer M., Joe Karaganis, and Brianna L. Schofield. "Notice and Takedown: Online Service Provider and Rightsholder Accounts of Everyday Practice." *Journal of the Copyright Society of the USA* 64, no. 3 (2017): 371–ii.

Visher, Christy A., Laura Winterfield, and Mark B. Coggeshall. "Ex-offender Employment Programs and Recidivism: A Meta-Analysis." *Journal of Experimental Criminology* 1, no. 3 (2005): 295–316. 10.1007/s11292-005-8127-x.

Vollaard, Ben, and Pierre Koning. "The Effect of Police on Crime, Disorder and Victim Precaution: Evidence from a Dutch Victimization Survey." *International Review of Law and Economics* 29, no. 4 (2009): 336–348. 10.1016/j.irle.2009.03.003.

Wakefield, Jane. Neighbour Wins Privacy Row over Smart Doorbell and Cameras. Accessed January 21, 2023. https://www.bbc.com/news/technology-58911296.

Wan, Ke Steven. "Gatekeeper Liability versus Regulation of Wrongdoers." *Ohio Northern University Law Review* 34, no. 2 (2008): 483–522.

Webster, Colin, and Sarah Kingston. *Poverty and Crime Review*. April 2014.

Withers, Iain. HSBC fined $85 Mln for Anti-Money Laundering Failings. https://www.reuters.com/business/hsbc-fined-85-mln-anti-money-laundering-failings-2021-12-17/.

YouTube. Access for all, a balanced ecosystem, and powerful tools. https://blog.youtube/news-and-events/access-all-balanced-ecosystem-and-powerful-tools/.

YouTube. Copyright Transparency Report: H2 2022. Accessed June 30, 2023. https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-22_2022-7-1_2022-12-31_en_v1.pdf.

YouTube. Qualify for Content ID. Accessed June 30, 2023. https://support.google.com/youtube/answer/1311402?hl=en.