THE USE AND MISUSE OF COORDINATED PUNISHMENTS*

DANIEL BARRON AND YINGNI GUO

Communication facilitates cooperation by ensuring that deviators are collectively punished. We explore how players might misuse communication to threaten one another, and we identify ways that organizations can deter misuse and restore cooperation. In our model, a principal plays trust games with a sequence of short-run agents who communicate with each other. An agent can shirk and then extort pay by threatening to report that the principal deviated. We show that these threats can completely undermine cooperation. Investigations of agents' efforts, or dyadic relationships between the principal and each agent, can deter extortion and restore some cooperation. Investigations of the principal's action, on the other hand, typically do not help. Our analysis suggests that collective punishments are vulnerable to misuse unless they are designed with an eye toward discouraging it. *JEL* Codes: C73, D02, D70.

I. INTRODUCTION

Productive relationships thrive on the enthusiastic cooperation of their participants. In many settings, individuals cooperate because they expect opportunistic behavior to be punished (Malcomson 2012). Communication plays an essential role in coordinating these punishments, because it allows those who do not directly observe misbehavior to nevertheless punish the perpetrator. These coordinated punishments are central to cooperation among participants in online marketplaces (Hörner and Lambert 2018), as well as between managers and workers (Levin 2002), suppliers and customers (Greif, Milgrom, and Weingast 1994; Bernstein 2015), and members of communities (Ostrom 1990).

Once armed with the power to trigger coordinated punishments, individuals face a grave temptation: they can extort

*We thank Nageeb Ali, Charles Angelucci, Nemanja Antic, Alessandro Bonatti, Renee Bowen, Joyee Deb, Wouter Dessein, Matthias Fahn, Benjamin Friedrich, George Georgiadis, Marina Halac, Johannes Hörner, Peter Klibanov, Ilan Kremer, Nicolas Lambert, Stephan Lauermann, Jin Li, Elliot Lipnowski, Shuo Liu, Bentley MacLeod, David Miller, Joshua Mollner, Dilip Mookherjee, Arijit Mukherjee, Jacopo Perego, Michael Powell, Luis Rayo, Jonah Rockoff, Mark Satterthwaite, Andy Skrzypacz, Takuo Sugaya, Jeroen Swinkels, Joel Watson, and audiences at many conferences, workshops, and seminars. We thank the UCSD theory reading group for comments on a draft of this article, and Andres Espitia for excellent research assistance.

The Quarterly Journal of Economics (2021), 471–504. doi:10.1093/qje/qjaa035.
Advance Access publication on October 10, 2020.

concessions from their partners by threatening to falsely report opportunistic behavior (Gambetta 1993; Dixit 2003a, 2007). In this article, we explore how individuals might misuse coordinated punishments. We emphasize two overarching takeaways. First, we show that misuse is a serious vulnerability that can completely undermine cooperation. Second, we identify practical ways for organizations to restore cooperation in the face of this vulnerability.

The possibility of misuse is a serious concern for online platforms, where most interactions are short-lived and coordinated punishments are essential for ensuring cooperation. In an evocative recent example, an investigation of the lodging platform Airbnb uncovered a network of hosts who misused the platform's review system to extort guests (Conti 2019). These hosts reneged on their obligations by altering guests' accommodations at the last minute. While Airbnb allows guests to request a refund in these circumstances, the hosts deterred refund requests by writing scathing reviews of guests who complained. By misusing Airbnb's review system in this way, hosts ensured that they received payment despite reneging on their end of the deal. The resulting scandal was serious enough to prompt action from the FBI.

This type of misuse is hardly unique to Airbnb. In the early days of eBay, buyers and sellers could review each other based on the accuracy of product descriptions, the quality of delivery service, and the timeliness of payment. This review system led to a severe form of misuse, in which sellers would extort positive reviews from buyers by threatening to negatively review any buyer who complained. Klein, Lambertz, and Stahl (2016) show that these threats led to lower seller effort and lower buyer satisfaction. Review aggregators are similarly susceptible to misuse, as a wedding planning business learned when its reputation was sullied by a customer who posted vitriolic reviews in an attempt to extort services (Proctor 2018). The phenomenon of extortionary reviews is widespread enough that platforms like TripAdvisor and Etsy have instituted explicit antiextortion policies. Etsy's policy, for example, forbids a buyer from leaving "a negative review in an attempt to force the seller into providing a refund" or additional items.[1] The possibility of misuse

1. TripAdvisor's antiextortion policy is at https://www.tripadvisor.com/TripAdvisorInsights/w592; Etsy's policy is at https://www.etsy.com/legal/policy/extortion/239966959186. These policies note that the platform has limited ability to combat extortion that occurs outside its official messaging system.

has also spurred responses from regulatory authorities and lawmakers.[2]

To explore how coordinated punishments can be used and misused, we consider a model of a long-run principal who interacts with a sequence of short-run agents. Each agent exerts costly effort to benefit the principal, who can then choose to pay him. Agents observe only their own interactions but can communicate with one another. To capture the idea that extortion entails action-contingent threats—that is, "pay me or else I will punish you"—we allow each agent to make a threat when he chooses his effort. This threat, which is observed by the principal but not by other agents, associates a message to each possible payment. Agents then follow through on their threats.

In this model, misuse completely undermines cooperation. The principal is willing to pay an agent only if she would otherwise be punished by future agents. Communication is therefore essential for cooperation. Once endowed with a message that triggers punishments, however, an agent can extort the principal by shirking and then threatening to send that message unless the principal pays him. Because this threat is enough to induce the principal to pay a hard-working agent, it is also enough to induce her to pay a shirking agent. Thus, the pay that an agent can demand is essentially independent of his effort. The stark implication of this logic is that agents do not exert any effort.

After establishing this impossibility result, we explore how organizations can deter extortion and encourage cooperation. We focus on two instruments that are available in many cooperative endeavors: investigations, which we model as public signals of either the agents' efforts or the principal's transfers, and dyadic relationships, which we model as a coordination game played by the principal and each agent.

The unifying idea of these instruments is that an agent is willing to exert effort only if doing so increases his leverage over the principal, which we define as the harshest punishment that the agent can trigger with a message. An agent can extort any transfer that is smaller than his leverage. If an agent's leverage is independent of his effort, as it is in any equilibrium of our

2. In the United Kingdom, the Competition and Markets Authority has noted that customers sometimes use the threat of negative reviews to demand discounts (Peachey 2015). The Washington state legislature has considered a bill to deter extortionary online reviews.

baseline model, then he has no incentive to exert effort. If an agent's leverage is increasing in his effort, on the other hand, then he might exert effort to increase his leverage, so that he can demand higher pay. An instrument is valuable exactly when it ties leverage to effort in this way.

Building on this idea, we show that investigations into agents' efforts typically improve cooperation, whereas investigations into the principal's transfers typically do not. Effort signals are useful for deterring extortion, not because agents are directly rewarded or punished on the basis of these signals but because these signals can tie leverage to effort. They do so by ensuring that harder-working agents can trigger harsher coordinated punishments. Agents are then willing to exert effort to obtain higher leverage and demand higher pay. In contrast, transfer signals can reveal whether the principal paid an agent but not whether that pay was deserved. Hence, such signals typically cannot tie leverage to effort and so cannot improve cooperation. The only exception is that under stringent conditions, transfer signals can make the principal indifferent between transfers in a way that leads to some effort. Even then, the extent of cooperation is limited by the need for occasional on-path punishments.

Next, we study how dyadic relationships between the principal and each agent can tie leverage to effort. Unlike online platforms, where short-lived interactions are the norm, manager—worker relationships are typically long-lived. As with short-lived relationships, institutions that coordinate punishments, such as unions, have the potential to improve cooperation and lead to exceptional productivity in long-lived relationships (Freeman and Medoff 1979; Levin 2002). To do so, however, these institutions must first deter agents from misusing the coordinated punishments that they make possible.

We show that dyadic relationships can guard against misuse by rewarding the principal for refusing to pay a shirking agent. Dyadic relationships therefore complement coordinated punishments: organizations with strong dyadic relationships can implement severe coordinated punishments without opening the door to extortion, and those with weak dyadic relationships give their agents little leverage and so result in low effort. In the latter case, misuse remains a real threat to coordinated punishments, which is consistent with General Motors' experience at its plant in Fremont, California, in the 1980s. Workers at that plant misused the threat of grievances to get away with "shirking"

behaviors like absenteeism and drug use during working hours, leading to low productivity that eventually resulted in the plant's closure (Glass and Langfitt 2015).[3]

The premise of our analysis is that while organizations can potentially benefit from coordinated punishments, they cannot perfectly control how their members use these punishments. Our model illustrates a stark version of this premise, in which misuse completely undermines cooperation. In practice, and as we explore in Section VI, we do not expect every agent in a particular context to engage in misuse. Rather, our point is that coordinated punishments are vulnerable to misuse, and this vulnerability can seriously impair cooperation. Organizations ignore this vulnerability at their peril. Guarding against misuse demands a fundamentally different approach to designing incentive systems. In our setting, these systems are embedded in the rules or culture of an organization, as represented by an equilibrium of our game. Our lessons extend to other settings in which cheap-talk messages are used to motivate cooperation.

I.A. Related Literature

Our contribution is to explore how misuse undermines coordinated punishments and how organizations can combat it. Therefore, we build on the literature that studies how coordinated punishments support cooperation (Milgrom, North, and Weingast 1990; Greif, Milgrom, and Weingast 1994; Dixit 2003a, 2003b; Pei and Strulovici 2020; Strulovici 2020). Much of this literature has as its goal the identification of network structures or equilibrium strategies that are particularly conducive to cooperation (Lippert and Spagnolo 2011; Wolitzky 2013; Ali and Miller 2013, 2016; Ali, Miller, and Yang 2017). Especially related is Ali and Miller (2016), which shows that players might not report deviations if doing so reveals they are more willing to renege on their own promises. Extortion is a different but complementary obstacle to coordinated punishments.

Since extortion is inherently action-contingent—that is, "pay me or else I will punish you"—our analysis is related to a growing literature on action-contingent threats and promises. Like us,

3. Another example comes from the recent Wells Fargo scandal. Wells Fargo faced allegations that it punished employees who spoke up about fraudulent practices by falsely reporting them to FINRA for unethical behavior (Arnold and Smith 2016). FINRA then shared this information with other prospective employers.

some of these publications assume that players commit to threats to allow for action-contingent deviations (Wolitzky 2012; Ortner and Chassang 2018; Chassang and Padro i Miquel 2019).[4] In our setting, we can also reinterpret commitment as an equilibrium refinement of the game without commitment, which is related to the approach taken in Zhu (2018, 2020). We contribute to this literature by studying how misuse undermines coordinated punishments and exploring new ways for organizations to deter it.

Most of the literature on cooperation focuses on the use of coordinated punishments rather than the potential for misuse. Dixit (2003a, 2007) was perhaps the first to formally model the misuse of coordinated punishments, albeit in a setting with centralized enforcers rather than decentralized communication. Bowen, Kreps, and Skrzypacz (2013), who study local adaptation in communities, consider a type of misuse that is not action-contingent. In contrast to that paper, our agents make action-contingent threats.

In our setting, an agent essentially threatens the principal with a bad "outside option" unless she pays him. Our article is therefore connected to the literature on bargaining and renegotiation in repeated games. Particularly related are papers that allow players to bargain over surplus in equilibrium (Baker, Gibbons, and Murphy 2002; Halac 2012, 2015, Miller and Watson 2013; Goldlücke and Kranz 2020; Miller, Olsen, and Watson 2020) and the literature on coalitional deviations (Ali and Liu 2018; Liu 2019). By focusing on communication across agents, our article studies a setting in which the principal's "outside option" depends on how messages affect future equilibrium play.[5]

More broadly, our framework builds on the relational contracting literature (Bull 1987; MacLeod and Malcomson 1989; Baker, Gibbons, and Murphy 1994; Levin 2003), especially those papers that study coordinated punishments (e.g., Levin 2002). We study how extortion undermines such punishments. Recent papers have explored relational contracts in the presence of limited transfers (Fong and Li 2017; Barron, Li, and Zator 2019), asymmetric information (Halac 2012; Malcomson 2016),

4. Indeed, Ortner and Chassang (2018) have an appendix that studies extortion. However, that appendix assumes that reports lead to exogenous and fixed punishments. The point of our analysis is to show how to optimally link messages to punishments.

5. We formalize the connection between our model and the literatures on bargaining and coalitional deviations in Online Appendix B.

or both (Li, Matouschek, and Powell 2017; Guo and Hörner 2018; Lipnowski and Ramos 2020). We focus on a monitoring friction—agents do not observe one another's relationships—which implies that cooperation must rely on communication. Other papers that study relational contracts with bilateral monitoring, including Board (2011), Andrews and Barron (2016), and Barron and Powell (2019), do not allow agents to communicate. We complement these papers by identifying a reason communication might be ineffective at sustaining cooperation.

II. Model

Our baseline model is the following extortion game. A long-run principal ("she") interacts with a sequence of short-run agents (each "he"). In each period $t \in \{0, 1, 2, \dots\}$, the principal and agent t play a trust game: agent t exerts effort, then the principal chooses how much to pay him. This interaction is observed only by the principal and agent t, but agent t can send a public message at the end of period t. Our key assumption is that before transfers are paid, agent t makes a threat, which is a mapping from the transfer he receives to the message he sends and is observed by the principal but not by other agents. Agent t then follows through on this threat.

Formally, the stage game in period t is:

i. Agent t chooses his effort $e_t \in \mathbb{R}_+$ and a threat $\mu_t : \mathbb{R} \to M$, where M is a large, finite message space.[6] Both e_t and μ_t are observed by the principal but not by any other agent.
ii. The principal makes a transfer to agent t, $s_t \geqslant 0$, which is observed by agent t but not by other agents.[7]
iii. The message $m_t = \mu_t(s_t)$ is realized and observed by all players.

The principal's period-t payoff and agent t's utility are $(e_t - s_t)$ and $(s_t - c(e_t))$, respectively, where $c(\cdot)$ is strictly increasing, strictly convex, and twice continuously differentiable, as well as satisfying $c(0) = c'(0) = 0$. We assume that there exists a first-best

6. The assumption that M is finite simplifies the proofs (by ensuring that various maxima and minima exist) but is not essential for the results.

7. For almost all of our results, the assumption that agents do not pay the principal is without loss. The exception is Section V; we allow agents to pay the principal in that section.

effort, e^{FB}, such that $c'(e^{FB}) = 1$. The principal has discount factor $\delta \in [0, 1)$, with corresponding normalized discounted payoffs $\Pi_t = (1-\delta)\sum_{t'=t}^{\infty} \delta^{t'-t}(e_{t'} - s_{t'})$. Players observe a public randomization device (notation for which is suppressed) in every step of the stage game.

The principal observes everything, whereas agents observe only their own interactions with the principal and all messages. Our solution concept is perfect Bayesian equilibrium (PBE).[8] Some of our results focus on principal-optimal equilibria, which maximize the principal's ex ante expected payoff among all equilibria.

In the context of Airbnb, the agents are hosts who exert effort (e_t) to ensure that their properties are safe, comfortable, and described accurately. The principal is a guest who rewards such efforts by treating a property well, following house rules, and not demanding an undeserved refund (s_t). To give the guest an incentive to follow through on these rewards, the platform allows hosts to review guests (m_t), and negative reviews make it harder for the guest to rent from other hosts in the future. As Conti (2019) discovered, some hosts took advantage of this review system to "shirk" on quality, knowing that they could use the threat of a negative review (μ_t) to force guests to pay them. In Section III, we show that agents have an incentive to similarly misuse coordinated punishments in the extortion game. Sections IV and V build on this result to explore how organizations can deter misuse.[9]

The threat, μ_t, is a transparent way to show how agents can misuse coordinated punishments, one that has precedent in the approaches taken by Dixit (2003a), Wolitzky (2012), Chassang and Padro i Miquel (2019), and Ortner and Chassang (2018). Other modeling approaches would result in a similar kind of misuse. We study several of these alternative models in Online Appendix B. There, we show that a similar kind of misuse arises when the principal and each agent Nash bargain over that agent's message. This bargaining model is similar in spirit to Halac (2012,

8. See the definition of plain PBE in Watson (2017).

9. In some of our examples, e_t and s_t have slightly different interpretations than effort and transfer. In particular, e_t sometimes includes a transfer paid by agent t, whereas s_t sometimes includes a productive action taken by the principal. The model can be generalized to account for this interpretation. In particular, we could make e_t a payment and s_t a productive action, or even make both e_t and s_t productive, without changing our argument for why misuse undermines cooperation.

2015), Miller and Watson (2013), and Miller, Olsen, and Watson (2020) and is related to the literature on coalitional deviations (Ali and Liu 2018; Liu 2019). We also prove that we can reinterpret commitment to μ_t as the result of agents having preferences for either reciprocity (Fehr, Powell, and Wilkening 2020) or keeping one's word (Vanberg 2008), or as an equilibrium refinement similar to Dewatripont (1987), Tranaes (1998), and Zhu (2018, 2020). In these alternative models, the main lessons from our analysis hold: coordinated punishments are vulnerable to misuse, and tying an agent's leverage to his effort deters misuse.

Online Appendix C considers alternative communication structures, including models in which the principal can send messages or commit to threats, as well as ones where agents can make repeated threats. In most of these variants, extortion continues to undermine cooperation. We also identify particular communication structures that can lead to cooperation in equilibrium, although these positive results typically come with substantial caveats.

We occasionally compare our results to a benchmark without extortion. Define the no-extortion game as identical to the extortion game, except that each agent t chooses m_t at the end of period t rather than being committed to μ_t. In the no-extortion game, agents cannot shirk and then make action-contingent threats, so they cannot misuse communication.

III. Threats Undermine Equilibrium Cooperation

This section shows how coordinated punishments are used and misused in equilibrium. We first illustrate how coordinated punishments sustain cooperation in the no-extortion game. Then, we show that misuse leads cooperation to completely unravel. This impossibility result uncovers the economics of misuse and forms the foundation for the rest of our analysis.

Cooperation requires agents to communicate with one another, because without communication an agent would have no way to punish the principal for deviating. In the no-extortion game, this type of communication is enough to sustain cooperation.

Proposition 1. In the no-extortion game, $e_t = e^*$ and $s_t = c(e^*)$ in each $t \geqslant 0$ of every principal-optimal equilibrium, where e^* equals the minimum of e^{FB} and the largest e that satisfies $c(e) = \delta e$.

Proof. We first argue that total equilibrium surplus is at most $e^* - c(e^*)$. By definition of e^{FB}, equilibrium surplus is at most $e^{FB} - c(e^{FB})$. If $c(e^{FB}) \leqslant \delta e^{FB}$, then $e^* = e^{FB}$ and the result follows. If $c(e^{FB}) > \delta e^{FB}$, then let $\overline{\Pi}$ be the principal's maximum ex ante equilibrium payoff. In any period $t \geqslant 0$ of any equilibrium, $(1-\delta)s_t \leqslant \delta\overline{\Pi}$ and $s_t - c(e_t) \geqslant 0$ must hold, because otherwise the principal or agent t could profitably deviate from s_t or e_t, respectively. Therefore, $(1-\delta)c(e_t) \leqslant \delta\overline{\Pi}$. Let $\bar{e}$ be the effort that maximizes $e - c(e)$ among those efforts that are attained in any period of any equilibrium. Then, $(1-\delta)c(\bar{e}) \leqslant \delta\overline{\Pi} \leqslant \delta(\bar{e} - c(\bar{e}))$ and so $c(\bar{e}) \leqslant \delta\bar{e}$. We conclude that $\bar{e} \leqslant e^* < e^{FB}$, so equilibrium surplus is at most $e^* - c(e^*)$.

Consider the following strategy profile for each period $t \geqslant 0$: if $m_{t'} = C$ in all $t' < t$, then agent t chooses $e_t = e^*$; the principal chooses $s_t = c(e^*)$ if $e_t = e^*$ and $s_t = 0$ otherwise; and agent t chooses $m_t = C$ if neither player deviates and $m_t = D$ otherwise. If $m_{t'} \neq C$ in at least one $t' < t$, then $e_t = s_t = 0$ and $m_t = D$.

Once $m_{t'} \neq C$ in some $t' < t$, this strategy profile specifies the stage-game equilibrium and players cannot profitably deviate. If $m_{t'} = C$ in all $t' < t$, then agent t has no profitable deviation because he earns 0 on-path and no more than 0 from deviating. The principal has no profitable deviation because $(1-\delta)s_t \leqslant \delta(e^* - c(e^*))$ is implied by $c(e^*) \leqslant \delta e^*$. This strategy is therefore an equilibrium. It is principal-optimal because it generates total surplus $e^* - c(e^*)$, which is the maximum equilibrium surplus, and it holds agents at their min-max payoffs. Moreover, every principal-optimal equilibrium gives the principal a payoff of $e^* - c(e^*)$ and so must entail $e_t = e^*$ in every period. □

The proof of Proposition 1 relies on the following equilibrium construction. On the equilibrium path, each agent sends the message C if the principal pays him and D otherwise. Future agents min-max the principal if they observe the message D. Off the equilibrium path, a shirking agent sends a message that is independent of the principal's transfer, so the principal pays him nothing. The principal would rather pay a hard-working agent a transfer than be punished, and each agent would rather exert effort than shirk and forgo the transfer, so this construction can motivate effort.

Proposition 1 summarizes a core idea from much of the literature on coordinated punishments: the principal pays a hard-working agent because that agent would otherwise send a message that triggers future punishments. Implicit in this

construction, and in much of the literature on coordinated punishments, is the requirement that shirking agents do not make similar threats, so that the principal refrains from paying a shirking agent. As our introduction makes clear, actual behavior does not always conform to this requirement. For instance, some Airbnb hosts shirk ($e_t = 0$) and then threaten to leave negative feedback ($m_t = D$) unless they are paid ($s_t > 0$).

The extortion game allows shirking agents to make exactly this type of threat. Our next result, which serves as the foundation for our analysis, shows that these threats destroy cooperation.

PROPOSITION 2. In the extortion game, every equilibrium entails $e_t = s_t = 0$ in every $t \geqslant 0$.

Proof. Fix a history of messages, $m^{t-1} = (m_0, m_1, \ldots, m_{t-1})$, and let

$$\overline{\Pi} = \max_{m \in M} \left\{ \mathbb{E}\left[\Pi_{t+1} | m^{t-1}, m_t = m \right] \right\}$$

be the principal's maximum continuation surplus that can be induced by some message, which we denote $m_t = C$. Let $\underline{\Pi}$ be the similarly defined minimum continuation payoff, with corresponding message $m_t = D$.

Suppose that agent t chooses some $e_t > 0$. He is willing to do so only if $s_t \geqslant c(e_t)$; the principal is willing to pay s_t only if

$$-(1-\delta)s_t + \delta\overline{\Pi} \geqslant \delta\underline{\Pi}. \tag{1}$$

For small $\epsilon > 0$, consider the following deviation by agent t. He chooses zero effort and makes the threat:

$$\mu_t(s) = \begin{cases} C & s = s_t - \epsilon \\ D & \text{otherwise.} \end{cases} \tag{2}$$

Since equation (1) holds weakly at s_t, it holds strictly for $s_t - \epsilon$ and so the principal's unique best response to this deviation is to pay $s_t - \epsilon$. Agent t's payoff from this deviation is therefore $s_t - \epsilon$, which is strictly larger than $s_t - c(e_t)$ for sufficiently small ϵ. Hence, agent t can profitably deviate from any $e_t > 0$. Every equilibrium therefore has $e_t = 0$ for all $t \geqslant 0$, in which case $\overline{\Pi} = \underline{\Pi} = 0$ and so $s_t = 0$. □

Whenever $s_t > 0$ on the equilibrium path, agent t can shirk and threaten to send a message that punishes the principal

unless she pays him slightly less than s_t. Because the principal is willing to pay s_t to avoid this punishment, she strictly prefers to pay a smaller amount. Agent t can therefore shirk and still guarantee nearly the same transfer as if he had exerted effort. This deviation is so tempting that no agent will work.[10]

Before moving on, we reflect on what Proposition 2 reveals about the economics of misuse. Any equilibrium specifies a mapping from agent t's messages to the principal's continuation payoffs. Let $\overline{\Pi}$ and $\underline{\Pi}$ be, respectively, the largest and smallest continuation payoffs in the image of this mapping. Agent t's gain from extortion depends on his leverage over the principal, defined as the normalized difference between these continuation payoffs,

$$L \equiv \frac{\delta}{1-\delta}\left(\overline{\Pi} - \underline{\Pi}\right). \tag{3}$$

In the no-extortion game, the principal pays $s_t = 0$ following any deviation and pays some $s_t \leqslant L$ on the equilibrium path. Increasing agent t's leverage therefore unambiguously increases the scope for cooperation. In the extortion game, on the other hand, agent t is paid $s_t \approx L$ regardless of his effort. He therefore exerts zero effort, because his leverage, L, is independent of his effort.

This argument suggests that agent t would have the incentive to exert effort if doing so increased his leverage and hence the pay he could demand. The next two sections explore this idea: to deter extortion, tie leverage to effort. As we will show, tying leverage to effort requires a fundamentally different approach to designing coordinated punishments.

IV. Investigations

This section considers public signals of efforts or transfers. In the no-extortion game, such signals would be irrelevant; transfer signals would be redundant with the agents' messages, while

10. Proposition 2 would continue to hold even if the principal was protected by limited liability, which would impose the constraint that $s_t \leqslant e_t$ in each $t \geqslant 0$. Proving this result requires a slightly modified argument. In particular, one can show that unless $s_t = e_t$, agent t can profitably decrease his effort and make the threat of equation (2). Thus, the principal's continuation payoff equals 0, which means that $s_t = 0$ and so $e_t = 0$ in every $t \geqslant 0$.

effort signals would be redundant with what the principal, the only player who can directly punish shirking, already observes.[11]

In contrast, these signals do have the potential to deter misuse in the extortion game. We first show that effort signals can tie an agent's leverage to his effort, which can induce effort in equilibrium. However, deterring extortion in this way requires agents to earn rent, creating a tension between the surplus created in equilibrium and the surplus captured by the principal. Then, we show that transfer signals usually cannot tie leverage to effort. Therefore, transfer signals improve cooperation only under stringent conditions.

In the context of Airbnb, our analysis suggests that Airbnb should investigate the actions of hosts, rather than just those of guests. As we will show, negative reviews by a host should optimally trigger harsher punishments when that investigation suggests that he has exerted more effort. Tying a host's leverage to his effort in this way leads to better outcomes for both hosts and guests.[12] Moreover, we should observe hosts exerting more effort in settings where doing so improves their ability to trigger coordinated punishments.

IV.A. Effort Investigations

The extortion game with effort signals is similar to the baseline extortion game, except that an effort-dependent signal, y_t, is publicly observed after s_t. We focus on a simple, binary signal structure: $y_t \in \{0, 1\}$ with $\Pr\{y_t = 1|e_t\} = \gamma(e_t)$ for $\gamma(\cdot)$ strictly increasing and twice continuously differentiable. Agent t's threat can be any mapping from his pay and this signal to a message, so that (with an abuse of notation) $\mu_t : \mathbb{R}^2 \to M$ and $m_t = \mu_t(s_t, y_t)$. Payoffs are the same as in the extortion game.

The signal y_t can deter extortion by making an agent's expected leverage an increasing function of his effort. Because signals are noisy, a shirking agent typically retains some leverage and hence can extort some pay. Agents refrain from extortion

11. Formally, the effort level in Proposition 1 is the highest attainable effort even if we drop all equilibrium constraints except for the principal's dynamic enforcement constraint and the agents' participation constraints. Those two sets of constraints would be unaffected by signals.

12. A practical caveat: this monitoring must be made immune to manipulation by the guest, since she has the incentive to fabricate evidence of shirking to ensure that the host's review is ignored.

only if they earn an equilibrium rent. We show how the tension between total surplus and the agents' rents determines effort in a principal-optimal equilibrium.[13]

PROPOSITION 3. Consider an equilibrium of the game with effort signals. If $e_t = e$ on the equilibrium path, then agent t's equilibrium payoff is at least $\bar{u}(e)$, where

$$\bar{u}(e) \equiv \max\left\{0, \frac{c'(e)}{\gamma'(e)}\gamma(e) - c(e)\right\}.$$

Suppose $\gamma(\cdot)$ is weakly concave. Then, $\bar{u}(\cdot)$ is strictly increasing, and in any $t \geqslant 0$ of any principal-optimal equilibrium, on-path effort solves

$$e_t \in \arg\max_e \left\{e - c(e) - \bar{u}(e)\right\}$$

subject to the constraint

$$\frac{c'(e)}{\gamma'(e)} \leqslant \frac{\delta}{1-\delta}(e - c(e) - \bar{u}(e)). \tag{4}$$

Proof. See Appendix A. □

To prove Proposition 3, let $\overline{\Pi}(y)$ and $\underline{\Pi}(y)$ be the largest and smallest continuation payoffs induced by some message when the signal equals y. We can define an agent's leverage, $L(y)$, analogously to equation (3). Then expected leverage, $\mathbb{E}\left[L(y)|e\right]$, depends on effort. As in Proposition 2, agent t can extort any transfer that is smaller than his expected leverage, so he chooses e_t to solve

$$e_t \in \arg\max_e \left\{\mathbb{E}\left[L(y)|e\right] - c(e)\right\}. \tag{5}$$

Because $L(\cdot) \geqslant 0$, this incentive constraint is identical to that of a static moral hazard problem with limited liability; agent t's leverage is the analogue of the contractual payment, which can depend on y. As is typical in such models, agent t earns a rent, which equals $\bar{u}(e_t)$ for this signal structure.

13. Effort signals lead to higher effort by restoring a degree of pairwise identifiability to the signaling structure (see, e.g., Fudenberg, Levine, and Maskin 1994). In particular, effort signals allow future agents to statistically distinguish between (i) the principal deviating in s_t, and (ii) an agent deviating in e_t and μ_t. Deviating in μ_t is profitable only if an agent also deviates in e_t, so effort signals are enough to restore at least some cooperation.

As in a static moral-hazard problem with limited liability, it is optimal to set $L(0) = 0$; that is, agent t's message affects the principal's continuation payoff only if $y_t = 1$. If $\gamma(\cdot)$ is concave, then we can replace equation (5) with its first-order condition, $L(1) = \frac{c'(e)}{\gamma'(e)}$. Calculating the principal-optimal equilibrium payoff therefore reduces to maximizing total surplus minus the agent's rent, subject to the constraint that $L(1)$ cannot exceed the principal's continuation payoff. Since the principal's on-path continuation payoff optimally equals her maximum equilibrium payoff, $L(1) = \frac{c'(e)}{\gamma'(e)}$ must satisfy the dynamic enforcement constraint, equation (4).

One immediate consequence of Proposition 3 is that there exists a principal-optimal equilibrium that is stationary on the equilibrium path. A second consequence is that agent t's maximum leverage is limited by the fact that future agents earn rent in equilibrium. That is, the right-hand side of equation (4) is decreasing in $\bar{u}(\cdot)$, which implies that each agent's rent-seeking behavior imposes a negative externality on the principal's relationships with other agents.

In practice, agents might have some sway over the signal distribution, as, for instance, when a union decides how to investigate grievances. Both the principal and agents prefer some kind of investigation to none, but they disagree on the optimal signal structure. In particular, agents would like the signal to maximize their rent, while the principal would like the signal to maximize total surplus net of that rent. Because an agent's rent in a principal-optimal equilibrium, $\bar{u}(\cdot)$, is equal to his rent from an optimal limited-liability contract, both his and the principal's preferences over signal structures are similar to those in a static contracting environment with limited liability.[14] For fixed effort e, $\bar{u}(e)$ is increasing in $\frac{\gamma(e)}{\gamma'(e)}$, so agents tend to prefer a signal distribution that puts weight on "false positives": $y_t = 1$ occurs frequently and with a probability that is (locally) not very responsive to effort.

IV.B. Transfer Investigations

We now turn to public signals of transfers. In contrast to Section IV.A, transfer signals are not a reliable remedy to

14. For an analysis of optimal signal structures in that static setting, see, for example, Starmans (2020). The sole difference between our analysis and the static contracting environment is the presence of a dynamic enforcement constraint, equation (4), which becomes slack as $\delta \to 1$.

extortion. The reason is that such signals reveal nothing about effort, so they usually cannot tie leverage to effort. The only exception is that certain signal distributions can be used to make the principal exactly indifferent between two different transfers when faced with an agent's optimal threat. Only under the stringent conditions that allow this indifference do equilibria with positive effort exist.

The extortion game with transfer signals is identical to the extortion game except that in each period $t \geqslant 0$, a public signal $x_t \in \mathbb{R}$ is realized after s_t and observed by everyone. Agent t's threat maps each (s_t, x_t) to a message m_t, so $\mu_t : \mathbb{R}^2 \to M$ with $\mu_t(s_t, x_t) = m_t$. We again focus on binary signals, so that $x_t \in \{0, 1\}$ with $\Pr\{x_t = 1|s_t\} = \phi(s_t)$ for some strictly increasing and twice continuously differentiable $\phi(\cdot)$.

Our main result in this section is a set of conditions on $\phi(\cdot)$ that must hold for an equilibrium with positive effort to exist. To understand these conditions, consider play in some period t. Define $\Pi(m_t, x_t)$ as the principal's continuation payoff if agent t's message is m_t and the signal is x_t. After agent t chooses his threat μ_t, the principal chooses s_t to maximize her payoff:

$$\max_s -(1-\delta)s + \delta \mathbb{E}\left[\Pi(\mu_t(s, x), x)|s\right]. \tag{6}$$

Note that expression (6) is independent of agent t's effort. Therefore, if a unique transfer maximizes expression (6), then the principal will pay that transfer regardless of agent t's effort. This leads to our first necessary condition: agent t exerts positive effort only if the principal is exactly indifferent between at least two transfers when she faces the equilibrium threat. The second necessary condition requires that no alternative threat would induce the principal to pay agent t more than his equilibrium payoff. Only under these two conditions is agent t willing to exert effort, and even then, the effort cost cannot exceed the difference between the on-path transfer and the largest amount that a shirking agent t can extort.

These two requirements imply a set of stringent necessary conditions on $\phi(\cdot)$.

PROPOSITION 4. Consider an equilibrium of the game with transfer signals. If $e_t > 0$ on the equilibrium path, there exists $s^* > 0$

and $\hat{s} \in [0, s^*)$ such that (i) $c(e_t) \leqslant s^* - \hat{s}$, (ii) $\phi''(s^*) \leqslant 0$, and (iii)

$$\phi'(s^*) = \frac{\phi(s^*) - \phi(\hat{s})}{s^* - \hat{s}}. \tag{7}$$

In particular, if $\phi(\cdot)$ is strictly concave on $\mathbb{R}_+$, then $e_t = 0$ in each $t \geqslant 0$ of every equilibrium.

Proof. See Appendix A. □

Equation (7) combines the two conditions for $e_t > 0$ described above. First, the principal must be indifferent between paying the on-path transfer, s^*, and some other amount that is no less than $\hat{s}$, when faced with the equilibrium threat. Second, no threat can induce the principal to pay a transfer near s^*. The first of these conditions pins down the average slope of $\phi(\cdot)$ between $\hat{s}$ and s^*, while the second condition says that the derivative of $\phi(\cdot)$ near s^* equals the same number. Therefore, the average slope between $\hat{s}$ and s^* must equal the tangent slope at s^*, implying equation (7). Period-t effort must then satisfy $s^* - c(e_t) \geqslant \hat{s}$, since otherwise agent t could profitably shirk and extort $\hat{s}$.

Equation (7) cannot hold if $\phi(\cdot)$ is strictly concave, in which case every equilibrium entails $e_t = s_t = 0$ in each $t \geqslant 0$, just as in the extortion game without transfer signals. Thus, positive equilibrium effort is possible only if $\phi(\cdot)$ has both convex and concave regions. For particular examples of such signal structures, we can construct equilibria with positive effort. Such equilibria require the principal to be punished on the equilibrium path. For these reasons, we view transfer investigations as unreliable, in the sense that they do not improve cooperation for a wide variety of signal distributions, and inefficient, because even when they can motivate effort, the resulting equilibrium entails occasional on-path punishments.

V. Dyadic Relationships

In the extortion game, the principal can punish an agent only by withholding pay, while an agent can punish the principal only by communicating with future agents. Although this is a reasonable model of online platforms and other settings with short-lived interactions, other relationships, including those between managers and workers, are long-lived.

In this section, we explore how ongoing dyadic interactions between the principal and each individual agent can deter extortion. As is familiar from the literature on repeated games, dyadic relationships can be used to punish an agent for shirking or the principal for reneging on a hard-working agent. We now emphasize a third effect that is new to our setting: dyadic relationships can be used to punish the principal for acquiescing to extortion, which decreases the leverage of a shirking agent. By tying leverage to effort in this way, dyadic relationships facilitate coordinated punishments.

Consider the extortion game with dyadic relationships, which makes two changes to the extortion game. The first is minor: when the principal chooses s_t, we allow agent t to simultaneously make a transfer to the principal, $s_t^A \geqslant 0$, which is observed by the principal but not by other agents. Note that allowing such transfers would not change any of our other results. The second, more substantial change is that after agent t sends his message, m_t, the principal and agent t play a symmetric, simultaneous-move coordination game. The actions of this coordination game are observed by the two participants but not by any other agents. Although our analysis can be readily extended to general, asymmetric coordination games, we focus on the following simple game:

$$\begin{array}{ccc} & h & l \\ h & (v_H, v_H) & (v_L, v_L) \\ l & (v_L, v_L) & (v_L, v_L) \end{array},$$

where $v_H > v_L$. Letting v_t be the realized payoff from this coordination game, the principal's and agent t's payoffs are $e_t - s_t + s_t^A + v_t$ and $s_t - s_t^A - c(e_t) + v_t$, respectively.

The coordination game represents, in a simple way, any future interactions between the principal and an agent. We use this simple approach to demonstrate two lessons. First, dyadic relationships can deter misuse by tying leverage to effort. Second, unless these dyadic relationships are strong, the possibility of misuse still undermines cooperation.

In Online Appendix D, we show that these two lessons also hold in a setting with truly long-lived relationships. This appendix studies a repeated game, where in each period, one of a finite number of long-lived agents is randomly chosen to play the extortion game with the principal. We prove two results in this game. On the one hand, long-lived relationships can

deter extortion by tying leverage to effort. On the other hand, unless those relationships are strong, extortion continues to undermine effort in equilibrium. This latter finding is consistent with the experience of the GM Fremont plant, where long-lived relationships were not strong enough to prevent extortionary grievances from severely curtailing productivity.

Now we show that positive effort can be sustained in the extortion game with dyadic relationships. However, effort is constrained by the strength of each dyadic relationship, as measured by $(v_H - v_L)$.

PROPOSITION 5. In the extortion game with dyadic relationships, $c(e_t) \leqslant 3(v_H - v_L)$ in every $t \geqslant 0$ of any equilibrium. If e^* is the minimum of e^{FB} and the solution to $c(e^*) = 3(v_H - v_L)$, then there exists a $\bar{\delta} < 1$ such that for any $\delta \geqslant \bar{\delta}$, $e_t = e^*$ in every $t \geqslant 0$ on the equilibrium path in any principal-optimal equilibrium.

Proof. See Appendix A. □

The constraint $c(e_t) \leqslant 3(v_H - v_L)$ reflects the fact that dyadic relationships optimally encourage cooperation via three channels: (i) they punish agents for shirking, (ii) they punish the principal for refusing to pay a hard-working agent, and (iii) they reward the principal for refusing to pay a shirking agent. The first two of these channels are familiar. The third channel is new and shows how dyadic relationships enable coordinated punishments.

Adopting the notation from Section III, an agent's on-path leverage equals

$$L^* = \frac{\delta}{1-\delta}(\overline{\Pi} - \underline{\Pi}) + (v_H - v_L),$$

reflecting the fact that a principal who reneges is punished in both the period-t coordination game and the continuation equilibrium. If agent t shirks, then his leverage decreases to

$$\hat{L} = \max\left\{\frac{\delta}{1-\delta}(\overline{\Pi} - \underline{\Pi}) - (v_H - v_L), 0\right\},$$

since play in the coordination game rewards the principal for not paying a shirking agent. An agent can extort any transfer that is strictly less than his leverage, so his on-path transfer is $L^* - \hat{L}$ larger than the maximum amount he can extort. This difference

is maximized if

$$\frac{\delta}{1-\delta}\left(\overline{\Pi}-\underline{\Pi}\right)=v_H-v_L, \tag{8}$$

in which case $L^*-\hat{L}=2(v_H-v_L)$. Combining this difference in transfers with the fact that a shirking agent faces a direct punishment of v_H-v_L, we conclude that equilibrium effort must satisfy $c(e^*)\leqslant 3(v_H-v_L)$.

Other papers that focus on extortionary incentives in dyadic relationships, including Basu (2003), Dixit (2003a), and Myerson (2004), focus on how those relationships can encourage cooperation by increasing on-path leverage, L^*, and by directly punishing a shirking agent. In contrast, we focus on a third channel: dyadic relationships can complement coordinated punishments by decreasing the leverage of a shirking agent, $\hat{L}$. Decreasing $\hat{L}$ would be irrelevant in a setting without extortion, because in that case, the value of coordinated punishments depends only on how they affect on-path leverage, L^*. In the extortion game, however, what matters is how leverage varies with effort, $L^*-\hat{L}$. Decreasing $\hat{L}$ means that each agent can be given access to coordinated punishments without misusing them. Consequently, as represented by equation (8), stronger dyadic relationships (measured by v_H-v_L) optimally expand the scope for coordinated punishments (measured by $\overline{\Pi}-\underline{\Pi}$).

Stepping back from the formal analysis, how might a firm cultivate dyadic relationships that deter extortion? The first step is to create manager—worker relationships with multiple equilibrium payoffs, so that v_H-v_L is large. The firm's formal contracts must be structured in a way that supports this multiplicity (see, e.g., Che and Yoo 2001). This was not the case at GM Fremont, where managerial incentives were based heavily on formal contracts that left little room for relational contracts (Glass and Langfitt 2015).[15] The firm's culture must select among these equilibria in a way that deters extortion. Our analysis suggests that by implementing the right formal incentives and fostering the right culture, an organization can encourage strong dyadic relationships that support effective coordinated punishments.

15. Both managers and workers were incentivized to keep the production line running at all times, so they had little incentive to cooperate on, for example, fixing production mistakes or improving quality.

VI. Extortion on the Equilibrium Path

This section enriches our baseline model so that some, but not all, agents commit to their threats. We show that in the resulting principal-optimal equilibria, cooperation and extortion coexist on the equilibrium path. Extortion spills over onto nonextorting relationships in two ways. First, the expectation of future extortion makes the principal less willing to pay transfers today. Second, to make extortion less attractive, principal-optimal equilibria entail weaker coordinated punishments that lead to less effort from nonextorting agents.

Consider the following costly extortion game. Suppose that at the start of every period $t \in \{0, 1, \dots\}$, agent t privately observes a cost $k_t \geqslant 0$, $k_t \sim G(\cdot)$, and then chooses whether to invest. If he invests, then his payoff decreases by k_t and he plays the extortion game with the principal; otherwise, he plays the no-extortion game with the principal. Only the principal observes agent t's investment decision; other agents observe only m_t.

We interpret k_t as agent t's cost of committing to his threat. An agent might incur this cost by signaling that he is willing to follow through on extortionary threats.[16] The extortion and the no-extortion games are special cases of this game where $k_t = 0$ or k_t is large, respectively. In this section, we focus on distributions over k_t such that agents invest with an interior probability.[17]

We characterize principal-optimal equilibria in the costly extortion game in terms of the leverage given to each agent.

Proposition 6. Consider the costly extortion game. In every period t of any principal-optimal equilibrium, on-path play is determined by an L_t that solves

$$L_t \in \arg\max_{L \geqslant 0} \left\{ \left(1 - G(L)\right) c^{-1}(L) - L \right\}$$

16. In the context of online platforms, an agent might incur this cost by using a nonofficial communication system to hide attempted extortion from the platforms, or it might represent the expected cost associated with the tail risk of getting caught.

17. Another special case is a model in which each agent can commit to threats with a fixed probability, λ. This corresponds to a cost distribution where $k_t = 0$ with probability λ and otherwise k_t is large.

subject to the constraint

$$L \leqslant \frac{\delta}{1-\delta}\big(\big(1-G(L)\big)c^{-1}(L)-L\big). \tag{9}$$

Agent t invests whenever $k_t < G(L_t)$. If agent t invests, then he chooses $e_t = 0$ and is paid $s_t = L_t$, whereas if he does not invest, then he chooses $e_t = c^{-1}(L_t)$ and is paid $s_t = L_t$.

Proof. See Appendix A. □

Using the notation from Section III, define

$$L \equiv \frac{\delta}{1-\delta}(\overline{\Pi} - \underline{\Pi})$$

as agent t's leverage. If agent t invests, then as in the proof of Proposition 2, his unique equilibrium strategy is to shirk and extort as much as possible, so $s_t = L$ and $e_t = 0$. If agent t does not invest, then as in the proof of Proposition 1, he is willing to choose e_t only if $c(e_t) \leqslant s_t$, while the principal is willing to pay s_t only if $s_t \leqslant L$. Agent t invests whenever the costs of doing so, k_t, are smaller than the gains, $L - (s_t - c(e_t))$.

As in Proposition 3, principal-optimal equilibria are sequentially principal-optimal. In each period of such an equilibrium, L ensures that the principal is exactly willing to compensate each agent for his effort, given that (i) an agent who invests exerts zero effort and is paid L, and (ii) L is no more than the principal's equilibrium continuation payoff. These two conditions lead to constraint (9).

Increasing an agent's leverage increases both his temptation to invest and the effort he is willing to exert if he does not. In a principal-optimal equilibrium, agent t extorts with probability $G(L)$ and otherwise exerts effort $e_t = c^{-1}(L)$. Thus, higher L has opposing effects on equilibrium cooperation: it leads to a higher prevalence of extortion and higher payments to extorting agents, but it also leads to higher effort among those agents who do not extort. The optimal L balances these forces and so is typically lower than it would be without the possibility of misuse.[18]

18. Note that this equilibrium is also an equilibrium of the no-extortion game. As in Proposition 2, the value of the commitment assumption is to be transparent about agents' incentives to misuse coordinated punishments, and about how misuse undermines cooperation.

The dynamic enforcement constraint (9) illustrates a further negative spillover from extorting to nonextorting relationships. The right side of this constraint equals the principal's on-path continuation payoff. Future nonextorting agents contribute $c^{-1}(L) - L > 0$ to this payoff, and future extorting agents contribute $-L < 0$. Thus, even in nonextorting relationships, the expectation that future agents will extort undermines cooperation.

As Proposition 6 shows, an agent who invests in extortion disproportionately benefits from severe coordinated punishments. Consequently, we might expect extorting agents to be disproportionately attracted to platforms and organizations that rely on coordinated punishments but fail to guard against misuse. Conti (2019) makes this point in the context of Airbnb, arguing that it was susceptible to this kind of negative selection because users could easily make multiple accounts. Conversely, organizations with restricted and stable memberships would be less susceptible to negative selection and misuse.

As in Sections IV and V, organizations can deter misuse in the costly extortion game by tying leverage to effort. We have seen that effort investigations and dyadic relationships can limit misuse when agents can costlessly commit to threats, which means that they can a fortiori do so in the costly extortion game. The principal might do even better by using these instruments in ways that would not be possible when extortion is costless. For instance, she might use these instruments to discourage investment in extortion without eliminating it entirely, in which case extortion would still occur on the equilibrium path.

VII. Conclusion

This article exposes a vulnerability in coordinated punishments: agents can misuse messages intended to report deviations. We also explore practical ways to restore cooperation, all of which build on the same core intuition: to deter misuse, tie an agent's leverage over the principal to his effort.

Fundamentally, misuse undermines cooperation because it severs the link from effort to transfer and message. Other modeling approaches that similarly sever this link would lead to similar takeaways as this article. Online Appendix B analyzes several such alternatives, including: (i) allowing the principal and each agent to bargain over the message (Halac 2012, 2015; Goldlücke and Kranz 2020), (ii) endowing the agents with

preferences for keeping their word (Vanberg 2008) or preferences for reciprocity (similar to those documented in Fehr, Powell, and Wilkening 2020), and (iii) imposing an equilibrium refinement in the no-extortion game (Zhu 2018, 2020). In these alternatives, just as in our main analysis, agents exert effort only if doing so increases their leverage.

All of these approaches, together with our applications, point to the conclusion that misuse is a crucial obstacle to cooperation across a variety of contexts. We do not claim that misuse arises whenever coordinated punishments are used; indeed, Online Appendix B shows that agents who have strict preferences for telling the truth (including revealing their own deviations) would not engage in misuse. Rather, our main point is that agents have a powerful incentive to misuse coordinated punishments in a way that undermines their value. Our applications illustrate, and our analysis demonstrates, that coordinated punishments are vulnerable to this type of misuse, which can severely undermine cooperation. Organizations ignore this vulnerability at their peril.

While the principal and agents are asymmetric in our model, extortionary threats are also a feature in more symmetric interactions, as in, for example, communal enforcement (e.g., Dixit 2007; Ali and Miller 2016). In such settings, both sides can potentially extort one another. How do players cooperate in the presence of two-sided extortion? What networks best facilitate cooperation, and how are rents shared in those networks? How should business associations, communities, and firms structure communication channels to support strong relational contracts? We hope that our analysis provides a foundation for analyzing such questions.

Appendix A: Proofs

Proof of Proposition 3

Consider an equilibrium. Suppose $e_t = e$ at some on-path, period-t history, and let $\overline{\Pi}(y)$ and $\underline{\Pi}(y)$ be the principal's largest and smallest continuation payoffs following signal realization y, with corresponding messages $\overline{m}(y)$ and $\underline{m}(y)$. Define $L(y) \equiv \frac{\delta}{1-\delta}(\overline{\Pi}(y) - \underline{\Pi}(y))$.

For each effort e_t, agent t can choose

$$\mu_t(s, y) = \begin{cases} \overline{m}(y) & s_t \geqslant \hat{s} \\ \underline{m}(y) & \text{otherwise.} \end{cases}$$

Whenever

$$\hat{s} < \hat{s}(e_t) \equiv L(0) + \gamma(e_t)(L(1) - L(0)),$$

the principal's unique best response to this μ_t is to pay $\hat{s}$. On the other hand $s_t = 0$ is a best response to any $\hat{s} \geqslant \hat{s}(e_t)$. Thus, agent t's equilibrium effort, e, must satisfy

$$e \in \arg\max_{e'} \left\{\hat{s}(e') - c(e')\right\}.$$

If $e > 0$, then a necessary condition for agent t to choose $e_t = e$ is that

$$c'(e) = \hat{s}'(e) = \gamma'(e)(L(1) - L(0)). \tag{10}$$

Because $\gamma'(e) > 0$, we can solve for $L(1) - L(0)$ in equation (10) and plug into the definition of $\hat{s}(e_t)$ to yield

$$\hat{s}(e) = L(0) + \gamma(e)\frac{c'(e)}{\gamma'(e)}.$$

Agent t earns at least 0, so

$$s_t - c(e) \geqslant \max\left\{0, \hat{s}(e) - c(e)\right\} \geqslant \max\left\{0, \gamma(e)\frac{c'(e)}{\gamma'(e)} - c(e)\right\} \equiv \bar{u}(e),$$

as desired.

Now suppose $\gamma(\cdot)$ is concave. Since $c'(0) = c(0) = 0$, $\bar{u}(0) = 0$, and

$$\frac{d}{de}\left\{\gamma(e)\frac{c'(e)}{\gamma'(e)} - c(e)\right\} > 0,$$

so that $\bar{u}(\cdot)$ is strictly increasing. Moreover, the first-order condition equation (10) is both necessary and sufficient for agent t to exert effort $e_t = e$.

We now characterize principal-optimal equilibrium. Let Π^* be the principal's payoff in such an equilibrium. Note that on the equilibrium path, the principal's continuation payoff equals $\overline{\Pi}(y)$ following realization y, because otherwise agent t could demand a higher transfer using the promise of $\overline{\Pi}(y)$.

Suppose that $\overline{\Pi}(y) < \Pi^*$ for some $y \in \{0, 1\}$. In that case, we can increase both $\overline{\Pi}(y)$ and $\underline{\Pi}(y)$ by the same constant to keep $L(y)$, and hence agent t's incentives, unchanged. Doing

so strictly increases the principal's payoff. So the principal's on-path continuation payoff equals Π^* in each $t \geqslant 0$ of any principal-optimal equilibrium. Then $\Pi^* = (1-\delta)(e_t - s_t) + \delta\Pi^*$, so $\Pi^* = e_t - s_t$ in any $t \geqslant 0$ on the equilibrium path.

In a principal-optimal equilibrium with $\gamma''(e) \leqslant 0$, $s_t = \mathbb{E}[L(y)|e]$, where e_t solves

$$\max_{L(\cdot)\geqslant 0, e} e - \mathbb{E}[L(y)|e]$$

subject to equation (10) and

$$L(y) \leqslant \frac{\delta}{1-\delta}\Pi^*.$$

Thus, $L(0) = 0$, in which case $L(1) = \frac{c'(e)}{\gamma'(e)}$ and so $\mathbb{E}\left[L(y)|e\right] = \gamma(e)\frac{c'(e)}{\gamma'(e)} = \bar{u}(e) + c(e)$. Substituting these simplifications into this constrained maximization problem yields the constrained maximization problem in the statement of the Proposition. □

Proof of Proposition 4.

Fix a period t. Let $\Pi(m, x)$ be the principal's continuation payoff following message m and signal x. Let $\overline{\Pi}(x) = \max_m \Pi(m, x)$ and $\underline{\Pi}(x) = \min_m \Pi(m, x)$ with $\overline{m}(x)$ and $\underline{m}(x)$ being the corresponding maximizer and minimizer. We let π^D be the smallest payoff that the principal can guarantee herself,

$$\pi^D = \max_s -(1-\delta)s + \delta\mathbb{E}\left[\underline{\Pi}(x)|s\right]. \tag{11}$$

Define s_A as the smallest maximizer of equation (11). We argue that agent t's payoff is at least s_A. He can always choose $e_t = 0$ and

$$\mu_t(s, x) = \begin{cases} \overline{m}(x), & \text{if } s = s_A \\ \underline{m}(x), & \text{if } s \neq s_A. \end{cases}$$

Faced with this threat, the principal earns π^D from paying s_A and strictly less than π^D from paying $s < s_A$. Therefore, the principal will pay at least s_A.

Consider the set of transfers that can give the principal a higher payoff than π^D:

$$\{s : -(1-\delta)s + \delta\mathbb{E}\left[\overline{\Pi}(x)|s\right] > \pi^D\}. \tag{12}$$

If this set is nonempty, we let s_B be the supremum of this set. We argue that agent t can get a payoff arbitrarily close to s_B. In particular, he can choose $e_t = 0$ and

$$\mu_t(s, x) = \begin{cases} \overline{m}(x), & \text{if } s = s_B - \epsilon \\ \underline{m}(x), & \text{if } s \neq s_B - \epsilon. \end{cases}$$

Since $\phi(\cdot)$ is continuous, the principal's unique best response is to pay $s_t = s_B - \epsilon$ for small enough $\epsilon > 0$.

Now define $\hat{s} = \max\{s_A, s_B\}$ if the set (12) is nonempty, and $\hat{s} = s_A$ otherwise. Agent t can guarantee a payoff arbitrarily close to $\hat{s}$ if he shirks, so he chooses $e_t = e^*$ only if $s^* - c(e^*) \geqslant \hat{s}$, which is our first necessary condition. Moreover, we can show that

$$-(1-\delta)s^* + \delta \mathbb{E}\left[\overline{\Pi}(x)|s^*\right] = s^D \tag{13}$$

$$-(1-\delta)\hat{s} + \delta \mathbb{E}\left[\overline{\Pi}(x)|\hat{s}\right] = s^D. \tag{14}$$

To see why equation (13) holds, note that the principal is willing to pay s^* so the left side of equation (13) must be weakly higher than π^D. But either s_B does not exist, in which case equation (13) must hold with equality, or the supremum of set (12) must be strictly below s^*, so that again equation (13) holds with equality. Equality (14) follows from the continuity of $\phi(\cdot)$ and the definition of $\hat{s}$.

Combining equations (13) and (14), we have

$$s^* - \hat{s} = \frac{\delta}{1-\delta}\left(\phi(s^*) - \phi(\hat{s})\right)\left(\overline{\Pi}(1) - \overline{\Pi}(0)\right). \tag{15}$$

Given equations (13), $-(1-\delta)s + \delta \mathbb{E}\left[\overline{\Pi}(x)|s\right]$ must attain a local maximum at $s = s^*$, since otherwise set (12) would contain elements arbitrarily close to s^* and so $s^* \leqslant \hat{s}$. Thus,

$$\phi'(s^*)\left(\overline{\Pi}(1) - \overline{\Pi}(0)\right) = \frac{1-\delta}{\delta} \tag{16}$$

and $\phi''(s) \leqslant 0$. Combining equations (15) and (16) yields our final necessary condition:

$$\phi'(s^*) = \frac{\phi(s^*) - \phi(\hat{s})}{s^* - \hat{s}}.$$

If $\phi(\cdot)$ is strictly concave, it cannot satisfy this condition for $s^* > \hat{s}$. □

Proof of Proposition 5.

Consider period t of an equilibrium. Define $\overline{\Pi}$ and $\underline{\Pi}$ as the principal's largest and smallest continuation payoffs, respectively, with corresponding messages $\overline{m}$ and $\underline{m}$. Agent t can always deviate to $e_t = s_t^A = 0$ and

$$\mu_t(s) = \begin{cases} \overline{m} & s = \hat{s} \\ \underline{m} & \text{otherwise.} \end{cases}$$

Following this deviation, the principal's unique best response is $s_t = \hat{s}$ if

$$\hat{s} < v_L - v_H + \frac{\delta}{1-\delta}(\overline{\Pi} - \underline{\Pi}). \tag{17}$$

Similarly, if agent t does not deviate, the principal is willing to pay $s_t = s^*$ only if

$$s^* \leqslant v_H - v_L + \frac{\delta}{1-\delta}(\overline{\Pi} - \underline{\Pi}). \tag{18}$$

Agent t is willing to choose $e_t = e^*$ only if $s^* - c(e^*) + (v_H - v_L) \geqslant \hat{s}$ for *any* $\hat{s}$ satisfying expression (17). Given the bound (18) on s^*, we conclude that $e_t = e^*$ in equilibrium only if $3(v_H - v_L) \geqslant c(e^*)$.

Each agent must earn at least v_L, so the principal's equilibrium payoff cannot exceed $e^* - c(e^*) + 2v_H - v_L$, where $e^* = e^{FB}$ if $c(e^{FB}) \leqslant 3(v_H - v_L)$ and e^* satisfies $c(e^*) = 3(v_H - v_L)$ otherwise. To complete the proof, we construct an equilibrium that attains this bound. Play starts in the cooperative phase: in each $t \geqslant 0$, agent t chooses $e_t = e^*$ and

$$\mu_t(s) = \begin{cases} C & s = c(e^*) \\ D & \text{otherwise.} \end{cases}$$

Transfers equal $s_t = max\{0,\ c(e^*) - (v_H - v_L)\}$, $s_t^A = max\left\{0, (v_H - v_L) - c(e^*)\right\}$ if agent t does not deviate and $s_t = 0$, $s_t^A = (v_H - v_L)$ if he does. If either nobody deviates or agent t deviates from (e_t, μ_t) but then nobody deviates from (s_t, s_t^A), then $v_t = v_H$; otherwise, $v_t = v_L$. Play continues in the cooperative phase until $m_t = D$, at which point it transitions to

the punishment phase with probability α. In the punishment phase, $e_t = s_t = 0$ in each period. Let α satisfy

$$\max\left\{0, c(e^*) - 2(v_H - v_L)\right\} = \frac{\delta}{1-\delta}\alpha\left(e^* - c(e^*) + 2v_H - v_L\right).$$

For $\delta < 1$ sufficiently close to 1, $\alpha \in [0, 1]$.

The principal earns $e^* - c(e^*) + v_H + (v_H - v_L)$ surplus in each period of the cooperative phase. If agent t deviates in (e_t, μ_t), then he earns v_L by paying $s_t^A = v_H - v_L$ and $-s_t^A + v_L$ from deviating, so he has no profitable deviation from s_t^A. Regardless of μ_t, the principal has no profitable deviation from $s_t = 0$ following a deviation in (e_t, μ_t) if

$$\begin{aligned} v_H - v_L &\geqslant \frac{\delta}{1-\delta}\alpha(e^* - c(e^*) + 2v_H - v_L) \\ &= \max\left\{0, c(e^*) - 2(v_H - v_L)\right\}, \end{aligned}$$

which holds because $c(e^*) \leqslant 3(v_H - v_L)$. On the equilibrium path, if $c(e^*) - (v_H - v_L) \geqslant 0$, then the principal has no profitable deviation from s_t because

$$\begin{aligned} -c(e^*) + (v_H - v_L) + v_H + \frac{\delta}{1-\delta}(e^* - c(e^*) + 2v_H - v_L) &\geqslant v_L \\ + \frac{\delta}{1-\delta}(1-\alpha)(e^* - c(e^*) + 2v_H - v_L). & \end{aligned}$$

This is because, by definition of α,

$$\frac{\delta}{1-\delta}\alpha(e^* - c(e^*) + 2v_H - v_L) \geqslant c(e^*) - 2(v_H - v_L).$$

If $c(e^*) - (v_H - v_L) < 0$, then agent t has no profitable deviation from s_t^A because $c(e^*) - (v_H - v_L) + v_H \geqslant v_L$.

Given these transfers, agent t earns v_L from choosing the equilibrium (e_t, μ_t) and no more than v_L from deviating. So this strategy profile is an equilibrium. It is principal-optimal because it attains the upper bound on the principal's equilibrium payoff. □

Proof of Proposition 6.

Consider an equilibrium and a history at the start of period t. Define $\overline{\Pi}$ and $\underline{\Pi}$ as in the proof of Proposition 2, with

corresponding messages $\overline{m}$ and $\underline{m}$, and let

$$L_t \equiv \frac{\delta}{1-\delta}(\overline{\Pi} - \underline{\Pi}).$$

Suppose agent t invests. If $e_t > 0$ or $s_t < L_t$, then the deviation from the proof of Proposition 2 is profitable for $\epsilon > 0$ sufficiently small. Consequently, $e_t = 0$ and $s_t = L_t$ whenever agent t invests.

Suppose agent t does not invest. He must earn at least a payoff of 0 in equilibrium, so $s_t - c(e_t) \geqslant 0$. The principal must be willing to pay s_t, so $s_t \leqslant L_t$. For any e_t and s_t that satisfy these two constraints, consider the following strategy profile:

i. Agent t chooses e_t.
ii. The principal pays s_t if agent t has not deviated and pays nothing otherwise.
iii. Agent t sends $\overline{m}$ if no deviation has occurred and $\underline{m}$ otherwise.

If agent t chooses e_t, the principal is willing to pay s_t because $s_t \leqslant L_t$. If agent t deviates, then $m_t = \underline{m}$ regardless of the principal's action, so she pays nothing. Agent t is willing to choose e_t because $s_t \geqslant c(e_t)$. Thus, neither player has a profitable deviation from this strategy profile. We conclude that any (e_t, s_t) with $c(e_t) \leqslant s_t \leqslant L_t$ can be implemented in an equilibrium, as desired.

Because an investing agent exerts no effort and obtains a pay of L_t, from now on we use e_t, s_t for the effort exerted by, and the pay received by, agent t who didn't invest. Given this continuation play, agent t is willing to invest if and only if

$$L_t - (s_t - c(e_t)) \geqslant k_t,$$

where $L_t - (s_t - c(e_t))$ and k_t represent the gain from and cost of investment, respectively.

Now consider a principal-optimal equilibrium, and let Π^* equal the principal's maximum equilibrium payoff. We must have $\overline{\Pi} = \Pi^*$ in each period t, since agent t's incentive depends only on L_t so we can increase $\overline{\Pi}$, $\underline{\Pi}$ while keeping L_t fixed. The principal's payoff is

$$\begin{aligned} &\max_{L_t, s_t, e_t} G(L_t - (s_t - c(e_t)))\left\{\delta\Pi^* - (1-\delta)L_t\right\} \\ &\quad + (1 - G(L_t - (s_t - c(e_t))))\left\{(1-\delta)(e_t - s_t) + \delta\Pi^*\right\} \end{aligned} \tag{19}$$

subject to the constraint that $c(e_t) \leqslant s_t \leqslant L_t$. The constraint $s_t \leqslant L_t$ must bind, since the principal would like L_t to be as small as possible. Substituting $L_t = s_t$ into expression (19), the objective in expression (19) becomes:

$$G(c(e_t))\left\{\delta\Pi^* - (1-\delta)s_t\right\} + (1 - G(c(e_t)))\left\{(1-\delta)(e_t - s_t) + \delta\Pi^*\right\}.$$

The derivative of this objective with respect to s_t is $-1 + \delta$. Hence, it is optimal to choose $s_t = c(e_t)$. The objective in expression (19) becomes

$$\delta\Pi^* + (1-\delta)\left((1 - G(c(e_t)))e_t - c(e_t)\right).$$

Therefore, the optimal effort maximizes $(1 - G(c(e_t)))e_t - c(e_t)$ and Π^* is given by this maximum:

$$\Pi^* = \max_{e_t} (1 - G(c(e_t)))e_t - c(e_t).$$

The formula for Π^* is quite clear. The principal has to pay $c(e_t)$ to both an extorting agent and a nonextorting one. However, she only obtains e_t from the nonexorting agent, which occurs with probability $1 - G(c(e_t))$. □

NORTHWESTERN UNIVERSITY
NORTHWESTERN UNIVERSITY

SUPPLEMENTARY DATA

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

REFERENCES

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond, "Preferences for Truth-Telling," *Econometrica*, 87 (2019), 1115–1153.
Ali, S. Nageeb, and Ce Liu, "Conventions and Coalitions in Repeated Games," Penn State University Working Paper, 2018.
Ali, S. Nageeb, and David Miller, "Enforcing Cooperation in Networked Societies," Penn State University Working Paper, 2013.
———, "Ostracism and Forgiveness," *American Economic Review*, 106 (2016), 2329–2348.
Ali, S. Nageeb, David Miller, and David Yang, "Renegotiation-Proof Multilateral Enforcement," (2017), Penn State University Working Paper.
Andrews, Isaiah, and Daniel Barron, "The Allocation of Future Business: Dynamic Relational Contracts with Multiple Agents," *American Economic Review*, 106 (2016), 2742–2759.
Arnold, Chris, and Robert Smith, "Bad Form, Wells Fargo," *NPR*, 2016.

Baker, George, Robert Gibbons, and Kevin Murphy, "Subjective Performance Measures in Optimal Incentive Contracts," *The Quarterly Journal of Economics*, 109 (1994), 1125–1156.
———, "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics*, 117 (2002), 39–84.
Barron, Daniel, Jin Li, and Michal Zator, "Morale and Debt Dynamics" Northwestern University Working Paper, 2019.
Barron, Daniel, and Michael Powell, "Policies in Relational Contracts," *American Economic Journal: Microeconomics*, 11 (2019), 228–249.
Basu, Kaushik, *Analytical Development Economics: The Less Developed Economy Revisited* (Cambridge, MA: MIT Press, 2003).
Bernstein, Lisa, "Beyond Relational Contracts: Social Capital and Network Governance in Procurement Contracts," *Journal of Legal Analysis*, 7 (2015), 561–621.
Board, Simon, "Relational Contracts and the Value of Loyalty," *American Economic Review*, 101 (2011), 3349–3367.
Bowen T., Renee, David M. Kreps, and Andrzej Skrzypacz, "Rules with Discretion and Local Information," *The Quarterly Journal of Economics*, 128 (2013), 1273–1320.
Bull, Clive, "Existence of Self-Enforcing Implicit Contracts," *The Quarterly Journal of Economics*, 102 (1987), 147–159.
Chassang, Sylvain, and Gerard Padro i Miquel, "Crime, Intimidation, and Whistleblowing: A Theory of Inference from Unverifiable Reports," *Review of Economic Studies*, 86 (2019), 2530–2553.
Che, Yeon-Koo, and Seung-Weon Yoo, "Optimal Incentives for Teams," *American Economic Review*, 91 (2001), 525–541.
Conti, Allie, "I Accidentally Uncovered a Nationwide Scam on AirBnB," *Vice*, October 31, 2019. https://www.vice.com/en/article/43k7z3/nationwide-fake-host-scam-on-airbnb.
Dewatripont, Mathias, "The Role of Indifference in Sequential Models of Spatial Competition: An Example," *Economics Letters*, 23 (1987), 323–328.
Dixit, Avinash, "On Modes of Economic Governance," *Econometrica*, 71 (2003a), 449–481.
———, "Trade Expansion and Contract Enforcement," *Journal of Political Economy*, 111 (2003b), 1293–1317.
———, *Lawlessness and Economics: Alternative Modes of Governance*. (Princeton, NJ: Princeton University Press, 2007).
Fehr, Ernst, Michael Powell, and Tom Wilkening, "Behavioral Constraints on the Design of Subgame-Perfect Implementation Mechanisms," 2020, Zurich University Working Paper.
Fong, Yuk-Fai, and Jin Li, "Relational Contracts, Limited Liability, and Employment Dynamics," *Journal of Economic Theory*, 169 (2017), 270–293.
Freeman, Richard, and James Medoff, "The Two Faces of Unionism," *Public Interest*, 57 (1979), 69–93.
Fudenberg, Drew, David Levine, and Eric Maskin, "The Folk Theorem with Imperfect Public Monitoring," *Econometrica*, 62 (1994), 997–1039.
Gambetta, Diego, *The Sicilian Mafia: The Business of Private Protection* (Cambridge, MA: Harvard University Press, 1993).
Glass, Ira, and Frank Langfitt, "NUMMI 2015," This American Life, NPR, July 17, 2015.
Goldlücke, Susanne, and Sebastian Kranz, "Reconciliating Relational Contracting and Hold-up: A Model of Repeated Negotiations," University of Konstanz Working Paper, 2020.
Greif, Avner, Paul Milgrom, and Barry Weingast, "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild," *Journal of Political Economy*, 102 (1994), 745–776.
Guo, Yingni, and Johannes Hörner, "Dynamic Allocation without Money," Northwestern University Working Paper, 2018.

Halac, Marina, "Relational Contracts and the Value of Relationships," *American Economic Review*, 102 (2012), 750–779.
———, "Investing in a Relationship," *RAND Journal of Economics*, 46 (2015), 165–186.
Hörner, Johannes, and Nicolas Lambert, "Motivational Ratings," *Cowles Foundation Discussion Paper No. 2035*, 2018.
Klein, Tobias, Christian Lambertz, and Konrad Stahl, "Market Transparency, Adverse Selection, and Moral Hazard," *Journal of Political Economy*, 124 (2016), 1677–1713.
Levin, Jonathan, "Multilateral Contracting and the Employment Relationship," *Quarterly Journal of Economics*, 117 (2002), 1075–1103.
———, "Relational Incentive Contracts," *American Economic Review*, 93 (2003), 835–857.
Li, Jin, Niko Matouschek, and Michael Powell, "Power Dynamics in Organizations," *American Economic Journal: Microeconomics*, 9 (2017), 217–241.
Lipnowski, Elliot, and Joäo Ramos, "Repeated Delegation," *Journal of Economic Theory*, 188 (2020), 105040.
Lippert, Steffen, and Giancarlo Spagnolo, "Networks of Relations and Word-of-Mouth Communication," *Games and Economic Behavior*, 72 (2011), 202–217.
Liu, Ce, "Stability in Repeated Matching Markets," Michigan State University Working Paper, 2019.
MacLeod, Bentley, and James Malcomson, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57 (1989), 447–480.
Malcomson, James, "Relational Incentive Contracts," In *Handbook of Organizational Economics*, Robert Gibbons and John Roberts, eds. (Princeton, NJ: Princeton University Press, 2012), 1014–1065.
———, "Relational Contracts with Private Information," *Econometrica*, 84 (2016), 317–346.
Milgrom, Paul, Douglass North, and Barry Weingast, "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics*, 2 (1990), 1–23.
Miller, David, Trond Olsen, and Joel Watson, "Relational Contracting, Negotiation, and External Enforcement," *American Economic Review*, 110 (2020), 2153–2197.
Miller, David, and Joel Watson, "A Theory of Disagreement in Repeated Games with Bargaining," *Econometrica*, 81 (2013), 2303–2350.
Myerson, Roger B., "Justice, Institutions, and Multiple Equilibria," *Chicago Journal of International Law*, 5 (2004), 91–108.
Ortner, Juan, and Sylvain Chassang, "Making Corruption Harder: Asymmetric Information, Collusion, and Crime," *Journal of Political Economy*, 126 (2018), 2108–2133.
Ostrom, Elinor, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press, 1990).
Peachey, Kevin, "Online Reviews' Used as Blackmail'," *BBC News*, June 19, 2015. https://www.bbc.com/news/business-33184207
Pei, Harry, and Bruno Strulovici, "Crime Aggregation, Deterrence, and Witness Credibility," Northwestern University Working Paper, 2020.
Proctor, Jason, "Disgruntled Bride Ordered to Pay 115K after Defamatory Posts Ruin Chinese Wedding-Photo Business," CBC, March 1, 2018. https://www.cbc.ca/news/canada/british-columbia/chinese-wedding-weibo-defamation-1.4556433.
Starmans, Jan, "Technological Determinants of Financial Constraints," Stockholm School of Economics Working Paper, 2020.
Strulovici, Bruno, "Can Society Function without Ethical Agents? An Informational Perspective," Northwestern University Working Paper, 2020.
Tranaes, Torben, "Tie-Breaking in Games of Perfect Information," *Games and Economic Behavior*, 22 (1998), 148–161.

Vanberg, Christoph, "Why Do People Keep Their Promises? An Experimental Test of Two Explanations," *Econometrica*, 76 (2008), 1467–1480.
Watson, Joel, "A General, Practicable Definition of Perfect Bayesian Equilibrium," UC San Diego Working Paper, 2017.
Wolitzky, Alexander, "Career Concerns and Performance Reporting in Optimal Incentive Contracts," *B.E. Journal of Theoretical Economics (Contributions)*, 12 (2012), 1–32.
———, "Cooperation with Network Monitoring," *Review of Economic Studies*, 80 (2013), 395–427.
Zhu, John Y., "A Foundation for Efficiency Wage Contracts," *American Economic Journal: Microeconomics*, 10 (2018), 248–288.
———, "Better Monitoring... Worse Productivity?," University of Kansas Working Paper, 2020.