Reporting Sexual Misconduct in the #MeToo Era*

Ing-Haw Cheng† Alice Hsiaw‡

June 2020

Abstract

What deters individuals from reporting sexual misconduct, and what are the effects of #MeToo for reporting? We show that individuals under-report sexual misconduct if and only if a manager's misconduct is widespread. The reason is that individuals face strategic uncertainty over whether others will also report misconduct and corroborate a pattern of behavior. We apply our model to study a manager's decision to mentor subordinates, the coordinating effect of raising public awareness of misconduct, and the policy effects of confidential holding tanks for reports and rewards for whistleblowers. Overall, our study highlights several unintended and intended consequences of #MeToo.

Keywords: Misconduct, Reporting, Retaliation, Coordination

*The authors thank Bob Gibbons, Robin Greenwood, Oliver Hart, Joni Hersch, Louis Kaplow, David Laibson, Stephen Morris, Matthew Rabin, David Scharfstein, Kathryn Spier, Jeremy Stein, Wei Xiong, Muhamet Yildiz, Kwok Yu, Nina Zipser and seminar participants at Brandeis University, College of the Holy Cross, Harvard University, and MIT for helpful comments and discussions.

†Dartmouth College, Tuck School of Business, 100 Tuck Mall, Hanover, NH 03755 USA. Phone: (603) 646-6492. Email: ing-haw.cheng@tuck.dartmouth.edu.

‡Brandeis University, International Business School, 415 South Street, Waltham, MA 02453 USA. Phone: (781) 736-2251. Email: ahsiaw@brandeis.edu.

The #MeToo movement has brought the question of whether individuals under-report sexual misconduct to the fore of public attention. Major organizations, including Google and the American Economics Association, are under growing pressure from stakeholders to re-evaluate procedures for handling and encouraging the reporting of misconduct (Griffin et al., 2018; The Economist, 2018). Reporting is the impetus for any investigation or action, yet evidence suggests that individuals are reluctant to report (U.S. Equal Employment Opportunity Commission, 2016; Hersch, 2010; Cortina and Berdahl, 2008). What deters individuals from reporting misconduct, and how might #MeToo and broader policy changes affect reporting? A lack of a clear theoretical framework for why individuals might under-report has made it difficult for economists to provide compelling answers.

The first part of this paper shows that economic agents under-report sexual misconduct if and only if a manager's misconduct is widespread. The reason for under-reporting is that agents face a coordination problem and strategic uncertainty over whether others will also report misconduct. The *coordination problem* occurs because reports from multiple individuals corroborate a pattern of behavior. Corroboration provides "safety in numbers": It increases the chances that relevant outside parties will act on an agent's report and reduces the chances that a reporting agent will face costly retaliation, stigma, or reprisal. As we discuss, the coordination problem is pervasive in the reporting of sexual misconduct.

Models of coordination problems that place little structure on the information environment may be "unable to reach sufficiently sharp predictions" about outcomes (Angeletos and Lian, 2016, p.1111). In our context, if agents are certain that all other agents will report a manager's misconduct, they will report misconduct themselves. This situation precipitates an "all-report" equilibrium where all agents report a manager's misconduct. However, there is also a "no-report" equilibrium where no agents report misconduct because agents are certain that no other agents will report. Small perturbations can select either equilibrium outcome as the unique rationalizable outcome (Weinstein and Yildiz, 2007b).

We show that introducing an important and realistic information friction in the context of reporting sexual misconduct leads to under-reporting. In our model, agents have imperfect knowledge of whether other agents experience misconduct. Agents receive heterogeneous private experiences from a manager of unknown type. The manager's type determines the distribution of experiences across agents. An agent's experience need not be bad, and an agent and manager may interpret the same interaction differently. An agent has an intrinsic

motive to report an experience of misconduct to an outside party (e.g., human resources, law enforcement, or other authorities).[1] However, if the outside party does not sanction (e.g., terminate or discipline) the manager based on the report, the agent incurs a retaliatory cost. The coordination problem exists because the outside party is more likely to sanction the manager if other agents also report and corroborate a pattern of behavior.

The information friction generates *strategic uncertainty* for an agent about whether other agents will report misconduct, and we derive agents' endogenous beliefs about how many other agents will report in the unique equilibrium of the baseline model. In the equilibrium, agents employ "threshold strategies" and report if and only if their experience is worse than a unique reporting threshold. Our technique draws from work on global games and coordination problems in finance and macroeconomics (for early work, see Carlsson and van Damme, 1993a,b and Morris and Shin, 1998; for reviews, see Morris and Shin, 2003 and Angeletos and Lian, 2016).

Due to strategic uncertainty, the coordination problem leads to under-reporting if and only if misconduct is widespread. We define under-reporting as occurring whenever a lower reporting threshold Pareto-improves agents' payoffs over the equilibrium threshold, given the true distribution of agent experiences. We define misconduct as widespread when the average agent experience is severe misconduct or when most agents experience misconduct. In a stark example, a manager's misconduct can be an "open secret": all agents can have nearly identical bad experiences, know that all other agents have bad experiences, *and* know that the best outcome would be for all agents to report, but almost no agents report in equilibrium. The reason is that, despite knowing that almost all experiences are identically bad, agents remain uninformed over how many other agents will report. As a result, all agents would be better off playing a lower threshold. More broadly, we characterize when under-reporting occurs and establish the Pareto-optimal threshold for agents.

The key applied insight is that agents may not report misconduct even when misconduct is so widespread that reporting is in their broader interests as a group. This insight is a robust consequence of the model's core feature: The information friction in the coordination problem makes agents less willing to report due to strategic uncertainty and the possibility

[1] As a matter of convention, we denote an agent as "she" and the manager as "he." Most reports of workplace harassment originate from women, and women form the bulk of victims of sexual assault (U.S. Department of Justice, 2002; U.S. Equal Employment Opportunity Commission, 2019). However, our model does not assume any genders, and we do not preclude other misconduct.

of retaliation. We show this insight holds when we relax simplifying assumptions over the distribution of experiences, agent priors, and the actions of the outside party, even when multiple threshold strategy equilibria may occur.

The second part of the paper applies our framework to study the implications of the #MeToo movement and policies intended to encourage reporting. Despite under-reporting, the consequences of such policies are not yet well understood, in practice and theory.

First, we ask: Does #MeToo reduce opportunities for mentorship from senior managers, and does any reduction affect reporting of misconduct? Anecdotal and survey evidence suggests that men in senior positions have become more reluctant to mentor or meet alone with women in junior positions due to the possibility of subsequent accusations of improper behavior (Miller, 2017; Smith, 2018; Atwater et al., 2018). Concerns about reluctance to mentor have reached the highest levels of business leadership (Bower, 2019).

We show that managers' reluctance to mentor is theoretically grounded and that this reluctance has strategic spillovers onto agents' reporting strategies. We extend the model to allow the manager an ex-ante choice over whether to mentor. Agents then form reporting strategies and beliefs over strategic uncertainty based on the set of manager types who mentor. If types with a high propensity for misconduct opt out of mentoring, this has a trade-off. Agents become more reluctant to report any misconduct that does occur because corroboration is less likely when most other agents are not experiencing misconduct. Conversely, if low-misconduct types opt out, agents become more likely to report any misconduct that occurs, but mentorship from low-misconduct types is now lower. Either case can occur, and which case prevails depends on whether managerial utility from mentoring decreases or increases with the propensity for misconduct.

Second, we ask: Can publicizing broader awareness of misconduct encourage reporting? Another popular narrative is that #MeToo helped "give people a sense of the magnitude of the problem" (Khomami, 2017) and precipitated a cascade of credible allegations against prominent figures (Griffin et al., 2018). Our model provides a theoretical basis for how heightened public awareness of misconduct increases reporting of otherwise-hidden misconduct by coordinating beliefs over strategic uncertainty. If public beliefs about the experiences of women become worse than previously thought, any given agent will believe that other agents are more likely to have bad experiences. The public nature of this shift in beliefs is crucial because the agent also knows that other agents share this shift. The agent

is then more willing to report, even keeping her own experience constant, resulting in a lower equilibrium reporting threshold and more reporting.

Next, we study the effects of two proposed policies intended to encourage reporting: a holding tank for confidential reports and rewards for whistleblowers. A holding tank holds reports of misconduct confidentially unless the number of reports exceeds a certain bar, in which case the tank opens, and reports are released to the outside party. We show that having a holding tank does not unequivocally increase reporting and may even discourage it due to strategic uncertainty over whether enough other agents will file reports to open the tank. Whether the tank discourages or encourages reporting then depends crucially on agents' utility gain from filing a report that may never be released. If such gains are low, having a tank may discourage reporting by making it more difficult for reports to be released to the outside party even though the tank keeps reports confidential. Our analysis raises a surprising conundrum for holding tanks: a holding tank can encourage reporting only if agents gain sufficient utility from filing a report that is kept in the tank and never released to the outside party.

Rewards for whistleblowers mitigate the coordination problem induced by retaliation, which imposes a cost on reports that do not result in a sanction. We show that such rewards can be calibrated to a benchmark level where agents report as if this coordination problem did not exist; however, rewards greater than this calibrated level incentivize agents to report a broader range of behaviors than in this benchmark. We do not restrict the interpretation of such rewards: they could be monetary, a sense of vindication, or public recognition.

Both the coordination problem and information friction that we highlight are first-order in reality for the reporting of sexual misconduct. Section 1 highlights the evidence and role of both in high-profile cases, including allegations of misconduct at Ford, Nike, Uber, and CBS, as well as in the criminal cases against Harvey Weinstein and Bill Cosby. The coordination problem is pervasive: Outside parties are more likely to act when multiple reports corroborate a pattern of behavior for several reasons. First, corroboration combats skepticism of reported misconduct in "he-said/she-said" situations involving ambiguous behaviors. Second, it establishes a strong record that an outside party may need to produce before taking an action such as a reprimand or termination of employment. Third, corroboration makes it more costly for organizations to "look the other way" when claims can be dismissed as isolated incidents. The information friction is also pervasive because sexual misconduct often

occurs in private, individuals often do not publicly share their experiences, and individuals may not know about other allegations of misconduct due to non-disclosure agreements.

Our contribution is to introduce a realistic and important information friction in the context of reporting sexual misconduct, to show that the resulting strategic uncertainty leads to under-reporting, and to study the implications of strategic uncertainty for #MeToo and related policies. Our focus on coordination and strategic uncertainty complements other approaches in the literature. Basu (2003), Chen and Sethi (2018), and Hersch (2018) consider important ramifications of misconduct distinct from issues related to coordination. Chassang and Miquel (2019) study how to optimally elicit reports from a single whistleblower and also do not consider coordination. Pei and Strulovici (2019) consider a setting where principals with an incentive to commit crime strategically choose how many crimes to commit. Our model differs by taking the perspective that differences in the propensity to commit sexual misconduct among managers in the real world largely reflect differences in fixed manager types. Lee and Suen (2019) focus on the credibility of early versus late reporting when some accusers might be lying. Daughety and Reinganum (2011) also study a timing problem where agents have an incentive to file a lawsuit if they corroborate previous lawsuits. While timing frictions can generate coordination problems, neither of these two papers investigates how agents endogenously form beliefs about other agents' actions.

We conclude by discussing implications for future empirical and theoretical research on #MeToo, workplace misconduct, and broader contexts where strategic uncertainty may chill speech, reporting, or otherwise create a "culture of silence" where agents do not speak up or express concerns.

1 An Agent's Reporting Decision

Agents in our model receive private experiences from a manager of an unobserved type. Specifically, a continuum of risk-neutral agents with mass 1 indexed by $i \in (0,1)$ work for a manager at a firm. The manager has a type $\theta \in \mathbb{R}$, a fixed characteristic that generates heterogeneous *private experiences* $\{x_i \in \mathbb{R}\}$, for which θ is the average agent experience: $x_i = \theta + \sigma\epsilon_i$ for $\epsilon_i \sim N(0,1)$, where ϵ_i is i.i.d. across agents. We assume the manager exogenously generates experiences, which focuses attention on agents' reporting decisions taking their experiences as given.

Each experience x_i reflects agent i's interpretation of an interaction with the manager. We do not model the manager's experience, but the manager may interpret the same interaction differently, and agent i may truthfully disagree. For example, a large body of evidence suggests there are gender differences in the perception of sexual harassment (McDonald, 2012), with women perceiving a wider range of behaviors as harassing (Rotundo et al., 2001); Bénabou, Falk and Tirole (2019) propose a model of why different groups, such as men and women, may interpret the same behavior according to different moral standards.

The key friction in our model is that agents' experiences are private information. Experiences are heterogeneous ($\sigma > 0$), and a larger σ reflects a wider range of heterogeneity. Agents do not know the average experience θ but learn about it through their own x_i. To start, we assume agents have improper uniform priors over θ. This assumption is a simplification but reflects a plausible belief that, ex ante, an agent is completely uninformed and "never knows" how a person behaves in private.

The information friction is realistic and of first-order importance in the context of sexual misconduct for several reasons. First, the nature of sexual misconduct means that it occurs in private. Second, individuals often do not publicly share their experiences for several reasons, including shame, fear of reprisal and blame, and fear of not being believed (Hotelling, 1991; Fitzgerald et al., 1995; for popular accounts, see Willingham and Maxouris, 2018). Finally, individuals may not know about other allegations of misconduct due to non-disclosure agreements (Lobel, 2018; Prasad, 2018).

Several assumptions above and introduced below are simplifications that provide tractability and transparency to the workings of the model, but are also plausible and can often be relaxed. The assumption that agent experiences are exogenously generated reflects the realistic observation embedded in the advice to professors to "always keep the office door open": Agents' interpretations of interactions with a manager are not entirely under that manager's control, with some possibility that a manager and agent interpret an interaction in extremely different ways. As Section 1.1 notes, the assumption of normality of x_i and thus unbounded experience support also reflects this observation. However, Section 1.7 shows that unbounded support, as well as the improper prior assumption, can be relaxed, and Section 2 allows the manager a choice to invoke the "Pence rule" and avoid interacting with agents at all.[2]

[2]U.S. Vice President Michael Pence famously does not dine with women alone arguably because of the possibility that the interaction is taken the "wrong way," among other reasons.

1.1 Misconduct and the coordination problem in reporting

Before turning to further details, we first provide real-world context for several model features by discussing misconduct and the coordination problem in reporting misconduct.

Misconduct. Sexual misconduct generally refers to unwelcome and unreasonable sex-related conduct and includes a broad category of behavior not limited to legal definitions of harassment and assault (Hersch, 2010; Cortina and Berdahl, 2008). For example, organizations often implement policies that prohibit a broad category of sexually-related behavior in the workplace (e.g., Dartmouth College, 2019; Brandeis University, 2019).

Sexual misconduct may also violate civil or criminal law. In the United States, sexual harassment as defined by federal civil law includes acts such as unwelcome sexual advances and verbal or physical conduct that creates a hostile work environment. Sexual assault, a category that includes rape, is a crime that the U.S. Department of Justice defines as "any nonconsensual sexual act proscribed by Federal, tribal, or State law, including when the victim lacks capacity to consent" (U.S. Department of Justice, 2019).

As a matter of terminology, we say an agent experiences *sexual misconduct* if his or her experience x_i weakly exceeds a normalized value of zero, with higher values of x_i corresponding with progressively worse acts in terms of agent utility. Negative values of x_i correspond to good experiences. Very high-θ manager types frequently generate experiences of misconduct. As a modeling device, one can interpret managers who never commit misconduct in the real world within the model as $\theta \ll 0$ manager types who generate $x_i > 0$ with arbitrarily small probability.

However, due to unbounded support, normality of x_i implies that, for every manager type, some fraction of agents in the model interpret interactions as misconduct. As noted above and in Section 1.7, the normality assumption yields tractability, particularly for later applications, and can be relaxed to distributions with bounded support. However, the assumption does reflect the aforementioned realistic observation that professors should "always keep the office door open" due to the possibility that two parties interpret an interaction extremely differently. The experience x_i reflects the agent's interpretation of the interaction, and it is plausible that, even for $\theta \ll 0$ types, an arbitrarily small number of agents interpret interactions as misconduct. This is particularly true with ambiguous issues involving harassment and consent, as different views about whether consent was given drastically change the interpretation of sexual behavior.

The coordination problem in reporting misconduct. Individuals face a coordination problem when deciding whether to report misconduct: Agents would like to report if other agents are reporting. The reason is that more reports of misconduct corroborate a pattern of behavior and make it more likely for an outside party or authority to act on those reports. If an outside party fails to act on an agent's report, that agent is more likely to face additional costs such as retaliation.

Examples of the importance of corroboration in individuals' decisions to report are abundant. Victims of Harvey Weinstein repeatedly weighed the possibility that their claims would not be taken seriously in the absence of corroboration (Farrow, 2017). Bill Cosby's initial trial included the testimony from only one other accuser and resulted in a mistrial due to jury deadlock. He was convicted upon retrial, when five additional accusers provided similar accounts of sexual assault. Commentary from legal experts indicates that the corroboration from other accusers likely significantly bolstered the primary accuser's case (Bowley, 2018), and press accounts suggest these concerns are not limited to high-profile cases (Weiser et al., 2012). As The Economist (2020) writes, "Numbers Matter."

Corroboration establishes a pattern of behavior that increases the likelihood an outside party acts on reports of misconduct through several potential channels. First, a strong pattern of behavior combats skepticism over whether any alleged behavior that might have occurred constitutes misconduct. Such skepticism can occur because accusations of misconduct often come down to "he-said/she-said" situations around ambiguous behaviors (Anderson, 2004; Tracy et al., 2012). Women often have different views over what constitutes harassment and consent than men have, and disagreement over whether alleged behavior constitutes misconduct can happen even when all parties are reporting what they believe to be the truth (Jozkowski et al., 2014; Rotundo et al., 2001). Accusers also face suspicion over whether they are fabricating outright lies, although the incidence of deceitful fabrication appears to be relatively low in practice and lower than common conception.[3]

Second, a strong pattern of behavior establishes a substantial record that an outside party may need to produce in order to sanction a manager. Such a record may be important irrespective of what the outside party believes because of due process, procedural requirements,

[3]As one example from a criminal context, 6% of rape cases were deemed "unfounded" in the 2008 FBI Uniform Crime Report (UCR), and this statistic likely exceeds the true fabrication rate as it includes cases that authorities deem truthful yet do not meet the criteria of assault (Lisak et al., 2010; Lonsway, 2010; for evidence in Europe, see Kelly, Lovett and Regan, 2005; Lovett and Kelly, 2009).

or the need for the outside party to convince still others that sanction is justified.[4] In the U.S., employers often provide "just cause" (evidence of a policy violation) for sanction even though at-will employment is the legal default (Rudy, 2002); outside the U.S., just-cause requirements are more common (Porter, 2008). In a criminal context, prosecutors often find it advantageous to find ways of introducing evidence of a pattern of behavior to convince a jury to convict (Tracy et al., 2012; Feldman, 2020 notes this tactic in the Weinstein case).

Third, a strong pattern of behavior established through multiple reports of misconduct raises the costs of not dealing with those reports, especially in organizational cultures that tend to "look the other way." Management may easily ignore or dismiss claims of harassment from a few employees as isolated incidents, particularly if the accused manager is powerful or a key person in the firm (Cooper, 2017; Gino, 2018). However, a large group of allegations may make it more costly for an organization to not act on the claims, even for such a person. For example, according to press accounts, CBS was resistant to ousting CEO Les Moonves in response to an initial set of reports of sexual misconduct; however, after complaints increased, the board ousted him (Koblin, 2018; Stewart, 2018; see also allegations at Ford, Nike, and Uber: Chira and Einhorn, 2017; Creswell et al., 2018; Isaac, 2017).

If victims' claims are dismissed because of insufficient corroboration or other reasons, they often face additional costs, such as social stigma or retaliation from managers and co-workers (Cortina and Magley, 2003; McDonald, 2012; for popular accounts, see Engel, 2018; Farrow, 2017; Rikleen, 2018; Sheiber and Creswell, 2017). Although federal law prohibits workplace retaliation, allegations of retaliation are among the most common filed to the EEOC (U.S. Equal Employment Opportunity Commission, 2016) and continue to rise in the wake of #MeToo (Weber, 2018). In criminal misconduct cases, fear of reprisal in particular is the most-cited concern for why victims do not report (U.S. Department of Justice, 2013).

Overall, corroboration provides "safety in numbers" by increasing the chances that relevant parties act on a report of sexual misconduct and by mitigating the chances of retaliation.

[4] A famous line from the movie *A Few Good Men* summarizes this channel well when fictional lawyer Dan Kaffee (played by Tom Cruise) exclaims in exasperation: "It doesn't matter what I believe, it only matters what I can prove!"

1.2 Agent strategies and payoffs

In the model, an agent can report her experience x_i to an outside party, who can potentially sanction the manager. To keep the model flexible, we do not take a strong stand on the identity of the outside party, other than that they have an incentive to sanction managers for misconduct. The outside party could be an internal committee charged with evaluating claims of harassment, human resources, law enforcement, or even the court of public opinion.

Motivated by our discussion in Section 1.1, we focus on the central reduced form feature we wish to capture: outside parties are more likely to act if more agents report and provide corroboration of a pattern of behavior. Formally, a strategy for an agent is $s(x_i) = 1$ if an agent reports her experience and reveals her x_i to the outside party, and $s(x_i) = 0$ if an agent does not report. If a mass of r agents report misconduct by playing $s(x_i) = 1$ for $x_i \geq 0$, the outside party sanctions the manager with probability $\Gamma(r)$, where $\Gamma(r)$ increases with r.

The function $\Gamma(r)$ reflects the reduced form mapping for how reports of misconduct translate into action by the outside party. The reduced form flexibly reflects the different channels that underpin the need for corroboration discussed in Section 1.1. The outside party may be more likely to act for higher r because a higher r combats skepticism, establishes a stronger record to sanction misconduct, or is more costly for an organization to ignore. We allow for (but do not require) sanction to be probabilistic conditional on r and thus for residual uncertainty about whether an outside party will act.

The key assumption regarding $\Gamma(r)$ is that it increases in r. While real-world sanction functions depend on reported x_i's, the key economic feature that derives from these functions is that the probability of sanction still increases in the number of agents r who report. Indeed, we show in Section 1.7 that a general sanction function that weakly increases in reported x_i's can be re-formulated in equilibrium as a function of both which x_i's agents choose to report and the number of agents that report. Since such a function is still weakly increasing in r, our main insights are unchanged. Section 1.7 also discusses the outside party's posterior beliefs upon observing reports and elaborates on interpretations of $\Gamma(r)$.

In our baseline formulation, we assume for tractability that $\Gamma(r) = \gamma r$, where $\gamma \in [0, 1]$ is the sensitivity of sanction to reporting. In Section 1.7, we show our main insights are unchanged when considering other Γ functions that increase in r, including non-linear or even discontinuous step functions.

We assume that agents truthfully reveal their x_i to the outside party if they report. Our

Table 1: Agent's Payoffs. Sanction occurs with probability $\Gamma(r) = \gamma r$, where r is the number of agents reporting $x_i \geq 0$. The parameter ω is the intrinsic motivation to report experience x_i, while β is the retaliatory disincentive, and $\omega > \beta$. The parameter c is a fixed cost of reporting. $\mathbf{1}_{[x_i \geq 0]}$ is an indicator function that equals one if $x_i \geq 0$.

	Sanction	No Sanction
Report	$x_i\omega - c$	$x_i(\omega - \mathbf{1}_{[x_i \geq 0]}\beta) - c$
No Report	0	0

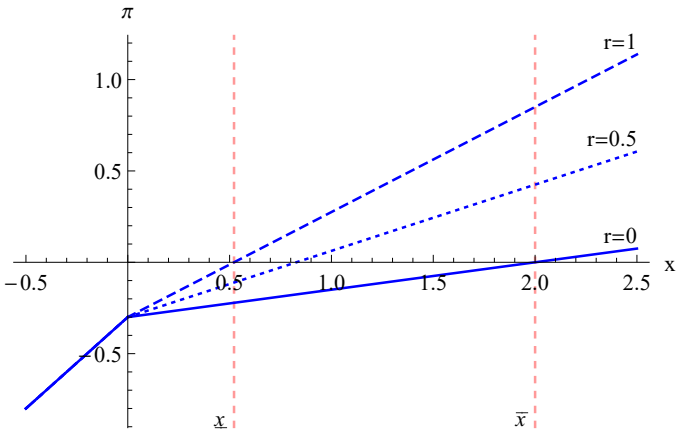
model thus takes the perspective of an agent who faces a coordination problem in reporting what she believes to be the truth, as faced by many victims in the real world. Importantly, we allow for the manager to view a reported experience quite differently and for the manager and agent to believe that the other is instead lying or exaggerating, consistent with our discussion in Section 1.1. For simplicity, the model does not include deceitful fabrication, or for agents to know x_i but strategically represent a different x to the outside party. The incidence of outright fabrication appears to be low (as Section 1.1 also notes), and thus would not meaningfully unravel the coordination problem.

We specify payoffs as follows. Agents have an *intrinsic motivation* to report misconduct that provides payoff ωx_i for reporting experience x_i, and reporting incurs a fixed cost c. In the event the outside party does not sanction the manager based on the report, agents incur a *retaliatory cost* equal to βx_i if $x_i \geq 0$. On net, the agent receives $\omega x_i - c$ from reporting and incurs additional cost βx_i if the outside party does not sanction the manager. Because the outside party is more likely to sanction when more agents report, an agent is less likely to incur the retaliatory cost when others also come forward. An agent who does not report receives a normalized payoff of zero, as the direct effect of the experience on her utility is sunk. Table 1 summarizes.

Our payoff specification appeals to the following intuitions. The intrinsic motivation $\omega > 0$ can be driven by moral obligation, a sense of justice, or civic duty.[5] The retaliatory cost $\beta > 0$ can be driven by outright retaliation by the manager, public humiliation, or other costs (Section 1.1). We assume that $\omega > \beta$: even if an agent knew an outside party would not sanction the manager, her incentive to report increases as her experience becomes worse (x_i increases). We also assume that the an agent incurs a retaliatory cost only for reports

[5]We assume agents incur ωx_i even if $x_i < 0$; one could alternatively assume $\omega = 0$ for $x < 0$ and results would be identical. We also assume that agents receive intrinsic utility from reporting instead of intrinsic disutility from non-reporting; the two formulations are identical.

Figure 1: Payoff gain function $\pi(r,x)$. This figure plots $\pi(r,x)$ for $r \in \left\{0, \frac{1}{2}, 1\right\}$ and $\gamma < 1$. The slope for $x < 0$ equals ω. The slope for $x > 0$ equals $\omega - (1-\gamma r)\beta$. The y-intercept equals $-c$.



of misconduct ($x_i > 0$). We assume that both benefit ω and cost β are proportional to x_i: Agents are intrinsically more motivated to report progressively worse experiences, but the retaliation is larger when agents report worse experiences that do not result in sanction.

A key quantity we consider is the *payoff gain* $\pi(r, x_i)$ from reporting, defined as the expected utility from reporting experience x_i given r agents reporting, minus the expected utility from not reporting: $\pi(r, x_i) = Eu(Report, r, x_i) - Eu(No\ Report, r, x_i)$, where the expectation is taken over the possibility of sanction given r. Because sanction occurs with probability $\Gamma(r) = \gamma r$, based on Table 1 we have:

$$\pi(r,x) = \begin{cases} x(\omega - (1-\gamma r)\beta) - c & \text{if } x \geq 0, \\ x\omega - c & \text{if } x < 0. \end{cases}$$

Figure 1 plots $\pi(r, x)$. Several observations are noteworthy.

First, note that the payoff gain is *monotone*: It strictly increases in x and weakly increases in r (strictly increases in r for $x > 0$). Worse experiences strengthen the incentive to report, and more reporting by other agents strengthens the incentive to report bad experiences.

Second, there are *dominance regions*. If $x_i < \underline{x} \equiv \frac{c}{\omega - \beta(1-\gamma)}$, it is strictly dominant for agent i to not report because $\pi(r, x_i) < 0$ for any r, even $r = 1$. Intuitively, the agent's experience is sufficiently good that there is no reason to report even she knew all other agents were reporting their experiences. Likewise, if $x_i > \overline{x} \equiv \frac{c}{\omega - \beta}$, it is strictly dominant for agent i to report because $\pi(r, x_i) > 0$ for any r, even $r = 0$. Intuitively, the agent's

experience is sufficiently bad that she wants to report even if she knew no other agents were reporting their experiences. Clearly, $0 < \underline{x} < \overline{x}$, although $\overline{x}$ can be arbitrarily large.

Finally, there is a *coordination problem*. For any $x_i \in (\underline{x}, \overline{x})$, the payoff gain can be either positive or negative, depending on the number of reporting agents r, due to the existence of the retaliatory cost $\beta > 0$. If no agents report, then $\pi(0, x) < 0$; if all agents report, then $\pi(1, x) > 0$. A higher retaliatory cost β widens the region of experiences $(\underline{x}, \overline{x})$ where there is a coordination problem, and also worsens the experiences by increasing both $\underline{x}$ and $\overline{x}$. The coordination problem distinguishes this game from games such as the prisoner's dilemma that feature only dominant strategies, and motivates the analysis in the rest of the paper.

1.3 Equilibria in the absence of information frictions

Without frictions, multiple equilibria occur because an agent's beliefs about other agents' actions are *self-fulfilling*. To illustrate, consider the following thought exercise. Suppose that the manager engages in misconduct so widespread that every agent experiences identical misconduct: $\sigma = 0$ so that $x_i = \theta \in (\underline{x}, \overline{x})$ $\forall i$ and all agents learn θ and every other agents' experience for certain. The best outcome for agents is for all agents to report because the outside party is most likely to act. These facts are common knowledge among all agents.

If an agent expects all other agents to report ($r = 1$), it is optimal to report because the payoff gain of reporting is positive: $\pi(1, x) > 0$ for $x \in (\underline{x}, \overline{x})$, as Figure 1 illustrates. If all agents share this belief, then all agents report, vindicating agents' initial beliefs in an "all-report" equilibrium. Conversely, if an agent expects no other agents to report ($r = 0$), then it is optimal to not report because $\pi(0, x) < 0$, precipitating a self-fulfilling "no-report" equilibrium.[6] The all-report equilibrium payoff-dominates the no-report equilibrium for any $\theta > \underline{x}$ because $\pi(1, x) > \pi(0, x)$ for any such θ.

Weinstein and Yildiz (2007b) show that there exist perturbations of multiple-equilibrium games that may yield any equilibrium outcome as the unique rationalizable outcome. We study the outcome of introducing an information friction that is realistic and important in the context of reporting sexual misconduct.

[6]There is also a third mixed strategy equilibrium where all agents report with probability $p = \frac{c-(\omega-\beta)x_i}{\gamma x_i \beta}$, where p solves $\pi(p, x) = 0$. This equilibrium suffers from counter-intuitive comparative statics that often occur in mixed strategy equilibria more broadly. For example, agents in this mixed strategy equilibrium are more likely to report when the cost of reporting increases or when retaliation increases.

1.4 Equilibrium with information frictions

Our model introduces the information friction where agents receive heterogeneous ($\sigma > 0$) experiences x_i that are private information. Agents are uncertain about θ but learn about it through their own experience. This approach provides structure for how expectations of other agents' actions are determined in equilibrium, which ultimately determines agents' actions. We start by stating the unique equilibrium in the baseline game where agents have improper uniform priors over θ.

Proposition 1 (Baseline Equilibrium)**.** *Suppose agents have improper uniform priors over θ.*

1. *Existence: There exists a unique symmetric threshold strategy equilibrium where all agents play a threshold strategy and report if and only if $x_i \geq x^*$:*

$$s(x_i) = \begin{cases} 1 & \text{if } x_i \geq x^* \\ 0 & \text{if } x_i < x^*, \end{cases}$$

where:

$$x^* = \frac{c}{\omega - \beta\left(1 - \frac{\gamma}{2}\right)}, \tag{1}$$

and $x^* > \underline{x} > 0$.

2. *Uniqueness: This threshold strategy is the unique strategy that survives the iterated deletion of strictly dominated strategies. In particular, the threshold equilibrium is the globally unique equilibrium, where uniqueness is defined up to either reporting or not-reporting when $x_i = x^*$.*

3. *Beliefs over r: In this equilibrium, the belief of the marginal agent who draws $x_i = x^*$ over the number of agents reporting r is uniformly distributed over $[0, 1]$.*

The result is a standard application of "global games" techniques, due to Carlsson and van Damme (1993a,b), and advanced by Morris and Shin (1998), Heinemann (2000), Morris and Shin (2000), Morris and Shin (2002), Frankel, Morris and Pauzner (2003), Morris and Shin (2004a,b), Goldstein and Pauzner (2005), Angeletos and Werning (2006), Angeletos, Hellwig and Pavan (2006, 2007), Goldstein and Huang (2018), and several others in finance and macroeconomics (see Morris and Shin, 2003 and Angeletos and Lian, 2016 for detailed reviews). We sketch a heuristic derivation here and provide the key details in Appendix A and Online Appendix B.1. Both heuristics and proofs follow Morris and Shin (2003).

The role of *strategic uncertainty* and beliefs about how many other agents will report play a central role in the intuition. Each agent must ask herself: Given my x, what is the probability that a proportion less than r of other agents have experiences worse than mine? If agents are playing threshold strategies around x, then the proportion of agents receiving experiences worse than x is given by $1-\Phi\left((x-\theta)/\sigma\right)$ given the true θ, where $\Phi(.)$ is the normal cumulative distribution function. Re-arranging, the proportion is less than r if $\theta < x - \sigma\Phi^{-1}(1-r)$. If the agent has no idea what θ equals ex-ante, her posterior belief over θ after experiencing x is normally distributed with mean x and standard deviation σ. Then, the probability the proportion falls below r equals $\Phi\left(\left(x-\sigma\Phi^{-1}(1-r)-x\right)/\sigma\right) = r$.

Thus, in any equilibrium, the marginal agent whose experience is just at the threshold $x_i = x^*$ has a Uniform$[0,1]$ belief over r (Part 3 of Proposition 1). Intuitively, at the threshold where she is indifferent between reporting and not reporting, the marginal agent is effectively uninformed about the actions of others. Given this belief, one can solve the unique threshold x^* by taking expectations of Equation 1 over r (Part 1). Given the monotonicity and the dominance regions of $\pi(r,x)$, one can further show the threshold equilibrium is dominance-solvable and thus globally unique (Part 2).

1.5 Under-reporting

We say there is *under-reporting* in equilibrium if a lower reporting threshold $\tilde{x} < x^*$ would Pareto-improve agent payoffs, given the true distribution of experiences among agents as determined by θ. Below, we show that under-reporting occurs if and only if misconduct is widespread because of strategic uncertainty stemming from the information friction in the model. We first illustrate this principle with a special case in Corollary 1.1 before turning to the general result in Proposition 2, where we define "widespread" precisely.

Corollary 1.1 revisits whether the payoff-dominant all-report or dominated no-report equilibrium emerges when agents have (nearly) identical experiences. The outcome contrasts starkly with the analysis without information frictions in Section 1.3.

Corollary 1.1 (Open-secret equilibrium)**.** *Let $\theta \in (\underline{x}, x^*)$. In the unique equilibrium, $\lim_{\sigma\to 0} r = 0$. However, in the limit as $\sigma \to 0$, $r = 1$ is the Pareto-optimal outcome and the Pareto-optimal threshold for reporting equals $\underline{x} < x^*$.*

Corollary 1.1 describes what we call an *open secret equilibrium*, an extreme form of under-reporting. For intuition, consider the case where σ is positive but vanishingly small and let

$\theta \in (\underline{x}, x^*)$ so that all agents have nearly identically bad experiences. All agents know that all other agents have nearly identical experiences (agents know σ is small), know that all other agents share nearly identical beliefs about θ (the manager's type is an "open secret"), *and* know that all-reporting is the best outcome for all agents ($\pi(1, x) > 0 > \pi(0, x)$ for almost all agents). Yet virtually no agents report because almost all $x_i < x^*$, leading to $r \approx 0$ as the unique outcome. In the limit as $\sigma \to 0$, we have $r \to 0$, but every agent would be weakly better off, with some agents strictly better off, if they had played a threshold strategy around a lower threshold of $\underline{x} < x^*$, as this would achieve $r \to 1$.

The open-secret equilibrium persists because strategic uncertainty persists whenever $\sigma > 0$, however small. If $\sigma = 0$ as in Section 1.3, every agent learns θ for certain: the manager's type is not an open secret, it is simply "out in the open." In that case, all agents are also certain of other agent's actions. In contrast, in the presence of information frictions, Part 3 of Proposition 1 says that the marginal agent is uninformed about how many other agents are reporting, even when all experiences are nearly identical, and agents know that all experiences are nearly identical. The key issue is that the experiences themselves are not common knowledge, creating "reverberant doubt" (Morris and Shin, 2002) and uncertainty over other agents' actions, even though there is little uncertainty over θ and agents anticipate all other agents' strategy functions $s(\cdot)$ in equilibrium.

More broadly, Proposition 2 shows that under-reporting occurs if and only if misconduct is *widespread*, as defined by either agents experiencing severe misconduct on average (condition 1) or most agents experiencing misconduct (condition 2). The Proposition characterizes when under-reporting occurs and the existence and uniqueness of the Pareto-optimal threshold $\tilde{x} < x^*$ that generates the maximal Pareto improvement. Corollary 1.1 is an application of condition (2).

Proposition 2 (Under-reporting with widespread misconduct)**.** *Given θ, there exists an $\tilde{x} < x^*$ such that playing a threshold strategy around $\tilde{x}$ generates a Pareto improvement in agent payoffs if and only if misconduct is* widespread *in that one of the following conditions are satisfied:*

1. *$\theta > x^*$, or*
2. *$\theta \in (\underline{x}, x^*]$ and $\sigma \leq \underline{\sigma}(\theta)$, where $\underline{\sigma}(\theta)$ is given in the proof of this Proposition.*

Furthermore, if such an $\tilde{x}$ exists, there exists a unique Pareto-optimal threshold $\tilde{x} < x^*$ (characterized in the proof) that generates the maximum improvement, with $\frac{\partial \tilde{x}}{\partial \sigma} > 0$ and $\lim_{\sigma \to 0} \tilde{x} = \underline{x}$.

With condition (1), misconduct is widespread in that agents experience severe misconduct on average ($\theta > x^*$). Intuitively, severe average misconduct will result in r realizing higher than $1/2$, the marginal agent's expectation of r. Some agents with x_i below x^* would then be better off by reporting despite their lower x_i because the level of corroboration r is high. The additional reporting by these agents would also make agents with $x_i > x^*$ better off through a higher r. A lower reporting threshold of $\tilde{x} < x^*$ thus results in a Pareto improvement.

With condition (2), misconduct is widespread in that most agents experience misconduct. Experiences are not as severe as in condition (1) on average but are still severe enough that a lower reporting threshold generates a Pareto improvement when most agents receive experiences close to the average. Intuitively, when the average experience θ is in $(\underline{x}, x^*]$, the number of agents reporting r will realize (weakly) less than $1/2$, the marginal agent's expectation of r. Thus, the argument for a Pareto improvement is distinct from that in condition (1). A lower threshold can only make agents with $x_i < x^*$ better off if it significantly increases aggregate reporting r. A small σ makes this large increase in r, and thus a Pareto improvement, possible. Indeed, whenever a Pareto improvement exists, a smaller σ leads to a smaller value of the threshold $\tilde{x}$ achieving the maximum Pareto improvement.

When misconduct is not widespread, no Pareto improvement is possible. If $\theta \in (\underline{x}, x^*]$ and σ is large, or $\theta < \underline{x}$, a lower threshold makes some agents with $x_i < x^*$ worse off because it makes them report experiences they previously did not want to report but does not significantly increase r. This observation provides the "only if" in Proposition 2.

The Pareto-optimal threshold of $\tilde{x}$ equals the threshold used in an equilibrium of a benchmark game where agents know θ ex-ante but $\sigma > 0$. Therefore, whenever under-reporting occurs, the model with information frictions features a Pareto-dominated threshold relative to the model without frictions.

Proposition 2 is a general statement that applies whenever agents play threshold strategies, not just to the baseline model. Section 1.7 considers variations of the model that lead to multiplicity, and Proposition 2 applies to every threshold strategy equilibrium whenever such multiplicity occurs.

Overall, Proposition 2 shows that under-reporting occurs precisely when misconduct is widespread, as the "open secret" equilibrium in Corollary 1.1 starkly illustrates. In both cases considered by the Proposition, agents fail to achieve the Pareto-higher payoffs in equilibrium due to the information friction whereby private experiences generate strategic uncertainty.

Table 2: Comparative statics.

Parameter	(1) Reporting Threshold x^*	(2) Aggregate Reporting $\hat{r}(\theta)$
ω: Intrinsic motive	-	+
c: Fixed reporting cost	+	-
γ: Sanction sensitivity	-	+
β: Retaliation cost	+	-
σ: Experience dispersion	0	+ if $\theta < x^*$ − if $\theta > x^*$ 0 if $\theta = x^*$

Information frictions thus provide a rationale for why the under-reporting of misconduct can persist even when large payoff gains for agents are possible from more reporting.

1.6 Comparative statics

We next consider the implications of encouraging reporting through the lens of the #MeToo movement. The first column of Table 2 reports comparative statics for the threshold x^*.

The #MeToo movement actively encourages people to "believe women." In the model, this most closely correponds with increasing γ, the outside party's sensitivity of sanction to reporting. One can also interpret increasing γ as an organization decreasing their tolerance for misconduct. Raising γ lowers the equilibrium reporting threshold through two effects. First, an agent's direct payoff from reporting increases because the agent is less likely to face retaliation. Second, she anticipates that others are more likely to report in equilibrium. Both of these effects increase her willingness to report, decreasing x^*.

Lowering β also lowers x^*, but in practice retaliation has proven difficult for policymakers to reduce. Although federal law prohibits workplace retaliation, there is ample evidence that various forms of retaliation are prevalent (Cortina and Magley, 2003; McDonald, 2012; U.S. Equal Employment Opportunity Commission, 2016). Increasing ω or lowering c also lower x^*. However, changing these values may prove difficult since these parameters more closely reflect agents' preferences rather than the outside party (γ) or the manager and firm (β).

To see how effects on x^* translate into the amount of equilibrium reporting, the second column of Table 2 considers the effect of changing parameters on *aggregate reporting* $\hat{r}(\theta) \equiv \int_0^\infty s(x) f(x \mid \theta)\, dx = \Phi\left(\frac{\theta - x^*}{\sigma}\right)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal density and cumulative

distribution functions. Aggregate reporting $\hat{r}$ equals the total number of agents that report experiences for manager type θ in equilibrium.

To continue our discussion of the sanction sensitivity γ, note that increasing γ increases the aggregate number of agents reporting $\hat{r}$ for every type θ, precisely because it lowers the reporting threshold x^*. Policymakers who encourage reporting should then not be surprised if the marginal reported behavior becomes relatively less severe (x^* falls) and the reporting rate for all manager types rises ($\hat{r}(\theta)$ increases). These effects increase payoffs among agents within the model, although it is important to note that we have not modeled the manager's utility. In Section 2, we consider a setting where managers can avoid sanction by avoiding agents, but first we re-visit several simplifying assumptions.

1.7 Robustness and further discussion

1.7.1 How important is the assumption of unbounded support?

Online Appendix B.2 shows that the Proposition 1 admits continuous densities with (possibly asymmetric) bounded support. The assumption of normality and unbounded support simplifies not only the proof of the Proposition itself, but makes the applications we consider in later sections tractable. Proposition 2 changes only in accounting for the bounds of ϵ_i in its two conditions.

1.7.2 How important are improper uniform priors?

A setting with improper uniform priors reflects a setting where, ex ante, agents have no idea about how the manager behaves in private. In Section 3, we consider a setting where agents begin with proper (or "informed") priors about θ to provide additional implications around the publicity effect of #MeToo. The section shows that the main insights about under-reporting remain unchanged with proper priors.

1.7.3 The outside party and $\Gamma(r)$

As noted earlier, $\Gamma(r)$ reflects the reduced form objective function of an outside party with an incentive to discipline misconduct, and the property that it is increasing in r captures the core real-world feature that the outside party is more likely to act if more agents report.

What are the beliefs of the outside party? Despite the reduced form, the model pins down beliefs in that the outside party learns the average agent experience θ with certainty ex-post. This is because the ex-post aggregate reporting $\hat{r}(\theta) = \Phi\left(\frac{\theta - x^*}{\sigma}\right)$ is a sufficient statistic for θ. Because the outside party uses its objective function to resolve the dispute in a manner that is more favorable to agents if r is higher, in equilibrium it is more likely to sanction when it learns that the average agent experience θ is higher.

Even though the outside party learns the average agent experience θ ex-post, under-reporting matters for outcomes because different levels of corroboration influence the likelihood that the outside party acts $\Gamma(r)$ for the realistic reasons discussed in Section 1.1. For example, consider a high-θ type identified by Proposition 2. If agents played a lower reporting threshold $\tilde{x} < x^*$, r would be higher, and the outside party would be more likely to act. Note that the outside party learns θ irrespective of whether agents play $\tilde{x}$ or x^* as long as it correctly anticipates agents' reporting thresholds. Realistically, however, even if the outside party knew that a manager was a high-θ type, a higher r may increase the likelihood that the outside party acts because a higher r establishes a stronger record of behavior or is more costly for an organization to ignore. Such considerations are likely of even greater importance if a manager perceives interactions differently than agents and counters agents' reports, which the model allows without loss of generality.

What about other $\Gamma(r)$ functions? Our analysis generalizes to broader sanction functions $\Gamma(r)$, so long as $\Gamma(r)$ is an increasing function. Because the marginal agent has a uniform belief over r, one would replace $\gamma/2$ in Equation 1 with $\int_0^1 \Gamma(r)\, dr$, and all implications follow. Examples include power functions such as $\Gamma(r) = r^n$ for $n \geq 0$, step functions such as $\Gamma(r) = 0$ for $r < \bar{r}$ and 1 for $r \geq \bar{r}$ for $\bar{r} \in (0, 1)$, or continuous non-differentiable functions such as $\Gamma(r) = \gamma r$ for $r < 1/\gamma$ and 1 for $r > 1/\gamma$ for $\gamma > 1$.

What if reported x_i's directly affect the probability of sanction? On the one hand, $\Gamma(r)$ incorporates information about reported x_i through the outside party's inference about θ because r is a sufficient statistic for θ. On the other hand, the outside party may care disproportionately about some values of reported x_i beyond their inference of θ.

Online Appendix B.3 writes down a general sanction function incorporating this possibility and shows that under-reporting occurs—that is, an analogous form of Proposition 2 holds—within any of the possibly-multiple threshold strategy equilibria. We sketch the outline here. Suppose the outside party, upon observing $n(x)$ number of each reported x

(and learning θ, as before), aggregates the reported x_i according to $\int_x \varphi(x)n(x)dx$ where $\varphi(x) \geq 0$ is a weakly increasing penalty function. The outside party then proceeds to sanction with probability $\Xi(\{n(x), x\}) \equiv B\left(\int_x \varphi(x)n(x)dx\right)$, where $B(.)$ is a weakly increasing sanction function (possibly discontinuous) mapping into $[0, 1]$. This formulation generalizes our previous sanction function: if $\varphi(x) \equiv 1$, then $\Xi(\{n(x), x\}) = B(r) = \Gamma(r)$. For more general $\varphi(x)$, we show that there is under-reporting in every threshold strategy equilibrium.

The key insight that we use is that one can always reformulate $\Xi(\{n(x), x\})$ in terms of an equivalent sanction function $\Gamma(x^*, r)$ that is increasing in r whenever agents play threshold strategies. This analysis illustrates that the feature of the sanction function driving our results is that the likelihood of sanction is increasing in r. As discussed in Section 1.1, this feature is both realistic and plausible.

1.7.4 What about over-reporting?

Analogous to our definition of under-reporting, we say there is over-reporting in equilibrium if a higher reporting threshold $\tilde{x} > x^*$ would Pareto-improve agent payoffs, given θ. This never occurs in the model because a higher reporting threshold always leads to less reporting, lower probability of sanction, and less expected utility for agents who report with $x_i > \bar{x}$.

1.7.5 What about other definitions of over/under-reporting?

Consider an alternative definition of over-/under-reporting that focuses on agents' willingness to report each x_i. Because we define misconduct as $x_i > 0$, a natural benchmark is for agents to report all $x_i > 0$ and not report $x_i \leq 0$. Because agents report if and only if $x_i \geq x^*$, agents then under-report any $x_i \in (0, x^*)$.

From the manager's perspective, over-reporting may occur for an experience he interprets as $x_i' < 0$ but that the agent interprets as $x_i > x^*$ and reports. To the extent that #MeToo may have led to a lower x^*, this insight provides a basis for understanding disagreement between managers and agents about the impact of #MeToo.

Separately, Corollary 1.1 supports the notion that equilibrium outcomes are constrained-inefficient for agents (Angeletos and Pavan, 2007; Angeletos and Lian, 2016). The reason is that, for $\sigma \to 0$, a planner without knowledge of θ could improve total agent payoffs without transferring information across agents by directing all agents to play threshold strategies around $\underline{x}$. Our definition of under-reporting is related but distinct: it conditions on θ

and thus the ex-post distribution of experiences, and requires that more reporting create a Pareto-improvement in agent payoffs, not just an improvement in the sum of payoffs.

Finally, we emphasize that our definition of under-reporting stems from considering only payoffs related to the intrinsic motive ω, retaliation β, and the fixed reporting cost c. Implicitly, we ask whether the information friction leads agents to play a Pareto-dominated threshold when weighing the possibility of retaliation against agents' own intrinsic motives, and Proposition 2 says the answer is yes if misconduct is widespread and no otherwise. However, the answer may also be no if one considers other outside payoffs. For example, high-θ-type managers might provide benefits to agents with $x_i < 0$ who do not experience misconduct through improved career opportunities. Our central insight would hold in that a lower reporting threshold would then lead to a Pareto improvement among agents who experienced misconduct ($x_i > 0$), but social welfare would depend on the relative weights placed on misconduct versus such outside benefits. We view our definition of under-reporting as a benchmark from which one can always expand payoffs if one is more interested in social welfare statements; in general, we do not take a stand on social welfare (see Section 1.7.7).

1.7.6 What if an agent's payoff when the manager is sanctioned is positive even if they do not report?

In reality, an agent's payoff from not reporting may be greater when the manager is sanctioned than when the manager is not sanctioned. This would introduce an additional incentive for agents to free-ride on others' reporting and not report themselves, exacerbating the coordination problem. Our model shows that the coordination problem and under-reporting exist even in the absence of an explicit free-riding incentive, highlighting the central role of strategic uncertainty.

1.7.7 What about social welfare?

While the existence of under-reporting in the model is a high bar to cross due to our requirement of a Pareto improvement, the impact of under-reporting on broader social welfare is ambiguous because we have not included manager utility. (One can analogously think of analyses of consumer surplus as important but insufficient for full welfare statements.) Social welfare analysis requires several inputs beyond the scope of our model that have little empirical guidance; a key such input is how to weigh the utility of agents who have experi-

enced misconduct versus other agents (see, e.g., Section 1.7.5) or the manager (as we discuss in Section 2.3).

1.7.8 Weinstein and Yildiz (2007b) and equilibrium uniqueness

Weinstein and Yildiz (2007b) points out that small perturbations can select any of the multiple-equilibrium outcomes of the full information game as the uniquely rationalizable outcome. Our contribution is to introduce a realistic and important information friction in the context of reporting sexual misconduct, to show that the resulting strategic uncertainty leads to under-reporting, and to study the implications of strategic uncertainty for #MeToo and related policies. The strength of the model's predictions originates from the first-order importance of the proposed information friction in the context we study.

Equilibrium uniqueness in Proposition 1 is a bonus of a parsimonious applied model that captures the role of the information friction in the specific problem of reporting sexual misconduct. Uniqueness is not a general property of global game models; Multiple equilibria may occur in global games more broadly (see, e.g., Angeletos et al., 2006, 2007). Morris, Shin and Yildiz (2016), building on insights from Weinstein and Yildiz (2007a,b), extends the global games framework to higher-dimensional uncertainty and cases where monotonicity fails. In general, multiplicity may occur, but they provide sufficient conditions for uniqueness and emphasize that the key property of the global games method is its emphasis on strategic uncertainty and higher-order beliefs.

As Angeletos and Lian (2016, p.1111-1112) note, the global game methodology is not "a panacea for getting rid of multiple equilibria"; instead, the "applied value of the global-games uniqueness result rests on elucidating the mechanics of coordination and on highlighting the importance of information...for the questions of interest." The methodology offers "useful applied lessons ... not possible in the context of the earlier [literature]" (p.1071).

The key applied lesson here is that, due to strategic uncertainty and the coordination problem, agents may not report misconduct even when misconduct is so widespread that reporting is in their broader interests as a group. This lesson holds even within two leading sources of multiplicity. First, as we discuss in Section 3, while uniqueness holds in the realistic case where priors are less precise than the informativeness of experiences, multiplicity occurs if priors are highly precise. Even if multiplicity occurs, however, Proposition 2 holds within every threshold strategy equilibrium. Second, as we discuss in Section 1.7.3, multiple

equilibria may occur if sanction is a function of reported x_i's. However, an analogous version of Proposition 2 holds in every threshold strategy equilibrium.

2 The Effects of Endogenous Mentorship

One popular narrative suggests that in the wake of the #MeToo movement, men in positions of authority have expressed increasing reluctance to work closely with, network with, or mentor female colleagues (Bennhold, 2019; Tan and Porzecanski, 2018; Smith, 2018; Ortiz, 2018). A decline in "soft opportunities" for networking and mentorship may be detrimental for the career trajectories of women (Kreiss, forthcoming). For brevity, we refer to such soft interactions under the umbrella term of mentorship.

We show a manager may avoid mentoring agents in order to avoid sanction, and that the mentoring decision has strategic spillovers onto agents' reporting strategies. Suppose a manager knows his own type θ, and chooses whether or not to mentor agents, $a \in \{0, 1\}$.[7] If $a = 1$, he mentors all agents, after which all agents realize experiences x_i and then make their reporting decisions. An agent who is mentored plays strategy $s(x_i, a)$, where $s(x_i, 1) = 1$ if the agent reports and $s(x_i, 1) = 0$ if the agent does not report. If $a = 0$, the manager mentors no agents, and no agents report: $s(x_i, 0) = 0$.[8] As in Section 1, the manager's type θ determines the average agent experience. We maintain the assumption that agents have improper uniform priors over θ for tractability.

We continue to assume that experiences are not entirely under a manager's control, so that conditional on $a = 1$, agent experiences are exogenously realized. For example, a manager may take as given that, should he choose to mentor, agents may interpret the manager's behavior in the "wrong way" as viewed from the manager's perspective, and this may deter the manager from mentoring. This scenario seems plausible in light of the real-world concerns outlined above; U.S. Vice President Michael Pence famously does not dine with women alone (the "Pence rule") arguably for this reason.

We assume the manager's payoff from mentoring (playing $a = 1$) equals $M(\theta) - \gamma r S$,

[7] The manager adjusts only his quantity of mentoring rather than the price of mentorship. This fits well within our context, because soft interactions and opportunities often occur outside of explicit markets.

[8] If the manager can choose $a \in [0, 1]$, our main insights that a manager may select out to avoid sanction and that selection generates a trade-off between mentoring and reporting still hold, though there can exist multiple separating equilibria with the same essential features.

where $M(\theta)$ is a function that captures the manager's utility from mentoring, γr is the probability of sanction conditional on the number of agents reporting r, and $S \geq 0$ is a fixed penalty of sanction. The manager's payoff from not mentoring (playing $a = 0$) is zero. We motivate the components of the manager's payoffs as follows.

The function $M(\theta)$ captures the manager's utility from mentoring, which may depend on his propensity for misconduct. This utility could come from a combination of organizational pressure, explicit incentives, and intrinsic personality characteristics. For example, many firms have made explicit efforts to incentivize the recruitment, retention, and advancement of women and minorities (Kwoh, 2012; Roose, 2012; Koenig, 2018); all else equal, these efforts increase M. Mentorship utility may also vary by θ. For example, low-θ types may include managers who derive utility from mentoring women because they believe there are altruistic benefits from doing so, while high-θ types may include managers who derive utility from sexual misconduct. We allow for different $M(\theta)$ and explore the implications of different properties of M on equilibrium behavior.

Finally, we assume the penalty S is fixed irrespective of θ. Although there are notable high-profile exceptions (Griffin et al., 2018), organizations commonly choose sanctions that do not vary with respect to the degree of substantiated misconduct (U.S. Equal Employment Opportunity Commission, 2016; Timmerman and Bajema, 1998).

2.1 Strategic spillovers

The manager mentors in equilibrium if and only if $M(\theta) - \gamma r S > 0$, given agents' equilibrium reporting strategies. Of course, if every manager type has mentorship utility that exceeds the expected penalty even when all agents report ($M(\theta) > \gamma S \ \forall \theta$), then all types mentor and there are no selection effects. Otherwise, some types θ may opt out of mentoring in equilibrium. This feeds back into reporting because agents' beliefs over strategic uncertainty and hence equilibrium reporting are based on the set of types θ who mentor.

Proposition 3 formalizes this observation and shows that this strategic spillover creates a trade-off between selection and reporting. If high-misconduct (high-θ) types opt out of mentorship, agents become less willing to report misconduct because corroboration is less likely when most other agents are not experiencing misconduct; if low-misconduct types opt out of mentorship, then agents become more willing to report misconduct. The proposition also shows that which effect prevails depends on whether $M(\theta)$ increases or decreases in θ.

Proposition 3 (Selection and strategic spillovers)**.** *Let $s(x_i, a)$ denote an agent's reporting strategy:*

$$s(x_i, 1) = \begin{cases} 1 & \text{if } x_i \geq x_S^* \\ 0 & \text{if } x_i < x_S^*. \end{cases}$$

1. *In any equilibrium where the manager mentors if and only if $\theta \leq \tilde{\theta}$ for a unique finite $\tilde{\theta}$, then $x_S^* > x^*$.*

   (a) *In any such equilibrium, the marginal agent's belief over r is uniformly distributed over $[0, \tilde{r}]$ for $\tilde{r} < 1$ determined in equilibrium.*

   (b) *A sufficient condition for the existence of a unique equilibrium that exhibits this effect is if $M(\theta)$ is a weakly decreasing function of θ with $\lim_{\theta \to -\infty} M(\theta) > 0$ and $S > \frac{\underline{M}}{\gamma}$, where $\underline{M} \equiv \min M(\theta)$. Then the manager chooses $a = 1$ if and only if $\theta \leq \tilde{\theta}$ for a unique finite $\tilde{\theta}$ in the unique equilibrium.*

2. *In any equilibrium where the manager mentors if and only if $\theta \geq \tilde{\theta}$ for a unique finite $\tilde{\theta}$, then $x_S^* < x^*$.*

   (a) *In any such equilibrium, the marginal agent's belief over r is uniformly distributed over $[\tilde{r}, 1]$ for $\tilde{r} > 0$ determined in equilibrium.*

   (b) *A sufficient condition for the existence of a unique equilibrium that exhibits this effect is as follows. Let $M(\theta)$ be a weakly increasing function that strictly increases at a finite θ' as follows:*

   $$M(\theta) = \begin{cases} g(\theta) & \text{if } \theta < \theta' \\ h(\theta) & \text{if } \theta \geq \theta', \end{cases}$$

   where $g(\theta') < 0$, $h(\theta') > \gamma S$, and $g(\theta)$ and $h(\theta)$ are weakly increasing functions of θ. Then the manager chooses $a = 1$ if and only if $\theta \geq \theta'$ in the unique equilibrium.

For both Parts 1 and 2, $\tilde{r}$ is determined in equilibrium by $\tilde{r} \equiv \hat{r}(\tilde{\theta}|a = 1) = \Phi\left(\frac{\tilde{\theta} - x_S^*}{\sigma}\right)$.

Part 1 considers the case where high-misconduct manager types opt out of mentoring and shows that reporting thresholds are lower than in the game without selection. Part 1(a) provides the intuition by relating this effect to strategic uncertainty. Intuitively, because a mentored agent observes $a = 1$ and anticipates that types with $\theta > \tilde{\theta}$ are not mentoring, she knows that, at most, the number of agents reporting is the number of agents who would report type $\tilde{\theta}$. This knowledge right-truncates her beliefs about how many agents are reporting. As a result, the marginal agent is still effectively uninformed about how many agents are

reporting, but only over the limited range of reporting r consistent with $\theta < \tilde{\theta}$. As a result, the equilibrium reporting threshold x_S^* exceeds x^*, and the effect of selection is to depress the motive to report misconduct.

Part 1(b) shows that high-misconduct types are likely to opt out when $M(\theta)$ is weakly decreasing, which could happen if low-misconduct types derive significant altruistic benefits from mentoring agents. If low-misconduct types receive positive utility from mentoring ($\lim_{\theta \to -\infty} M(\theta) > 0$) that is larger than the utility of high-misconduct types, and S is sufficiently large, then only high-misconduct types opt out due to their small mentoring utility and high expected sanction cost.

Part 2 provides the converse analysis: In any equilibrium in which only low-misconduct types opt out, agents become more willing to report misconduct. Part 2(a) provides the reason: mentored agents know that types $\theta < \tilde{\theta}$ are not mentoring, which left-truncates their beliefs about how many agents are reporting, strengthening their motive to report.

Part 2(b) shows that low-misconduct types are likely to opt out when $M(\theta)$ is increasing, which could happen because high-misconduct types are predatory and derive significant utility from misconduct. The sufficient condition starkly illustrates with an $M(\theta)$ that discontinuously jumps at θ'. In this case, low-misconduct types opt out as they do not receive positive mentoring utility ($M(\theta) < 0$ for $\theta < \theta'$), while high-misconduct types mentor as their $M(\theta)$ is so high that it exceeds expected penalties even if all agents report ($M(\theta) > \gamma S$ for $\theta > \theta'$). More broadly, if $M(\theta)$ is continuous and increasing in θ, but there exists some θ' such that $M(\theta) < 0$ for $\theta < \theta'$, a set of intermediate-θ types may also opt out in addition to the low-θ types.

Proposition 3 presents two starkly contrasting cases that illustrate the robust prediction that a manager's decision to mentor always generates a strategic spillover on reporting. In reality, the true shape of the $M(\theta)$ function is an open empirical question beyond the scope of this paper, but the strategic spillover is always present. Proposition 3 also holds if sanction S depends on the accusations x; the sanction function would affect which types θ opt out but not the described spillover effect. The Proposition provides a theoretical basis for the popular concern, supported by survey evidence (Atwater et al., 2018), that #MeToo could have the unintended effect of dampening the career prospects of women through fewer networking and mentoring opportunities.

2.2 Unintended effects of tools intended to encourage reporting

We now examine how tools that are intended to encourage reporting or sanction misconduct have unintended consequences on agents' reporting decisions when managers can endogenously choose whether to mentor. We study two interventions: increasing the sanction penalty S, and increasing the sensitivity of sanction to reporting γ. To fix ideas, we analyze the simplest case where mentorship utility does not change with manager types: $M(\theta) = m$, where $m > 0$ is constant. Proposition 4 characterizes the equilibrium of this environment.

Proposition 4 (Equilibrium when $M(\theta) = m$)**.** *Suppose agents have improper uniform priors over θ. If $S > m/\gamma$, there is a unique symmetric perfect Bayesian equilibrium in which the manager mentors ($a = 1$) if $\theta \leq \tilde{\theta}$ and does not mentor ($a = 0$) if $\theta > \tilde{\theta}$. The marginal type $\tilde{\theta}$ that mentors agents is given by:*

$$\tilde{\theta} = \frac{c}{\omega - \beta\left(1 - \frac{m}{2S}\right)} + \sigma\Phi^{-1}\left(\frac{m}{\gamma S}\right). \tag{2}$$

Agent's reporting strategies equal:

$$s(x_i, 1) = \begin{cases} 1 & \text{if } x_i \geq x_S^* \\ 0 & \text{if } x_i < x_S^*, \end{cases} \tag{3}$$

$$x_S^* = \frac{c}{\omega - \beta\left(1 - \frac{m}{2S}\right)}, \tag{4}$$

where $x_S^* > \underline{x} > 0$.

Uniqueness is up to either strategy being played by the manager with $\theta = \tilde{\theta}$ and the agent with $x_i = x_S^*$. The marginal agent's belief over r is uniformly distributed over $[0, \tilde{r}]$ where $\tilde{r} < 1$.

In this unique equilibrium, high-misconduct types are reluctant to mentor agents, so Part 1 of Proposition 3 applies. Because agents realize they are mentored by a $\theta < \tilde{\theta}$ type, they are more reluctant to report any misconduct because they are more pessimistic about how many others will report due to the right-truncation in beliefs over r: $x_S^* > x^*$. Table 3 summarizes comparative statics within the game considered by Proposition 4.

2.2.1 The effect of sanctions

Corollary 4.1 shows that increasing the sanction S depresses reporting and can have potentially adverse effects for agents through several channels.

Table 3: Comparative statics given manager selection.

Parameter	Reporting Threshold x_S^*	Aggregate Reporting $\hat{r}(\theta\|\theta < \tilde{\theta})$	Marginal Manager Type $\tilde{\theta}$
ω: Intrinsic motive	-	+	-
c: Reporting cost	+	-	+
γ: Sanction sensitivity	0	0	-
β: Retaliation	+	-	+
σ: Experience dispersion	0	+ if $\theta < x_S^*$ − if $\theta > x_S^*$ 0 if $\theta = x_S^*$	− if $\tilde{\theta} < x_S^*$ + if $\tilde{\theta} > x_S^*$ 0 if $\tilde{\theta} = x_S^*$
S: Sanction	+	-	- (unless medium S and low σ)
m: Mentorship utility	-	+	+ (unless medium S and low σ)

Corollary 4.1. *The following hold with respect to S:*

1. *Increasing sanction S raises the reporting threshold and lowers the reporting of the marginal type: $\frac{\partial x_S^*}{\partial S} > 0$, and $\frac{\partial \tilde{r}}{\partial S} < 0$.*

2. *Increasing sanction S can lower $\tilde{\theta}$: $\frac{\partial \tilde{\theta}}{\partial S} < 0$ when $S = \frac{m}{\gamma}$ or as $S \to \infty$. A sufficient condition for $\frac{\partial \tilde{\theta}}{\partial S} < 0$ is:*

$$\sigma > \frac{1}{\sqrt{2\pi}} \frac{c\gamma\beta}{2(\omega - \beta)^2}.$$

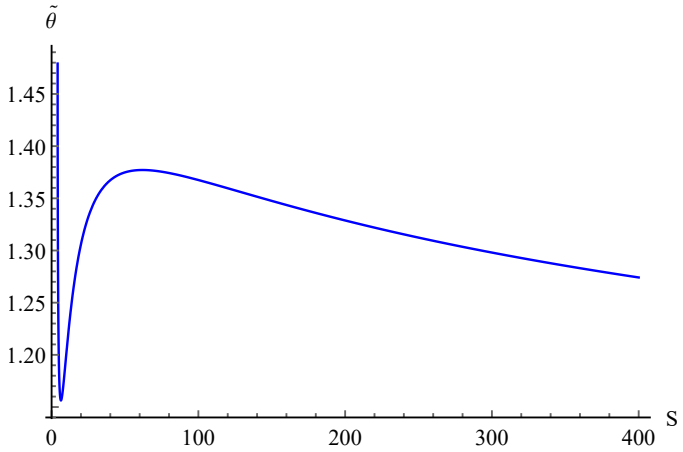
3. *Increasing sanction S can raise $\tilde{\theta}$: A sufficient condition for $\frac{\partial \tilde{\theta}}{\partial S} > 0$ when $\tilde{r} = 1/2$ is:*

$$\sigma \quad < \quad \frac{1}{\sqrt{2\pi}} \frac{c\gamma\beta}{2\left(\omega - \beta + \frac{1}{4}\gamma\beta\right)^2}.$$

Part 1 says that increasing the size of sanction S raises the equilibrium reporting threshold x_S^*. The reason is that, with a higher sanction cost, the amount of reporting $\tilde{r}$ for the marginal type $\tilde{\theta}$ must fall to keep him indifferent between mentoring and not-mentoring agents. This right-truncates agents' beliefs over r even further, raising the reporting threshold x_S^*.

Parts 2 and 3 together show that increasing sanctions have ambiguous effects on the identity of the marginal type $\tilde{\theta}$ that chooses to mentor. The intuition for the ambiguity is because $\tilde{r}$ can fall in two ways when S rises: 1) directly because x_S^* rises, and 2) indirectly because the marginal type $\tilde{\theta}$ changes. In Part 2, the higher x_S^* has a small effect on $\tilde{r}$, so $\tilde{\theta}$ must fall to induce the required fall in $\tilde{r}$ that compensates for an increase in S. In Part 3, the higher x_S^* may have such a large effect on $\tilde{r}$ that $\tilde{\theta}$ can increase. Figure 2 illustrates a

Figure 2: Equilibrium $\tilde{\theta}$. This figure plots a non-monotone example of equilibrium $\tilde{\theta}$ as a function of S.



case where $\tilde{\theta}$ is non-monotone with respect to S.[9]

This leads to the following striking conclusion: there can exist different levels of S that result in the same amount of mentoring and the same level of expected sanction for the marginal type, yet the *lowest* S among these leads to the lowest reporting threshold for agents and the highest aggregate reporting of any given manager. One can see this by drawing a horizontal line in Figure 2, as it can intersect the $\tilde{\theta}$ curve at up to three points. All three points correspond with identical amounts of mentoring. Because the expected sanction is also the same for all three points, the point with the lowest S must correspond with the highest probability of sanction and highest $\tilde{r}$.[10]

2.2.2 The effect of sanction sensitivity

Corollary 4.2 highlights that selection effects blunt the effectiveness of tools intended to encourage reporting through an increase in γ, the sensitivity of sanction to reporting.

Corollary 4.2. *In the presence of selection, agents' reporting strategies are less responsive to increases in γ compared to the case without selection:* $\frac{\partial x_S^*}{\partial \gamma} = 0 < \frac{\partial x^*}{\partial \gamma}$.

[9] In detail: For Part 2, $\tilde{r}$ is very large or very small when S is very small (m/γ) or large (∞), respectively. Any change in x_S^* thus has very small effect on $\tilde{r}$. Similarly, if S is intermediate and σ is large, then $\tilde{r}$ does not fall by much because any given change in the reporting threshold x_S^* does not have a large impact on aggregate reporting. For Part 3, if S is intermediate and σ is small, $\tilde{\theta}$ can rise and the marginal manager can become more willing to mentor because he anticipates that the increase in x_S^* will translate into a significant decrease in $(\tilde{\theta} - x_S^*)/\sigma$, leading to a large drop in $\tilde{r}$.

[10] Since the mentorship utility m has the opposite effect of sanction S, the inverse results apply to m. Increasing m leads to more reporting but can have a non-monotonic effect on manager selection.

Intuitively, without selection, an increase in γ increases an agent's incentive to report because it increases the probability of sanction for any given level of reporting r. However, in the presence of selection with constant $M(\theta)$, agents expect a lower r because they expect fewer agents will report in equilibrium (Part 1 of Proposition 3), weakening this effect. When $\Gamma(r) = \gamma r$, the selection effect is strong enough to reduce the effect of γ on reporting to zero. One can show analogous results for a class of non-linear functions, $\Gamma(r) = \gamma r^{\frac{1}{n}}$ where $n > 0$.

2.3 Social welfare and takeaways

Overall, our results show that encouraging reporting or punishing misconduct interacts with the manager's mentorship decision in potentially adverse ways for agents. A social welfare analysis in this environment is beyond the scope of this paper because it requires several inputs beyond our model that have little guidance from empirical evidence. These inputs include valuing any disutility from misconduct to agents, any positive utility from misconduct for managers, the exact utility function $M(\theta)$ of mentorship for managers, the benefits of mentorship for agents, and whether managerial sanctions are dead-weight loss. No less important, such an analysis also requires distributional assumptions on θ and welfare weights for agents and managers.

Instead, our theory offers guidance for how #MeToo and policies that encourage reporting affect agents and managers. First, policymakers should recognize that managers may opt out of mentorship, consistent with popular concern. Second, this opt-out affects agents' incentives to report misconduct. If high-misconduct managers opt out, fewer agents tend to experience severe misconduct, but those who do are more reluctant to report it. If low-misconduct managers opt out, then agents receive fewer positive mentoring opportunities, but are more willing to report bad experiences. Finally, tools intended to encourage reporting or punish misconduct may be less effective than supposed, due to these strategic spillovers. For example, as we show in the $M(\theta) = m$ case, increasing sanction penalties depresses reporting, and increasing the sensitivity of sanction to reporting may be ineffective at changing agents' reporting thresholds.

3 The Publicity Effect of #MeToo

The #MeToo movement was popularized by actress Alyssa Milano on Twitter, who wrote: "If all the women who have been sexually harassed or assaulted wrote 'Me too' as a status, we might give people a sense of the magnitude of the problem" (Khomami, 2017). Surveys confirm that women believe that there is value to heightened awareness of widespread problems (e.g., within the economics profession; see American Economic Association, 2019b; Casselman and Tankersley, 2019).

We show that heightened public awareness of misconduct increases reporting of otherwise-hidden misconduct by coordinating beliefs over strategic uncertainty. We associate changes in public information with changes in agents' common priors (Angeletos and Lian, 2016). In Section 1.4, we assumed that agents had common improper uniform priors over θ. Now suppose that agents share proper priors $p(\theta)$ that is normally distributed with mean y and standard deviation τ. Such priors could be shaped by public information about a specific manager or information about misconduct in the broader population of managers. Proposition 5 characterizes behavior in this environment. The proof strategy follows Morris and Shin (2004a); Appendix A provides all details.

Proposition 5 (Equilibrium with proper priors)**.** *Suppose agents have common proper priors that θ is distributed normally with mean y and standard deviation τ.*

1. *Existence: There exists at least one symmetric threshold strategy equilibrium $x_I^* \in (\underline{x}, \overline{x})$ where agents report if and only if $x_i \geq x_I^*$, where x_I^* is implicitly defined by:*

$$x_I^* = \frac{c}{\omega - \beta\left(1 - \gamma\Phi\left(\frac{y - x_I^*}{\kappa}\right)\right)}, \tag{5}$$

   for $\kappa \equiv \frac{\sigma^2+\tau^2}{\sigma}\sqrt{\frac{\sigma^2+2\tau^2}{\sigma^2+\tau^2}}$ and where $x_I^* > \underline{x} > 0$.

2. *Uniqueness (up to either reporting or not reporting when $x_i = x_I^*$):*

   (a) *The equilibrium x_I^* is a unique threshold equilibrium if $\kappa > \frac{1}{\sqrt{2\pi}}\frac{c\gamma\beta}{(\omega-\beta)^2}$.*

   (b) *Whenever $y = \frac{c}{\omega-\beta\left(1-\frac{\gamma}{2}\right)}$, a sufficient condition for non-unique equilibria is $\kappa < \frac{1}{\sqrt{2\pi}}\frac{\gamma\beta c}{\left(\omega-\beta\left(1-\frac{\gamma}{2}\right)\right)^2}$.*

   (c) *For $y \to \infty$, there is a unique threshold equilibrium x_I^* with $x_I^* \to \underline{x}$. For $y \to -\infty$, there is a unique threshold equilibrium x_I^* with $x_I^* \to \overline{x}$.*

(d) If there is a unique equilibrium in threshold strategies, then the equilibrium strategy is the only strategy that satisfies the iterated deletion of strictly dominated strategies. In particular, the unique threshold strategy equilibrium is the globally unique equilibrium.

3. *Beliefs over r: The marginal agent, who has experience $x_i = x_I^*$, has a belief over the incidence of reporting r characterized by the cumulative distribution function $\Psi_I(.)$:*

$$\Psi_I(r) \equiv \Phi\left(\frac{\sigma}{\tau}\frac{1}{\sqrt{\sigma^2+\tau^2}}(x_I^* - y) + \frac{1}{\tau}\sqrt{\sigma^2+\tau^2}\,\Phi^{-1}(r)\right), \tag{6}$$

and has expectation $E_I^*[r] = \Phi\left(\frac{y-x_I^*}{\kappa}\right)$.

3.1 Equilibrium uniqueness and under-reporting

There is a unique equilibrium when priors are sufficiently diffuse or when private experiences are informative for θ relative to priors. Specifically, Part 2(a) of Proposition 5 is satisfied for high τ or when σ is low relative to τ. The reason is that κ increases in τ; furthermore, κ decreases in σ when $\sigma^2/\tau^2 < \sqrt{2}$. Indeed, as $\tau \to \infty$ or $\sigma \to 0$, we converge to the unique equilibrium in Proposition 1, because $\kappa \to \infty$ and $\Psi_I(r) \to r$.

This observation comports with well-known results in the literature about global games: precise public information induces coordination and multiple equilibria, while private information hinders coordination because signals are not common knowledge (Angeletos and Lian, 2016; Morris and Shin, 2003, 2004a). Part 2(b) illustrates equilibrium multiplicity. Suppose $y = \frac{c}{\omega - \beta\left(1-\frac{\gamma}{2}\right)}$, so that $y = x_I^*$ and $E_I^*[r] = 1/2$. Multiplicity occurs if $\tau \approx 0$ and σ is not too large.[11] We discuss the relevance of Part 2(c) below.

With proper priors, under-reporting still occurs if and only if misconduct is widespread. The "open-secret" equilibrium of Corollary 1.1 is unchanged even when priors are proper. Proposition 2 holds in the unique equilibrium or in any threshold strategy equilibrium of Proposition 5 (Appendix A.2.2). Condition (1) holds so long as $\theta > x_I^* + \frac{\sigma}{\kappa}(y - x_I^*)$. Condition (2) holds for $\theta \in \left(\underline{x} + \frac{\sigma}{\kappa}(y - x_I^*), x_I^* + \frac{\sigma}{\kappa}(y - x_I^*)\right]$ and σ sufficiently small.

From here on, we assume the environment satisfies the condition for a unique threshold equilibrium in Part 2(a); the equilibrium is then globally unique by Part 2(d). This assump-

[11] Small τ is insufficient to guarantee small κ and equilibrium multiplicity because $\lim_{\tau\to 0}\kappa = \sigma$. If σ is large, κ can be large, potentially satisfying the sufficient condition for uniqueness in Part 2(a). Part 2(b) guarantees this does not happen.

Table 4: Comparative statics with proper priors. Comparative statics for ω, c, γ, and β are identical to those in Section 1.6. Note that $y > x_I^*$ if and only if $y > \frac{c}{\omega - \beta\left(1-\frac{\gamma}{2}\right)}$.

Parameter	(1) Reporting Threshold x^*	(2) Aggregate Reporting $\hat{r}(\theta)$
y: Public belief of average type	-	+
τ: Public belief of type dispersion (for $y > x_I^*$; flip signs if $>$; 0 if "=")	+	-

tion seems plausible because public information about any prior misconduct by a specific individual is often diffuse. Such information does not accumulate smoothly in public through time, often because the accused and accusers settle claims using non-disclosure agreements (NDAs). Information about whether such NDAs exist is disperse, and the threat of enforcement is effective at keeping information about prior allegations hidden (Lobel, 2018). For example, NDAs kept information about Harvey Weinstein's misconduct hidden for many years (for other anecdotes, see Benner, 2017, or the story of "LaDonna" in Episode #647 of "This American Life"). Moreover, information about misconduct in the broader population of managers is likely less informative about a specific manager's type than experiences from that manager, suggesting Part 2(a) is satisfied.

3.2 Implications for #MeToo

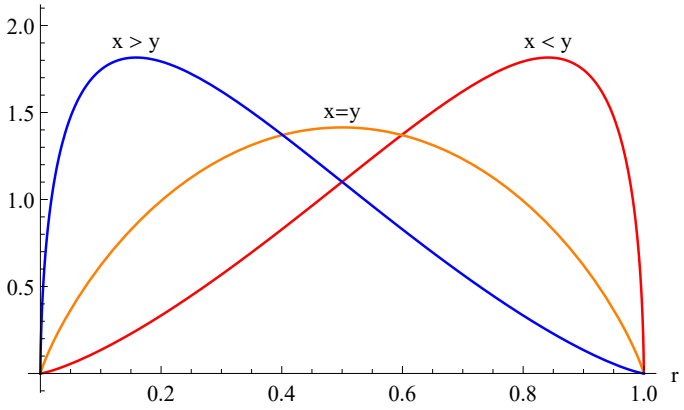
Table 4 summarizes how changes in the mean and standard deviation of the prior belief, y and τ, affect equilibrium reporting. Corollary 5.1 highlights the key implication relevant for the #MeToo movement.[12]

Corollary 5.1 (Publicity effect of #MeToo)**.** *Agents become more willing to report (x_I^* falls) when y increases. When y is high, agents also become more willing to report when τ falls.*

If y increases so that the public believes that managers as a whole are engaging in worse misconduct on average than previously believed, the reporting threshold x_I^* falls and

[12]The comparative statics for ω, c, γ, and β are identical to those in Section 1.6. For σ, distinct from the improper prior case in Section 1.6, an increase in σ can either increase or decrease the reporting threshold x_I^*, because the change in σ has two effects on beliefs over r. First, it affects the marginal agent's assessment of how likely her experience was relative to what other agents might be experiencing based on her priors, but it also affects how much she revises her belief about what experience she expects others to have when forming her posteriors. Which effect dominates depends on σ^2/τ^2 and whether y is low or high. For brevity, we omit details about σ as it seems less directly relevant to the #MeToo movement.

Figure 3: Probability density function $\psi_I(r)$. This figure plots the probability density function of the marginal agent's belief over the number of agents reporting for three different cases denoted in the figure. The cumulative distribution function for this belief is $\Psi_I(r)$ given in Proposition 5.



aggregate reporting $\hat{r}(\theta)$ rises for every θ. The reason is that an increase in y makes agents believe the average experience is worse and hence more agents are likely to report.

The intuition is easiest to see when we are starting from an equilibrium where $y = \frac{c}{\omega - \beta\left(1-\frac{\gamma}{2}\right)}$ so that $x_I^* = y$. The marginal agent believes $E_I^*[r] = \frac{1}{2}$, and Figure 3 plots the density of her belief over r. Note that, from Equation 6, the CDF of her belief shifts "to the right" when y increases, as $\Psi_I(r)$ decreases in y for all $r \in (0,1)$ and any fixed value of x_i. This implies that, in response to an increase in y, the marginal agent will think that more agents will report because they are having worse experiences, making her willing to report and no longer indifferent.

At the higher value of y, the marginal agent in the new equilibrium must be indifferent at a less-bad experience; at the new indifference point, $x_I^* < y$. Figure 3 plots the marginal agent's belief over r in this new equilibrium. When $x_I^* < y$, the marginal agent believes many other agents are reporting, giving her the confidence to come forward despite a less-bad experience than the marginal agent in the previous equilibrium. Analogous logic applies if y falls: at the new indifference point where $x_I^* > y$, the marginal agent believes fewer other agents are reporting, so she must have a worse experience to come forward.

How reporting responds to the reliability of the public signal τ depends on whether x_I^* is greater or less than y. Suppose $x_I^* < y$ and the marginal agent believes many other agents are reporting. In response to a lower τ, the marginal agent believes that she had an even lower-than-expected draw of x, and therefore that even more agents are reporting. She thus becomes more willing to report and is no longer indifferent, lowering the new equilibrium

threshold. Analogous logic applies if $x_I^* > y$: when τ decreases, the marginal agent believes she had an even higher-than-expected x and that even fewer agents are reporting, making her less willing to report and raising the equilibrium threshold.

Corollary 5.1 is a robust first-order prediction irrespective of underlying parameters: Part 2(c) from Proposition 5 guarantees that the equilibrium is unique when $y \to \infty$ and that $x_I^* \to \underline{x}$, irrespective of other parameters. Part 2(c) also shows that Corollary 5.1 has first-order effects on the magnitudes of x_I^*, as changes in y can move the equilibrium threshold x_I^* across the entire range of $(\underline{x}, \overline{x})$.

Overall, Corollary 5.1 is consistent with more agents coming forward with accusations in the wake of #MeToo, which publicized several major incidents of misconduct and arguably raised y (and perhaps decreased τ) by raising public awareness of sexual misconduct. Our model suggests that heightened awareness led directly to more reporting even though: 1) experiences may have remained unchanged, and 2) there was no direct impact on agent payoffs $\pi(r, x)$. In particular, agents with "hidden x_i" who were previously not reporting due to a low-y environment may come forward in a higher-y (possibly lower-τ) environment purely from a change in beliefs about whether other agents are reporting.

4 Policies Intended to Encourage Reporting

4.1 Holding tanks for reports

One proposal to encourage reporting of misconduct is to hold any reports of misconduct about a manager in a confidential "holding tank" unless the number of reports exceeds a certain bar. In this case, the tank opens, and reports are released to the outside party (e.g., similar to American Economic Association, 2019a). Importantly, the outside party is not privy to allegations unless the tank opens. The idea behind this policy is to encourage reporting by protecting agents from retaliation.

However, we show that a holding tank does not unequivocally encourage reporting: if agents do not receive much utility from filing a report that only ever stays in the tank, then a tank may inadvertently discourage reporting by making it more difficult for reports to be released to the outside party.

Suppose that a policymaker institutes a *release bar* $\overline{r} \in [0, 1]$ in which reports are held in

Table 5: Agent's Payoffs with Release Bar $\overline{r}$. The parameter δ measures utility from making an unreleased report. $\mathbf{1}_{[x_i \geq 0]}$ is an indicator function that equals one if $x_i \geq 0$.

		Released ($r \geq \overline{r}$)	
	Unreleased ($r < \overline{r}$)	Sanction	No Sanction
Report	$x_i\delta - c$	$x_i\omega - c$	$x_i(\omega - \mathbf{1}_{[x_i \geq 0]}\beta) - c$
Not Report	0	0	0

a holding tank as long as $r < \overline{r}$, and are released to the outside party only if $r \geq \overline{r}$. Given the threshold $\overline{r}$, the sanction function $\Gamma(r)$ becomes equivalent to:[13]

$$\Gamma(r) = \begin{cases} 0 & \text{if } r < \overline{r} \\ \gamma r & \text{if } r \geq \overline{r}. \end{cases} \tag{7}$$

Table 5 summarizes the agent's payoffs. If $r \geq \overline{r}$, then the agent's payoffs are identical to her payoffs in the base model from Section 1.2. If $r < \overline{r}$, then the agent's report is not released to the outside party, the manager cannot be sanctioned, and the agent receives $x_i\delta - c$ for $\delta \in [0, \omega]$. The parameter δ reflects the agent's utility from filing a report that is not released and has several non-mutually-exclusive interpretations. First, δ could reflect intrinsic utility from reporting misconduct and is thus naturally smaller from a report that is never released than from a report that is released to the outside party. Second, δ could reflect intrinsic utility net of (real or perceived) possible retaliation resulting from a "leak" from the holding tank. A lower δ would then reflect greater concerns from the agent that her confidentiality could be breached.

Given the sanction function and payoffs, an agent's payoff gain π_H from reporting equals:

$$\pi_H(r, x) = \begin{cases} x\delta - c & \text{if } x \geq 0 \text{ and } r < \overline{r} \\ x(\omega - (1 - \gamma r)\beta) - c & \text{if } x \geq 0 \text{ and } r \geq \overline{r} \\ x\omega - c & \text{if } x < 0. \end{cases} \tag{8}$$

Note that the sanction function, agent payoffs, and payoff gain function π_H nest the model from Section 1.2 and outcomes determined by Proposition 1 when $\overline{r} = 0$, which corresponds to the absence of a holding tank. We maintain the assumption that agents have improper

[13] To make a clear comparison to Section 1.2, we continue to assume that $\Gamma(r) = \gamma r$ once reports are released (if $r \geq \overline{r}$). But our conclusions apply to a general $\Gamma(r)$ that weakly increases in r if $r \geq \overline{r}$.

uniform priors over θ for tractability and comparability to Section 1.2.

Proposition 6 shows that raising the release bar $\overline{r}$ can either encourage or discourage reporting depending on δ, the payoff from filing a report that is not released. The reason is that raising $\overline{r}$ has two countervailing effects. First, a *protection effect* encourages reporting: raising $\overline{r}$ protects reporting agents from retaliation in the event that too few others report ($r < \overline{r}$). Second, a *raise-the-bar effect* discourages reporting: raising $\overline{r}$ reduces an agent's belief that an outside party investigates and that the manager is sanctioned ($r \geq \overline{r}$). Intuitively, the raise-the-bar effect exists because strategic uncertainty remains and because an agent is still uncertain whether or not others' reporting will spill over the release bar $\overline{r}$. If δ is too small, the raise-the-bar effect dominates, and raising $\overline{r}$ strictly discourages reporting.

Proposition 6. *Suppose agents have improper uniform priors over θ.*

1. *If $\delta \in [0, \omega - \beta)$, there exists a globally unique symmetric threshold strategy equilibrium in which agents use reporting threshold x_H^*:*

$$x_H^* = \frac{c}{\delta\overline{r} + \omega(1-\overline{r}) - \beta\left(1 - \overline{r} - \frac{1}{2}\gamma(1-\overline{r}^2)\right)}. \tag{9}$$

   Furthermore, $\frac{\partial x_H^*}{\partial \overline{r}} > 0$, so that the policy $\overline{r}_{min}$ that minimizes x_H^* is $\overline{r}_{min} = 0$.

2. *If $\delta \in [\omega - \beta, \omega]$, there exists a unique symmetric threshold equilibrium, with reporting threshold x_H^* described by Equation 9, when the following closed-form expression of π^* satisfies the single-crossing condition: $\pi^*(x, x_H^*) < 0$ for all $x < x_H^*$ and $\pi^*(x, x_H^*) > 0$ for all $x > x_H^*$:*

$$\begin{aligned}\pi^*(x,k) =& x(\omega - \beta(1-\gamma)) - x(\omega - \beta(1-\gamma) - \delta)\, F\left(\frac{k-x}{\sigma} - F^{-1}(1-\overline{r})\right) \\ & - x\beta\gamma\left[BvN\left(\frac{x-k}{\sigma} + F^{-1}(1-\overline{r}), \frac{1}{\sqrt{2}}\frac{k-x}{\sigma}; \frac{-1}{\sqrt{2}}\right)\right] - c,\end{aligned} \tag{10}$$

   where $BvN(h, l; \rho)$ is the bivariate normal cumulative distribution function with limits of integration h and l and correlation ρ (see Owen, 1980). By definition, Equation 10 satisfies $\pi^*(x_H^*, x_H^*) = 0$. As a sufficient condition, the function $\pi^*(x, k)$ satisfies the single-crossing requirement at the limit $\sigma \to \infty$.

   When such an equilibrium exists, then there exists a unique $\overline{r}_{min} \in (0, 1]$ that minimizes x_H^*, and $\frac{\partial x_H^*}{\partial \overline{r}} < 0$ for $r < \overline{r}_{min}$ and $\frac{\partial x_H^*}{\partial \overline{r}} > 0$ for $r > \overline{r}_{min}$. Furthermore, $\frac{\partial \overline{r}_{min}}{\partial \delta} > 0$ if $\overline{r}_{min} \in (0, 1)$.

Part 1 shows that instituting a holding tank deters reporting when δ is sufficiently low: the policy $\overline{r}_{min}$ that minimizes x_H^* is $\overline{r}_{min} = 0$. If δ is low, an agent does not sufficiently value

filing a report that stays in the tank and fails to lead to an investigation; she strictly prefers to risk exposure to retaliation in order to have the outside party investigate and potentially sanction the manager. Because agents place a low value on the protection afforded by the tank, raising $\overline{r}$ decreases the marginal agent's willingness to report and raises x_H^*, and the raise-the-bar effect always dominates. Imposing a holding tank thus deters reporting.

Part 2 shows that if δ is sufficiently high, then instituting a release bar $\overline{r} \in (0, 1]$ can encourage or discourage reporting relative to $\overline{r} = 0$. Agents capture sufficient utility from filing a report even if it stays in the tank, and the protection effect can dominate the raise-the-bar effect. Starting from a low value of $\overline{r}$ (potentially zero), the raise-the-bar effect is small: conditional on reports, it is likely the tank will open even with a slightly higher $\overline{r}$, yet a higher $\overline{r}$ affords agents valuable protection if the tank does not open. Thus, the protection effect dominates, and raising $\overline{r}$ makes the marginal agent more willing to report and lowers x_H^*, encouraging reporting. However, further increase of $\overline{r}$ will progressively make the release of reports less likely. At sufficiently high $\overline{r}$, the raise-the-bar effect dominates, and raising $\overline{r}$ discourages reporting and raises x_H^*. If $\overline{r}$ is raised sufficiently, the equilibrium reporting threshold x_H^* may rise above the threshold without a holding tank $x_H^*(\overline{r} = 0)$. However, there exists an interior release bar $\overline{r}_{min}$ that minimizes x_H^* with $x_H^*(\overline{r} = \overline{r}_{min}) < x_H^*(\overline{r} = 0)$. Overall, a holding tank with $\overline{r} = \overline{r}_{min}$ encourages reporting, but a release bar $\overline{r}$ that is too high can discourage reporting. The release bar $\overline{r}_{min}$ increases with δ because the relative value of protection increases with δ.[14]

Overall, our analysis raises a surprising conundrum for holding tanks: a holding tank can encourage reporting only if agents gain sufficient utility from filing a report that is kept in the tank and never released to the outside party.[15]

[14]The existence of a symmetric threshold strategy equilibrium is not guaranteed when δ is sufficiently high because a high δ perversely leads an agent near the margin to prefer that reports remain protected in the holding tank rather than be released. This is particularly the case when σ is small, as an agent who draws x below threshold x_H^* and does not report may infer that it is exceedingly unlikely that reports will be released, and thus may prefer to deviate and report so that she can safely capture δ without the tank opening. With σ high, such deviations are less likely, and the threshold equilibrium may be sustained. No agents have an incentive to deviate if the closed-form solution of agent's expected payoffs given x when other agents play reporting threshold k, given by $\pi^*(x, k)$ in Equation 10, satisfies the single crossing condition at $k = x_H^*$. The Proposition shows that this single crossing condition is satisfied at the limit of $\sigma \to \infty$.

[15]The conundrum is robust and not specific to our functional form. The economic intuition is that, if agents primarily gain utility from having reports released to the outside party ($\pi(r, x)$ is monotone in r), a holding tank precludes this possibility and is thus not desirable for agents. In this case, Part 1 of Proposition 6 applies. A holding tank can only be desirable if agents gain sufficient utility from having reports *not* released to the outside party. This makes $\pi(r, x)$ non-monotone in r and leads to Part 2 of Proposition 6. In our

Table 6: Agent's Payoffs with Whistleblower Rewards. The parameter α is the whistleblower reward. $\mathbf{1}_{[x_i \geq 0]}$ is an indicator function that equals one if $x_i \geq 0$.

	Sanction	No Sanction
Report	$x_i(\omega + \mathbf{1}_{[x_i \geq 0]}\alpha) - c$	$x_i(\omega - \mathbf{1}_{[x_i \geq 0]}\beta) - c$
Not Report	0	0

4.2 Rewarding whistleblowers

Eliminating retaliation ($\beta = 0$) would eliminate the coordination problem. In practice, however, eliminating retaliation is difficult; despite federal law, various forms of retaliation remain prevalent (Cortina and Magley, 2003; McDonald, 2012; U.S. Equal Employment Opportunity Commission, 2016). Furthermore, an agent who considers reporting retaliation suffers from the same coordination problem as from reporting harassment.

Given these difficulties, we next consider the effect of rewarding reports that result in sanction. Such a "whistleblower reward" for coordination success is a natural counterweight to a retaliatory cost for coordination failure. For example, in practice, victims of workplace sexual harassment can win emotional distress damages in Title VII lawsuits. More broadly, when the outside party acts, agents may also gain utility through social recognition, a sense of vindication, or less harassment going forward from the sanctioned manager.

We incorporate whistleblower rewards in the model by extending the payoff structure so that an agent receives a proportional reward αx_i if the third party acts on a claim of $x_i \geq 0$. Table 6 summarizes an agent's final payoffs. Given these payoffs, an agent's payoff gain function π_R equals:

$$\pi_R(r, x) = \begin{cases} x(\omega + \gamma r\alpha - (1 - \gamma r)\beta) - c & \text{if } x \geq 0 \\ x\omega - c & \text{if } x < 0 \end{cases}. \tag{11}$$

Note that the payoffs in Table 6 and the payoff gain function $\pi_R(r, x)$ nest the previous payoff structure with $\alpha = 0$. All previous results on existence, uniqueness, and comparative statics of other parameters in Sections 1-3 continue to apply with straightforward modification, and proofs of all results incorporate the extended payoff structure allowing for $\alpha \geq 0$. For brevity, we provide the formulas for the reporting and selection thresholds when $\alpha > 0$ in Appendix

model, the incentive for an agent to file a report that is never released when δ is very high can also perversely sustain an equilibrium with $\overline{r}_{min} = 1$.

A and omit them in the main text.

In this section, we focus on the role of α for equilibrium behavior. It is straightforward to show that α mitigates the retaliatory cost β because agents are more willing to report for higher α ($\partial x^*/\partial\alpha < 0$ and $\partial\hat{r}(\theta)/\partial\alpha > 0\,\forall\theta$). The following Proposition captures our main insight from implementing whistleblower rewards.

Proposition 7 (Required rewards)**.** *For every $\beta \in (0,\omega)$, there exists a unique* required reward $\alpha^F > 0$ *such that the equilibrium reporting threshold equals the "no-retaliation benchmark," $x^F = c/\omega$. If $\alpha < \alpha^F$, the equilibrium threshold exceeds x_F, so that there is less reporting than in the no-retaliation benchmark. Conversely, if $\alpha > \alpha^F$, the equilibrium threshold is less than x_F, so that there is more reporting.*

1. *In the reporting game with improper uniform priors (Section 1), $\alpha^F = \beta\left(\frac{2-\gamma}{\gamma}\right)$.*
2. *In the constant-m selection game with improper uniform priors (Section 2), $\alpha^F = \beta\left(\frac{2S-m}{m}\right)$.*
3. *In the reporting game with proper priors (Section 3), $\alpha^F = \beta\left(\frac{1-\gamma\Phi(\frac{\omega y-c}{\omega\kappa})}{\gamma\Phi(\frac{\omega y-c}{\omega\kappa})}\right)$.*

Furthermore, in each of the above cases, there also exists an $\alpha^E \in (0,\alpha^F)$ such that the equilibrium reporting threshold equals $x^E \equiv \underline{x} = c/(\omega - \beta(1-\gamma))$.

Proposition 7 suggests that rewards for coordination success have the potential to significantly and even fully offset the retaliatory cost of coordination failure. A sufficiently high α can induce each agent to report as though she were sure that others were all reporting: that is, so that agents play the reporting threshold $x^E \equiv \underline{x}$ in equilibrium.[16] Even at such a threshold, however, an agent faces possible retaliation if $\gamma < 1$. An even higher α can induce each agent to report as though she were sure she would not face retaliation: that is, so that agents play the threshold $x^F = c/\omega$ in equilibrium, where $x^F < x^E$.

The Proposition also suggests that a reward of $\alpha > \alpha^F$ leads to an equilibrium reporting threshold that is lower than x^F and hence to more reporting than in the no-retaliation benchmark. In this case, α itself creates a coordination motive: agents report behavior that they otherwise would not report even in a world where retaliation did not exist. To see this, consider the extreme case where $\beta = 0$ so that $\alpha^F = 0$, but where $\alpha > 0$. Agents in this case are motivated to report when other agents report only due to the possibility of the reward,

[16]The threshold x^E is also the greatest lower bound of all possible Pareto-optimal thresholds $\tilde{x}$ from Proposition 2 and is plausibly relevant for a policymaker who does not know θ.

which in this case mitigates the fixed cost of reporting c.[17] With $\alpha = 0$, changes in β or γ cannot generate an equilibrium threshold less than x^E, let alone x^F.

We highlight three other observations from Proposition 7. Part 1 suggests that there is a potential substitution across different policy tools: The required reward α^F is lower when γ is high or β is low. Part 2 suggests that a higher reward is needed to counteract the strategic spillover effect of manager selection on reporting when $M(\theta)$ is constant. With the same improper prior, α^F is higher in the game with selection than without selection, since $S > m/\gamma$. Likewise, the required reward increases with the size of the sanction and decreases with the size of mentorship utility.

Finally, Part 3 suggests that a lower required reward α^F is needed when public beliefs shift toward a worse average experience through an increase in y. Higher y makes agents more willing to report, thus the required reward decreases accordingly. This suggests that as public awareness of widespread problems increases in the wake of #MeToo, the required reward to induce reporting consistent with the no-retaliation benchmark falls.

5 Conclusion

This paper shows that, due to strategic uncertainty, the coordination problem leads agents to under-report sexual misconduct if and only if misconduct is widespread. It highlights several policy-relevant unintended and intended effects of #MeToo.

Our model suggests several lines of future empirical research. First, empirical work should quantify the effects of several model parameters on the frequency and severity of reporting, including the effect of retaliation (β) the sensitivity of sanction to reports (γ). Second, empirical work should attempt to tease out the role of strategic uncertainty by testing several of the model's applied predictions. First, the model predicts that managers should test whether managers systematically opt out of soft mentorship, and that this spills over onto reporting strategies for agents who are still mentored. Second, the model predicts that raising public awareness (y) should increase reporting. Third, the model predicts that a holding tank encourages reporting only if δ is high and may discourage reporting otherwise, and that rewards for whisleblowers α should encourage reporting.

[17] Whether one interprets this as desirable depends on whether one views c as reflecting a friction (such as fixed retaliation costs) or an agent's personal preferences.

The model also suggests future research in conceptually related areas, including not only more general forms of workplace misconduct, but also contexts where strategic uncertainty may chill speech, reporting, or otherwise create a "culture of silence" where agents do not speak up or express concerns. A common feature of these environments is that agents may fail to coordinate and achieve higher payoffs because any individual agent is uncertain whether other agents will act and fears retaliation or reprisal if they act alone. Coordination problems and strategic uncertainty in these contexts are promising areas for future research.

References

American Economic Association, "AEA Ombudsperson Frequently Asked Questions," `https://www.aeaweb.org/about-aea/aea-ombudsperson/faq` 2019. (accessed September 19, 2019).

_ , "AEA Professional Climate Survey: Main Findings," September 15 2019. `https://www.aeaweb.org/resources/member-docs/final-climate-survey-results-sept-2019` (accessed September 19, 2019).

Anderson, Michelle J., "The Legacy of the Prompt Complaint Requirement, Corroboration Requirement, and Cautionary Instructions on Campus Sexual Assault," *Boston University Law Review*, October 2004, *84* (4), 945–1022.

Angeletos, George-Marios and Alessandro Pavan, "Efficient Use of Information and Social Value of Information," *Econometrica*, 2007, *75* (4), 1103–1142.

_ **and Chen Lian**, "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2A, Elsevier, 2016.

_ **and Ivan Werning**, "Crises and Prices: Information Aggregation, Multiplicity, and Volatility," *The American Economic Review*, 2006, *96* (5), 1720–1736.

_ **, Christian Hellwig, and Alessandro Pavan**, "Signaling in a Global Game: Coordination and Policy Traps," *Journal of Political Economy*, 2006, *114* (3), 452–484.

_ **,** _ **, and** _ , "Dynamic Global Games of Regime Change: Learning, Multiplicity, and the Timing of Attacks," *Econometrica*, 2007, *75* (3), 711–756.

Atwater, Leanne E., Allison M. Tringale, Rachel E. Sturm, Scott N. Taylor, and Phillip W. Braddy, "Looking Ahead: How What We Know About Sexual Harassment Now Informs Us of the Future," *Organizational Dynamics*, 2018.

Basu, Kaushik, "The Economics and Law of Sexual Harassment in the Workplace," *Journal of Economic Perspectives*, 2003, *17* (3), 141–157.

Bénabou, Roland, Armin Falk, and Jean Tirole, "Narratives, Imperatives, and Moral Persuasion," September 2019. Working paper, Princeton University.

Benner, Katie, "Abuses Hide in the Silence of Nondisparagement Agreements," *The New York Times*, "July 21" 2017. `https://www.nytimes.com/2017/07/21/technology/silicon-valley-sexual-harassment-non-disparagement-agreements.html` (accessed September 6, 2019).

Bennhold, Katrin, "Another Side of #MeToo: Male Managers Fearful of Mentoring Women," *The New York Times*, January 28 2019. https://www.nytimes.com/2019/01/27/world/europe/metoo-backlash-gender-equality-davos-men.html (accessed February 11, 2019).

Bower, Tim, "The #MeToo Backlash," *Harvard Business Review*, September-October 2019. https://hbr.org/2019/09/the-metoo-backlash?

Bowley, Graham, "Judge Says Five More Women Can Testify Against Bill Cosby," *The New York Times*, March 15 2018. https://nyti.ms/2GvxYan (accessed August 8, 2019.

Brandeis University, "Policy Against Discrimination, Harassment, and Sexual Misconduct," September 2019. https://www.brandeis.edu/equal-opportunity/policies/pdfs/discrimination-harassment-misconduct.pdf (accessed September 3, 2019).

Carlsson, Hans and Eric van Damme, "Equilibrium Selection in Stag Hunt Games," in Ken Binmore, Alan Kirman, and Piero Tani, eds., *Frontiers of Game Theory*, MIT Press, 1993.

__ **and** __ , "Global Games and Equilibrium Selection," *Econometrica*, 1993, *61* (5), 989–1018.

Casselman, Ben and Jim Tankersley, "Women Face High Levels Of Sex Abuse In Economics," *The New York Times*, March 18 2019, p. B1.

Chassang, Sylvain and Gerard Padró I Miquel, "Crime, Intimidation, and Whistleblowing: A Theory of Inference from Unverifiable Reports," *Review of Economic Studies*, 2019, *0*, 1–24.

Chen, Daniel L and Jasmin K Sethi, "Insiders, Outsiders, and Involuntary Unemployment: Sexual Harassment Exacerbates Gender Inequality," 2018. TSE Working Paper.

Chira, Susan and Catrin Einhorn, "How Tough Is It to Change a Culture of Harassment? Ask Women at Ford," *New York Times*, December 19 2017. https://www.nytimes.com/interactive/2017/12/19/us/ford-chicago-sexual-harassment.html (accessed January 23, 2020).

Cooper, Marianne, "The 3 Things That Make Organizations More Prone to Sexual Harassment," *The Atlantic*, November 27 2017. https://www.theatlantic.com/business/archive/2017/11/organizations-sexual-harassment/546707/ (accessed January 23, 2020).

Cortina, Lilia M and Jennifer L Berdahl, "Sexual harassment in organizations: A decade of research in review," *Handbook of organizational behavior*, 2008, *1*, 469–497.

__ **and Vicki J Magley**, "Raising Voice, Risking Retaliation: Events Following Interpersonal Mistreatment in the Workplace.," *Journal of Occupational Health Psychology*, 2003, *8* (4), 247.

Creswell, Julie, Kevin Draper, and Rachel Abrams, "At Nike, Revolt Led by Women Leads to Exodus of Male Executives," *The New York Times*, April 28 2018. https://www.nytimes.com/2018/04/28/business/nike-women.html (accessed January 23, 2020).

Dartmouth College, "Dartmouth College Sexual and Gender-Based Misconduct Policy," September 2019. https://sexual-respect.dartmouth.edu/policy/dartmouth-college-sexual-and-gender-based-misconduct-policy (accessed September 3, 2019).

Daughety, Andrew F. and Jennifer F. Reinganum, "A Dynamic Model of Lawsuit Joinder and Settlement," *RAND Journal of Economics*, 2011, *42*, 471–494.

Engel, Beverly, "Stop Shaming Victims of Sexual Assault for Not Reporting," *Psychology Today*, September 23 2018. https://www.psychologytoday.com/us/blog/the-compassion-chronicles/201809/stop-shaming-victims-sexual-assault-not-reporting (accessed July 31, 2019).

Farrow, Ronan, "From Aggressive Overtures to Sexual Assault: Harvey Weinsteins Accusers Tell Their Stories," *The New Yorker*, October 23 2017. https://www.newyorker.com/news/news-desk/from-aggressive-overtures-to-sexual-assault-harvey-weinsteins-accusers-tell-their-stories (accessed July 31, 2019).

Feldman, Noah, "Harvey Weinstein's Half-Conviction Is a Full Win for Prosecutors," *Bloomberg*, February 24 2020. https://www.bloomberg.com/opinion/articles/2020-02-24/harvey-weinstein-rape-verdict-half-conviction-is-full-win (accessed February 25, 2020).

Fitzgerald, Louise F., Suzanne Swan, and Karla Fischer, "Why Didn't She Just Report Him? The Psychological and Legal Implications of Women's Responses to Sexual Harassment," *Journal of Social Issues*, 1995, *51*, 117–138.

Frankel, David M., Stephen Morris, and Ady Pauzner, "Equilibrium Selection in Global Games with Strategic Complementarities," *Journal of Economic Theory*, 2003, *108*, 1–44.

Gino, Francesca, "Why It's So Hard to Speak Up Against a Toxic Culture," *Harvard Business Review*, May 21 2018.

Goldstein, Itay and Ady Pauzner, "Demand-Deposit Contracts and the Probability of Bank Runs," *The Journal of Finance*, 2005, *60* (3), 1293–1327.

__ **and Chong Huang**, "Credit Rating Inflation and Firms' Investments," 2018. SSRN Working Paper #3082428.

Griffin, Riley, Hannah Recht, and Jeff Green, "#MeToo: One Year Later," *Bloomberg*, October 5 2018. https://www.bloomberg.com/graphics/2018-me-too-anniversary/ (accessed September 19, 2019).

Heinemann, Frank, "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks: Comment," *The American Economic Review*, 2000, *90* (1), 316–318.

Hersch, Joni, "Sexual Harassment in the Workplace," *IZA World of Labor*, October 2010, *188*, 1–10.

__ , "Valuing the Risk of Workplace Sexual Harassment," *Journal of Risk and Uncertainty*, Oct 2018, *57* (2), 111–131.

Hotelling, Kathy, "Sexual harassment: A problem shielded by silence," *Journal of Counseling & Development*, 1991, *69* (6), 497–501.

Isaac, Mike, "Inside Uber's Aggressive, Unrestrained Workplace Culture," *New York Times*, February 22 2017. https://www.nytimes.com/2017/02/22/technology/uber-workplace-culture.html (accessed January 23, 2020).

Jozkowski, Kristen N, Zoe D Peterson, Stephanie A Sanders, Barbara Dennis, and Michael Reece, "Gender differences in heterosexual college students' conceptualizations and indicators of sexual consent: Implications for contemporary sexual assault prevention education," *The Journal of Sex Research*, 2014, *51* (8), 904–916.

Kelly, Liz, Jo Lovett, and Linda Regan, "A Gap or a Chasm? Attrition in Reported Rape Cases," February 2005.

Khomami, Nadia, "#MeToo: How a Hashtag Became a Rallying Cry Against Sexual Harassment," *The Guardian*, October 20 2017. https://www.theguardian.com/world/2017/oct/20/women-worldwide-use-hashtag-metoo-against-sexual-harassment (accessed July 26, 2019).

Koblin, John, "Les Moonves Put CBS on Top. Then It All Came Crashing Down.," *New York Times*, September 10 2018. https://www.nytimes.com/2018/09/10/business/media/leslie-moonves-cbs-harassment-allegations.html (accessed January 23, 2020).

Koenig, Rebecca, "How Diversity Officers Change Corporate Culture," *U.S. News & World Report*, November 21 2018. https://money.usnews.com/careers/company-culture/articles/how-diversity-officers-change-corporate-culture (accessed August 8, 2019).

Kreiss, Anthony Michael, "Defensive Glass Ceiling," *George Washington Law Review*, forthcoming, *99.*

Kwoh, Leslie, "Firms Hail New Chiefs (of Diversity)," *The Wall Street Journal*, January 5 2012. https://www.wsj.com/articles/SB10001424052970203899504577129261732884578 ((accessed August 8, 2019).

Lee, Frances Xu and Wing Suen, "Credibility of Crime Allegations," *American Economic Review: Microeconomics*, 2019.

Lisak, David, Lori Gardinier, Sarah C. Nicksa, and Ashley M. Cote, "False Allegations of Sexual Assault: An Analysis of Ten Years of Reported Cases," *Violence Against Women*, 2010, *16* (12), 1318–1334.

Lobel, Orly, "NDAs Are Out of Control. Heres What Needs to Change," *Harvard Business Review*, "January 30" 2018. https://hbr.org/2018/01/ndas-are-out-of-control-heres-what-needs-to-change.

Lonsway, Kimberly A., "Trying to Move the Elephant in the Living Room: Responding to the Challenge of False Rape Reports," *Violence Against Women*, 2010, *16* (12), 1356–1371.

Lovett, Jo and Liz Kelly, "Different Systems, Similar Outcomes? Tracking Attrition in Reported Rape Cases Across Europe," 2009. ISBN 978-0-9544803-9-4.

McDonald, Paula, "Workplace Sexual Harassment 30 Years on: A Review of the Literature," *International Journal of Management Reviews*, 2012, *14*, 1–17.

Miller, Claire Cain, "Unintended Consequences of Sexual Harassment Scandals," *New York Times*, October 9 2017. https://www.nytimes.com/2017/10/09/upshot/as-sexual-harassment-scandals-spook-men-it-can-backfire-for-women.html (accessed September 19, 2019).

Morris, Stephen and Hyun Song Shin, "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *The American Economic Review*, 1998, *88* (3), 587–597.

__ **and** __ , "Rethinking Multiple Equilibria in Macroeconomic Modeling," *NBER Macroeconomics Annual*, 2000, *15*, 139–161.

__ **and** __ , "Measuring Strategic Uncertainty," 2002. Mimeo, London School of Economics.

__ **and** __ , "Global Games: Theory and Applications," in Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in Economics and Econometrics*, Cambridge University Press, 2003.

__ **and** __ , "Coordination Risk and the Price of Debt," *European Economic Review*, 2004, *48*, 133–153.

__ **and** __ , "Liquidity Black Holes," *Review of Finance*, 2004, *8* (1), 1–18.

__ **,** __ **, and Muhamet Yildiz**, "Common Belief Foundations of Global Games," *Journal of Economic Theory*, 2016, *163*, 826–848.

Ortiz, Jorge, "Will #MeToo turn into #NotHer? Movement May Come with Unintended Workplace Consequences," *USA Today*, October 4 2018. `https://www.usatoday.com/story/news/2018/10/04/metoo-movement-unintended-career-consequences-women/1503516002/` (accessed July 24, 2019).

Owen, D. B., "A Table of Normal Integrals," *Communications in Statistics - Simulation and Computation*, 1980, *9* (4), 389–419.

Patel, Jagdish K. and Campbell B. Read, *Handbook of the Normal Distribution*, 2 ed., New York: Marcel Dekker, Inc., 1996.

Pei, Harry and Bruno Strulovici, "Crime Entanglement, Deterrence, and Witness Credibility," 2019. Working paper.

Porter, Nicole B., "The Perfect Compromise: Bridging the Gap Between At-Will Employment and Just Cause," *Nebraska Law Review*, 2008, *87* (1), 62–124.

Prasad, Vasundhara, "If Anyone Is Listening, #MeToo: Breaking the Culture of Silence Around Sexual Abuse Through Regulating Non-Disclosure Agreements and Secret Settlements," *Boston College Law Review*, 2018, *59* (7), 2507–2549.

Rikleen, Lauren Stiller, "Fear-fueled Silence, Power Imbalance Perpetuate Bad Behavior at Law Firms," *ABA Journal*, August 21 2018. `http://www.abajournal.com/voice/article/fear-fueled_silence_power_imbalance_perpetuate_bad_behavior_at_law_firms` (accessed January 23, 2020).

Roose, Kevin, "A New Diversity Monitor for the S.E.C.," *The Wall Street Journal*, January 4 2012. `https://dealbook.nytimes.com/2012/01/04/a-new-diversity-monitor-for-the-s-e-c/?searchResultPosition=18` (accessed August 8, 2019).

Rotundo, Maria, Dung-Hanh Nguyen, and Paul R. Sackett, "A Meta-Analytic Review of Gender Differences in Perceptions of Sexual Harassment," *Journal of Applied Psychology*, 2001, *86*, 914–922.

Rudy, Jesse, "What They Don't Know Won't Hurt Them: Defending Employment-at-Will in Light of Findings That Employees Believe They Possess Just Cause Protection," *Berkeley Journal of Employment & Labor Law*, 2002, *23* (2), 307–367.

Sheiber, Noam and Julie Creswell, "Sexual Harassment Cases Show the Ineffectiveness of Going to H.R.," *The New York Times*, December 12 2017.

Smith, Kyle, "A Male Backlash Against #MeToo is Brewing," *New York Post*, February 3 2018. `https://nypost.com/2018/02/03/a-male-backlash-against-metoo-is-brewing/` (accessed July 24, 2019).

Stewart, James, "Threats and Deception: Why CBS's Board Turned Against Les Moonves," *New York Times*, September 12 2018. `https://www.nytimes.com/2018/09/12/business/cbs-les-moonves-board.html` (accessed January 23, 2020).

Tan, Gillian and Katia Porzecanski, "Wall Street Rule for the #MeToo Era: Avoid Women at All Cost," *Bloomberg*, December 3 2018. https://www.bloomberg.com/news/articles/2018-12-03/a-wall-street-rule-for-the-metoo-era-avoid-women-at-all-cost (accessed July 24, 2019).

The Economist, "American Business and #MeToo," *The Economist*, September 27 2018. https://www.economist.com/business/2018/09/27/american-business-and-metoo (accessed July 31, 2019).

__ , "Why So Few Rapists Are Convicted," *The Economist*, January 4 2020. https://www.economist.com/international/2020/01/04/why-so-few-rapists-are-convicted (accessed January 20, 2020).

Timmerman, G and C Bajema, "Sexual harassment in the workplace in the European Union," *Manuscript for the European Commission for employment, industrial relations and social affairs*, 1998.

Tracy, Carol E., Terry L. Fromson, Jennifer Gentile Long, and Charlene Whitman, "Rape and sexual assault in the legal system," *National Research Council of the National Academies Panel on Measuring Rape and Sexual Assault in the Bureau of Justice Statistics Household Surveys Committee on National Statistics*, 2012. http://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_080060.pdf (accessed September 3, 2019).

U.S. Department of Justice, "Rape and Sexual Assault: Reporting to Police and Medical Attention, 1992-2000," August 2002. NCJ #194530.

__ , "Female Victims of Sexual Violence, 1994-2010," March 2013. NCJ #240655.

__ , "Sexual Assault," 2019. https://www.justice.gov/ovw/sexual-assault. (accessed September 3, 2019).

U.S. Equal Employment Opportunity Commission, "Select Task Force on the Study of Harassment in the Workplace," January 2016.

__ , "EEOC & FEPA Charges Filed Alleging Sexual Harassment, by State & Gender, FY 1997 - FY 2018," 2019. https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment_fepas_by_state.cfm. (accessed September 30, 2019).

Weber, Lauren, "After #MeToo, Those Who Report Harassment Still Risk Retaliation," *The Wall Street Journal*, December 12 2018. https://www.wsj.com/articles/after-metoo-those-who-report-harassment-still-risk-retaliation-11544643939 (accessed July 31, 2019).

Weinstein, Jonathan and Muhamet Yildiz, "Impact of Higher-order Uncertainty," *Games and Economic Behavior*, 2007, *60* (1), 200 – 212.

__ **and** __ , "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica*, 2007, *75* (2), 365–400.

Weiser, Benjamin, Ali Watkins, and Joseph Goldstein, "Jeffrey Epstein, R. Kelly and a Change in How Prosecutors Look at Sexual Assault," *The New York Times*, July 25 2012. https://nyti.ms/2YtEzxs (accessed August 9, 2019).

Willingham, AJ and Christina Maxouris, "#WhyIDidntReport: These Tweets Show Why People Don't Report Sexual Assaults," *CNN*, September 21 2018. https://www.cnn.com/2018/09/21/health/why-i-didnt-report-tweets-trnd/index.html (accessed July 31, 2019).

Appendix A Proofs of Main Propositions

We prove all statements in the text including the possibility that payoffs include α, so that:

$$\pi(r,x) = \begin{cases} x(\omega + \gamma r\alpha - (1-\gamma r)\beta) - c & \text{if } x \geq 0 \\ x\omega - c & \text{if } x < 0 \end{cases}.$$

A.1 Proof of Proposition 1

Lemmas A1, A2, and A3 set the stage from which Proposition 1 follows. The latter two lemmas are a direct application of Proposition 2.1 of Morris and Shin (2003) and Lemma A.2 of Morris and Shin (2004a), with technical modification to account for the fact that r only includes the number of agents reporting $x_i > 0$. We spell out the Lemmas for completeness and because we will use Lemma A3 in proving Proposition 5, which is not a direct application. We refer to the reader to Morris and Shin (2003) and Morris and Shin (2004a) for their proofs but also provide step-by-step proofs with the appropriate technical modifications to our setting in Online Appendix B.1.

Note that here we do not require ϵ is normally distributed, only that it has a continuous density with support over the real line. We denote by F the cumulative distribution function associated with ϵ, and f its probability density function.

A.1.1 Intermediate results

Lemma A1. *The function $\pi(r,x)$ satisfies the following:*

P1. Action monotonicity. *$\pi(r,x)$ weakly increases in r, and strictly increases in r for $x > 0$.*

P2. State monotonicity. *$\pi(r,x)$ strictly increases in x.*

P3. Unique threshold solution. *There exists a unique x^* solving $\int_0^1 \pi(r,x^*)\,dr = 0$.*

P4. Limit dominance. *There exist $\underline{x} \in \mathbb{R}$ and $\overline{x} \in \mathbb{R}$ such that: [1] $\pi(r,x) < 0$ for all $r \in [0,1]$ and $x \leq \underline{x} - \varepsilon$; and [2] $\pi(r,x) > 0$ for all $r \in [0,1]$ and $x \geq \overline{x} + \varepsilon$, for any $\varepsilon > 0$.*

P5. Continuity. *$\int_0^1 g(r)\pi(r,x)\,dr$ is continuous with respect to signal x and density g.*

Proof. [P1]. Let x be given. If $x \leq 0$, then $\partial\pi/\partial r = 0$. If $x > 0$, $\partial\pi/\partial r = x\gamma(\alpha+\beta) > 0$.

[P2]. Let $r \in [0,1]$ be given. We have:

$$\frac{\partial \pi}{\partial x}(r,x) = \begin{cases} \omega - \beta + \gamma r(\alpha+\beta) & \text{if } x \geq 0 \\ \omega & \text{if } x < 0 \end{cases} \geq \begin{cases} \omega - \beta & \text{if } x \geq 0 \\ \omega & \text{if } x < 0 \end{cases} > 0.$$

[P3]. Conjecture that there exists a $x^* > 0$ that solves $\int_0^1 \pi(r,x^*)\,dr = 0$. Using the definition of π over the positive domain,

$$\int_0^1 x^*(\omega - \beta + \gamma r(\alpha+\beta)) - c\,dr = x^*\left(\omega - \beta + \frac{\gamma}{2}(\alpha+\beta)\right) - c.$$

Evidently, $x^* = \frac{c}{\omega - \beta + \frac{\gamma}{2}(\alpha+\beta)}$ is the unique positive solution to $\int_0^1 \pi(r,x^*)\,dr = 0$.

To show that this is the unique solution, conjecture that there exists at least one other solution $x^- < 0$ that solves $\int_0^1 \pi(r,x^-)\,dr = 0$. Using the definition of π over its negative domain,

$$\int_0^1 x\omega - c\,dr = x\omega - c,$$

so $x^- = \frac{c}{\omega}$. But then $x > 0$, a contradiction.

[P4]. Fix some $\varepsilon > 0$. Reporting is a dominant strategy if $x \geq \bar{x} + \varepsilon$ where $\bar{x} \equiv \frac{c}{\omega - \beta}$ since $\pi(r,x) > 0$ for all $(r,x) \in [0,1] \times (\bar{x}, \infty)$. Not reporting is payoff dominant if $x < \underline{x} - \varepsilon$ where $\underline{x} \equiv \frac{c}{\omega + \gamma\alpha - (1-\gamma)\beta}$, since $\pi(r,x) < 0$ for all $(r,x) \in [0,1] \times (-\infty, \underline{x})$.

[P5]. To show continuity with respect to x, let probability density $g(r)$ be given. $\pi(r,x)$ is continuous over both its positive and negative domain and is also continuous at $x = 0$ for any r, with $\lim_{x \uparrow 0} \pi(r,x) = \lim_{x \downarrow 0} \pi(r,x) = \pi(r,0) = -c$. This implies $\int_0^1 g(r)\,\pi(r,x)\,dr$ is continuous in x.

To show continuity with respect to density g, fix x and let a sequence of cumulative distribution functions $G_n \to G$ be given, i.e., $G_n(r) \to G(r)$ for every r at which G is continuous. Note that, fixing x, $\pi(r,x)$ is continuous and bounded in r. Then $\int_0^1 g_n(r)\,\pi(r,x)\,dr = \int_0^1 \pi(r,x)\,dG_n(r) \to \int_0^1 \pi(r,x)\,dG(r) = \int_0^1 g(r)\,\pi(r,x)\,dr$ by the Portmanteau theorem. ■

Let $r(\theta; k)$ equal the proportion of other agents drawing $x > k$ for type θ for any $k > 0$, or $r(\theta;k) = \int_k^\infty \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) dx = 1 - F\left(\frac{k-\theta}{\sigma}\right)$. For any such $k > 0$, there is a one-to-one mapping from r into θ:

$$r(\theta;k) = 1 - F\left(\frac{k-\theta}{\sigma}\right) \quad \Leftrightarrow \quad \theta(r;k) = k - \sigma F^{-1}(1-r). \tag{A.1}$$

Recall that $\pi(r,x)$ equals the payoff gain to reporting for an agent who draws experience x when r other agents are reporting. Let $\pi^*(x,k) : \mathbb{R} \times [0,\infty) \to \mathbb{R}$ be the *expected* payoff gain from reporting when a player's signal is x and other players are playing threshold strategies of $k \geq \underline{x}$. It suffices to consider $k \geq \underline{x}$ by iterated deletion of strictly dominated strategies: from the definition of π, not reporting strictly dominates reporting for any $x < \underline{x}$ irrespective of what other agents play, so playing a cutoff of $k = \underline{x}$ strictly dominates playing a cutoff strategy of $k < \underline{x}$. The one-to-one mapping from r to θ for $k \geq \underline{x} > 0$ allows us to write:

$$\pi^*(x,k) \quad = \quad \int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) \pi\left(1 - F\left(\frac{k-\theta}{\sigma}\right), x\right) d\theta,$$

where $\frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$ is the posterior density of θ upon observing x for an agent with improper prior over θ.

Lemma A2. *The following properties hold:*

[1] $\pi^*(x,k)$ *is strictly increasing in* x*, weakly decreasing in* k *(strictly decreasing for* $x > 0$ *and* $k > 0$*), and continuous in* x *and* k*.*

[2] *The sequence* $\left\{\xi^1, \xi^2 ... \xi^n ...\right\}$ *defined as the solutions to the equations:*

$$\begin{aligned} \pi^*\left(\xi^1, 0\right) &= 0 \\ \pi^*\left(\xi^2, \xi^1\right) &= 0 \ \ldots \\ \ldots\ \pi^*\left(\xi^{n+1}, \xi^n\right) &= 0 \ \ldots \end{aligned}$$

is a well-defined increasing sequence, with $\lim_{n\to\infty} \xi^n = \xi$*, and with* ξ *being the smallest solution to* $\pi^*(\xi,\xi) = 0$*. Analogously, the sequence* $\left\{\bar{\xi}^1, \bar{\xi}^2 ... \bar{\xi}^n ...\right\}$ *defined as the solutions to:*

$$\begin{aligned} \pi^*\left(\bar{\xi}^1, \infty\right) &= 0 \\ \pi^*\left(\bar{\xi}^2, \bar{\xi}^1\right) &= 0 \ \ldots \\ \ldots\ \pi^*\left(\bar{\xi}^{n+1}, \bar{\xi}^n\right) &= 0 \ \ldots \end{aligned}$$

is a well-defined decreasing sequence, with $\lim_{n\to\infty} \bar{\xi}^n = \bar{\xi}$*, and with* $\bar{\xi}$ *being the largest solution to* $\pi^*(\xi,\xi) = 0$*. Any such solution to* $\pi^*(\xi,\xi) = 0$ *is a threshold equilibrium.*

[3] *There is a unique threshold strategy equilibrium where the threshold is given by the unique solution to* $\pi^*(\xi,\xi) = 0$*. Uniqueness is up to the action at the threshold* ξ*.*

Proof. We know $\pi(r, x)$ satisfies Properties P1-P5 in Lemma A1. By Proposition 2.1 from Morris and Shin (2003), the Lemma follows. Appendix B.1.1 contains the detailed step-by-step proof appropriately modified for our setting. ■

Lemma A3. *If s is a strategy that survives n rounds of iterated deletion of interim-strictly dominated strategies, then* $s(\xi) = \begin{cases} 0 \text{ [do not report]} & \text{if } \xi < \underline{\xi}^n \\ 1 \text{ [report]} & \text{if } \xi > \bar{\xi}^n \end{cases}$.

Proof. This Lemma follows from Lemma A.2 of Morris and Shin (2004a) and is also subsumed in Proposition 2.1 of Morris and Shin (2003). Appendix B.1.2 contains the step-by-step proof appropriately modified for our setting. ■

A.1.2 Proposition 1

Proof. For Part 1, let ξ be the unique solution to $\pi^*(\xi, \xi) = 0$ associated with the unique equilibrium in threshold strategies from Lemma A2. From Property P3 in Lemma A1, $x^* = \xi$.

For Part 2, note that, by Lemma A2, $x^* = \xi = \lim_{n\to\infty} \underline{\xi}^n = \lim_{n\to\infty} \bar{\xi}^n$, so by Lemma A3 the only strategy which survives the iterated deletion of dominated strategies is the x^*-threshold strategy. This implies that the x^*-threshold equilibrium is the globally unique equilibrium.

For Part 3, the proof of Lemma A2 Part 3 contains this derivation. ■

A.2 Proof of Proposition 2

A.2.1 Main proof

Proof. Let $H(x) \equiv x\left(\omega - \beta + \gamma\Phi(\frac{\theta - x}{\sigma})(\alpha + \beta)\right) - c$. Then $H(x)$ is the expected payoff of the marginal agent given that all agents use reporting threshold x and given the true θ. First, note that there can never be a Pareto improvement by selecting a threshold strategy around a reporting threshold that is greater than x^*, since all agents with $x \geq x''$, where x'' is defined by $\pi(r(\theta|x^*), x'') = 0$, would be worse off due to less aggregate reporting with a higher reporting threshold than x^*.

1. Let $\theta > x^*$. Note that $H(x^*) > 0$ if $\theta > x^*$. Moreover,

$$\begin{aligned}\frac{\partial H^2}{\partial x^2} &= \gamma(\alpha+\beta)\phi(\frac{\theta-x}{\sigma})(-\frac{1}{\sigma}) - (\frac{1}{\sigma})\gamma(\alpha+\beta)\phi(\frac{\theta-x}{\sigma}) - (\frac{1}{\sigma})\gamma(\alpha+\beta)(-\frac{\theta-x}{\sigma})\phi(\frac{\theta-x}{\sigma})(-\frac{1}{\sigma}) \\ &= -\frac{\gamma(\alpha+\beta)}{\sigma}\phi(\frac{\theta-x}{\sigma})\left(2 + \frac{x(\theta-x)}{\sigma}\phi(\frac{\theta-x}{\sigma})\right).\end{aligned}$$

   Since $\theta > x^*$, then $\frac{\partial H^2}{\partial x^2} < 0$ for all $x \in (\underline{x}, x^*]$. Since $H(\underline{x}) < 0$ and $H(x^*) > 0$, then this implies that $\frac{\partial H}{\partial x} > 0$ for all $x \in (\underline{x}, x^*]$.

   Let $\tilde{x}$ be the threshold such that $H(\tilde{x}) = 0$. Given that $H(\underline{x}) < 0$ and $H(x^*) > 0$, such a $\tilde{x} < x^*$ exists. Since $\frac{\partial H}{\partial x} > 0$ for all $x \in (\underline{x}, x^*]$, this $\tilde{x}$ is unique.

   The above properties of $H(x)$ imply that the maximal Pareto improvement would be achieved by directing agents to play a threshold strategy around $\tilde{x}$. Let $W(\theta|x')$ be total welfare for a given reporting threshold x': $W(\theta|x') \equiv \int_{x'}^{\infty} [x(\omega - \beta + \gamma r(\theta|x')(\alpha+\beta)) - c] f(x|\omega)dx$. Since $\frac{\partial r(\theta|x')}{\partial x'} < 0$ and

$\frac{\partial}{\partial x}[x(\omega-\beta+\gamma r(\theta|x')(\alpha+\beta))-c]>0$, then for any $x'\in[\tilde{x},x^*)$,

$$\begin{aligned}
W(\theta|x') &= \int_{x'}^{\infty}[x(\omega-\beta+\gamma r(\theta|x')(\alpha+\beta))-c]f(x|\theta)dx \\
&= \int_{x'}^{x^*}[x(\omega-\beta+\gamma r(\theta|x')(\alpha+\beta))-c]f(x|\omega)dx + \int_{x^*}^{\infty}[x(\omega-\beta+\gamma r(\theta|x')(\alpha+\beta))-c]f(x|\theta)dx \\
&> \int_{x'}^{x^*}[x(\omega-\beta+\gamma r(\theta|x')(\alpha+\beta))-c]f(x|\omega)dx + \int_{x^*}^{\infty}[x(\omega-\beta+\gamma r(\theta|x^*)(\alpha+\beta))-c]f(x|\theta)dx \\
&= \int_{x'}^{x^*}[x(\omega-\beta+\gamma r(\theta|x')(\alpha+\beta))-c]f(x|\omega)dx + W(\theta|x^*) \\
&> W(\theta|x^*),
\end{aligned}$$

and each agent with signal $x_i\in(x',x^*)$ is strictly better off and an agent with signal $x_i=x'$ is equally well off. Since $\frac{\partial H}{\partial x}>0$ for all $x\in(\underline{x},x^*]$, then clearly this Pareto improvement is maximized by directing agents to play $x'=\tilde{x}$.

Moreover, by the implicit function theorem, $\frac{\partial\tilde{x}}{\partial\sigma}>0$. By direct calculation, $\lim_{\sigma\to0}\tilde{x}=\underline{x}$.

2. Let $\theta\leq\underline{x}$. Note that $H(\theta)<0$ where $\frac{\partial H}{\partial\sigma}|_{x=\theta}=0$. Further, $H(x^*)<0$ for all finite σ, with $\lim_{\sigma\to\infty}H(x^*)=0$. Since $\lim_{\sigma\to\infty}\frac{\partial H}{\partial x}=\omega-\beta+\gamma(\alpha+\beta)>0$, then this implies $H(x)<0$ for all $x\in[\theta,x^*)$ when $\sigma\to\infty$. Thus, there exists no $\tilde{x}\in(\underline{x},x^*)$ such that $H(\tilde{x})=0$ when $\sigma\to\infty$.

   By direct differentiation, $\frac{\partial H}{\partial\sigma}>0$ for all $x>\theta$. Thus, for all $\sigma\in[0,\infty)$ $H(x)<0$. Thus, for all $\sigma\in[0,\infty)$, there exists no $\tilde{x}\in(\underline{x},x^*)$ such that $H(\tilde{x})=0$.

3. Let $\theta\in(\underline{x},x^*]$. Note that $\frac{\partial H^2}{\partial x^2}<0$ for all $x\in(\underline{x},\theta]$ when $\sigma>0$.

   Also, $\lim_{\sigma\to\infty}\frac{\partial H}{\partial x}=\omega-\beta+\gamma(\alpha+\beta)$. By direct differentiation, $\frac{\partial H}{\partial\sigma}>0$ for all $x>\theta$. When $\theta\leq x^*$, then $H(x^*)\leq0$ and $H(\theta)\leq0$ with strict equality if and only if $\theta=x^*$. This implies that $H(x)<0$ for all $x\in[\theta,x^*]$ when $\theta<x^*$, and $H(x^*)=0$ if and only if $\theta=x^*$. Thus there cannot exist a $\tilde{x}\in[\theta,x^*)$. This $\tilde{x}$ must lie in the interval $[\underline{x},\theta)$ and satisfy the following conditions:

$$\tilde{x}\left(\omega-\beta+\gamma\Phi(\frac{\theta-\tilde{x}}{\sigma})(\alpha+\beta)\right)-c=0 \tag{A.2}$$

$$\omega-\beta+\gamma(\alpha+\beta)\Phi(\frac{\theta-\tilde{x}}{\underline{\sigma}})-\frac{\tilde{x}\gamma(\alpha+\beta)\phi(\frac{\theta-\tilde{x}}{\underline{\sigma}})}{\sigma}\geq0. \tag{A.3}$$

   Since $H(\underline{x})<0$, $H(x^*)\leq0$, $\frac{\partial H}{\partial\sigma}<0$ when $x<\theta$, and $\frac{\partial H^2}{\partial x^2}<0$ for all $x\in(\underline{x},\theta)$ when $\sigma>0$, then such a $\tilde{x}<x^*$ exists and is unique when $\sigma\leq\underline{\sigma}$, where $\underline{\sigma}$ is determined by the value of σ at which $\max H(x)=0$ for $x\in(\underline{x},\theta]$. This condition is satisfied for the pair $(\underline{\tilde{x}},\underline{\sigma})$ that satisfy Equations A.4 and A.5:

$$\underline{\tilde{x}}(\omega-\beta+\gamma(\alpha+\beta)\Phi(\frac{\theta-\underline{\tilde{x}}}{\underline{\sigma}}))-c=0 \tag{A.4}$$

$$\omega-\beta+\gamma(\alpha+\beta)\Phi(\frac{\theta-\tilde{x}}{\underline{\sigma}})-\frac{\tilde{x}\gamma(\alpha+\beta)\phi(\frac{\theta-\tilde{x}}{\underline{\sigma}})}{\sigma}=0, \tag{A.5}$$

   where Equation A.4 is the requirement that $H(x',\underline{\sigma})=0$ and Equation A.5 is the requirement that $H(x',\underline{\sigma})$ is a local maximum. We know that a solution to $(\tilde{x},\underline{\sigma})$ exists for any $\theta\in(\underline{x},x^*]$ due to the following. First, $\lim_{\sigma\to0}H(\underline{x})=0$ and $\lim_{\sigma\to0}\frac{\partial H}{\partial x}>0$ for all $x\in(\underline{x},\theta)$, implying that $\lim_{\sigma\to0}H(x)>0$ for all $x\in(\underline{x},\theta)$. Second, $\lim_{\sigma\to\infty}H(x)<0$ for all $x\in(\underline{x},\theta)$. Since $\frac{\partial H}{\partial\sigma}<0$ and $\frac{\partial H^2}{\partial x^2}<0$ when $x<\theta$, then by continuity of H such a $\underline{\sigma}$ exists and is unique.

Note that if $\theta = x^*$, the explicit solution for $\underline{\sigma}$ is $\underline{\sigma} = \frac{c\gamma(\alpha+\beta)}{\sqrt{2\pi}\left(\omega-\beta+\frac{1}{2}\gamma(\alpha+\beta)\right)^2}$.

Let $\theta \leq x^*$ and $\sigma < \underline{\sigma}$. Since $H(\underline{x}) < 0$, $H(\tilde{x}) = 0$, and $\frac{\partial H^2}{\partial x^2} < 0$ when $x < \theta$, then $H(x) < 0$ for $x \in (\underline{x}, \tilde{x})$ and we can apply the analogous argument for Pareto improvement using threshold $\tilde{x}$ as in the above $\theta > x^*$ case.

Moreover, by the implicit function theorem, when such an $\tilde{x}$ exists, then $\frac{\partial \tilde{x}}{\partial \sigma} > 0$. By direct calculation, $\lim_{\sigma \to 0} \tilde{x} = \underline{x}$.

■

A.2.2 Application to any threshold equilibrium, including Proposition 5

The proof of Proposition 2 applies to any equilibrium in which $\Gamma(r) = \gamma r$ and agents use symmetric threshold strategies x^*.

The following Lemma establishes that, given payoff functions and any arbitrary sanction function, all agents play the same strategy in equilibrium. The Lemma thus implies that we can drop the "symmetric" qualifier and simply state that Proposition 2 applies to any equilibrium in which $\Gamma(r) = \gamma r$ and agents use threshold strategies x^*. This includes any threshold strategy equilibrium of Proposition 5, not just the unique equilibrium identified by Proposition 1. Section B.3 also uses this Lemma when discussing general sanction functions.

Lemma A4. *Let any sanction function be given. In any equilibrium, all agents use the same strategies. Therefore, every equilibrium is symmetric across agents.*

Proof. Suppose there is an equilibrium in which at least two agents i and j use different strategies, so there exists some x' such that agent i reports if $x_i = x'$ but agent j does not report if $x_j = x'$. Given a sanction function and profile of equilibrium reporting strategies by all agents, this results in some equilibrium sanction probability denoted p^*.

For any given agent with $x \geq 0$, her expected payoff gain from reporting is then

$$\begin{aligned} E(\pi|x) &= E\left(p^*(\omega x - c) + (1-p^*)(\omega x - \beta x - c)\right) \\ &= x(\omega - (1 - E(p^*|x))\beta) - c. \end{aligned}$$

Suppose agents i and j receive signal x'. For agent i's threshold strategy to hold in equilibrium, it must be that $E(\pi|x_i = x') > 0$. For agent j's threshold strategy to hold in equilibrium, it must be that $E(\pi|x_j = x') < 0$. But $E(p^*|x_i = x') = E(p^*|x_j = x')$, so $E(\pi|x_i = x') = E(\pi|x_j = x')$. Thus agents i and j cannot be using different equilibrium strategies in equilibrium. ■

A.3 Proof of Proposition 3

[1] Suppose an $M(\theta)$ such that there is an equilibrium in which types $\theta \leq \tilde{\theta}$ choose $a = 1$ and types $\theta > \tilde{\theta}$ all choose $a = 0$. Once the manager has selected a, an agent's decision only differs from that of Proposition 1 in her beliefs about r. All of the properties from Appendix A.1 therefore apply and we do not repeat their analogous verification here.

We first construct the agent's beliefs and reporting threshold x_S^*. Since the agent's belief about the density of θ is uniform over $(-\infty, \tilde{\theta}]$, then her improper prior must be

$$f(\theta|\theta < \tilde{\theta}) = \begin{cases} 1 & \text{for } \theta \in (-\infty, \tilde{\theta}] \\ 0 & \text{for } \theta \in (\tilde{\theta}, \infty). \end{cases}$$

Thus her posterior distribution $f(\theta|x, \theta < \tilde{\theta})$ is $f(\theta|x, \theta < \tilde{\theta}) = \frac{f(x|\theta)f(\theta|\theta<\tilde{\theta})}{\int_{-\infty}^{\infty} f(x|\theta)f(\theta|\theta<\tilde{\theta})d\theta} = \frac{f(x|\theta)}{\int_{-\infty}^{\tilde{\theta}} f(x|\theta)d\theta}$.

For an arbitrary r that corresponds to some θ where $\theta = k - \sigma F^{-1}(1-r)$, we have

$$\begin{aligned}
\Psi(r; x, k, \theta < \tilde{\theta}) &= \int_{\theta=-\infty}^{\theta} f(\theta|x, \theta < \tilde{\theta})d\theta \\
&= \int_{\theta=-\infty}^{k-\sigma F^{-1}(1-r)} \left(\frac{f(x|\theta)}{\int_{-\infty}^{\tilde{\theta}} f(x|\theta)d\theta} \right) d\theta \\
&= \frac{\int_{z=\frac{x-k}{\sigma}+F^{-1}(1-r)}^{\infty} f(z)dz}{\int_{z=\frac{x-k}{\sigma}+F^{-1}(1-\tilde{r})}^{\infty} f(z)dz} \\
&= \frac{1 - F(\frac{x-k}{\sigma} + F^{-1}(1-r))}{1 - F(\frac{x-k}{\sigma} + F^{-1}(1-\tilde{r}))}
\end{aligned}$$

where we define $\tilde{r}$ to be the reporting that corresponds to our upper bound $\tilde{\theta}$ where $\tilde{\theta} = k - \sigma F^{-1}(1-\tilde{r})$ and we perform a change of variables $z = \frac{x-\theta}{\sigma}$ and $dz = -d\theta$.

For the marginal agent who has $x = k$, we have

$$\Psi(r; x = k, k, \theta < \tilde{\theta}) = \frac{1 - F(F^{-1}(1-r))}{1 - F(F^{-1}(1-\tilde{r}))} = \frac{r}{\tilde{r}}.$$

Thus the marginal agent's density function of r is $\psi_\sigma^*(r; x = k, k, \theta < \tilde{\theta}) = \frac{1}{\tilde{r}}$ over $[0, \tilde{r}]$.

Because there is a one-to-one mapping of r and θ (because $r(\theta; k) = 1 - F(\frac{k-\theta}{\sigma})$), then the agent's reporting threshold satisfies:

$$\begin{aligned}
\pi^*(x, k, \theta < \tilde{\theta}) &= \int_{-\infty}^{\tilde{\theta}} \frac{1}{\sigma} f(\frac{x-\theta}{\sigma}) \pi \left(1 - F(\frac{k-\theta}{\sigma}), x \right) d\theta \\
&= \int_0^{\tilde{r}} \psi_\sigma^*(r; x, k, \theta < \tilde{\theta}) \pi(r, x) dr
\end{aligned}$$

For the marginal agent, this is $\pi^*(x, k, \theta < \tilde{\theta}) = \int_0^{\tilde{r}} (\frac{1}{\tilde{r}}) \pi(r, x) dr = 0$. To find the solution - denote the cutoff $x_i = k \equiv x_S^*$. Given the marginal agent's beliefs, x_S^* is given by Equation A.6:

$$\begin{aligned}
\pi^*(x, k, \theta < \tilde{\theta}) &= 0 \\
\int_{r=0}^{\tilde{r}} (\frac{1}{\tilde{r}}) \left(x_S^*[\omega + \gamma r\alpha - (1 - \gamma r)\beta] - c \right) dr &= 0 \\
x_S^* &= \frac{c}{\omega - \beta + \frac{1}{2}\gamma(\alpha + \beta)\tilde{r}}. \qquad \text{(A.6)}
\end{aligned}$$

Since x_S^* is decreasing in $\tilde{r}$, then $x_S^* > x^*$.

To characterize a sufficient condition for this form of separating equilibrium, we construct an $M(\theta)$ that guarantees that types choose $a = 1$ if and only if $\theta \leq \tilde{\theta}$. Let $M(\theta)$ be a weakly decreasing function of θ with $\lim_{\theta \to -\infty} M(\theta) > 0$ and $S > \frac{M}{\gamma}$. For any given threshold x_S^*, $-\gamma S\hat{r}(\theta)$ strictly decreases in θ. Since $M(\theta)$ is weakly decreasing in θ and $S > \frac{M}{\gamma}$, type $\theta \to \infty$ receives a strictly negative payoff from mentoring because he is sure that his $\hat{r}(\theta) = 1$. Since $\lim_{\theta \to -\infty} M(\theta) > 0$, type $\theta \to \infty$ receives a strictly positive payoff from mentoring since he is sure that his $\hat{r}(\theta) = 0$. Since $M(\theta)$ weakly decreases in θ and $-\gamma S\hat{r}(\theta)$ strictly decreases in θ, then there exists a unique $\tilde{\theta}$ such that $M(\tilde{\theta}) - \gamma S\hat{r}(\tilde{\theta}) = 0$. Thus types choose $a = 1$ if and

only if $\theta \leq \tilde{\theta}$. The above arguments also clearly rule out pooling equilibria, so this separating equilibrium is unique. As shown above, the equilibrium $\tilde{\theta}$ and x_S^* are thus given by Equations A.6 and $M(\tilde{\theta}) - \gamma S\hat{r}(\tilde{\theta}) = 0$.

[2] Suppose an $M(\theta)$ such that there is an equilibrium in which types $\theta \leq \tilde{\theta}$ choose $a = 1$ and types $\theta > \tilde{\theta}$ choose $a = 0$. Once the manager has selected a, an agent's decision only differs from that of Proposition 1 in her beliefs about r. All of the properties from Appendix A.1 therefore apply and we do not repeat their analogous verification here. Since the construction is analogous to that of Part 1 above, we provide only the key steps below for brevity.

We first construct the agent's beliefs and reporting threshold x_S^*. Since the agent's belief about the density of θ is uniform over $(-\infty, \tilde{\theta}]$, then her improper prior must be

$$f(\theta|\theta < \tilde{\theta}) = \begin{cases} 0 & \text{for } \theta \in (-\infty, \tilde{\theta}] \\ 1 & \text{for } \theta \in (\tilde{\theta}, \infty). \end{cases}$$

Thus her posterior distribution $f(\theta|x, \theta < \tilde{\theta})$ is $f(\theta|x, \theta > \tilde{\theta}) = \frac{f(x|\theta)}{\int_{\tilde{\theta}}^{\infty} f(x|\theta)d\theta}$.

For an arbitrary r that corresponds to some θ where $\theta = k - \sigma F^{-1}(1 - r)$, we have

$$\Psi(r; x, k, \theta > \tilde{\theta}) = \int_{\theta=-\infty}^{\theta} f(\theta|x, \theta > \tilde{\theta})d\theta = \frac{1 - F(\frac{x-k}{\sigma} + F^{-1}(1 - r))}{F(\frac{x-k}{\sigma} + F^{-1}(1 - \tilde{r}))}.$$

For the marginal agent who has $x = k$, we have $\Psi(r; x = k, k, \theta < \tilde{\theta}) = \frac{r}{1-\tilde{r}}$. Thus the marginal agent's density function of r is $\psi_\sigma^*(r; x = k, k, \theta < \tilde{\theta}) = \frac{1}{1-\tilde{r}}$ over $[\tilde{r}, 1]$.

Because there is a one-to-one mapping of r and θ, then the agent's reporting threshold satisfies: $\pi^*(x, k, \theta > \tilde{\theta}) = \int_{\tilde{r}}^{1} \psi_\sigma^*(r; x, k, \theta > \tilde{\theta})\pi(r, x)dr$. For the marginal agent, this is $\pi^*(x, k, \theta > \tilde{\theta}) = \int_{\tilde{r}}^{1} (\frac{1}{1-\tilde{r}})\pi(r, x)dr$. To find the solution - denote the cutoff $x_i = k \equiv x_S^*$. Given the marginal agent's beliefs, x_S^* is given by Equation A.7:

$$\pi^*(x, k, \theta > \tilde{\theta}) = 0$$

$$\int_{r=\tilde{r}}^{1} (\frac{1}{1-\tilde{r}}) \left(x_S^*[\omega + \gamma r\alpha - (1 - \gamma r)\beta] - c\right) dr = 0$$

$$x_S^* = \frac{c}{\omega - \beta + \frac{1}{2}\gamma(\alpha + \beta)(1 + \tilde{r})}. \tag{A.7}$$

Since x_S^* is decreasing in $\tilde{r}$, then $x_S^* < x^*$.

To characterize a sufficient condition for this form of separating equilibrium, we construct an $M(\theta)$ that guarantees that types choose $a = 1$ if and only if $\theta \geq \tilde{\theta}$. Let $M(\theta)$ be a weakly increasing function such that:

$$M(\theta) = \begin{cases} g(\theta) & \text{if } \theta < \theta' \\ h(\theta) & \text{if } \theta \geq \theta', \end{cases}$$

where $\theta' \in (-\infty, \infty)$, $g(\theta') < 0$, $h(\theta') > \gamma S$, and $g(\theta) : \mathbb{R} \to \mathbb{R}$ and $h(\theta) : \mathbb{R} \to \mathbb{R}$ are weakly increasing functions of θ. For any given threshold x_S^*, $-\gamma S\hat{r}(\theta)$ strictly decreases in θ. Since $g(\theta)$ is weakly increasing in θ and $g(\theta') < 0$, any $\theta < \theta'$ receives a strictly negative payoff from mentoring. Thus any type $\theta < \theta'$ would choose $a = 0$ in any equilibrium. Since $h(\theta') > \gamma S$ and $h(\theta)$ is weakly increasing in θ, any type $\theta \geq \theta'$ receives a strictly positive payoff from mentoring, even if $\hat{r}(\theta) = 1$. Thus types $\theta \geq \theta'$ would choose $a = 1$ in any equilibrium. Therefore the unique separating equilibrium is that equilibrium such that types choose $a = 1$ if and only if $\theta \geq \theta'$. As shown above, the equilibrium reporting threshold x_S^* is thus given by Equation A.7 where $\tilde{r} = \Phi(\frac{\theta' - x_S^*}{\sigma})$ and $\tilde{\theta} = \theta'$.

A.4 Proof of Proposition 4

We first show that the unique perfect Bayesian equilibrium is an equilibrium in which types $\theta \leq \tilde{\theta}$ choose $a = 1$ and types $\theta > \tilde{\theta}$ all choose $a = 0$.

[1] If $m - \gamma S \geq 0$, the unique equilibrium is $a(\theta) = 1$ for all θ.

Since the probability of investigation is γr, the maximum expected sanction that a manager can pay is γS. If $m - \gamma S \geq 0$, then no type has an incentive to select out. In this case, it is weakly better for any type to mentor agents even if he were to incur the maximal expected sanction. Thus, the agents' equilibrium reporting strategy is given by $s(x_i, 1)$ using threshold x^*, as defined in Proposition 1.

[2] If $m - \gamma S < 0$: we rule out all of the following forms of equilibria.

[*a*] No pooling on $a(\theta) = 0$ for all θ: Suppose that all types pool on $a = 0$, so each types's payoff is 0. We know that it is strictly dominant for an agent to not report if $\pi(r = 1) = x_i(\omega - \beta + \gamma(\alpha + \beta)) - c < 0$, which is true whenever $x_i < \frac{c}{\omega - \beta + \gamma(\alpha+\beta)}$. This means that the maximal off-equilibrium reporting is if all agents report when $x_i \geq k'$ where $k' = \frac{c}{\omega - \beta + \gamma(\alpha+\beta)}$. Suppose a manager of type θ' deviates to $a = 1$. Then the highest probability of sanction is $\gamma r' = \gamma(1 - \Phi(\frac{k' - \theta'}{\sigma})) = \gamma\Phi(\frac{\theta' - k'}{\sigma})$. So he will deviate if $m - \gamma\Phi(\frac{\theta' - k'}{\sigma})(S) > 0$, where r' is increasing in θ' and $\lim_{\theta' \to -\infty} r' = 0$. Thus, there exists some type θ' sufficiently small that he would prefer to deviate to $a = 1$ rather than pool on $a = 0$. Intuitively, there always exists a type good enough that he is quite sure he would not send high enough x_i signals to be punished.

[*b*] No pooling on $a(\theta) = 1$ for all θ: Suppose all types pool on $a = 1$, which implies that agents use threshold x^* as in Proposition 1. Since any type's payoff is $m - \gamma\hat{r}(\theta)S$ and $\lim_{\theta \to \infty} \hat{r}(\theta) = 1$, then there must exist some type θ' sufficiently large that he prefers to deviate to $a = 0$ rather than pool on $a = 1$ if $S < \frac{m}{\gamma}$. Intuitively, there always exists a type bad enough that he is quite sure he would send high enough x_i signals to be punished. Thus if S is sufficiently high, this type does not mentor.

This leaves a unique equilibrium in which types $\theta \leq \tilde{\theta}$ choose $a = 1$ and types $\theta > \tilde{\theta}$ all choose $a = 0$. Given this selection, the proof of Part 1 of Proposition 3 applies. The equilibrium is the solution to Equations A.8 and A.9, which are the indifference conditions for the marginal agent and manager, respectively:

$$\int_{r=0}^{\tilde{r}} (\frac{1}{\tilde{r}}) \left(x_S^*[\omega + \gamma r\alpha - (1 - \gamma r)\beta] - c\right) dr = 0 \tag{A.8}$$

$$m - \gamma\Phi(\frac{\tilde{\theta} - x_S^*}{\sigma})S = 0. \tag{A.9}$$

Note that Equation A.9 implies that $\tilde{r} = \frac{m}{\gamma S}$, where $m - \gamma S < 0$ implies that $\frac{m}{\gamma S} < 1$. Using $\tilde{r} = \frac{m}{\gamma S}$ in Equation A.8, we obtain $x_S^* = \frac{c}{\omega - \beta + (\frac{m}{2S})(\alpha+\beta)}$ and $\tilde{\theta} = \frac{c}{\omega - \beta + (\frac{m}{2S})(\alpha+\beta)} + \sigma\Phi^{-1}(\frac{m}{\gamma S})$. Clearly, there is no incentive for any type to deviate given x_S^* and $\tilde{\theta}$.

A.4.1 Corollary 4.1

[1] By direct differentiation, $\frac{\partial \tilde{r}}{\partial S} = -\frac{m}{\gamma S^2} < 0$ and $\frac{\partial x_S^*}{\partial S} = \frac{-c\gamma(\alpha+\beta)\frac{\partial \tilde{r}}{\partial S}}{2\left(\omega - \beta + \frac{1}{2}\tilde{r}\gamma(\alpha+\beta)\right)^2} > 0$.

[2] What is $\frac{\partial \tilde{\theta}}{\partial S}$? $\frac{\partial \tilde{\theta}}{\partial S} = \frac{\partial x_S^*}{\partial S} + \frac{\partial \tilde{r}}{\partial S}\frac{1}{\phi\left(\frac{\tilde{\theta} - x_S^*}{\sigma}\right)}$ with $\frac{\partial \tilde{r}}{\partial S} = -\frac{m}{\gamma S^2} < 0$ and $\frac{\partial x_S^*}{\partial S} = \frac{-c\gamma(\alpha+\beta)\frac{\partial \tilde{r}}{\partial S}}{2\left(\omega - \beta + \frac{1}{2}\tilde{r}\gamma(\alpha+\beta)\right)^2} > 0$. Combining:

$$\frac{\partial \tilde{\theta}}{\partial S} = \left(\sigma\frac{1}{\phi\left(\frac{\tilde{\theta} - x_S^*}{\sigma}\right)} - \frac{c\gamma\left(\alpha + \beta\right)\frac{\partial \tilde{r}}{\partial S}}{2\left(\omega - \beta + \frac{1}{2}\tilde{r}\gamma\left(\alpha + \beta\right)\right)^2}\right)\frac{\partial \tilde{r}}{\partial S}. \tag{A.10}$$

So $\frac{\partial \tilde{\theta}}{\partial S} < 0$ if and only if:

$$\begin{aligned}
\sigma \frac{1}{\phi\left(\frac{\tilde{\theta}-x_S^*}{\sigma}\right)} &> \frac{c\gamma(\alpha+\beta)}{2\left(\omega-\beta+\frac{1}{2}\tilde{r}\gamma(\alpha+\beta)\right)^2} \\
\sigma &> \phi\left(\frac{\tilde{\theta}-x_S^*}{\sigma}\right)\frac{c\gamma(\alpha+\beta)}{2\left(\omega-\beta+\frac{1}{2}\tilde{r}\gamma(\alpha+\beta)\right)^2} \\
\sigma &> \phi\left(\Phi^{-1}(\tilde{r})\right)\frac{c\frac{1}{2}\gamma(\alpha+\beta)}{\left(\omega-\beta+\frac{1}{2}\tilde{r}\gamma(\alpha+\beta)\right)^2}. \qquad \text{(A.11)}
\end{aligned}$$

Thus a sufficient condition for $\frac{\partial \tilde{\theta}}{\partial S} < 0$ is $\sigma > \frac{1}{\sqrt{2\pi}}\frac{c\frac{1}{2}\gamma(\alpha+\beta)}{(\omega-\beta)^2}$. When $S = \frac{m}{\gamma}$, $\tilde{r} = 1$ so from Equation A.10 we have $\frac{\partial \tilde{\theta}}{\partial S} < 0$ when $S = \frac{m}{\gamma}$. As $S \to \infty$, $\tilde{r} \to 0$ so $x_S^* \to \underline{x}$ and $\tilde{\theta} \to -\infty$. From Equation A.11, we have $\frac{\partial \tilde{\theta}}{\partial S} < 0$ as $S \to \infty$.
[3] A sufficient condition for $\frac{\partial \tilde{\theta}}{\partial S} > 0$ when $\tilde{r} = 1/2$ is $\sigma < \frac{1}{\sqrt{2\pi}}\frac{c\gamma(\alpha+\beta)}{2\left(\omega-\beta+\frac{1}{4}\gamma(\alpha+\beta)\right)^2}$.

A.5 Proof of Proposition 5

Suppose the prior belief of θ with density $p(\theta)$ is normally distributed with mean y and standard deviation τ, and experiences are $x = \theta + \sigma\epsilon$ where $\epsilon \sim N(0,1)$. Let $t = 1/\tau^2$ and $u = 1/\sigma^2$ denote the precisions of the prior and x. The posterior density $f(\theta \mid x)$ is a normal density with:

$$\begin{aligned}
\text{mean } \lambda &= \frac{\sigma^2 y + \tau^2 x}{\sigma^2+\tau^2} = \frac{ty+ux}{t+u}, \\
\text{standard deviation } v &= \sqrt{\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} = \frac{1}{\sqrt{t+u}}, \\
\text{precision } h &= t+u = \frac{\sigma^2+\tau^2}{\sigma^2\tau^2}.
\end{aligned}$$

Define $r(\theta, k)$ as in the proof of Proposition 1. Recall that we can consider $k \geq 0$ and that for such k there is a one-to-one map of r into θ: $r(\theta; k) = 1 - \Phi\left(\frac{k-\theta}{\sigma}\right)$ so $\theta(r; k) = k - \sigma\Phi^{-1}(1-r)$, where $\Phi(.)$ denotes the normal CDF; let $\phi(.)$ denote the normal PDF. The expected payoff gain of reporting for agent drawing x when other agents are playing threshold strategies around $k \geq 0$ when $x > 0$ equals:

$$\begin{aligned}
\pi^*(x,k) &= \int_{-\infty}^{\infty} f(\theta \mid x)\,\pi(r(\theta;k), x)\,d\theta \\
&= x\left(\omega - \beta + \gamma(\alpha+\beta)\int_{-\infty}^{\infty} f(\theta \mid x)\left(1-\Phi\left(\frac{k-\theta}{\sigma}\right)\right)d\theta\right) - c. \qquad \text{(A.12)}
\end{aligned}$$

For $x \leq 0$, the payoff gain equals $\pi^*(x,k) = x\omega - c < 0$.

A.5.1 Intermediate results

Lemma A5. *Any solution x^* to $\pi^*(x,x) = 0$ satisfies the implicit equation:*

$$x^* = \frac{c}{\omega - \beta + \gamma(\alpha+\beta)\Phi\left(\frac{y-x^*}{\kappa}\right)},$$

where we drop the I subscript on x_I^* for notational brevity.

Proof. We have $f(\theta \mid x) = \frac{1}{v}\phi\left(\frac{\theta-\lambda}{v}\right)$. Then Equation A.12 becomes:

$$\pi^*(x,k) = x\left(\omega - \beta + \gamma(\alpha+\beta)\int_{-\infty}^{\infty}\frac{1}{v}\phi\left(\frac{\theta-\lambda}{v}\right)\left(1-\Phi\left(\frac{k-\theta}{\sigma}\right)\right)d\theta\right) - c.$$

We know the following general relationship (Patel and Read, 1996, p.36):

$$\int_{-\infty}^{\infty}\Phi(a+bz)\,\phi(z)\,dx = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right).$$

Letting $z = \frac{\theta-\lambda}{v}$ and $a + bz = \frac{k-\lambda}{\sigma} - \frac{v}{\sigma}z$, we have:

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{1}{v}\phi\left(\frac{\theta-\lambda}{v}\right)\left(1-\Phi\left(\frac{k-\theta}{\sigma}\right)\right)d\theta &= 1-\int_{-\infty}^{\infty}\phi(z)\,\Phi\left(\frac{k-\lambda}{\sigma}-\frac{v}{\sigma}z\right)dz\\ &= 1-\Phi\left(\frac{\frac{k-\lambda}{\sigma}}{\sqrt{1+\frac{v^2}{\sigma^2}}}\right)\\ &= \Phi\left(\frac{\lambda-k}{\sqrt{\sigma^2+v^2}}\right).\end{aligned}$$

For any agent who draws x, the payoff gain is then:

$$\begin{aligned}\pi^*(x,k) &= x\left(\omega-\beta+\gamma(\alpha+\beta)\,\Phi\left(\frac{\lambda-k}{\sqrt{\sigma^2+v^2}}\right)\right)-c \text{ if } x>0,\\ &= x\omega - c \text{ if } x\le 0.\end{aligned} \tag{A.13}$$

For the marginal agent with $x = k$, we have $\lambda - k = \lambda - x = \frac{v^2}{\tau^2}(y-x)$. Then:

$$\begin{aligned}\Phi\left(\frac{\lambda-x}{\sqrt{\sigma^2+v^2}}\right) &= \Phi\left(\frac{v^2}{\tau^2}\frac{y-x}{\sqrt{\sigma^2+v^2}}\right)\\ &= \Phi\left(\frac{y-x}{\kappa}\right) \text{ for } \kappa \equiv \frac{\tau^2}{v^2}\sqrt{\sigma^2+v^2}.\end{aligned}$$

The payoff gain for the marginal agent (assuming $x > 0$, which we verify) then equals:

$$\pi^*(x,x) = 0 = x\left(\omega-\beta+\gamma(\alpha+\beta)\,\Phi\left(\frac{y-x}{\kappa}\right)\right)-c,$$

which gives us the implicit equation for x^*. Note that any solution must be positive, as required. ■

Lemma A6. *There exists at least one symmetric threshold strategy equilibrium $x_I^* \in (\underline{x}, \overline{x})$ where agents report for $x \ge x_I^*$ and do not report for $x < x_I^*$.*

Proof. From Lemma A5, any potential equilibria must solve $G(x^*) = 0$, where:

$$\begin{aligned}G(x) = \pi^*(x,x) &= x(\omega-\beta+\gamma E(x)(\alpha+\beta))-c,\\ E(x) &= \Phi\left(\frac{y-x}{\kappa}\right).\end{aligned}$$

For notational brevity, we drop the I subscript in x_I^*.

First, we claim there exists a $x^* \in (\underline{x}, \overline{x})$ that is a solution to $G(x^*) = 0$ where $G(x^*)$ is increasing. Observe that $E(x)$ is continuous, which implies $G(x)$ is continuous. Notice that $G\left(\frac{c}{\omega-\beta}\right) > 0$, and

$G\left(\frac{c}{\omega+\gamma\alpha-(1-\gamma)\beta}\right) < 0$, and $\frac{c}{\omega-\beta} > \frac{c}{\omega+\gamma\alpha-(1-\gamma)\beta}$. By the intermediate value theorem, there exists at least one solution $G(x^*) = 0$, where $G(x^*)$ is increasing.

Second, we show that any such solution x^* constitutes a symmetric threshold equilibrium. Recall from Equation A.13 in Lemma A5 that the payoff from reporting when an agent's signal is $x > 0$, conditional on other players playing threshold strategies around k, equals:

$$\pi^*(x,k) = x\left(\omega-\beta+\gamma(\alpha+\beta)\Phi\left(\frac{\frac{\sigma^2 y+\tau^2 x}{\sigma^2+\tau^2}-k}{\sqrt{\sigma^2+v^2}}\right)\right)-c.$$

This is a strictly increasing function in x. Since $G(x^*) = \pi^*(x^*,x^*) = 0$, we have $\pi^*(x,x^*) > 0$ for $x > x^*$ and $\pi^*(x,x^*) < 0$ for $x < x^*$, so that a threshold strategy around x^* is a best response to other players playing the same threshold strategy. If $x < 0 < x^*$, $\pi^*(x,x^*) < 0$. ■

Lemmas A7 and A8 substitute for Lemma A2 in the informed prior case.

Lemma A7. *The following properties hold:*

[1] $\pi^*(x,k)$ *is strictly increasing in* x*, weakly decreasing in* k *(strictly decreasing for* $x > 0$*), and continuous in both* x *and* k*. Furthermore, for any* $k \geq 0$*,* $\pi^*(x,k)$ *maps onto* $\mathbb{R}$*.*

[2] *Let* ξ *solve* $\pi^*(\xi,\xi) = 0$*. The sequence* $\left\{\underline{\xi}^1, \underline{\xi}^2 ... \underline{\xi}^n ...\right\}$ *defined as the solutions to the equations:*

$$\begin{aligned} \pi^*\left(\underline{\xi}^1, 0\right) &= 0 \\ \pi^*\left(\underline{\xi}^2, \underline{\xi}^1\right) &= 0 \ \ldots \\ \ldots\ \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) &= 0 \ \ldots \end{aligned}$$

is a well-defined increasing sequence, bounded from above by ξ *and below by 0, with* $\lim_{n\to\infty}\underline{\xi}^n = \underline{\xi}$*, where* $\underline{\xi}$ *is the smallest solution to* $\pi^*(\xi,\xi)$*. Analogously, the sequence* $\left\{\bar{\xi}^1, \bar{\xi}^2 ... \bar{\xi}^n ...\right\}$ *defined as the solutions to:*

$$\begin{aligned} \pi^*\left(\bar{\xi}^1, \infty\right) &= 0 \\ \pi^*\left(\bar{\xi}^2, \bar{\xi}^1\right) &= 0 \ \ldots \\ \ldots\ \pi^*\left(\bar{\xi}^{n+1}, \bar{\xi}^n\right) &= 0 \ \ldots \end{aligned}$$

is a well-defined decreasing sequence, bounded from below by $\xi > 0$*, with* $\lim_{n\to\infty}\bar{\xi}^n = \bar{\xi}$*, where* $\bar{\xi}$ *is the largest solution to* $\pi^*(\xi,\xi)$*.*

Proof. [1] These properties are evident from Equation A.13 in Lemma A5.

[2] Property [1] implies that all $\underline{\xi}^n$ and $\bar{\xi}^n$ are well-defined. We know that not reporting is dominant for $x < \underline{x}$, so $\pi^*(x,0) < 0$ for all $x < \underline{x}$. But we also know that $\pi^*(x,0) > 0$ for all $x > \bar{x}$. Define $\underline{\xi}^0 \equiv 0$. By continuity in x, there exists at least one solution x with $\pi^*\left(x, \underline{\xi}^0\right) = 0$, where $x \in [\underline{x}, \bar{x}]$. Call $\underline{\xi}^1$ the smallest such solution. Note that $\bar{\xi}^1 > 0$. Furthermore, note that $\underline{\xi}^1 < \xi \in (\underline{x}, \bar{x})$: if not, then $0 = \pi^*\left(\underline{\xi}^1, 0\right) \geq \pi^*(\xi,0) > \pi^*(\xi,\xi) = 0$, a contradiction.

To show that $\underline{\xi}^n$ is an increasing sequence, proceed by induction. Our starting point is to show that, because $\pi^*\left(\underline{\xi}^1, 0\right) = \pi^*\left(\underline{\xi}^2, \underline{\xi}^1\right) = 0$, we have $\underline{\xi}^1 < \underline{\xi}^2$. To see why, proceed by contradiction. Suppose $\underline{\xi}^1 \geq \underline{\xi}^2$. Then $\pi^*\left(\underline{\xi}^1, 0\right) \geq \pi^*\left(\underline{\xi}^2, 0\right)$ because π^* is increasing in x, but $\pi^*\left(\underline{\xi}^2, 0\right) > \pi^*\left(\underline{\xi}^2, \underline{\xi}^1\right)$ because π^* is decreasing in k. Thus, $\pi^*\left(\underline{\xi}^1, 0\right) > \pi^*\left(\underline{\xi}^2, \underline{\xi}^1\right)$, a contradiction. Note that $\underline{\xi}^2 < \xi$: if not, then $0 = \pi^*\left(\underline{\xi}^2, \underline{\xi}^1\right) \geq \pi^*\left(\xi, \underline{\xi}^1\right) > \pi^*(\xi,\xi) = 0$, a contradiction.

The inductive hypothesis is that $\underline{\xi}^{n-1} < \underline{\xi}^n$ with $\underline{\xi}^n < \xi$; we claim $\underline{\xi}^n < \underline{\xi}^{n+1}$ with $\underline{\xi}^{n+1} < \xi$. Proceed again by contradiction. By definition, $\pi^*\left(\underline{\xi}^n, \underline{\xi}^{n-1}\right) = \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right)$. Suppose that $\underline{\xi}^n \geq \underline{\xi}^{n+1}$. Then $\pi^*\left(\underline{\xi}^n, \underline{\xi}^{n-1}\right) > \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^{n-1}\right)$ because π^* is increasing in x, but $\pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^{n-1}\right) > \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right)$ because π^* is decreasing in k. Thus, $\pi^*\left(\underline{\xi}^n, \underline{\xi}^{n-1}\right) > \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right)$, a contradiction. Note that $\underline{\xi}^{n+1} < \xi$: if not, then $0 = \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) \geq \pi^*\left(\xi, \underline{\xi}^n\right) > \pi^*(\xi,\xi) = 0$, a contradiction.

Because $\{\underline{\xi}^n\}$ is a bounded increasing sequence, there exists a $\underline{\xi}$ with $\lim_{n\to\infty}\underline{\xi}^n = \underline{\xi}$. Note that $\lim_{n\to\infty}\pi^*\left(\underline{\xi}^{n+1},\underline{\xi}^n\right) = 0$ so by construction and continuity of π^*, we must have $\pi^*\left(\underline{\xi},\underline{\xi}\right) = 0$ and that $\underline{\xi}$ is the smallest such solution to $\pi^*\left(\xi,\xi\right) = 0$.

An analogous argument works identically to show that $\left\{\bar{\xi}^n\right\}$ is a bounded decreasing sequence, that there exists a $\bar{\xi}$ with $\lim_{n\to\infty}\bar{\xi}^n = \bar{\xi}$, and that $\bar{\xi}$ is the largest such solution to $\pi^*\left(\xi,\xi\right) = 0$. ■

Lemma A8. *Uniqueness of equilibrium (allowing for either strategy to be played at x^*):*

[a] The equilibrium x^* is a unique threshold equilibrium if:

$$\kappa > \frac{1}{\sqrt{2\pi}}\frac{c\gamma\left(\alpha+\beta\right)}{\left(\omega-\beta\right)^2}.$$

[b] Whenever $y = \frac{c}{\omega-\beta+\frac{1}{2}\gamma(\alpha+\beta)}$, a sufficient condition for non-unique equilibria is:

$$\kappa < \frac{1}{\sqrt{2\pi}}\frac{\gamma\left(\alpha+\beta\right)c}{\left(\omega-\beta+\frac{1}{2}\gamma\left(\alpha+\beta\right)\right)^2}.$$

[c] For $y\to\infty$, there is a unique threshold equilibrium x^* with $x^*\to\underline{x}$. For $y\to-\infty$, there is a unique threshold equilibrium x^* with $x^*\to\bar{x}$.

[d] If there is a unique equilibrium in threshold strategies, then the equilibrium strategy is the only strategy that satisfies the iterated deletion of strictly dominated strategies, and in particular, the unique threshold strategy equilibrium is the globally unique equilibrium.

Proof. Because $G\left(\underline{x}\right) < G\left(\bar{x}\right)$ with at least one solution in between where G is increasing from Lemma A6, and because G is differentiable for all x, a necessary and sufficient condition for uniqueness is that $G'\left(x^*\right) > 0$ for all $x^*\in\left(\underline{x},\bar{x}\right)$ such that $G\left(x^*\right) = 0$. We have

$$\begin{aligned} G'\left(x\right) &= \gamma\left(\alpha+\beta\right)xE'\left(x\right) + \left(\omega-\beta+\gamma\left(\alpha+\beta\right)E\left(x\right)\right) && \text{(A.14)}\\ &= \omega-\beta+\gamma\left(\alpha+\beta\right)\left(xE'\left(x\right)+E\left(x\right)\right), && \text{(A.15)}\end{aligned}$$

and also:

$$G'\left(x\right) = \gamma\left(\alpha+\beta\right)xE'\left(x\right)+\frac{c}{x} \quad \text{(A.16)}$$

Substituting in $E'\left(x\right) = -\frac{1}{\kappa}\phi\left(\frac{y-x}{\kappa}\right)$ into Equation A.16, a necessary and sufficient condition for uniqueness is, for all solutions x^*:

$$\gamma\left(\alpha+\beta\right)\left(x^*\right)^2\phi\left(\frac{y-x^*}{\kappa}\right)\frac{1}{\kappa} < c. \quad \text{(A.17)}$$

Recall from Lemmas A5 and A6 that any solution to $G(x)$ constitutes a symmetric threshold equilibrium and that there exists at least one such equilibrium. We now provide conditions under which such an equilibrium is unique or not unique.

[a] Using the fact that $\phi\left(z\right) < \frac{1}{\sqrt{2\pi}}$ and $x^* < \bar{x} = \frac{c}{\omega-\beta}$, a sufficient condition for uniqueness from Equation A.17 is:

$$\gamma\left(\alpha+\beta\right)\frac{1}{\sqrt{2\pi}}\frac{c}{\left(\omega-\beta\right)^2}\frac{1}{\kappa} < 1 \quad\Leftrightarrow\quad \frac{1}{\sqrt{2\pi}}\frac{c\gamma\left(\alpha+\beta\right)}{\left(\omega-\beta\right)^2} < \kappa.$$

[b] For $y = \frac{c}{\omega-\beta+\frac{1}{2}\gamma(\alpha+\beta)}$, note that $x^* = y$, $E\left(y\right) = 1/2$, $E'\left(y\right) = -\frac{1}{\kappa}\frac{1}{\sqrt{2\pi}}$, and we have the following

expression for $G'(y)$ from Equation A.14:

$$G'(y) = \gamma(\alpha+\beta) y E'(y) + (\omega - \beta + \gamma(\alpha+\beta) E(y))$$

But $y = \frac{c}{\omega-\beta+\frac{1}{2}\gamma(\alpha+\beta)}$, so $G'(y) < 0$ if and only if $\frac{1}{\sqrt{2\pi}} \frac{\gamma(\alpha+\beta)c}{\left(\omega-\beta+\frac{1}{2}\gamma(\alpha+\beta)\right)^2} > \kappa$.

[c] For $y \to \infty$, $\lim_{y\to\infty} E(x) = 1$, so $\lim_{y\to\infty} G(x) = x(\omega - \beta + \gamma(\alpha+\beta)) - c$. By continuity, any solutions x^* are arbitrarily close to $\underline{x}$.

For $y \to -\infty$, $\lim_{y\to-\infty} E(x) = 0$, so $\lim_{y\to-\infty} G(x) = x(\omega - \beta) - c$. By continuity, any solutions x^* are arbitrarily close to $\overline{x}$.

To show uniqueness in both cases, observe from Equation A.15 that:

$$\begin{aligned} \lim_{y\to\infty} G'(x) &= \omega - \beta + \gamma(\alpha+\beta) > 0, \\ \lim_{y\to-\infty} G'(x) &= \omega - \beta > 0. \end{aligned}$$

Continuity of $G'(x)$ in y and x implies $G'(x) > 0$ for arbitrarily large (positive or negative) y, in particular $G'(x^*) > 0$ for any solution x^*.

[d] If $G(x)$ has a unique solution x^*, then from Lemma A7 we have $x^* = \xi = \underline{\xi} = \overline{\xi}$. By Lemma A3, the only strategy which survives the iterated deletion of dominated strategies is the x^*-threshold strategy. This implies that the x^*-threshold equilibrium is the globally unique equilibrium. ■

Lemma A9. *The marginal agent has beliefs over r summarized by the cumulative distribution function $\Phi\left(\frac{t}{\sqrt{t+u}}(x-y) + \frac{\sqrt{t+u}}{\sqrt{u}}\Phi^{-1}(r)\right)$, with expectation $E_I^*[r] = \Phi\left(\frac{y-x^*}{\kappa}\right)$.*

Proof. Given an agent's signal x, what is her assessment of the cumulative distribution function of r when others are playing cutoff strategies around k, $\Psi(\tilde{r}; x, k)$? We can follow the same logic as in Lemma A2. For any $\tilde{r}$, the probability that $r < \tilde{r}$ equals the probability that $\theta < k - \sigma F^{-1}(1-\tilde{r})$. In words, the probability $\Psi(\tilde{r}; x, k) \equiv \Pr(r < \tilde{r} \mid x)$ that the true proportion of players reporting is less than $\tilde{r}$ equals the probability that the true θ satisfies $r(\theta; k) = 1 - F\left(\frac{k-\theta}{\sigma}\right) < \tilde{r}$, or equivalently that θ is such that fewer than $\tilde{r}$ players observe a signal greater than k; in turn, this equals the probability that the true θ is less than $k - \sigma F^{-1}(1-\tilde{r})$, integrated against the conditional density $f(\theta \mid x)$. With some abuse of notation,

$$\begin{aligned} \Psi(r; x, k) &= \int_{-\infty}^{k-\sigma\Phi^{-1}(1-r)} f(\theta \mid x)\, d\theta \\ &= \int_{-\infty}^{k-\sigma\Phi^{-1}(1-r)} \frac{1}{v}\phi\left(\frac{\theta-\lambda,}{v}\right) d\theta \\ &= \int_{-\infty}^{z=\frac{k-\lambda}{v}-\frac{\sigma}{v}\Phi^{-1}(1-r)} \phi(z)\, dz \text{ for } z = \frac{\theta-\lambda}{v},\ dz = \frac{1}{v}d\theta \\ &= \Phi\left(\frac{k-\lambda}{v} - \frac{\sigma}{v}\Phi^{-1}(1-r)\right). \end{aligned}$$

For the marginal agent with $x = k$,

$$\frac{k-\lambda}{v} = \frac{(t+u)x - ty - ux}{\sqrt{t+u}} = \frac{t}{\sqrt{t+u}}(x-y).$$

Combining this insight with $\sigma/v = \sqrt{t+u}/\sqrt{u}$ yields:

$$\Psi(r; x, x) = \Phi\left(\frac{t}{\sqrt{t+u}}(x-y) + \frac{\sqrt{t+u}}{\sqrt{u}}\Phi^{-1}(r)\right),$$

where we use $\Phi^{-1}(r) = -\Phi^{-1}(1-r)$ in the derivation.

Let $\psi(r; x, x)$ denote the probability density function associated with $\Psi(r; x, x)$. Because there is a one-to-one mapping of r and θ, we can re-write $\pi^*(x, k)$ as $\pi^*(x, k) = \int_0^1 \psi(r; x, k)\, \pi(r, x)\, dr$, so for the marginal agent (assuming $x > 0$, which we verify):

$$\pi^*(x, x) = \int_0^1 \psi(r; x, x) \left[x\left(\omega - \beta + \gamma r\left(\alpha + \beta\right)\right) - c\right] dr.$$

Because the marginal agent must be indifferent between reporting and not reporting, the equilibrium condition is then $\pi^*(x, x) = 0$. This gives the implicit function:

$$x^* = \frac{c}{\omega - \beta + \gamma E_I^*[r]\left(\alpha + \beta\right)},$$

where $E_I^*[r] = \int_0^1 r\, \psi(r; x, x)\, dr$. But then by Lemma A5, $E_I^*[r] = \Phi\left(\frac{y - x^*}{\kappa}\right)$. ■

A.5.2 Proposition 5

Proof. Part 1 follows from Lemma A6. Part 2 follows from Lemma A8. Part 3 follows from Lemma A9. ■

A.6 Proof of Proposition 6

1. It is straightforward to verify that Lemmas A1, A2, and A3 hold when $\delta \in [0, \omega - \beta)$. In particular, $\pi_H(r, x)$ satisfies action monotonicity (P1) for all $\overline{r} \in [0, 1]$ only if $x\delta - c < x\left(\omega - (1 - \gamma(0))\,\beta\right) - c$, which implies $\delta \in [0, \omega - \beta)$. The bound on the lower dominance region is still $\underline{x} = \frac{c}{\omega - \beta(1-\gamma)}$. When $\delta \in [0, \omega - \beta)$, the bound on the upper dominance region is $\overline{x} = \frac{c}{\delta}$.

   Thus, all previous results on existence, uniqueness, and comparative statics of other parameters in Sections 1-3 continue to apply with straightforward modification, and proofs of all results incorporate the extended payoff structure, since $\Gamma(r)$ is still weakly increasing in r. The symmetric equilibrium threshold strategy given by Equation 9 is the solution to $\int_0^1 \pi(x_H^*, r)dr = 0$. It is straightforward to verify that $x_H^* \in (\underline{x}, \overline{x})$ when $\delta \in [0, \omega - \beta)$.

   Differentiating x_H^* with respect to $\overline{r}$ yields

$$\begin{aligned} \frac{\partial x_H^*}{\partial \overline{r}} &= \frac{-c[\delta - \omega + \beta(1 - \gamma\overline{r})]}{\left(\delta\overline{r} + \omega(1 - \overline{r}) - \beta[(1 - \overline{r}) - \frac{1}{2}\gamma - \frac{1}{2}\gamma\overline{r}^2]\right)^2} \\ &> 0 \text{ if } \delta < \omega - \beta(1 - \gamma\overline{r}) \\ &= 0 \text{ if } \delta = \omega - \beta(1 - \gamma\overline{r}) \\ &< 0 \text{ if } \delta > \omega - \beta(1 - \gamma\overline{r}). \end{aligned} \tag{A.18}$$

   Thus, $\frac{\partial x_H^*}{\partial \overline{r}} > 0$ for all $\overline{r} \in [0, 1]$ when $\delta \in [0, \omega - \beta]$. This implies that the release threshold $\overline{r}_{min}$ that minimizes x_H^* is $\overline{r}_{min} = 0$.

2. If $\delta \in [\omega - \beta, \beta]$, then $\pi_H(r, x)$ fails action monotonicity (P1) for some $\overline{r} \in [0, 1]$, and a symmetric equilibrium threshold strategy is not guaranteed. However, when such an equilibrium exists, it must be the unique solution to $\int_0^1 \pi(x_H^*, r)dr = 0$, which is characterized by Equation 9. For all $\delta \in [0, \omega]$, the bound on the lower dominance region becomes $\underline{x} = \min\{\frac{c}{\omega - \beta(1-\gamma)}, \frac{c}{\delta}\}$. When $\delta > \omega - \beta(1 - \gamma\overline{r})$, the bound on the upper dominance region satisfies $\pi(\overline{x}^H, \overline{r}) = 0$, which is $\overline{x}^H = \frac{c}{\omega - \beta(1-\gamma\overline{r})}$. Thus for all $\delta \in [0, \omega]$, then the bound on the upper dominance region becomes $\overline{x} = \max\{\frac{c}{\omega - \beta(1-\gamma\overline{r})}, \frac{c}{\delta}\}$.

   We can verify that $x_H^* \in [\underline{x}, \overline{x}]$ for all $\overline{r} \in [0, 1]$.

Given Equation A.18 above, it is straightforward to show that $\frac{\partial x_H^*}{\partial \overline{r}}|_{\overline{r}=0} < 0$ for all $\delta \in (\omega - \beta, \omega]$ and that there is a unique $\overline{r}_{min} \in (0, 1]$ that minimizes x_H^* for all $\delta \in (\omega - \beta, \omega]$. Since $\frac{\partial x_H^*}{\partial \overline{r}} = 0$ when $\delta = \omega - \beta(1 - \gamma\overline{r})$, then $\overline{r}_{min} = \min\{\frac{\delta-(\omega-\beta)}{\beta\gamma}, 1\}$ and $x_H^*(\overline{r}_{min}) < x_H^*(0)$. In particular, $\overline{r}_{min} \in (0, 1)$ when $\delta \in (\omega-\beta, \omega-\beta(1-\gamma))$ and $\overline{r}_{min} = 1$ when $\delta \in [\omega-\beta(1-\gamma), \omega]$. Moreover, note that $\frac{\partial \overline{r}_{min}}{\partial \delta} > 0$ if $\overline{r}_{min} \in (0, 1)$.

When does such an equilibrium exist? Suppose all agents play threshold strategies. As before, the key quantity to understand is $\pi^*(x, k)$, which is the expected payoff gain to reporting for a player who has observed signal x and anticipates that all the other players will not report if they observe signals less than k.

Define $r(\theta; k)$ as the proportion of other agents drawing $x > k$ for type θ, or $r(\theta; k) = \int_k^\infty \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) dx = 1 - F\left(\frac{k-\theta}{\sigma}\right)$ where F is the cumulative distribution function of x. With improper priors, $f(\theta \mid x) = \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$. For simplicity, assume Gaussian noise at this point. Observe that:

$$\int_{-\infty}^{z} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) d\theta = \int_{-\infty}^{z} \frac{1}{\sigma} f\left(\frac{\theta-x}{\sigma}\right) d\theta \tag{A.19}$$

$$= F\left(\frac{z-x}{\sigma}\right). \tag{A.20}$$

Therefore we can write the expected payoffs as:

$$\begin{aligned} \pi^*(x, k) &= x\delta F\left(\frac{k-x}{\sigma} - F^{-1}(1-\overline{r})\right) \\ &\quad + x(\omega-\beta)\left(1 - F\left(\frac{k-x}{\sigma} - F^{-1}(1-\overline{r})\right)\right) \\ &\quad + x\beta\gamma \int_{k-\sigma F^{-1}(1-\overline{r})}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)\left(1 - F\left(\frac{k-\theta}{\sigma}\right)\right) d\theta \\ &\quad - c. \end{aligned} \tag{A.21}$$

$$\begin{aligned} &= x\delta F\left(\frac{k-x}{\sigma} - F^{-1}(1-\overline{r})\right) \\ &\quad + x(\omega-\beta)\left(1 - F\left(\frac{k-x}{\sigma} - F^{-1}(1-\overline{r})\right)\right) \\ &\quad + x\beta\gamma \int_{k-\sigma F^{-1}(1-\overline{r})}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) d\theta \\ &\quad - x\beta\gamma\left(\int_{k-\sigma F^{-1}(1-\overline{r})}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) F\left(\frac{k-\theta}{\sigma}\right) d\theta\right) \\ &\quad - c. \end{aligned} \tag{A.22}$$

Consider the "little integral":

$$\text{little integral} = \int_{k-\sigma F^{-1}(1-\overline{r})}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) d\theta.$$

Use a u-substitution:

$$\begin{aligned}
z &= \frac{x-\theta}{\sigma},\ dz = -\frac{1}{\sigma}d\theta \\
\frac{k-\theta}{\sigma} &= \frac{x-\theta}{\sigma} - \frac{x-k}{\sigma} = z - \frac{x-k}{\sigma} = \frac{k-x}{\sigma} + z \\
\underline{\theta} &= k - \sigma F^{-1}\left(1-\overline{r}\right) \\
\underline{z} &= \frac{x-\underline{\theta}}{\sigma} = \frac{x-k+\sigma F^{-1}\left(1-\overline{r}\right)}{\sigma} = \frac{x-k}{\sigma} + F^{-1}\left(1-\overline{r}\right) \\
\overline{\theta} &= \infty \\
\overline{z} &= \frac{x-\underline{\theta}}{\sigma} = -\infty,
\end{aligned}$$

and $F\left(-x\right) = 1 - F\left(x\right)$ to obtain:

$$\text{little integral} = 1 - F\left(\frac{k-x}{\sigma} - F^{-1}\left(1-\overline{r}\right)\right).$$

Now consider the "big integral":

$$\text{big integral} = \int_{k-\sigma F^{-1}(1-\overline{r})}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) F\left(\frac{k-\theta}{\sigma}\right) d\theta.$$

Use the same u-substitution to get:

$$\text{big integral} = \int_{-\infty}^{\frac{x-k}{\sigma}+F^{-1}(1-\overline{r})} f\left(z\right) F\left(\frac{k-x}{\sigma} + z\right) dz.$$

From Owen (1980),

$$\int_{-\infty}^{h} F'\left(x\right) F\left(\frac{l-\rho x}{\sqrt{1-\rho^2}}\right) dx = BvN(h, l; \rho),$$

where BvN is the bivariate normal cumulative distribution function. We can apply the formula to the "big integral" with:

$$\begin{aligned}
\rho =& -\frac{1}{\sqrt{2}}, \\
l =& \frac{1}{\sqrt{2}}\frac{k-x}{\sigma}, \\
h =& \frac{x-k}{\sigma} + F^{-1}\left(1-\overline{r}\right),
\end{aligned}$$

so:

$$\text{big integral} = BvN\left(\frac{x-k}{\sigma} + F^{-1}\left(1-\overline{r}\right), \frac{1}{\sqrt{2}}\frac{k-x}{\sigma}; \frac{-1}{\sqrt{2}}\right).$$

The expected payoff then becomes:

$$\begin{aligned}\pi^*(x,k) =& x(\omega-\beta(1-\gamma)) - x(\omega-\beta(1-\gamma)-\delta)F\left(\frac{k-x}{\sigma}-F^{-1}(1-\overline{r})\right)\\ &- x\beta\gamma\left[BvN\left(\frac{x-k}{\sigma}+F^{-1}(1-\overline{r}),\frac{1}{\sqrt{2}}\frac{k-x}{\sigma};\frac{-1}{\sqrt{2}}\right)\right]-c,\end{aligned}$$

which is Equation 10 in the text.

Note that x_H^* is a candidate threshold equilibrium since $\pi^*(x^*, k = x_H^*) = 0$. To see this, recall the definition of x_H^*:

$$x_H^* = \frac{c}{\overline{r}\delta + (1-\overline{r})(\omega-\beta) + \frac{1}{2}\beta\gamma\left(1-\overline{r}^2\right)}.$$

Plug this into the formula for $\pi^*(x,k)$:

$$\begin{aligned}\pi^*(x_H^*,x_H^*) =& x_H^*(\omega-\beta(1-\gamma)) - x^*(\omega-\beta(1-\gamma)-\delta)\overline{r}\,[\text{using } F^{-1}(1-r)=-F^{-1}(r)]\\ &- x_H^*\beta\gamma\left[BvN\left(F^{-1}(1-\overline{r}),0;\frac{-1}{\sqrt{2}}\right)\right]-c\\ =& x_H^*(\omega-\beta(1-\gamma)) - x^*(\omega-\beta(1-\gamma)-\delta)\overline{r} - x_H^*\beta\gamma\frac{1}{2}(1-\overline{r})^2 - c\\ =& x_H^*\left[\delta\overline{r}+(1-\overline{r})(\omega-\beta)+\frac{1}{2}\beta\gamma\left(1-\overline{r}^2\right)\right]-c\\ =& 0.\end{aligned}$$

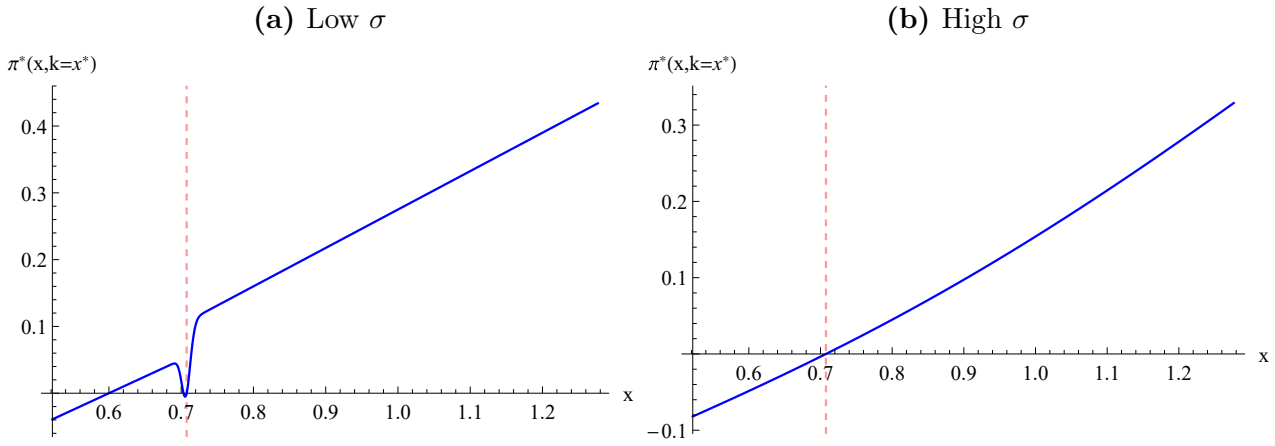
The step solving BvN above is as follows. Letting $T(a,h)$ denote Owen's T-function (Owen, 1980):

$$\begin{aligned}BvN\left(F^{-1}(1-\overline{r}),0;\frac{-1}{\sqrt{2}}\right) =& T\left(F^{-1}(1-\overline{r}),0\right)+T(0,\infty)-T\left(F^{-1}(1-\overline{r}),\frac{\frac{1}{\sqrt{2}}F^{-1}(1-\overline{r})}{F^{-1}(1-\overline{r})\frac{1}{\sqrt{2}}}\right)\\ &-T(0,\infty)+F\left(F^{-1}(1-\overline{r})\right)F(0), \text{ using Owen (1980) Equation 3.2}\\ =& -T\left(F^{-1}(1-\overline{r}),1\right)+\frac{1}{2}(1-\overline{r}), \text{ using Owen (1980) Equation 2.1}\\ =& -F\left(F^{-1}(1-\overline{r})\right)\left[1-F\left(F^{-1}(1-\overline{r})\right)\right]/2+\frac{1}{2}(1-\overline{r}),\\ &\text{using Owen (1980) Equation 2.3}\\ =& \frac{1}{2}(1-\overline{r})^2.\end{aligned}$$

Thus, the symmetric threshold equilibrium exists at x_H^* when $\pi^*(x, x_H^*) < 0$ for all $x < x_H^*$ and $\pi^*(x, x_H^*) > 0$ for all $x > x_H^*$, and it must be unique (among symmetric threshold equilibria) since there are no other crossings.

In the limit as $\sigma \to \infty$, Equation 10 satisfies the single-crossing condition. To see this, note that

Figure 4: Expected payoff gain function $\pi^*(x, k = x_H^*)$. This figure plots $\pi^*(x, k = x_H^*)$ for $x \in (\underline{x}, \overline{x})$ with $\{\beta, c, \omega, \gamma, \delta, \overline{r}\} = \{0.85, 0.3, 1, 0.5, 0.5, 0.2\}$. The vertical line denotes x_H^*. In Panel (a), σ=0.005 and x_H^* is not an equilibrium. In Panel (b), $\sigma = 0.5$ and x_H^* is the unique symmetric threshold equilibrium.



$\pi^*(x, k)$ is a continuous function. Consider σ large. Then for any x and k:

$$\begin{aligned}
\pi^*(x,k) =& x(\omega - \beta(1-\gamma)) - x(\omega - \beta(1-\gamma) - \delta) F\left(0 - F^{-1}(1-\overline{r})\right) \\
& - x\beta\gamma \left[BvN\left(0 + F^{-1}(1-\overline{r}), \frac{1}{\sqrt{2}} 0; \frac{-1}{\sqrt{2}}\right)\right] - c \\
=& x(\omega - \beta(1-\gamma)) - x(\omega - \beta(1-\gamma) - \delta)\overline{r} \\
& - x\beta\gamma \left[BvN\left(F^{-1}(1-\overline{r}), 0; \frac{-1}{\sqrt{2}}\right)\right] - c \\
=& x\left[\delta\overline{r} + (1-\overline{r})(\omega - \beta) + \frac{1}{2}\beta\gamma\left(1-\overline{r}^2\right)\right] - c, \text{ using same logic as above.}
\end{aligned}$$

But then $\pi^*(x, k) < 0$ for $x < x_H^*$ and $\pi^*(x, k) > 0$ for $x > x_H^*$ by the definition of x_H^*. Figure 4 illustrates when the equilibrium does and does not exist for low and high σ, consistent with the limit result.

A.7 Proof of Proposition 7

[1] The α^F such that $x^* = x^F$ must satisfy $\omega + \frac{1}{2}\gamma\alpha - (1 - \frac{1}{2}\gamma)\beta = \omega$. Thus $\alpha^F = \beta\left(\frac{2-\gamma}{\gamma}\right)$.

The α^E such that $x^* = x^E$ must satisfy $\omega + \frac{1}{2}\gamma\alpha - (1 - \frac{1}{2}\gamma)\beta = \omega - (1-\gamma)\beta$. Thus $\alpha^E = \beta$.

[2] The α^F such that $x_S^* = x^F$ must satisfy $\omega - \beta + \left(\frac{m}{2S}\right)(\alpha + \beta) = \omega$. Thus $\alpha^F = \beta\left(\frac{2S-m}{m}\right)$.

The α^E such that $x_S^* = x^E$ must satisfy $\omega - \beta + \left(\frac{m}{2S}\right)(\alpha + \beta) = \omega - (1-\gamma)\beta$. Thus $\alpha^E = \beta\left(\frac{2S\gamma - m}{m}\right)$.

[3] The α^F such that $x_{I,R}^* = x^F$ must satisfy $\omega - \beta + \gamma(\alpha + \beta)\Phi(\frac{y - x_{I,R}^*}{\kappa}) = \omega$. Thus $\alpha^F = \beta\left(\frac{1 - \gamma\Phi(\frac{\omega y - c}{\omega\kappa})}{\gamma\Phi(\frac{\omega y - c}{\omega\kappa})}\right)$.

Clearly, α^F and α^E are unique and positive for any given β.

To obtain the comparative statics:

[a] By direct differentiation, $\frac{\partial \alpha^F}{\partial y} = -\left(\frac{\beta}{\gamma\kappa}\right)\phi\left(\frac{\omega y - c}{\omega\kappa}\right)\left[\Phi\left(\frac{\omega y - c}{\omega\kappa}\right)\right]^{-2} < 0$.

[b] By direct differentiation:

$$\frac{\partial \alpha^F}{\partial \tau} = \left(\frac{\beta}{\gamma\kappa^2}\right)\phi\left(\frac{\omega y - c}{\omega\kappa}\right)\left[\Phi\left(\frac{\omega y - c}{\omega\kappa}\right)\right]^{-2}\left(y - \frac{c}{\omega}\right)\left(\frac{\partial \kappa}{\partial \tau}\right).$$

Since $\frac{\partial \kappa}{\partial \tau} > 0$, this implies that $\frac{\partial \alpha^F}{\partial \tau} \begin{cases} > 0 \text{ if } y > \frac{c}{\omega} \\ = 0 \text{ if } y = \frac{c}{\omega} \\ < 0 \text{ if } y < \frac{c}{\omega}. \end{cases}$

The α^E such that $x^*_{I,R} = x^E$ must satisfy $\omega - \beta + \gamma(\alpha + \beta)\Phi(\frac{y - x^*_{I,R}}{\kappa}) = \omega - (1-\gamma)\beta$. Thus $\alpha^E = \beta\left(\frac{1 - \Phi\left(\frac{y(\omega-(1-\gamma)\beta)-c}{\kappa(\omega-(1-\gamma)\beta)}\right)}{\Phi\left(\frac{y(\omega-(1-\gamma)\beta)-c}{\kappa(\omega-(1-\gamma)\beta)}\right)}\right)$.

Appendix B Online Appendix

B.1 Supplemental detail for main proofs

We will need the following standard result. For any two densities g and h for a random variable z that ranges over $(-\infty, \infty)$, we say *g stochastically dominates h* ($g \succeq h$) if $G(z) = \int_{-\infty}^{z} g(s)\, ds \leq \int_{-\infty}^{z} h(s)\, ds = H(z)\ \forall z$. If g stochastically dominates h, then for any weakly increasing function u, the expected value of u under the former weakly exceeds that of the latter.

Lemma B1. *If $g \succeq h$ and $u(z)$ is a weakly increasing differentiable function of z, then:*

$$\int_{-\infty}^{\infty} u(z)\, g(z)\, dz \geq \int_{-\infty}^{\infty} u(z)\, h(z)\, dz.$$

If $u(z)$ is a weakly decreasing function of z, then the inequality is reversed.

B.1.1 Proof of Lemma A2

Proof. We know $\pi(r, x)$ satisfies Properties P1-P5 in Lemma A1.

[1] Stochastic dominance arguments and Property P2 implies $\pi^*(x, k)$ is strictly increasing in x. To see this, implement a change of variables with $z = -\theta$. Note that $f_z(z) = \frac{1}{\sigma} f\left(\frac{x+z}{\sigma}\right)$, and $F_z(z; x_1) = \int_{-\infty}^{z} \frac{1}{\sigma} f\left(\frac{x_1+z}{\sigma}\right) dz < \int_{-\infty}^{z} \frac{1}{\sigma} f\left(\frac{x_2+z}{\sigma}\right) dz = F_z(z; x_2)$ for any z and $x_1 < x_2$. That is, z under x_1 stochastically dominates z under x_2, because the former has more probability mass "shifted to the right." Observe that:

$$\begin{aligned}
\pi^*(x, k) &= \int_{\theta=-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) \pi\left(1 - F\left(\frac{k-\theta}{\sigma}\right), x\right) d\theta \\
&= \int_{z=-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x+z}{\sigma}\right) \pi\left(1 - F\left(\frac{k+z}{\sigma}\right), x\right) dz
\end{aligned}$$

Note that $\pi\left(1 - F\left(\frac{k+z}{\sigma}\right), x_1\right)$ is a differentiable weakly decreasing function of z, and increasing function of x. Therefore, under Lemma B1, for $x_1 < x_2$,

$$\begin{aligned}
\pi^*(x_1, k) &= \int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x_1+z}{\sigma}\right) \left[\pi\left(1 - F\left(\frac{k+z}{\sigma}\right), x_1\right)\right] dz \\
&\leq \int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x_2+z}{\sigma}\right) \left[\pi\left(1 - F\left(\frac{k+z}{\sigma}\right), x_1\right)\right] dz \\
&< \int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x_2+z}{\sigma}\right) \left[\pi\left(1 - F\left(\frac{k+z}{\sigma}\right), x_2\right)\right] dz \\
&= \pi^*(x_2, k).
\end{aligned}$$

Property P1 implies $\pi^*(x, k)$ is weakly decreasing in k, and Property P5 implies $\pi^*(x, k)$ is continuous in x and k. Note that for $x > 0$, $\pi^*(x, k)$ is strictly decreasing in k for $k > 0$.

[2] We show that $\{\underline{\xi}^n\}$ and $\{\bar{\xi}^n\}$ are well-defined increasing and decreasing sequences, respectively, through induction. From Property P4, we know that not reporting is dominant for $x < \underline{x}$, so $\pi^*(x, 0) < 0$ for all $x < \underline{x}$. But we also know that $\pi^*(x, 0) > 0$ for all $x > \bar{x}$. Define $\underline{\xi}^0 \equiv 0$ and $\bar{\xi}^0 \equiv \infty$. By continuity in x, there exists at least one solution x with $\pi^*\left(x, \underline{\xi}^0\right) = 0$, where $x \in [\underline{x}, \bar{x}]$. Call $\underline{\xi}^1$ the smallest such solution. Define $\bar{\xi}^1 \in [\underline{x}, \bar{x}]$ analogously to be the largest such solution with $\pi^*\left(x, \bar{\xi}^0\right) = 0$. Note that $\underline{\xi}^0 < \underline{\xi}^1 < \bar{\xi}^1 < \bar{\xi}^0$; if the inside inequality did not hold, then $0 = \pi^*\left(\underline{\xi}^1, \underline{\xi}^0\right) \geq \pi^*\left(\bar{\xi}^1, \underline{\xi}^0\right) > \pi^*\left(\bar{\xi}^1, \bar{\xi}^0\right) = 0$, a contradiction.

Our starting point for the induction is as follows. Given $\underline{\xi}^1$ and $\bar{\xi}^1$ with $\underline{\xi}^0 < \underline{\xi}^1 < \bar{\xi}^1 < \bar{\xi}^0$, $\pi^*\left(\underline{\xi}^1, \underline{\xi}^0\right) = 0$, and $\pi^*\left(\bar{\xi}^1, \bar{\xi}^0\right) = 0$, we claim there exists a smallest solution $\underline{\xi}^2$ of $\pi\left(\underline{\xi}^2, \underline{\xi}^1\right) = 0$ and a largest solution $\bar{\xi}^2$

of $\pi\left(\bar{\xi}^2,\bar{\xi}^1\right)=0$, and that $\xi^1<\xi^2<\bar{\xi}^2<\bar{\xi}^1$. We know $\pi^*\left(\xi^1,\xi^0\right)=0>\pi^*\left(\xi^1,\xi^1\right)$, and $\pi^*\left(\bar{\xi}^1,\bar{\xi}^0\right)=0<\pi^*\left(\bar{\xi}^1,\xi^1\right)$. Note for the latter inequality that $\bar{\xi}^0>\bar{\xi}^1>\xi^1$. By continuity, there exists a smallest solution $\xi^2\in\left(\xi^1,\bar{\xi}^1\right)$ with $\pi^*\left(\xi^2,\xi^1\right)=0$. Analogously, we know $\pi^*\left(\xi^1,\xi^0\right)=0>\pi^*\left(\xi^1,\bar{\xi}^1\right)$, and $\pi^*\left(\bar{\xi}^1,\bar{\xi}^0\right)=0<\pi^*\left(\bar{\xi}^1,\xi^1\right)$; by continuity there exists a largest solution $\bar{\xi}^2\in\left(\xi^1,\bar{\xi}^1\right)$ with $\pi^*\left(\bar{\xi}^2,\bar{\xi}^1\right)=0$. Note that $\xi^1<\xi^2<\bar{\xi}^2<\bar{\xi}^1$; if the inside inequality did not hold, then $0=\pi^*\left(\xi^2,\xi^1\right)\geq\pi^*\left(\bar{\xi}^2,\xi^1\right)>\pi^*\left(\bar{\xi}^2,\bar{\xi}^1\right)=0$, a contradiction.

The inductive hypothesis is that, given ξ^n and $\bar{\xi}^n$ with $\xi^{n-1}<\xi^n<\bar{\xi}^n<\bar{\xi}^{n-1}$, $\pi^*\left(\xi^n,\xi^{n-1}\right)=0$, and $\pi^*\left(\bar{\xi}^n,\bar{\xi}^{n-1}\right)=0$, there exists a smallest solution ξ^{n+1} of $\pi\left(\xi^{n+1},\xi^n\right)=0$ and a largest solution $\bar{\xi}^{n+1}$ of $\pi\left(\bar{\xi}^{n+1},\bar{\xi}^n\right)=0$, and that $\xi^n<\xi^{n+1}<\bar{\xi}^{n+1}<\bar{\xi}^n$. We know $\pi^*\left(\xi^n,\xi^{n-1}\right)=0>\pi^*\left(\xi^n,\xi^n\right)$, and $\pi^*\left(\bar{\xi}^n,\bar{\xi}^{n-1}\right)=0<\pi^*\left(\bar{\xi}^n,\xi^n\right)$. Note for the latter inequality that $\bar{\xi}^{n-1}>\bar{\xi}^n>\xi^n$. By continuity, there exists a smallest solution $\xi^{n+1}\in\left(\xi^n,\bar{\xi}^n\right)$ with $\pi^*\left(\xi^{n+1},\xi^n\right)=0$. Similarly, we know $\pi^*\left(\xi^n,\xi^{n-1}\right)=0>\pi^*\left(\xi^n,\bar{\xi}^n\right)$, and $\pi^*\left(\bar{\xi}^n,\bar{\xi}^{n-1}\right)=0<\pi^*\left(\bar{\xi}^n,\bar{\xi}^n\right)$; by continuity there exists a largest solution $\bar{\xi}^{n+1}\in\left(\xi^n,\bar{\xi}^n\right)$ with $\pi^*\left(\bar{\xi}^{n+1},\bar{\xi}^n\right)=0$. Note that $\xi^n<\xi^{n+1}<\bar{\xi}^{n+1}<\bar{\xi}^n$; if the inside inequality did not hold, then $0=\pi^*\left(\xi^{n+1},\xi^n\right)\geq\pi^*\left(\bar{\xi}^{n+1},\xi^n\right)>\pi^*\left(\bar{\xi}^{n+1},\bar{\xi}^n\right)=0$, a contradiction.

Note that $\{\xi^n\}$ is bounded from above by construction. Because it is also an increasing sequence, there exists a ξ with $\lim_{n\to\infty}\xi^n=\xi$. Note that $\lim_{n\to\infty}\pi^*\left(\xi^{n+1},\xi^n\right)=0$ so by construction and continuity of π^*, we must have $\pi^*(\xi,\xi)=0$ and that ξ is the smallest such solution to $\pi^*(\xi,\xi)=0$. Analogously, there exists a $\bar{\xi}$ with $\lim_{n\to\infty}\xi^n=\xi$ and $\pi^*\left(\bar{\xi},\bar{\xi}\right)=0$ and that $\bar{\xi}$ is the smallest such solution to $\pi^*(\xi,\xi)=0$. This shows, among other things, that there exists at least one threshold equilibrium ξ. One can see that any such solution ξ is an equilibrium because $x_1<\xi<x_2$ implies $\pi^*(x_1,\xi)<\pi^*(\xi,\xi)=0<\pi^*(x_2,\xi)$.

[3] Note that we can write:

$$\pi^*(x,k) \quad = \quad \int_{-\infty}^{\infty}\psi(r;x,k)\,\pi(r,x)\;dr$$

Given the agent's signal x, what is her assessment of the cumulative distribution function of r, $\Psi(\tilde{r};x,k)$? For any $\tilde{r}$, the probability that $r<\tilde{r}$ equals the probability that $\theta<k-\sigma F^{-1}(1-\tilde{r})$. In words, the probability $\Psi(\tilde{r};x,k)\equiv\Pr(r<\tilde{r}\mid x)$ that the true proportion of players reporting is less than $\tilde{r}$ equals the probability that the true θ satisfies $r(\theta;k)=1-F\left(\frac{k-\theta}{\sigma}\right)<\tilde{r}$, or equivalently that θ is such that fewer than $\tilde{r}$ players observe a signal greater than k; in turn, this equals the probability that the true θ is less than $k-\sigma F^{-1}(1-\tilde{r})$, integrated against the conditional density $f(\theta\mid x)$. With some slight abuse of notation, we thus have:

$$\begin{aligned}
\Psi(r;x,k) &= \int_{-\infty}^{k-\sigma F^{-1}(1-r)} f(\theta\mid x)\;d\theta \\
&= \int_{-\infty}^{k-\sigma F^{-1}(1-r)} \frac{1}{\sigma}f\left(\frac{x-\theta}{\sigma}\right)\;d\theta \\
&= \int_{z=\frac{x-k}{\sigma}+F^{-1}(1-r)}^{\infty} f(z)\;dz \text{ for } z=\frac{x-\theta}{\sigma},\; dz=-\frac{1}{\sigma}d\theta \\
&= 1-F\left(\frac{x-k}{\sigma}+F^{-1}(1-r)\right).
\end{aligned}$$

For the marginal agent, $x=k$, so $\Psi(r;x,x)=r$. The density function of r is then $\psi(r;x,x)=1$ over $[0,1]$.

But then $\pi^*(x,x)=\int_0^1\pi(r,x)\;dr$. By Property P3, there is exactly one such solution ξ. From [2], it must be that $\xi=\underline{\xi}=\bar{\xi}$ and that this is the unique threshold equilibrium. ■

B.1.2 Proof of Lemma A3

Proof. Let Σ be the strategy profile used by all players other than i, and denote by $\tilde{\pi}^i(\xi, \Sigma)$ the payoff gain of reporting for player i, conditional on ξ when other players play Σ. We proceed by induction.

If everyone with $x > 0$ reports, player i's payoff is the highest, and if no one reports, player i's payoff is the lowest. Therefore:

$$\pi^*(\xi, \infty) \leq \tilde{\pi}^i(\xi, \Sigma) \leq \pi^*(\xi, 0).$$

From the definition of $\underline{\xi}^1$ and monotonicity of $\pi^*(x, k)$ in x,

$$\xi < \underline{\xi}^1 \Rightarrow \text{for any } \Sigma,\ \tilde{\pi}^i_\sigma(\xi, \Sigma) \leq \pi^*(\xi, 0) < \pi^*\left(\underline{\xi}^1, 0\right) = 0.$$

In words, not-reporting strictly dominates reporting ($\tilde{\pi}^i_\sigma(\xi, \Sigma) < 0$) whenever $\xi < \underline{\xi}^1$, irrespective of other players' strategies. Similarly, from the definition of $\bar{\xi}^1$ and monotonicity,

$$\xi > \bar{\xi}^1 \Rightarrow \text{for any } \Sigma,\ \tilde{\pi}^i_\sigma(\xi, \Sigma) \geq \pi^*(\xi, \infty) > \pi^*\left(\bar{\xi}^1, \infty\right) = 0.$$

In words, reporting strictly dominates not reporting ($\tilde{\pi}^i_\sigma(\xi, \Sigma) > 0$) whenever $\xi > \bar{\xi}^1$, irrespective of other players' strategies. Thus, if $s(\xi)$ survives the first round of deletion of dominated strategies, we must have:

$$s(\xi) = \begin{cases} 0 \text{ [do not report]} & \text{if } \xi < \underline{\xi}^1 \\ 1 \text{ [report]} & \text{if } \xi > \bar{\xi}^1 \end{cases}.$$

The inductive hypothesis is that if $s(\xi)$ survives the n-th round of deletion of dominated strategies, we must have:

$$s(\xi) = \begin{cases} 0 \text{ [do not report]} & \text{if } \xi < \underline{\xi}^n \\ 1 \text{ [report]} & \text{if } \xi > \bar{\xi}^n \end{cases}$$

Let S^n denote the set of strategies that survives this n-rounds of deletion. Our claim is that if player i faces a strategy profile Σ^n consisting of those drawn from S^n, then the set of strategies that survives the next round of deletion of dominated strategies S^{n+1} satisfies:

$$s(\xi) = \begin{cases} 0 \text{ [do not report]} & \text{if } \xi < \underline{\xi}^{n+1} \\ 1 \text{ [report]} & \text{if } \xi > \bar{\xi}^{n+1} \end{cases}.$$

If everyone else is playing a $\underline{\xi}^n$-threshold strategy (reporting for $\xi > \underline{\xi}^n$), player i's payoff is maximized. Therefore:

$$\xi < \underline{\xi}^{n+1} \Rightarrow \text{for any } \Sigma,\ \tilde{\pi}^i_\sigma(\xi, \Sigma^n) \leq \pi^*(\xi, \underline{\xi}^n) < \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) = 0,$$

so that not-reporting strictly dominates reporting ($\tilde{\pi}^i_\sigma(\xi, \Sigma^n) < 0$) whenever $\xi < \underline{\xi}^{n+1}$, irrespective of other players' strategies. Conversely, if everyone else is playing a $\bar{\xi}^n$-threshold strategy (reporting for $\xi > \bar{\xi}^n$), player i's payoff is minimized. Therefore:

$$\xi > \bar{\xi}^{n+1} \Rightarrow \text{for any } \Sigma,\ \tilde{\pi}^i_\sigma(\xi, \Sigma^n) \geq \pi^*(\xi, \underline{\xi}^n) > \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) = 0,$$

so that reporting strictly dominates not-reporting ($\tilde{\pi}^i_\sigma(\xi, \Sigma^n) > 0$) whenever $\xi > \underline{\xi}^{n+1}$, irrespective of other players' strategies, from which the claim follows. ■

B.2 Distributions with bounded support

The key result that needs revisiting with bounded support is Lemma A2. We consider the case where ϵ has bounded support and is possibly asymmetric.

Suppose ϵ has bounded support with cumulative distribution function (CDF) F and probability density function (PDF) f. In particular, suppose ϵ has CDF representation:

$$F_\epsilon(\epsilon) = \begin{cases} 0 & \epsilon < -\underline{l} \\ \tilde{F}_\epsilon(\epsilon) & \epsilon \in \left[-\underline{l}, \bar{l}\right] \\ 1 & \epsilon > \bar{l} \end{cases},$$

where $\tilde{F}_\epsilon(\epsilon)$ is weakly increasing and has $\tilde{F}_\epsilon(-\underline{l}) = 0$ and $\tilde{F}_\epsilon(\underline{l}) = 1$, and $\underline{l}, \bar{l} > 0$. The PDF representation is:

$$f_\epsilon(\epsilon) = \begin{cases} 0 & \epsilon < -\underline{l} \\ \tilde{f}_\epsilon(\epsilon) & \epsilon \in \left[-\underline{l}, \bar{l}\right] \\ 1 & \epsilon > \bar{l} \end{cases},$$

for a density function $\tilde{f}_\epsilon$ that is continuous, positive and integrates to 1 over $\left[-\underline{l}, \bar{l}\right]$. Other than requiring ϵ to have zero mean, we place no other restrictions on $\tilde{f}$.

The PDF for $x_i = \theta + \sigma\epsilon_i$ given θ is then:

$$f_x(x \mid \theta) = \begin{cases} 0 & x < \theta - \sigma\underline{l} \\ \tilde{f}_x(x \mid \theta) & x \in \left[\theta - \sigma\underline{l}, \theta + \sigma\bar{l}\right] \\ 0 & x > \theta + \sigma\bar{l} \end{cases},$$

for:

$$\tilde{f}_x(x \mid \theta) = \frac{1}{\sigma}\tilde{f}_\epsilon\left(\frac{x-\theta}{\sigma}\right).$$

The CDF for x_i is:

$$\begin{aligned} F_x(x \mid \theta) &= \begin{cases} 0 & x < \theta - \sigma\underline{l} \\ \tilde{F}_x(x \mid \theta) & x \in \left[\theta - \sigma\underline{l}, \theta + \sigma\bar{l}\right] \\ 1 & x > \theta + \sigma\bar{l} \end{cases} \\ &= \begin{cases} 0 & x < \theta - \sigma\underline{l} \\ \tilde{F}_\epsilon\left(\frac{x-\theta}{\sigma}\right) & x \in \left[\theta - \sigma\underline{l}, \theta + \sigma\bar{l}\right] \\ 1 & x > \theta + \sigma\bar{l} \end{cases}, \end{aligned}$$

for:

$$\begin{aligned} \tilde{F}_x(x \mid \theta) &= \int_{\theta - \sigma\underline{l}}^{x} \frac{1}{\sigma}\tilde{f}_\epsilon\left(\frac{s-\theta}{\sigma}\right) ds \\ &= \tilde{F}_\epsilon\left(\frac{x-\theta}{\sigma}\right). \end{aligned}$$

The posterior in θ conditional on x equals:

$$\begin{aligned} f_\theta(\theta \mid x) &= \frac{f_x(x \mid \theta) f_\theta(\theta)}{\int_x f_x(x \mid \theta) f_\theta(\theta) d\theta} \\ &= f_x(x \mid \theta) \\ &= \begin{cases} 0 & \theta < x - \sigma\bar{l} \\ \tilde{f}_x(x \mid \theta) & \theta \in \left[x - \sigma\bar{l}, x + \sigma\underline{l}\right] \\ 0 & \theta > x + \sigma\underline{l} \end{cases} \\ &= \begin{cases} 0 & \theta < x - \sigma\bar{l} \\ \frac{1}{\sigma}\tilde{f}_\epsilon\left(\frac{x-\theta}{\sigma}\right) & \theta \in \left[x - \sigma\bar{l}, x + \sigma\underline{l}\right] \\ 0 & \theta > x + \sigma\underline{l} \end{cases}, \end{aligned}$$

under the improper prior assumption. Thus:

$$\begin{aligned} F_\theta(\theta \mid x) &= \begin{cases} 0 & \theta < x - \sigma\bar{l} \\ \int_{x-\sigma\bar{l}}^{\theta} \tilde{f}_x(x \mid \theta) d\theta & \theta \in \left[x - \sigma\bar{l}, x + \sigma\underline{l}\right] \\ 1 & \theta > x + \sigma\underline{l} \end{cases} \\ &= \begin{cases} 0 & \theta < x - \sigma\bar{l} \\ 1 - \tilde{F}_\epsilon\left(\frac{x-\theta}{\sigma}\right) & \theta \in \left[x - \sigma\bar{l}, x + \sigma\underline{l}\right] \\ 1 & \theta > x + \sigma\underline{l} \end{cases}. \end{aligned}$$

B.2.1 One-to-one map

Let k be the cutoff that agents play. Given k, $r(\theta; k) = \int_k^\infty f_x(x \mid \theta) dx = 1 - F_x(k \mid \theta)$. Then:

$$r(\theta; k) = \begin{cases} 0 & \theta < k - \sigma\bar{l} \\ 1 - \tilde{F}_\epsilon\left(\frac{k-\theta}{\sigma}\right) & \theta \in \left[k - \sigma\underline{l}, k + \sigma\bar{l}\right] \\ 1 & \theta > k + \sigma\underline{l} \end{cases}.$$

It follows that:

$$r(\theta; k) = 1 - \tilde{F}_\epsilon\left(\frac{k-\theta}{\sigma}\right) \quad \Leftrightarrow \quad \theta(r; k) = k - \sigma\tilde{F}_\epsilon^{-1}(1 - r)$$

is a bijection for any $r \in (0, 1)$ to $\theta \in \left(k - \sigma\underline{l}, k + \sigma\bar{l}\right)$, where the open intervals are important. The 1-1 map fails if $r = 1$, since then $\theta \geq k + \sigma\underline{l}$ and if $r = 0$, since then $\theta \leq k - \sigma\bar{l}$.

B.2.2 Lemma A2, revisited

The statement of the Lemma is unchanged with the exception of Part 1, which should now state that "π^* weakly decreases in k (strictly decreasing for $x > 0$ and $k > 0$ for $k \in \left(x - \sigma(\underline{l} + \bar{l}), x + \sigma(\underline{l} + \bar{l})\right)$."

Proof. The payoff gain $\pi(r, x)$ continues to satisfy Properties P1-P5 in Lemma A1.

[1]: $\pi^*(x,k)$ increases in x: As before, implement a change in variables $z=-\theta$. Then:

$$\begin{aligned}\pi^*(x,k) &= \int_{\theta=x-\sigma\bar{l},}^{\theta=x+\sigma\underline{l}} \frac{1}{\sigma} f_\epsilon\left(\frac{x-\theta}{\sigma}\right)\pi\left(1-F_\epsilon\left(\frac{k-\theta}{\sigma}\right),x\right)\,d\theta \\ &= \int_{z=-x-\sigma\underline{l},}^{z=-x+\sigma\bar{l}} \frac{1}{\sigma} f_\epsilon\left(\frac{x+z}{\sigma}\right)\pi\left(1-F_\epsilon\left(\frac{k+z}{\sigma}\right),x\right)\,dz\end{aligned}$$

It is trivial to show that, for $x_1<x_2$, $F_\epsilon(x_1+\epsilon)\leq F_\epsilon(x_2+\epsilon)\ \forall\epsilon$. Therefore z under x_1 stochastically dominates z under x_2, and the original proof flows.

$\pi^*(x,k)$ weakly decreases in k follows from Property P1 and that r weakly decreases in k. To obtain strictly decreasing over $x>0$ and $k>0$, we also need $k\in\left(x-\sigma\left(\underline{l}+\bar{l}\right),x+\sigma\left(\underline{l}+\bar{l}\right)\right)$. Observe:

$$\begin{aligned}\pi^*(x,k) &= \int_{\theta=x-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k-\theta}{\sigma}\right),x\right)\,d\theta \\ &= \begin{cases} \int_{\theta=x-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi(0,x)\,d\theta=\pi(0,x) & x<k-\sigma\left(\underline{l}+\bar{l}\right) \\ \int_{\theta=x-\sigma\bar{l}}^{\theta=k-\sigma\bar{l}} f_\theta(\theta\mid x)\,\pi(0,x)\,d\theta & \\ \quad+\int_{\theta=k-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k-\theta}{\sigma}\right),x\right)\,d\theta & x\in\left(k-\sigma\left(\underline{l}+\bar{l}\right),k\right) \\ \int_{\theta=x-\sigma\bar{l}}^{\theta=k+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k-\theta}{\sigma}\right),x\right)\,d\theta & \\ \quad+\int_{\theta=k+\sigma\underline{l},}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi(1,x)\,d\theta & x\in\left(k,k+\sigma\left(\underline{l}+\bar{l}\right)\right) \\ \int_{\theta=x-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi(1,x)\,d\theta=\pi(1,x) & x>k+\sigma\left(\underline{l}+\bar{l}\right)\end{cases}\end{aligned}$$

Let $k_1<k_2$ be given. Re-write as:

$$\begin{aligned}\pi^*(x,k) &= \int_{\theta=x-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k-\theta}{\sigma}\right),x\right)\,d\theta \\ &= \begin{cases} F_1(x,k)\equiv\int_{\theta=x-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi(0,x)\,d\theta=\pi(0,x) & k>x+\sigma\left(\underline{l}+\bar{l}\right) \\ F_2(x,k)\equiv\int_{\theta=x-\sigma\bar{l}}^{\theta=k-\sigma\bar{l}} f_\theta(\theta\mid x)\,\pi(0,x)\,d\theta & \\ \quad+\int_{\theta=k-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k-\theta}{\sigma}\right),x\right)\,d\theta & k\in\left(x,x+\sigma\left(\underline{l}+\bar{l}\right)\right) \\ F_3(x,k)\equiv\int_{\theta=x-\sigma\bar{l}}^{\theta=k+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k-\theta}{\sigma}\right),x\right)\,d\theta & \\ \quad+\int_{\theta=k+\sigma\underline{l},}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi(1,x)\,d\theta & k\in\left(x-\sigma\left(\underline{l}+\bar{l}\right),x\right) \\ F_4(x,k)\equiv\int_{\theta=x-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi(1,x)\,d\theta=\pi(1,x) & k<x-\sigma\left(\underline{l}+\bar{l}\right)\end{cases}\end{aligned}$$

Notice that $F_1\left(x,k^1\right)<F_2\left(x,k^2\right)<F_3\left(x,k^3\right)<F_4\left(x,k^4\right)$ for every x for any $k^1>k^2>k^3>k^4$ satisfying the conditions of k for each function. So if $k_2>k_1$ in any way that crosses these regions, $\pi^*(x,k_2)<\pi^*(x,k_1)$.

If $k_2>k_1$ but each both lie within a single region, evidently $\pi^*(x,k_2)=\pi^*(x,k_1)$ in regions 1 and 4.

In Region 2:

$$\begin{aligned}
F_2(x,k_2)-F_2(x,k_1) &= \int_{\theta=x-\sigma\bar{l}}^{\theta=k_2-\sigma\bar{l}} f_\theta(\theta\mid x)\,\pi(0,x)\;d\theta \\
&\quad+\int_{\theta=k_2-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right),x\right)\;d\theta \\
&\quad-\int_{\theta=x-\sigma\bar{l}}^{\theta=k_1-\sigma\bar{l}} f_\theta(\theta\mid x)\,\pi(0,x)\;d\theta \\
&\quad-\int_{\theta=k_1-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right),x\right)\;d\theta \\
&= \int_{\theta=k_1-\sigma\bar{l}}^{\theta=k_2-\sigma\bar{l}} f_\theta(\theta\mid x)\,\pi(0,x)\;d\theta \\
&\quad+\int_{\theta=k_2-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right),x\right)\;d\theta \\
&\quad-\int_{\theta=k_1-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right),x\right)\;d\theta \\
&\leq \int_{\theta=k_1-\sigma\bar{l}}^{\theta=k_2-\sigma\bar{l}} f_\theta(\theta\mid x)\,\pi(0,x)\;d\theta \\
&\quad+\int_{\theta=k_2-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right),x\right)\;d\theta \\
&\quad-\int_{\theta=k_1-\sigma\bar{l}}^{\theta=x+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right),x\right)\;d\theta \\
&= \int_{\theta=k_1-\sigma\bar{l}}^{\theta=k_2-\sigma\bar{l}} f_\theta(\theta\mid x)\,\pi(0,x)\;d\theta \\
&\quad-\int_{\theta=k_1-\sigma\bar{l}}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta\mid x)\,\pi\left(1-F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right),x\right)\;d\theta \\
&< 0,
\end{aligned}$$

since π is an increasing function of r and $1 - F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right) > 0$ for $\theta > k_1 - \sigma\bar{l}$. In Region 3:

$$\begin{aligned}
F_3(x,k_2) - F_3(x,k_1) &= \int_{\theta=x-\sigma\bar{l}}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi\left(1 - F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right), x\right) d\theta \\
&\quad + \int_{\theta=k_2+\sigma\underline{l},}^{\theta=x+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi(1,x)\; d\theta \\
&\quad - \int_{\theta=x-\sigma\bar{l}}^{\theta=k_1+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi\left(1 - F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right), x\right) d\theta \\
&\quad - \int_{\theta=k_1+\sigma\underline{l},}^{\theta=x+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi(1,x)\; d\theta \\
&= -\int_{\theta=k_1+\sigma\underline{l},}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi(1,x)\; d\theta \\
&\quad + \int_{\theta=x-\sigma\bar{l}}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi\left(1 - F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right), x\right) d\theta \\
&\quad - \int_{\theta=x-\sigma\bar{l}}^{\theta=k_1+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi\left(1 - F_\epsilon\left(\frac{k_1-\theta}{\sigma}\right), x\right) d\theta \\
&\leq -\int_{\theta=k_1+\sigma\underline{l},}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi(1,x)\; d\theta \\
&\quad + \int_{\theta=x-\sigma\bar{l}}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi\left(1 - F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right), x\right) d\theta \\
&\quad - \int_{\theta=x-\sigma\bar{l}}^{\theta=k_1+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi\left(1 - F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right), x\right) d\theta \\
&= \int_{\theta=k_1+\sigma\underline{l}}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi\left(1 - F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right), x\right) d\theta \\
&\quad - \int_{\theta=k_1+\sigma\underline{l},}^{\theta=k_2+\sigma\underline{l}} f_\theta(\theta \mid x)\,\pi(1,x)\; d\theta \\
&< 0,
\end{aligned}$$

since π is an increasing function of r and $1 - F_\epsilon\left(\frac{k_2-\theta}{\sigma}\right) < 1$ for $\theta < k_2 + \sigma\underline{l}$.

Continuity in x and k should follow from Property P5, which is unchanged.

[2] This portion of the proof follows very similarly from before, with a few additional arguments to account for the fact that $\pi^*(x,k)$ strictly decreases in k only locally when k is close to x.

We show that $\{\underline{\xi}^n\}$ and $\{\bar{\xi}^n\}$ are well-defined increasing and decreasing sequences, respectively, through induction. From Property P4, we know that not reporting is dominant for $x < \underline{x}$, so $\pi^*(x,0) < 0$ for all $x < \underline{x}$. But we also know that $\pi^*(x,0) > 0$ for all $x > \bar{x}$. Define $\underline{\xi}^0 \equiv 0$ and $\bar{\xi}^0 \equiv \infty$. By continuity in x, there exists at least one solution x with $\pi^*\left(x, \underline{\xi}^0\right) = 0$, where $x \in [\underline{x}, \bar{x}]$. Call $\underline{\xi}^1$ the smallest such solution. Define $\bar{\xi}^1$ analogously to be the largest such solution with $\pi^*\left(x, \bar{\xi}^0\right) = 0$. Note that $\underline{\xi}^0 < \underline{\xi}^1 < \bar{\xi}^1 < \bar{\xi}^0$; if the inside inequality did not hold and $\underline{\xi}^1 \geq \bar{\xi}^1$, then $0 = \pi^*\left(\underline{\xi}^1, \underline{\xi}^0\right) > \pi^*\left(\underline{\xi}^1, \underline{\xi}^1\right) \geq \pi^*\left(\bar{\xi}^1, \underline{\xi}^1\right) \geq \pi^*\left(\bar{\xi}^1, \bar{\xi}^0\right) = 0$, a contradiction. The inequalities are because $\pi^*\left(\underline{\xi}^1, \underline{\xi}^1\right)$ strictly decreases in k locally, the contradiction assumption with π^* strictly increasing in x, and that π^* is globally weakly decreasing with $\underline{\xi}^1 < \bar{\xi}^0 = \infty$, respectively.

Our starting point for the induction is as follows. Given $\underline{\xi}^1$ and $\bar{\xi}^1$ with $\underline{\xi}^0 < \underline{\xi}^1 < \bar{\xi}^1 < \bar{\xi}^0$, $\pi^*\left(\underline{\xi}^1, \underline{\xi}^0\right) = 0$, and $\pi^*\left(\bar{\xi}^1, \bar{\xi}^0\right) = 0$, we claim there exists a smallest solution $\underline{\xi}^2$ of $\pi\left(\underline{\xi}^2, \underline{\xi}^1\right) = 0$ and a largest solution $\bar{\xi}^2$ of $\pi\left(\bar{\xi}^2, \bar{\xi}^1\right) = 0$, and that $\underline{\xi}^1 < \underline{\xi}^2 < \bar{\xi}^2 < \bar{\xi}^1$. We know $\pi^*\left(\underline{\xi}^1, \underline{\xi}^0\right) = 0 > \pi^*\left(\underline{\xi}^1, \underline{\xi}^1\right)$, and $\pi^*\left(\bar{\xi}^1, \bar{\xi}^0\right) = 0 < \pi^*\left(\bar{\xi}^1, \bar{\xi}^1\right) \leq \pi^*\left(\bar{\xi}^1, \underline{\xi}^1\right)$, where the first inequality is strict because $\pi^*\left(\bar{\xi}^1, \bar{\xi}^1\right)$ is strictly decreasing in

k. By continuity, there exists a smallest solution $\underline{\xi}^2 \in \left(\underline{\xi}^1, \bar{\xi}^1\right)$ with $\pi^*\left(\underline{\xi}^2, \underline{\xi}^1\right) = 0$. Analogously, we know $\pi^*\left(\underline{\xi}^1, \underline{\xi}^0\right) = 0 > \pi^*\left(\underline{\xi}^1, \underline{\xi}^1\right) \geq \pi^*\left(\underline{\xi}^1, \bar{\xi}^1\right)$, and $\pi^*\left(\bar{\xi}^1, \bar{\xi}^0\right) = 0 < \pi^*\left(\bar{\xi}^1, \bar{\xi}^1\right)$; by continuity there exists a largest solution $\bar{\xi}^2 \in \left(\underline{\xi}^1, \bar{\xi}^1\right)$ with $\pi^*\left(\bar{\xi}^2, \bar{\xi}^1\right) = 0$. Note that $\underline{\xi}^1 < \underline{\xi}^2 < \bar{\xi}^2 < \bar{\xi}^1$; if the inside inequality did not hold and $\underline{\xi}^2 \geq \bar{\xi}^2$, then $0 = \pi^*\left(\underline{\xi}^2, \underline{\xi}^1\right) > \pi^*\left(\underline{\xi}^2, \underline{\xi}^2\right) \geq \pi^*\left(\bar{\xi}^2, \underline{\xi}^2\right) \geq \pi^*\left(\bar{\xi}^2, \bar{\xi}^1\right) = 0$, a contradiction. The inequalities are because $\pi^*\left(\underline{\xi}^2, \underline{\xi}^2\right)$ strictly decreases in k locally, the contradiction assumption with π^*strictly increasing in x, and that π^* is globally weakly decreasing with $\underline{\xi}^2 \in \left(\underline{\xi}^1, \bar{\xi}^1\right)$, respectively.

The inductive hypothesis is that, given $\underline{\xi}^n$ and $\bar{\xi}^n$ with $\underline{\xi}^{n-1} < \underline{\xi}^n < \bar{\xi}^n < \bar{\xi}^{n-1}$, $\pi^*\left(\underline{\xi}^n, \underline{\xi}^{n-1}\right) = 0$, and $\pi^*\left(\bar{\xi}^n, \bar{\xi}^{n-1}\right) = 0$, there exists a smallest solution $\underline{\xi}^{n+1}$ of $\pi\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) = 0$ and a largest solution $\bar{\xi}^{n+1}$ of $\pi\left(\bar{\xi}^{n+1}, \bar{\xi}^n\right) = 0$, and that $\underline{\xi}^n < \underline{\xi}^{n+1} < \bar{\xi}^{n+1} < \bar{\xi}^n$. We know $\pi^*\left(\underline{\xi}^n, \underline{\xi}^{n-1}\right) = 0 > \pi^*\left(\underline{\xi}^n, \underline{\xi}^n\right)$, and $\pi^*\left(\bar{\xi}^n, \bar{\xi}^{n-1}\right) = 0 < \pi^*\left(\bar{\xi}^n, \bar{\xi}^n\right) \leq \pi^*\left(\bar{\xi}^n, \underline{\xi}^n\right)$. By continuity, there exists a smallest solution $\underline{\xi}^{n+1} \in \left(\underline{\xi}^n, \bar{\xi}^n\right)$ with $\pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) = 0$. Similarly, we know $\pi^*\left(\underline{\xi}^n, \underline{\xi}^{n-1}\right) = 0 > \pi^*\left(\underline{\xi}^n, \underline{\xi}^n\right) \geq \pi^*\left(\underline{\xi}^n, \bar{\xi}^n\right)$, and $\pi^*\left(\bar{\xi}^n, \bar{\xi}^{n-1}\right) = 0 < \pi^*\left(\bar{\xi}^n, \bar{\xi}^n\right)$; by continuity there exists a largest solution $\bar{\xi}^{n+1} \in \left(\underline{\xi}^n, \bar{\xi}^n\right)$ with $\pi^*\left(\bar{\xi}^{n+1}, \bar{\xi}^n\right) = 0$. Note that $\underline{\xi}^n < \underline{\xi}^{n+1} < \bar{\xi}^{n+1} < \bar{\xi}^n$; if the inside inequality did not hold and $\underline{\xi}^{n+1} \geq \bar{\xi}^{n+1}$, then $0 = \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) > \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^{n+1}\right) \geq \pi^*\left(\bar{\xi}^{n+1}, \underline{\xi}^{n+1}\right) \geq \pi^*\left(\bar{\xi}^{n+1}, \bar{\xi}^n\right) = 0$, a contradiction. The inequalities are because $\pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^{n+1}\right)$ strictly decreases in k locally, the contradiction assumption with π^*strictly increasing in x, and that π^* is globally weakly decreasing with $\underline{\xi}^{n+1} \in \left(\underline{\xi}^n, \bar{\xi}^n\right)$, respectively.

Note that $\{\underline{\xi}^n\}$ is bounded from above by construction. Because it is also an increasing sequence, there exists a $\underline{\xi}$ with $\lim_{n\to\infty} \underline{\xi}^n = \underline{\xi}$. Note that $\lim_{n\to\infty} \pi^*\left(\underline{\xi}^{n+1}, \underline{\xi}^n\right) = 0$ so by construction and continuity of π^*, we must have $\pi^*\left(\underline{\xi}, \underline{\xi}\right) = 0$ and that $\underline{\xi}$ is the smallest such solution to $\pi^*\left(\xi, \xi\right) = 0$. Analogously, there exists a $\bar{\xi}$ with $\lim_{n\to\infty} \underline{\xi}^n = \underline{\xi}$ and $\pi^*\left(\bar{\xi}, \bar{\xi}\right) = 0$ and that $\bar{\xi}$ is the smallest such solution to $\pi^*\left(\xi, \xi\right) = 0$. This shows, among other things, that there exists at least one threshold equilibrium $\underline{\xi}$. One can see that any such solution $\underline{\xi}$ is an equilibrium because $x_1 < \underline{\xi} < x_2$ implies $\pi^*\left(x_1, \underline{\xi}\right) < \pi^*\left(\underline{\xi}, \underline{\xi}\right) = 0 < \pi^*\left(x_2, \underline{\xi}\right)$.

[3] Given the agent's signal x, what is her assessment of the cumulative distribution function of r, $\Psi\left(\tilde{r}; x, k\right)$? For $\tilde{r} \in (0,1)$, the argument follows from before, which we restate here. The probability that $r < \tilde{r}$ equals the probability that $\theta < k - \sigma F_\epsilon^{-1}\left(1 - \tilde{r}\right)$. In words, the probability $\Psi\left(\tilde{r}; x, k\right) \equiv \Pr\left(r < \tilde{r} \mid x\right)$ that the true proportion of players reporting is less than $\tilde{r}$ equals the probability that the true θ satisfies $r\left(\theta; k\right) = 1 - F_\epsilon\left(\frac{k-\theta}{\sigma}\right) < \tilde{r}$, or equivalently that θ is such that fewer than $\tilde{r}$ players observe a signal greater than k; in turn, this equals the probability that the true θ is less than $k - \sigma F_\epsilon^{-1}\left(1 - \tilde{r}\right)$, integrated against the conditional density $f_\theta\left(\theta \mid x\right)$.

Importantly, conditional on x, we must have $\theta \in \left[x - \sigma\bar{l}, x + \sigma\underline{l}\right]$.

Given x, what is the agent's probability assessment that $r = 0$? This must equal the posterior probability that $\theta < k - \sigma\bar{l}$ given x. If $k - \sigma\bar{l} \in \left[x - \sigma\bar{l}, x + \sigma\underline{l}\right]$, then $x \in \left[k - \sigma\left(\bar{l} + \underline{l}\right), k\right]$ then:

$$\begin{aligned}
\Psi\left(r; x, k\right) &= \int_{x-\sigma\bar{l}}^{k-\sigma\bar{l}} f_\theta\left(\theta \mid x\right)\, d\theta \\
&= \int_{x-\sigma\bar{l}}^{k-\sigma\bar{l}} \frac{1}{\sigma} f_\epsilon\left(\frac{x-\theta}{\sigma}\right) d\theta \\
&= \int_{z=\frac{x-k}{\sigma}+\bar{l}}^{z=\bar{l}} f_z\left(z\right)\, dz \text{ for } z = \frac{x-\theta}{\sigma},\ dz = -\frac{1}{\sigma} d\theta \\
&\quad \text{noting that } z\left(k - \sigma\bar{l}\right) = \frac{x - \left(k - \sigma\bar{l}\right)}{\sigma} = \frac{x-k}{\sigma} + \bar{l} \\
&\quad \text{and that } z\left(x - \sigma\bar{l}\right) = \frac{x - \left(x - \sigma\bar{l}\right)}{\sigma} = \bar{l} \\
&= 1 - F_\epsilon\left(\frac{x-k}{\sigma} + \bar{l}\right).
\end{aligned}$$

If $k - \sigma\bar{l} < x - \sigma\bar{l}$ then $x > k$ and the probability is zero by the definition of $f_\theta\left(\theta\right)$. If $k - \sigma\bar{l} > x + \sigma\underline{l}$ or

equivalently if $\frac{k-x}{\sigma} > \bar{l} + \underline{l}$ then the probability is 1. So:

$$\Psi(0; x, k) = \begin{cases} 0 & x \geq k \\ 1 - F_\epsilon\left(\frac{x-k}{\sigma} + \bar{l}\right) & x \in \left(k - \sigma\left(\bar{l} + \underline{l}\right), k\right) \\ 1 & x \leq k - \sigma\left(\bar{l} + \underline{l}\right) \end{cases}.$$

Given x, what is the agent's probability assessment that $r \leq 1$? Trivially, this must be 1, since this equals the probability that $r < 1$, which is the probability that $\theta < k + \sigma\underline{l}$, plus the probability that $r = 1$, which is the probability that $\theta \geq k + \sigma\underline{l}$. Thus, $\Psi(1) = 1$.

Given $\tilde{r} \in (0, 1)$, what is the agent's probability assessment that $r < \tilde{r}$? Given k, we know $\theta(r; k) = k - \sigma\tilde{F}_\epsilon^{-1}(1 - r) \in \left(k - \sigma\underline{l}, k + \sigma\bar{l}\right)$. We also must have $\theta \in \left[x - \sigma\bar{l}, x + \sigma\underline{l}\right]$ in the posterior distribution of θ. Some useful facts to reference later:

$$\begin{aligned}
k - \sigma\tilde{F}_\epsilon^{-1}(1 - r) > x - \sigma\bar{l} &\Leftrightarrow r > 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) \\
x \geq k + \sigma\left(\bar{l} + \underline{l}\right) &\Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) = 1 \\
x < k + \sigma\left(\bar{l} + \underline{l}\right) &\Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) < 1 \\
x < k &\Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) = 0, \\
k - \sigma\tilde{F}_\epsilon^{-1}(1 - r) < x + \sigma\underline{l} &\Leftrightarrow r < 1 - \tilde{F}_\epsilon^{-1}\left(\frac{k-x}{\sigma} - \underline{l}\right) \\
x \leq k - \sigma\left(\bar{l} + \underline{l}\right) &\Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} - \underline{l}\right) = 0 \\
x > k - \sigma\left(\bar{l} + \underline{l}\right) &\Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} - \underline{l}\right) > 0 \\
x > k &\Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} - \underline{l}\right) = 1.
\end{aligned}$$

Suppose first $x \in \left(k - \sigma\left(\bar{l} + \underline{l}\right), k + \sigma\left(\bar{l} + \underline{l}\right)\right)$. Then $1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} - \underline{l}\right) > 0$ and $1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) < 1$. Several cases can occur:

$$\begin{aligned}
k - \sigma\tilde{F}_\epsilon^{-1}(1 - r) > x + \sigma\underline{l} &\Rightarrow r \in \left(1 - \tilde{F}_\epsilon^{-1}\left(\frac{k-x}{\sigma} - \underline{l}\right), 1\right) \\
k - \sigma\tilde{F}_\epsilon^{-1}(1 - r) \in \left(x - \sigma\bar{l}, x + \sigma\underline{l}\right) &\Rightarrow r \in \left(1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right), 1 - \tilde{F}_\epsilon^{-1}\left(\frac{k-x}{\sigma} - \underline{l}\right)\right) \\
k - \sigma\tilde{F}_\epsilon^{-1}(1 - r) < x - \sigma\bar{l} &\Rightarrow r \in \left(0, 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right)\right)
\end{aligned}$$

If $k-\sigma\tilde{F}_{\epsilon}^{-1}(1-r)>x-\sigma\bar{l}$, then $r\in\left(1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}+\bar{l}\right),1\right)$, and furthermore:

$$\begin{aligned}
\Psi(r;x,k) &= \int_{x-\sigma\bar{l}}^{k-\sigma\tilde{F}_{\epsilon}^{-1}(1-r)} f_{\theta}(\theta\mid x)\,d\theta \\
&= \int_{x-\sigma\bar{l}}^{k-\sigma\tilde{F}_{\epsilon}^{-1}(1-r)} \frac{1}{\sigma}\tilde{f}_{\epsilon}\left(\frac{x-\theta}{\sigma}\right)d\theta \\
&= \int_{z=\frac{x-k}{\sigma}+\tilde{F}_{\epsilon}^{-1}(1-r)}^{z=\bar{l}} f_{z}(z)\,dz \text{ for } z=\frac{x-\theta}{\sigma},\ dz=-\frac{1}{\sigma}d\theta \\
&\quad \text{noting that } z\left(k-\sigma\tilde{F}_{\epsilon}^{-1}(1-r)\right)=\frac{x-\left(k-\sigma\tilde{F}_{\epsilon}^{-1}(1-\tilde{r})\right)}{\sigma}=\frac{x-k}{\sigma}+\tilde{F}_{\epsilon}^{-1}(1-r) \\
&\quad \text{and that } z\left(x-\sigma\bar{l}\right)=\frac{x-\left(x-\sigma\bar{l}\right)}{\sigma}=\bar{l} \\
&= 1-\tilde{F}_{\epsilon}\left(\frac{x-k}{\sigma}+\tilde{F}_{\epsilon}^{-1}(1-r)\right)
\end{aligned}$$

If $k-\sigma\tilde{F}_{\epsilon}^{-1}(1-r)<x-\sigma\bar{l}$, then $r\in\left(0,1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}+\bar{l}\right)\right)$ and furthermore by the definition of f_{θ}:

$$\begin{aligned}
\Psi(r;x,k) &= -\int_{k-\sigma F_{\epsilon}^{-1}(1-r)}^{x-\sigma\bar{l}} f_{\theta}(\theta\mid x)\,d\theta \\
&= 0.
\end{aligned}$$

Suppose next $x\geq k+\sigma\left(\bar{l}+\underline{l}\right)$. Then $1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}+\bar{l}\right)=1$ and $1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}-\underline{l}\right)>0$. From before, $1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}+\bar{l}\right)=1$ implies $k-\sigma\tilde{F}_{\epsilon}^{-1}(1-r)<x-\sigma\bar{l}$ since $r<1$. But then, following a similar argument,

$$\begin{aligned}
\Psi(r;x,k) &= -\int_{k-\sigma F_{\epsilon}^{-1}(1-r)}^{x-\sigma\bar{l}} f_{\theta}(\theta\mid x)\,d\theta \\
&= 0.
\end{aligned}$$

Suppose finally $x\leq k-\sigma\left(\bar{l}+\underline{l}\right)$. Then $1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}-\underline{l}\right)=0$ and $1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}+\bar{l}\right)<1$. From before, $1-\tilde{F}_{\epsilon}\left(\frac{k-x}{\sigma}-\underline{l}\right)=0$ implies $k-\sigma\tilde{F}_{\epsilon}^{-1}(1-r)>x+\sigma\underline{l}$ since $r>0$. But then:

$$\begin{aligned}
\Psi(r;x,k) &= \int_{x-\sigma\bar{l}}^{x+\sigma\underline{l}} f_{\theta}(\theta\mid x)\,d\theta \\
&= 1.
\end{aligned}$$

To summarize:

$$\begin{aligned}
\Psi\left(r; x \geq k + \sigma\left(\bar{l} + \underline{l}\right)\right) &= \begin{cases} 0 & r = 0 \\ 0 & r \in (0,1) \\ 1 & r = 1, \end{cases} \\
\Psi\left(r; x \in \left(k, k + \sigma\left(\bar{l} + \underline{l}\right)\right)\right) &= \begin{cases} 0 & r = 0 \\ 0 & r \in \left(0, 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right)\right) \\ 1 - \tilde{F}_\epsilon\left(\frac{x-k}{\sigma} + \tilde{F}_\epsilon^{-1}(1-r)\right) & r \in \left(1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right), 1\right) \\ 1 & r = 1, \end{cases} \\
\Psi\left(r; x \in \left(k - \sigma\left(\bar{l} + \underline{l}\right), k\right)\right) &= \begin{cases} 1 - F_\epsilon\left(\frac{x-k}{\sigma} + \bar{l}\right) & r = 0 \\ 1 - \tilde{F}_\epsilon\left(\frac{x-k}{\sigma} + \tilde{F}_\epsilon^{-1}(1-r)\right) & r \in (0,1) \\ 1 & r = 1, \end{cases} \\
\Psi\left(r; x \leq k - \sigma\left(\bar{l} + \underline{l}\right)\right) &= \begin{cases} 1 & r = 0 \\ 1 & r \in (0,1) \\ 1 & r = 1, \end{cases} \\
\Psi\left(r; x = k\right) &= \begin{cases} 0 & r = 0 \\ r & r \in (0,1) \\ 1 & r = 1. \end{cases}
\end{aligned}$$

Therefore, the marginal agent has a uniform belief over r.
Given this belief, we can solve for the threshold x^*. Recall the expected payoff gain equals:

$$\pi^*(x,k) = \int_{-\infty}^{\infty} f(\theta \mid x)\, \pi\left(1 - F\left(\frac{k-\theta}{\sigma}\right), x\right) d\theta \tag{B.1}$$

$$\Rightarrow \pi^*(x,k) = \int_{\theta = x - \sigma\bar{l},}^{\theta = x + \sigma\underline{l}} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) \pi\left(1 - F\left(\frac{k-\theta}{\sigma}\right), x\right) d\theta. \tag{B.2}$$

For the marginal agent with $x = k$, this payoff equals:

$$\pi^*(x,x) = \int_{x - \sigma\bar{l},}^{x + \sigma\underline{l}} \frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) \pi\left(1 - F\left(\frac{x-\theta}{\sigma}\right), x\right) d\theta. \tag{B.3}$$

Given the marginal agent's posterior belief over θ translates into a uniform belief over r, we can write:

$$\pi^*(x,x) = \int_0^1 \pi(r,x)\, dr.$$

■

B.3 General sanction function

The outside party observes $\mathbb{X} \equiv \{(n(x), x) | \forall x \in \mathbb{R}\}$, the frequency and value of each reported x. Let $\Xi(\mathbb{X}|\mathbb{X}) : \mathbb{R}^{\mathfrak{c}} \times \mathbb{R}^{\mathfrak{c}} \to [0,1]$ represent the sanction probability given the set of all reports and the value of those reports. Consider a class of sanction functions $\Xi(\mathbb{X}) = B(\int_x \varphi(x) n(x) dx)$, where $\varphi(x)$ is a weight for each reported x and $n(x) \geq 0$ is the frequency of reports for each x. Let $\varphi(x)$ be weakly monotone increasing and differentiable, where $\varphi(x) = 0$ if $x < 0$ and $\varphi(x) > 0$ for some $x > 0$. Let $B(s) : [0, \infty) \to [0,1]$ be weakly

monotone increasing, and strictly increasing at some s. Note that $B(s)$ does not have to be continuous. This describes a natural class of sanction functions in which the sanction probability weakly increases in the number of reports and in the severity of misconduct.

1. We can show that when $\varphi(x) = 1$ for all $x > 0$, then $\Xi(\mathbb{X}) = B(r) = \Gamma(r)$. Thus, the sanction probability is a weakly increasing function of r alone.

   Proof. By Lemma A4, all agents use the same strategy in equilibrium. Suppose agents use threshold strategy x^*. Then $n(x) = \mathbf{1}_{[x>x^*]}\frac{1}{\sigma}\phi(\frac{x-\theta}{\sigma})$, giving

   $$\begin{aligned}\Xi(\mathbb{X}) &= B(\int_x \varphi(x)n(x)dx) \\ &= B\left(\int_{x^*}^{\infty} \frac{1}{\sigma}\varphi(x)\phi\left(\frac{x-\theta}{\sigma}\right)dx\right).\end{aligned}$$

   The outside party observes $\hat{r}(\theta) = \int_{x^*}^{\infty} \frac{1}{\sigma}\phi(x \mid \theta)\,dx = \Phi\left(\frac{\theta - x^*}{\sigma}\right)$, so $\theta = x^* + \sigma\Phi^{-1}(r)$.

   Suppose $\varphi(x) = 1$. Then

   $$\begin{aligned}\Xi(\mathbb{X}) &= B\left(\int_{x^*}^{\infty} \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)dx\right) \\ &= B\left(1 - \Phi\left(\frac{x^* - \theta}{\sigma}\right)\right) \\ &= B\left(1 - \Phi\left(\frac{x^* - \left[x^* + \sigma\Phi^{-1}(r)\right]}{\sigma}\right)\right) \\ &= B\left(1 - \Phi\left(-\Phi^{-1}(r)\right)\right) \\ &= B\left(1 - \Phi\left(\Phi^{-1}(1-r)\right)\right) \\ &= B(r).\end{aligned}$$

   Thus, $\Xi(\mathbb{X}) = B(r) = \Gamma(r)$. Thus, the sanction probability is a weakly increasing function of r alone. ■

   Notable special cases: If $B(s) = \gamma s$, then $\Xi(\mathbb{X}) = \gamma r$. If $B(s) = \begin{cases} 0 \text{ if } s \in [0, \overline{r}) \\ 1 \text{ if } s \in [\overline{r}, 1] \end{cases}$, then $\Xi(\mathbb{X}) = \begin{cases} 0 \text{ if } s \in [0, \overline{r}) \\ 1 \text{ if } s \in [\overline{r}, 1] \end{cases}$ where $\overline{r} \in [0,1]$ is a constant.

2. For more general $\varphi(x)$, the sanction function can be reformulated as $\Xi(\mathbb{X}) = \Gamma(x^*, r)$ (though there may be multiple equilibria in which agents use threshold strategies).

   By Lemma A4, all agents use the same strategy in equilibrium. Suppose agents use threshold strategy x^*. Then

   $$\begin{aligned}\Xi(\mathbb{X}) &= B\left(\int_{x^*}^{\infty} \frac{\varphi(x)}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right)dx\right) \\ &= B\left(\int_{x^*}^{\infty} \frac{\varphi(x)}{\sigma}\phi\left(\frac{x-x^*}{\sigma} - \Phi^{-1}(r)\right)dx\right) \\ &= \Gamma(x^*, r),\end{aligned}$$

since $\theta = x^* + \sigma\Phi^{-1}(r)$ when agents use threshold x^*. Since each agent, including the marginal agent, takes x^* as given, then the marginal agent solves:

$$\int_0^1 \pi(r, x^*)dr = 0,$$

which has the implicit solution

$$x^* = \frac{c}{\omega - \beta + \left(\int_0^1 \Gamma(x^*, r)dr\right)(\alpha + \beta)}. \tag{B.4}$$

To show the existence of such an x^*, note that we can still define the dominance regions in which the agent never reports if $x \leq \underline{x}$ (i.e., even if everyone reports and the manager is definitely sanctioned) and always reports if $x \geq \overline{x}$ (i.e., even if no one reports so the manager is definitely not sanctioned), where

$$\underline{x} = \frac{c}{\omega + \alpha}$$
$$\overline{x} = \frac{c}{\omega - \beta}.$$

Let $G(x) = x\left(\omega - \beta + \left(\int_0^1 \Gamma(x, r)dr\right)(\alpha + \beta)\right) - c$, so x^* satisfies $G(x^*) = 0$. Note that $G(\overline{x}) > 0$ if and only if $\int_0^1 \Gamma(x^*, r)dr > 0$. Given that $f(x)$ is weakly monotone increasing where $f(x) = 0$ if $x < 0$ and $f(x) > 0$ for some $x > 0$, and $B(s) : [0, \infty) \to [0, 1]$ is weakly monotone increasing (and strictly increasing at some s), then $\int_0^1 \Gamma(x^*, r)dr > 0$. Note that $G(\underline{x}) < 0$ if and only if $\int_0^1 \Gamma(x^*, r)dr < 1$. Since $\int_0^1 \Gamma(x^*, r)dr = 1$ if and only if $B(s) = 1$ for all s, then $\int_0^1 \Gamma(x^*, r)dr < 1$ because $B(s) : [0, \infty) \to [0, 1]$ is weakly monotone increasing, and strictly increasing at some s. Thus, such an x^* exists, and $x^* \in (\underline{x}, \overline{x})$. Given such an x^*, it is straightforward to verify that agents with $x_i < x^*$ do not report and agents with $x_i > x^*$, since $x_i\left(\omega - \beta + \left(\int_0^1 \Gamma(x^*, r)dr\right)(\alpha + \beta)\right)$ monotonically increases in x_i when agents use threshold x^*. Thus, all agents using threshold x^* is an equilibrium.

Although we have established the existence of an equilibrium in which agents use threshold x^*, which must satisfy Equation B.4, multiplicity of equilibria is entirely possible. Nonetheless, in any such equilibrium, the existence of an "open secret" in which there is under-reporting, as described by Corollary 1.1, still holds. This is because when, $\sigma \to 0$, for any $\theta \in (\underline{x}, x^*)$ we still have $\pi(1, x) > 0 > \pi(0, x)$. Since $x^* \in (\underline{x}, \overline{x})$, this implies there will be under-reporting for any $\theta \in (\underline{x}, x^*)$.

3. Even when the sanction function includes a general $\varphi(x)$, we show that under-reporting occurs in any equilibrium that involves threshold strategies, under qualitatively similar conditions as Proposition 2. When θ is sufficiently high, there always exists some $\hat{x} < x^*$ such that there is a Pareto improvement in agent payoffs. When θ is intermediate, there exists some $\hat{x} < x^*$ such that there is a Pareto improvement in agent payoffs if σ is sufficiently small.

Lemma B2. *Let g and h be density functions, where $g \succeq h$ if $\int_{-\infty}^z g(s)ds \leq \int_{-\infty}^z h(s)ds$. If $g \succeq h$ and $u(z)$ is a weakly increasing and weakly positive, differentiable function of z, then $\int_a^\infty u(z)g(z)dz \geq \int_a^\infty u(z)h(z)dz$.*

Proof. Define $G(z) = \int_{-\infty}^z g(s)ds$ and $H(z) = \int_{-\infty}^z h(s)ds$. Note that if $g \succeq h$, then $G(z) \leq H(z)$.

Using integration by parts,

$$\begin{aligned}\int_a^\infty u(z)g(z)dz &= \int_a^\infty u(z)G'(z)dz \\ &= u(z)G(z)|_a^\infty - \int_a^\infty G(z)u'(z)dz \\ \int_a^\infty u(z)h(z)dz &= \int_a^\infty u(z)H'(z)dz \\ &= u(z)H(z)|_a^\infty - \int_a^\infty H(z)u'(z)dz.\end{aligned}$$

Since $G(\infty) = H(\infty) = 1$, then

$$\int_a^\infty u(z)g(z)dz - \int_a^\infty u(z)h(z)dz = u(a)[H(a) - G(z)] + \int_a^\infty u'(z)[H(z) - G(z)]dz \geq 0.$$

■

Let θ' be the minimum value of θ such that the marginal agent's expected probability of sanction is less than or equal to the realized probability of sanction:

$$\theta' = \min\{\theta : \int_0^1 B\left(\int_{x^*}^\infty \frac{\varphi(x)}{\sigma}\phi\left(\frac{x-x^*}{\sigma} - \Phi^{-1}(r)\right)dx\right)dr \leq B\left(\int_{x^*}^\infty \frac{\varphi(x)}{\sigma}\phi\left(\frac{x-\theta'}{\sigma}\right)\right)\}$$

Recall that $\Xi(\mathbb{X}|\theta, x^*) = B\left(\int_{x^*}^\infty \frac{1}{\sigma}\varphi(x)\phi\left(\frac{x-\theta}{\sigma}\right)dx\right)$. By Lemma B2, we know that $\Xi(\mathbb{X}|\theta, \hat{x})$ is weakly increasing in θ, so clearly such a θ' exists.

(a) Suppose $\theta > \theta'$. There exists some $\hat{x} < x^*$ such that total welfare is higher due to a Pareto improvement.

Recall that $\Xi(\mathbb{X}|\theta, x^*) = B\left(\int_{x^*}^\infty \frac{1}{\sigma}\varphi(x)\phi\left(\frac{x-\theta}{\sigma}\right)dx\right)$. For any $\hat{x} < x^*$, total agent payoffs equal:

$$\begin{aligned}W(\theta \mid \hat{x}) &= \int_{x^*}^\infty \left[x\left(\omega - \beta + \Xi(\mathbb{X}|\theta, \hat{x})(\alpha + \beta)\right) - c\right] f(x \mid \theta)\, dx \\ &\quad + \int_{\hat{x}}^{x^*} \left[x\left(\omega - \beta + \Xi(\mathbb{X}|\theta, \hat{x})(\alpha + \beta)\right) - c\right] f(x \mid \theta)\, dx.\end{aligned}$$

We know that equilibrium total payoffs are such that:

$$\begin{aligned}W(\theta \mid x^*) &= \int_{x^*}^\infty \left[x\left(\omega - \beta + \Xi(\mathbb{X}|\theta, x^*)(\alpha + \beta)\right) - c\right] f(x \mid \theta)\, dx \\ &< \int_{x^*}^\infty \left[x\left(\omega - \beta + \Xi(\mathbb{X}|\theta, \hat{x})(\alpha + \beta)\right) - c\right] f(x \mid \theta)\, dx,\end{aligned}$$

for any $\hat{x} < x^*$.

The remaining claim is that there exists a $\hat{x} < x^*$ such that:

$$K \equiv \int_{\hat{x}}^{x^*} \left[x\left(\omega - \beta + \Xi(\mathbb{X}|\theta)(\alpha + \beta)\right) - c\right] f(x \mid \theta)\, dx > 0.$$

By Lemma B2, we know that $\Xi(\mathbb{X}|\theta,\hat{x})$ is weakly increasing in θ, which implies that when $\theta > \theta'$:

$$\begin{aligned} 0 &= x^* \left(\omega - \beta + \left(\int_0^1 \Gamma(x^*, r)dr\right)(\alpha+\beta)\right) - c \\ &< x^* \left(\omega - \beta + \Xi(\mathbb{X}|\theta, x^*)(\alpha+\beta)\right) - c. \end{aligned} \tag{B.5}$$

Let $H(x) \equiv x\left(\omega - \beta + \Xi(\mathbb{X}|\theta, x)(\alpha+\beta)\right) - c$. Note that $H(x^*) > 0$ from Equation B.5.
Suppose H is continuous at x^*. If $H'(x^*) < 0$, then there exists a $\hat{x} < x^*$ such that:

$$\begin{aligned} 0 &< x^* \left(\omega - \beta + \Xi(\mathbb{X}|\theta, x^*)(\alpha+\beta)\right) - c \\ &< \hat{x} \left(\omega - \beta + \Xi(\mathbb{X}|\theta, \hat{x})(\alpha+\beta)\right) - c. \end{aligned}$$

If $H'(x^*) > 0$, there exists some $\varepsilon > 0$ such that, for $\hat{x} = x^* - \varepsilon$,

$$\begin{aligned} 0 &< \hat{x} \left(\omega - \beta + \Xi(\mathbb{X}|\theta, \hat{x})(\alpha+\beta)\right) - c \\ &< x^* \left(\omega - \beta + \Xi(\mathbb{X}|\theta, x^*)(\alpha+\beta)\right) - c. \end{aligned}$$

If $H'(x^*) = 0$, observe that because $H^{(n)}(x^*) \neq 0$ for some $n > 0$, we can apply a similar argument to find an $\hat{x} < x^*$ such that $H(\hat{x}) > 0$. Either way, there exists some $\hat{x} < x^*$ such that $H(\hat{x}) > 0$.
Suppose H is not continuous at x^* (if $B(\cdot)$ is not continuous at x^*). By direct computation, $\frac{\partial}{\partial \hat{x}} \int_{\hat{x}}^{\infty} \frac{1}{\sigma}\varphi(x)\phi\left(\frac{x-\theta}{\sigma}\right) dx < 0$. Since $B(s)$ increases in s, then H is either left-continuous at x^* or strictly decreasing in x. Either way, there exists some $\hat{x} < x^*$ such that $H(\hat{x}) > 0$.
But then:

$$\begin{aligned} K &= \int_{\hat{x}}^{x^*} \left[x\left(\omega - \beta + \Xi(\mathbb{X}|\theta,\hat{x})(\alpha+\beta)\right) - c\right] f(x \mid \theta)\, dx \\ &> \int_{\hat{x}}^{x^*} \left[\hat{x}\left(\omega - \beta + \Xi(\mathbb{X}|\theta,\hat{x})(\alpha+\beta)\right) - c\right] f(x \mid \theta)\, dx \\ &= \left[\hat{x}\left(\omega - \beta + \Xi(\mathbb{X}|\theta,\hat{x})(\alpha+\beta)\right) - c\right] \times \left[\Phi\left(\frac{x^*-\theta}{\sigma}\right) - \Phi\left(\frac{\hat{x}-\theta}{\sigma}\right)\right] \\ &> 0. \end{aligned}$$

Note that this is a Pareto improvement because for any $x \in (\hat{x}, x^*)$,

$$x\left(\omega - \beta + \Xi(\mathbb{X}|\theta,\hat{x})(\alpha+\beta)\right) - c \;>\; H(\hat{x}) > 0,$$

whereas these agents all receive 0 when playing a threshold strategy around x^*.

(b) Suppose $\underline{\underline{x}} < \theta \leq \theta'$, where $\underline{\underline{x}}$ is defined below. If σ is sufficiently low, there exists some $\hat{x} \in (\underline{\underline{x}}, \theta)$ such that total welfare is higher due to a Pareto improvement.
If $\theta \leq \theta'$, then $H(x^*) \leq 0$. Note that by Lemma B2, for all $\hat{x} < \theta$, $B\left(\int_{\hat{x}}^{\infty} \frac{\varphi(x)}{\sigma}\phi\left(\frac{x-\theta'}{\sigma}\right)\right)$ increases as σ decreases. By direct computation, $\frac{\partial}{\partial \hat{x}} \int_{\hat{x}}^{\infty} \frac{1}{\sigma}\varphi(x)\phi\left(\frac{x-\theta}{\sigma}\right) dx < 0$. Define $\underline{\underline{x}}$ as the threshold such that

$$\underline{\underline{x}}\left(\omega - \beta + \lim_{\sigma\to 0} B\left(\int_{\underline{x}}^{\infty} \frac{\varphi(x)}{\sigma}\phi\left(\frac{x-\theta'}{\sigma}\right)\right)(\alpha+\beta)\right) - c = 0.$$

By construction, $\lim_{\sigma\to 0} H(\underline{\underline{x}}) = 0$ and $x^* > \underline{\underline{x}} \geq \underline{x}$.
Suppose $H(\cdot)$ is continuous on the interval $[\underline{\underline{x}}, \theta)$ for all σ. By direct differentiation, $\lim_{\sigma\to 0} \frac{\partial H}{\partial \hat{x}} > 0$

for all $\hat{x} < \theta$. Thus, $\lim_{\sigma \to 0} H(\hat{x}) > 0$ for all $\hat{x} \in (\underline{\underline{x}}, \theta)$. By Lemma B2, $H(\hat{x})$ is decreasing as σ increases. But by continuity of H, for any $\hat{x} \in (\underline{\underline{x}}, \theta)$, there exists some $\sigma > 0$ sufficiently small that $H(\hat{x}) > 0$.

Suppose $H(x)$ is not continuous on the interval $[\underline{\underline{x}}, \theta)$ for all σ. Since $\underline{\underline{x}}$ is determined by the case of all agents reporting, then it is sufficient to consider the following to find a Pareto-improving $\hat{x} < x^*$. Let $Z = \{z : \lim_{x \uparrow z} B(x) \neq \lim_{x \downarrow z} B(x)\}$. If f is discontinuous and has a finite number of discontinuities, then Z is a finite non-empty set consisting of real numbers and $z_0 = \min(z \in Z)$ is well-defined. Furthermore, $y_0 = B(z_0)$ is also well-defined. Let $\overline{\sigma}$ satisfy:

$$B\left(\int_{\underline{\underline{x}}}^{\infty} \frac{\varphi(x)}{\overline{\sigma}} \phi\left(\frac{x-\theta}{\overline{\sigma}}\right) dx\right) = y_0.$$

Since for all $\hat{x} < \theta$, $B\left(\int_{\hat{x}}^{\infty} \frac{\varphi(x)}{\sigma} \phi\left(\frac{x-\theta'}{\sigma}\right)\right)$ increases as σ decreases, and $B(s)$ is increasing in s, then there exists some $\epsilon > 0$ such that $H(x)$ is continuous on the interval $[\underline{\underline{x}}, \underline{\underline{x}} + \epsilon)$ for all $\sigma < \overline{\sigma}$ and we can apply the preceding argument. Thus for any $\hat{x} \in (\underline{\underline{x}}, \underline{\underline{x}} + \epsilon)$, there exists some $\sigma > 0$ sufficiently small that $H(\hat{x}) > 0$.