Abstract

We model the reporting of sexual misconduct. Individuals under-report misconduct due to strategic uncertainty over whether others will report and corroborate a pattern of behavior. Under-reporting occurs if and only if misconduct is widespread. Making sanctions more responsive to reports, raising public awareness of misconduct, implementing confidential holding tanks, and appropriately calibrating damage awards can encourage reporting. However, we also show when such policies are ineffective or backfire. Managers may avoid mentoring subordinates, spilling over into reporting. A holding tank may discourage reporting by raising the bar to access reports. Overall, we highlight several unintended and intended consequences of #MeToo.

Keywords: Misconduct, Reporting, Retaliation, Coordination
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 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?

To answer such questions, this paper builds a model of the decision to report sexual misconduct and applies it to analyze the effectiveness of encouraging reporting. The model features agents who are reluctant to report misconduct by a manager because they face strategic uncertainty over whether other agents will also report and corroborate a pattern of behavior. In the context of sexual misconduct, corroboration is critical for an outside party to sanction a manager and to mitigate possible retaliation, stigma, or reprisal for an agent. The need for corroboration leads to a coordination problem among agents. However, agents face strategic uncertainty due to a realistic information friction: any one agent has imperfect information over whether other agents have privately experienced misconduct.

More specifically, agents in the model 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). 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 outside party is more likely to sanction the manager if other agents also report, resulting in a coordination problem. Agents form beliefs about how many other agents might report misconduct based on their own experience. In the equilibrium, agents employ threshold strategies, reporting if and only if their experience is worse than a unique reporting threshold.

We show that agents under-report misconduct if and only if misconduct is widespread. We

\footnote{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.}
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 experiences severe misconduct or when most agents experience misconduct. If misconduct is widespread, a lower threshold benefits some agents who previously did not want to report by providing significant corroboration. If misconduct is not widespread, a lower threshold makes those agents worse off by providing little additional corroboration. Our results characterize the Pareto-optimal threshold for agents.

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.

Under-reporting is a robust consequence of the model’s core feature: The information friction makes agents less willing to report due to strategic uncertainty in the coordination problem. This insight holds when we relax assumptions over the distribution of experiences, agent priors, and the outside party’s actions, and when multiple threshold strategy equilibria may occur.

We apply our framework to study the #MeToo movement’s implications 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 show that making sanctions more responsive to reports and raising public awareness of misconduct, two key elements of #MeToo, can encourage reporting by mitigating the coordination problem. Increasing sanction responsiveness makes agents more optimistic that the outside party will act on reports, and raising public awareness makes agents more optimistic about the number of agents that will report.

Next, 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. If manager 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. These spillover effects can cause unintended effects for policies designed to encourage reporting.

We also study whether a holding tank for confidential reports encourages reporting. 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 depends crucially on agents’ utility gain from filing a report that may never be released. Holding tanks face a surprising conundrum: 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.

Finally, we study awards for damages, whereby the outside party awards victims if it sanctions a manager for misconduct. In practice, the legal system provides for compensatory and punitive damages for victims of sexual harassment, although we do not restrict the interpretation of awards in the model. We show that awards mitigate the coordination problem and can be calibrated to a benchmark level where agents report as if this coordination problem did not exist; however, awards greater than this calibrated level incentivize agents to report a broader range of behaviors than in this benchmark.

Our central contribution is to show that information frictions can explain the under-reporting of sexual misconduct and to analyze the effectiveness of various policies for mitigating these frictions and encouraging reports. We show how these frictions and strategic uncertainty generate new applied insights for how public awareness of misconduct, endoge-
nous selection of managers into mentorship, confidential holding tanks, and damage awards affect reporting incentives. Our approach relies on the global games framework developed and advanced by Carlsson and van Damme (1993a,b) and Morris and Shin (1998).

Our approach complements others in the literature that study coordination and reporting. Pei and Strulovici (2019) study how adjudication criteria affect a strategic principal’s incentive to commit multiple crimes and the informativeness of reports. 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. Naess (2020) studies reporting incentives when the distribution of crime is common knowledge, rather than unknown to agents, and criminal behavior is exogenous. 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.\footnote{Several papers study misconduct outside of the context of coordination. Basu (2003), Chen and Sethi (2018), and Hersch (2018) study the ramifications of misconduct, and Chassang and Miquel (2019) study how to optimally elicit reports from a single whistleblower.}

Our insights apply not only to sexual misconduct but also to other types of misconduct where strategic uncertainty and fears of retaliation may chill speech, reporting, or otherwise create a “culture of silence” where agents do not speak up or express concerns. Abusive workplace behavior, certain discriminatory behavior, and misconduct that depends on agents’ private perceptions of behavior come to mind. Our insights may also apply to settings where outside parties need corroboration from multiple reports before beginning a more thorough investigation of forensically-verifiable misbehavior such as embezzlement or corruption. The implications of strategic uncertainty for dealing with different types of misconduct are fruitful areas for future research.

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 $\theta$ is the average agent experience:

$$x_i = \theta + \sigma \epsilon_i \text{ for } \epsilon_i \sim N(0, 1),$$

where $\epsilon_i$ is i.i.d. across agents. A larger $\sigma > 0$ reflects a wider
range of heterogeneity. We assume the manager exogenously generates agent experiences, which focuses attention on agents’ reporting decisions taking their experiences as given.

Experiences differ across agents because different agents have different interactions with the manager. Each experience $x_i$ reflects agent $i$’s interpretation of their interaction with the manager. We do not model the manager’s experiences with agents, but the manager may interpret an interaction with agent $i$ differently than the agent does, 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 (2020) propose a model of why different groups, such as men and women, may interpret the same behavior according to different moral standards.

Agents do not know the average experience $\theta$ but learn about it through their own $x_i$. To start, we assume agents have improper uniform priors over $\theta$. 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 key friction in our model is that agents’ experiences are private information. 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 Online Appendix shows that the assumptions of normality (and unbounded support) for $\epsilon_i$, the improper uniform prior over $\theta$, and the continuum of players can be relaxed. Section 2 allows the manager a choice to avoid interacting with agents at all.

3The manager’s intent does not affect our analysis of how policies affect agents’ reporting behaviors as agents truthfully believe they experienced $x_i$ and therefore have utility that depends only on $x_i$.

4While primarily a modeling device, unbounded support does reflect the plausible observation that there is at least some possibility that two parties interpret an interaction extremely differently.
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 break the law.\(^5\)

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 agent experiences. Negative values of \(x_i\) correspond to good experiences. Very high-\(\theta\) manager types frequently generate experiences of misconduct. Within the model, one can interpret managers who never commit misconduct in the real world as \(\theta \ll 0\) manager types who generate \(x_i > 0\) with arbitrarily small probability.

**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. More reports corroborate a pattern of behavior, and corroboration provides “safety in numbers” by making it more likely for an outside party to act on those reports and by mitigating possible retaliation or reprisal. Real-world examples of the importance of corroboration in the decision to report misconduct abound.\(^6\)

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,

---

\(^5\) U.S. civil law defines sexual harassment to include 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 of “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).

\(^6\) As The Economist (2020) writes, “Numbers Matter.” 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 jury deadlock. A key factor in his conviction upon retrial was the additional testimony of five accusers (Bowley, 2018).
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 (Lisak et al., 2010; Lonsway, 2010; Kelly et al., 2005; Lovett and Kelly, 2009).

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, or the need for the outside party to convince still others that sanction is justified. 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 (e.g., allegations at CBS, Ford, Nike, and Uber; Koblin, 2018; Stewart, 2018; Chira and Einhorn, 2017; Creswell et al., 2018; Isaac, 2017).

If an outside party fails to act on an agent’s report, that agent is more likely to face costs such as retaliation from managers and co-workers or social stigma (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; Weber, 2018). In criminal misconduct cases, fear of reprisal is the most-cited concern for why victims do not report (U.S. Department of Justice, 2013).

---

A famous line from the movie *A Few Good Men* summarizes this channel well when fictional lawyer Dan Kaffee exclaims: “It doesn’t matter what I believe, it only matters what I can prove!”
1.2 Agent strategies and payoffs

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. For example, 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. 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.\(^8\)

If a mass of \( r \) agents report misconduct by playing \( s(x_i) = 1 \) for \( x_i > 0 \), the outside party sanctions the manager with probability \( \Gamma(r) \). The function \( \Gamma(r) \) reflects the reduced form mapping for how reports of misconduct translate into action by the outside party. We allow for 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 \( r \). This assumption captures, in reduced form, the importance of corroboration through any of the channels described in Section 1.1. In the Online Appendix, we provide a few stylized models that endogenize a weakly increasing \( \Gamma(r) \) through these channels and show that our main results are qualitatively similar. 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. For general \( \Gamma(r) \), one would replace \( \gamma/2 \) in Equation 1 with \( \int_0^1 \Gamma(r) \, dr \), and our main insights would remain unchanged.

We specify payoffs as follows. Agents have an intrinsic motivation to report misconduct that provides payoff \( \omega x_i \) 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 \( \beta x_i \) if \( x_i > 0 \). On net, the agent receives \( \omega x_i - c \) from reporting and incurs additional cost \( \beta x_i \) if the outside party does not sanction the manager.\(^9\) An agent

---

\(^8\)We assume that agents truthfully reveal their \( x_i \) to the outside party if they report. Our 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. Although the presence of fabricating agents could affect equilibrium levels of truthful reporting, the coordination problem would still exist amongst the truthful agents due to the stronger incentive to report when others report.

\(^9\)It is plausible in the context of sexual misconduct that an agent receives reporting utility based on \( x_i \).
Table 1: Agent’s Payoffs. Sanction occurs with probability $\Gamma(r) = \gamma r$, where $r$ is the number of agents reporting $x_i > 0$. The parameter $\omega$ is the intrinsic motivation to report experience $x_i$, while $\beta$ is the retaliatory disincentive, and $\omega > \beta$. The parameter $c$ is a fixed cost of reporting. $1_{[x_i>0]}$ is an indicator function that equals one if $x_i > 0$.

<table>
<thead>
<tr>
<th>Sanction</th>
<th>No Sanction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report</td>
<td>$x_i\omega - c$</td>
</tr>
<tr>
<td>No Report</td>
<td>0</td>
</tr>
</tbody>
</table>

who does not report receives a normalized payoff of zero, as the direct effect of the experience on her utility is sunk.\(^{10}\) 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. 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 an agent incurs a retaliatory cost only for reports of misconduct ($x_i > 0$). We assume that both benefit $\omega$ and cost $\beta$ 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: $\pi(r, x_i) = Eu(\text{Report}, r, x_i) - Eu(\text{No Report}, r, x_i)$, where the expectation is taken over the possibility of sanction given $r$. Based on Table 1, we have:

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

rather than $\theta$. Take, for example, a subordinate who reports that the CEO engaged in sexual misconduct with her ($x_i > 0$). The CEO might have different interactions with other subordinates that those agents perceive as good ($\theta < 0$), but even if the agent knew this, she might truthfully believe she privately experienced misconduct and receive reporting utility based on that experience.

\(^{10}\)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 reduce the incentive to report because it introduces an incentive for agents to free-ride on others’ reporting and not report themselves. However, as long as the agent still prefers to report when others report (i.e., the payoff gain function is monotone in $r$), then the qualitative insights remain because the coordination problem and strategic uncertainty remain. The Online Appendix contains details.
The payoff gain is monotone in \( x \) and \( r \). The monotonicity in \( r \) is the key source of the coordination problem. An agent is less likely to incur the retaliatory cost when \( r \) is higher because sanction is more likely when more agents report. Sufficiently high \( r \) makes reporting optimal \((\pi(r, x) > 0)\) while sufficiently low \( r \) makes reporting not optimal \((\pi(r, x) < 0)\) for agents with \( x_i \in (\underline{x}, \bar{x}) \), where \( \underline{x} \equiv \frac{c}{\omega - \beta (1 - \gamma)} \) and \( \bar{x} \equiv \frac{c}{\omega - \beta} \). Not reporting is a dominant strategy for agents with \( x_i < \underline{x} \), and reporting is a dominant strategy for agents with \( x_i > \bar{x} \).

### 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 \( x_i = \theta \in (\underline{x}, \bar{x}) \) \( \forall i \) and all agents learn \( \theta \) 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, it is optimal to report because the payoff gain of reporting is positive: \( \pi(1, x) > 0 \) for \( x \in (\underline{x}, \bar{x}) \). 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, then it is optimal to not report because \( \pi(0, x) < 0 \), precipitating a self-fulfilling “no-report” equilibrium. The all-report equilibrium payoff-dominates the no-report equilibrium because \( \pi(1, x) > \pi(0, x) \) for any \( \theta > \underline{x} \).

### 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 \( \theta \) but learn about it through their own experience \( x_i \). Proposition 1 states the unique equilibrium in the baseline game where agents have improper uniform priors over \( \theta \).

**Proposition 1** (Baseline Equilibrium). *Suppose agents have improper uniform priors over \( \theta \). There exists a globally unique symmetric threshold strategy equilibrium where all agents play a threshold strategy and report \( (s(x_i) = 1) \) if and only if \( x_i \geq x^* \), where*

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

(*1*)
and \( x^* > x > 0 \). The threshold strategy is the unique strategy that survives the iterated deletion of strictly dominated strategies. In 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) and several others in finance and macroeconomics (see Morris and Shin, 2003 and Angeletos and Lian, 2016 for detailed reviews). We provide the key details in Appendix A and the Online Appendix following Morris and Shin (2003) with minor modification.

The key insight is that the information friction leads to strategic uncertainty over how many other agents will report. An agent anticipates other agents’ threshold strategies \( s(\cdot) \), but does not know other agents’ \( x_i \) and thus the number of agents \( r \) that will report. She forms posterior beliefs about \( r \) by updating on \( \theta \) given her own \( x_i \). Nevertheless, the marginal agent whose experience is just at the threshold \( x_i = x^* \) has a Uniform\([0, 1]\) belief over \( r \) and is effectively uninformed about the actions of others.\(^{11}\) Taking expectations of Equation 1 over \( r \) given this belief pins down \( x^* \). The equilibrium is dominance-solvable and globally unique due to the monotonicity and the dominance regions of \( \pi(r, x) \).

### 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 \( \theta \). Below, we show that under-reporting occurs if and only if misconduct is widespread. Under-reporting occurs due to 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.

\(^{11}\)Heuristically (following Morris and Shin, 2003), 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 \( 1 - \Phi((x - \theta)/\sigma) \) given the true \( \theta \), where \( \Phi(\cdot) \) is the normal cumulative distribution function. Re-arranging, the proportion is less than \( r \) if \( \theta < x - \sigma\Phi^{-1}(1 - r) \). If an agent has an improper uniform prior over \( \theta \), then her posterior belief over \( \theta \) is normally distributed with mean \( x \) and standard deviation \( \sigma \). Then, the probability the proportion falls below \( r \) equals \( \Phi\left((x - \sigma\Phi^{-1}(1 - r) - x)/\sigma\right) = r \).
**Corollary 1.1** (Open-secret equilibrium). Let \( \theta \in (x, x^*) \). 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 \( \bar{x} < x^* \).

Corollary 1.1 describes what we call an open secret equilibrium, an extreme form of under-reporting. For intuition, consider the case where \( \sigma \) is positive but vanishingly small and let \( \theta \in (x, x^*) \) so that all agents have nearly identically bad experiences. All agents know that all other agents have nearly identical experiences (agents know \( \sigma \) is small), know that all other agents share nearly identical beliefs about \( \theta \) (the manager’s type is an open secret”), and know that all-reporting is the best outcome for all agents \( (\pi(1, x) > 0 \geq \pi(0, x) \) 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 \( x < x^* \), 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 \( \theta \) 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, 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 are not common knowledge, creating uncertainty over other agents’ actions even though there is little uncertainty over \( \theta \) 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 \( \bar{x} < x^* \) that generates the maximal Pareto improvement. Corollary 1.1 is an application of condition (2).

**Proposition 2** (Under-reporting with widespread misconduct). Given \( \theta \), there exists an \( \bar{x} < x^* \) such that playing a threshold strategy around \( \bar{x} \) 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. $\theta \in (x, x^*)$ and $\sigma \leq \sigma(\theta)$ for a finite $\sigma(\theta)$.

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 $\frac{\partial x}{\partial \sigma} > 0$ and $\lim_{\sigma \to 0} \tilde{x} = x$.

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 $\theta$ is in $(\bar{x}, x^*)$, 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 $\sigma$ makes this large increase in $r$, and thus a Pareto improvement, possible. Indeed, whenever a Pareto improvement exists, a smaller $\sigma$ 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 $\theta \in (\bar{x}, x^*)$ and $\sigma$ is large, or $\theta < \bar{x}$, 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 $\theta$ 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.

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.
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.\footnote{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 $\theta$. This never occurs because a higher reporting threshold always leads to less expected utility for agents who report with $x_i > \bar{x}$.} \footnote{Equilibrium uniqueness is not a general property of global games models. Weinstein and Yildiz (2007) points out that small perturbations can select any of the multiple-equilibrium outcomes of the full information game as the uniquely rationalizable outcome. The key applied lesson of Proposition 2 is that, due to strategic uncertainty, agents may not report misconduct even when misconduct is so widespread that reporting is in their broader interests as a group. Appendix A.2.2 shows that the Proposition is a general statement that applies to any threshold strategy equilibrium and holds even in variations of the model that may lead to multiple threshold strategy equilibria.}

1.6 Effects of #MeToo on reporting

The #MeToo movement actively encourages people to “believe women”; in the model, this most closely corresponds with increasing $\gamma$, the sensitivity of sanction to reporting. One can also interpret increasing $\gamma$ as decreasing an organization’s tolerance for misconduct. Increasing $\gamma$ lowers the equilibrium reporting threshold $x^*$ and makes the marginal reported behavior less severe. A lower reporting threshold directly increases the equilibrium number of reports, or aggregate reporting, $\hat{r}(\theta) \equiv \int_0^\infty s(x) f(x | \theta) dx = \Phi\left(\frac{\theta - x^*}{\sigma}\right)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ are the normal density and cumulative distribution functions.

The #MeToo movement also raises public awareness of sexual misconduct. The 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).

The model predicts that heightened public awareness of misconduct also lowers reporting thresholds and increases aggregate reporting. The Online Appendix considers a model extension where agents have a normally-distributed common prior over $\theta$ with mean $y$ and standard deviation $\tau$ instead of an improper uniform prior. Heightened awareness of misconduct most closely corresponds with an increase in $y$, or a shift in the common prior towards a belief that the average agent experience is worse. In this extension, agents play symmetric threshold strategies in equilibrium, and an increase in $y$ decreases the equilibrium reporting threshold, increases reporting, and mitigates under-reporting. Intuitively, if $y$ increases and public beliefs about the experiences of women become worse than previously thought, the marginal agent becomes more confident that other agents will come forward. She is then
more willing to report, even keeping her experience constant, resulting in a lower equilibrium reporting threshold. The public nature of the shift in priors is crucial because it implies the agent knows that all agents share this shift.\footnote{The symmetric threshold strategy equilibrium is globally unique and dominance-solvable when agent experiences are fairly informative relative to priors ($\sigma$ is low relative to $\tau$). The Online Appendix provides the exact uniqueness condition and how the condition is plausibly satisfied due to the widespread use of non-disclosure agreements (NDAs) in settlements. Nevertheless, even if multiple equilibria exist, there exists a symmetric threshold strategy equilibrium where the threshold falls when $y$ increases.}

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, 2020). For brevity, we refer to such soft interactions under the umbrella term of mentorship.

Suppose a manager knows his own type $\theta$, and chooses whether or not to mentor agents, $a \in \{0, 1\}$.\footnote{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.} If $a = 1$, he mentors all agents, after which all agents realize exogenous experiences $x_i$ and then make their reporting decisions. An mentored agent 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: $s(x_i, 0) = 0$.\footnote{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.} As before, the manager’s type $\theta$ determines the average agent experience. We continue to assume agents have improper uniform priors over $\theta$.

We assume the manager’s payoff from mentoring (playing $a = 1$) equals $M(\theta) - \gamma r S$, where $M(\theta)$ is a function that captures the manager’s utility from mentoring, $\gamma r$ 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 $\theta$. For example, low-$\theta$ types may
include managers who derive utility from mentoring women because they believe there are
altruistic benefits from doing so, while high-$\theta$ types may include managers who derive utility
from sexual misconduct.

Finally, we assume the penalty $S$ is fixed irrespective of $\theta$. 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

2.1 Strategic spillovers

The manager mentors in equilibrium if and only if $M(\theta) - \gamma r S > 0$, given agents’ equilibrium
reporting strategies. Proposition 3 shows that the equilibrium decision to mentor creates a
strategic spillover between the selection of manager types who mentor and the equilibrium
level of reporting by agents. If high-misconduct (high-$\theta$) 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 $\theta$.

**Proposition 3** (Selection and strategic spillovers). Let $x^*_S$ denote the reporting threshold
such that agents report $(s(x_i, 1) = 1)$ if and only if $x_i \geq x^*_S$.

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

   (a) 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 and uniqueness of such an equilibrium is
if $M(\theta)$ is a weakly decreasing function of $\theta$ with $\lim_{\theta \to -\infty} M(\theta) > 0$ and $S > \frac{M}{\gamma}$,
where $M \equiv \min M(\theta)$.
2. In any equilibrium where the manager mentors if and only if \( \theta \geq \tilde{\theta} \) for a unique finite \( \tilde{\theta} \), the threshold satisfies \( x^*_S < x^* \).

(a) 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 and uniqueness of such an equilibrium is for \( M(\theta) \) to be a weakly increasing function that strictly increases at a finite \( \theta' \) and satisfies \( M(\theta) < 0 \) for \( \theta < \theta' \) and \( M(\theta) \geq \gamma S \) for \( \theta \geq \theta' \).

For both Parts 1 and 2, \( \tilde{r} \) is determined in equilibrium by \( \tilde{r} \equiv \Phi \left( \frac{\tilde{\theta} - x^*_S}{\sigma} \right) \).

Part 1 considers the case where high-misconduct manager types opt out of mentoring and shows that agents are less willing to report misconduct 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 illustrates with an \( M(\theta) \) that discontinuously jumps at \( \theta' \). 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 \( \theta \), but there exists some \( \theta' \) such that \( M(\theta) < 0 \) for \( \theta < \theta' \), a set of intermediate-\( \theta \) types may also opt out in addition to the low-\( \theta \) types.

Overall, Proposition 3 illustrates 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.

### 2.2 Unintended effects of tools intended to encourage reporting

We now examine how policies 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 \( \gamma \). 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 \( \theta \). If \( S > m/\gamma \), there is a unique symmetric perfect Bayesian equilibrium in which the manager mentors if and only if \( \theta \leq \tilde{\theta} \). The marginal type \( \tilde{\theta} \) that mentors agents is given by:

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

The equilibrium reporting threshold for agents equals:

\[
x^*_S = \frac{c}{\omega - \beta \left(1 - \frac{m}{2S}\right)},
\]  

where \( x^*_S > \underline{x} > 0 \). The reporting for the marginal manager type who mentors equals \( \bar{r} = \frac{m}{\gamma S} \).

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 2 summarizes comparative statics within the game considered by Proposition 4.
Table 2: Comparative statics.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No Selection</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reporting Threshold</td>
<td>Aggregate Reporting</td>
</tr>
<tr>
<td></td>
<td>$x^*$</td>
<td>$r(\theta &lt; \tilde{\theta})$</td>
</tr>
<tr>
<td>$\omega$: Intrinsic motive</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>$c$: Reporting cost</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>$\gamma$: Sanction sensitivity</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>$\beta$: Retaliation</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>$\sigma$: Experience dispersion</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$S$: Sanction</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>$m$: Mentorship utility</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>

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.

**Corollary 4.1.** Increasing sanction $S$:

1. **Raises the reporting threshold for agents:** $\frac{\partial x^*_S}{\partial S} > 0$,

2. Can either lower or raise $\tilde{\theta}$:

$$\frac{\partial \tilde{\theta}}{\partial S} < 0 \text{ if and only if } \sigma > \phi \left( \Phi^{-1} \left( \frac{m}{\gamma S} \right) \right) \frac{c\gamma \beta}{2(\omega - \beta + \frac{3m}{2S})^2}.$$

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 $x^*_S$.

Part 2 shows that increasing sanction $S$ has ambiguous effects on the identity of the marginal type $\tilde{\theta}$ that chooses to mentor. An increase in $S$ makes the equilibrium value of $\tilde{r} = \frac{m}{\gamma S}$ fall and $x^*_S$ increase, and these effects change $\tilde{\theta}$ since $\tilde{\theta} = \sigma \Phi^{-1}(\tilde{r}) + x^*_S$ in equilibrium. When $\sigma$ is large, increasing $S$ leads $\tilde{\theta}$ to fall through the first term. Intuitively, greater sanctions lead fewer high-misconduct manager types to mentor. Conversely, when $\sigma$
Figure 1: Equilibrium $\tilde{\theta}$. This figure plots a non-monotone example of equilibrium $\tilde{\theta}$ as a function of $S$. Parameter values equal $\{\beta, c, \omega, \gamma, \sigma, S, m\} = \{0.85, 0.3, 1, 0.5, 2, 5, 2\}$.

is small, increasing $S$ can lead $\tilde{\theta}$ to rise through the second term. Intuitively, greater sanctions generate sufficiently higher reporting thresholds that more high-misconduct manager types mentor.

The possibility of non-monotonicity of $\tilde{\theta}$ in $S$ leads to the following striking conclusion: there can exist different levels of $S$ that result in the same level of expected sanction $\tilde{r}S$ and same marginal type $\tilde{\theta}$, yet the lowest $S$ among these leads to the lowest reporting threshold for agents and the highest aggregate reporting. One can see this by drawing a horizontal line in Figure 1, 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 equilibrium amount of reporting $\tilde{r}$.\textsuperscript{17}

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 $\gamma$, the sensitivity of sanction to reporting.

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

Intuitively, without selection, an increase in $\gamma$ increases an agent’s incentive to report because it increases the probability of sanction for any given level of reporting $r$. However,\textsuperscript{17}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.
Table 3: Agent’s Payoffs with Release Bar \( r \). The parameter \( \delta \) measures utility from making an unreleased report. \( 1_{[x_i > 0]} \) is an indicator function that equals one if \( x_i > 0 \).

<table>
<thead>
<tr>
<th></th>
<th>Released (( r \geq \bar{r} ))</th>
<th>Unreleased (( r &lt; \bar{r} ))</th>
<th>Sanction</th>
<th>No Sanction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report</td>
<td>( x_i \delta - c )</td>
<td>( x_i \omega - c )</td>
<td>( x_i(\omega - 1_{[x_i &gt; 0]}\beta) - c )</td>
<td></td>
</tr>
<tr>
<td>Not Report</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

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 \( \gamma \) on reporting to zero. One can show analogous results for a class of non-linear functions, \( \Gamma(r) = \gamma r^\frac{1}{n} \) where \( n > 0 \).

3 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 intent of 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 \( \bar{r} \in [0, 1] \) in which reports are held in a holding tank as long as \( r < \bar{r} \), and are released to the outside party only if \( r \geq \bar{r} \). Given the threshold \( \bar{r} \), the sanction function \( \Gamma(r) \) becomes equivalent to:\(^{18}\)

\[
\Gamma(r) = \begin{cases} 
0 & \text{if } r < \bar{r} \\
\gamma r & \text{if } r \geq \bar{r}.
\end{cases}
\]  

\(^{18}\)To make a clear comparison to Section 1, we continue to assume that \( \Gamma(r) = \gamma r \) once reports are released (if \( r \geq \bar{r} \)). But our conclusions apply to a general \( \Gamma(r) \) that weakly increases in \( r \) if \( r \geq \bar{r} \).
Table 3 summarizes the agent’s payoffs. If \( r \geq \tau \), then the agent’s payoffs are identical to her payoffs in the base model from Section 1.2. If \( r < \tau \), then the agent’s report is not released to the outside party, the manager cannot be sanctioned, and the agent receives \( x_\delta - c \) for \( \delta \in [0, \omega] \). The parameter \( \delta \) reflects the agent’s utility from filing a report that is not released. It has several non-mutually-exclusive interpretations. First, \( \delta \) could reflect intrinsic utility from reporting misconduct and is thus naturally weakly less than \( \omega \): a report that is never released delivers less intrinsic utility than a report that is released to the outside party.\(^{19}\) Second, \( \delta \) could reflect intrinsic utility net of (real or perceived) possible retaliation resulting from a “leak” from the holding tank. A lower \( \delta \) 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 \( \pi_H \) from reporting equals:

\[
\pi_H (r, x) = \begin{cases} 
  x\delta - c & \text{if } x > 0 \text{ and } r < \tau \\
  x(\omega - (1 - \gamma r)\beta) - c & \text{if } x > 0 \text{ and } r \geq \tau \\
  x\omega - c & \text{if } x \leq 0.
\end{cases}
\]

(5)

Note that the sanction function, agent payoffs, and payoff gain function \( \pi_H \) nest the model from Section 1 and outcomes determined by Proposition 1 when \( \tau = 0 \), which corresponds to the absence of a holding tank. We maintain the assumption that agents have improper uniform priors over \( \theta \) for tractability and comparability to Section 1.

Proposition 5 shows that raising the release bar \( \tau \) can either encourage or discourage reporting depending on \( \delta \). The reason is that raising \( \tau \) has two countervailing effects. First, a protection effect encourages reporting: raising \( \tau \) protects reporting agents from retaliation in the event that too few others report \( (r < \tau) \). Second, a raise-the-bar effect discourages reporting: raising \( \tau \) reduces an agent’s belief that an outside party investigates her report \( (r \geq \tau) \). Intuitively, the raise-the-bar effect exists because strategic uncertainty remains for an agent over whether or not other agents’ reporting will spill over the release bar \( \tau \). If \( \delta \) is too small, the raise-the-bar effect dominates, and raising \( \tau \) strictly discourages reporting.

**Proposition 5.** Suppose agents have improper uniform priors over \( \theta \).

1. If \( \delta \in [0, \omega - \beta] \), there exists a globally unique symmetric threshold strategy equilibrium

\(^{19}\)It is also possible that the reporting cost could be lower when a report remains in the holding tank than when a report is released. This is qualitatively similar to raising \( \delta \).
where agents use reporting threshold \( x^*_{H} \):

\[
x^*_{H} = \frac{c}{\delta \tau + \omega (1 - \tau) - \beta \left( 1 - \tau - \frac{1}{2} \gamma (1 - \tau^2) \right)}.
\]  

(6)

The policy \( \tau_{min} \) that minimizes \( x^*_{H} \) is \( \tau_{min} = 0 \) since \( \frac{\partial x^*_{H}}{\partial \tau} > 0 \).

2. If \( \delta \in [\omega - \beta, \omega] \), there exists a unique symmetric threshold equilibrium, with \( x^*_{H} \) described by Equation 6, when the closed-form expression \( \pi(x) \) defined by Equation A.17 in the Appendix satisfies a single-crossing condition. As a sufficient condition, the condition is satisfied when \( \sigma \) is sufficiently large.

When such an equilibrium exists, there exists a release bar \( \tau_{min} \in (0, 1) \) such that a holding tank with \( \tau = \tau_{min} \) minimizes \( x^*_{H} \) and \( x^*_{H}(\tau = \tau_{min}) < x^*_{H}(\tau = 0) \). This release bar is unique if \( \delta \in [\omega - \beta, \omega - \beta(1 - \gamma)] \) and equals 1 if \( \delta \in (\omega - \beta(1 - \gamma), \omega] \).

However, implementing a holding tank with sufficiently high \( \tau > \tau_{min} \) whenever \( \delta \in [\omega - \beta, \omega - \beta(1 - \gamma/2)] \) results in \( x^*_{H}(\tau) > x^*_{H}(\tau = 0) \).

Part 1 shows that instituting a holding tank deters reporting when \( \delta \) is sufficiently low: the policy \( \tau_{min} \) that minimizes \( x^*_{H} \) is \( \tau_{min} = 0 \). If \( \delta \) 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 \( \tau \) 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 \( \delta \) is sufficiently high, instituting a holding tank with release bar \( \tau = \tau_{min} \) encourages more reporting than if \( \tau = 0 \) and there were no holding tank. Intuitively, if there is no holding tank, the marginal agent knows that any reports will automatically go to the outside party, and introducing a holding tank with a small \( \tau \) has a negligible effect on raising the bar for release but non-negligible protection payoff given the sufficiently high \( \delta \). Thus, the protection effect dominates the raise-the-bar effect, and increasing \( \tau \) up to \( \tau_{min} \) makes the marginal agent more willing to report and lowers \( x^*_{H} \). The release bar \( \tau_{min} \) increases with \( \delta \) because the relative value of protection increases with \( \delta \).20

20The existence of a symmetric threshold strategy equilibrium is not guaranteed when \( \delta \) is sufficiently high because a high \( \delta \) 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 \( \sigma \) 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 \( \delta \) without the tank opening. With \( \sigma \) high, such deviations are less likely, and the threshold equilibrium may be sustained. No
Table 4: Agent’s Payoffs with Damage Awards. The parameter \( \alpha \) is the damage award. \( 1_{[x_i > 0]} \) is an indicator function that equals one if \( x_i > 0 \).

<table>
<thead>
<tr>
<th>Sanction</th>
<th>No Sanction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report</td>
<td>( x_i(\omega + 1_{[x_i &gt; 0]} \alpha) - c )</td>
</tr>
<tr>
<td>Not Report</td>
<td>0</td>
</tr>
</tbody>
</table>

However, Part 2 also shows that a holding tank with sufficiently high release bar \( r > r_{min} \) may discourage reporting relative to \( r = 0 \). Intuitively, increasing \( r \) past \( r_{min} \) progressively raises the bar and makes agents less willing to report, potentially leading the reporting threshold \( x_H^* \) to exceed the threshold that would prevail if \( r = 0 \).

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.\(^{21}\)

4 Awards for Damages

We next consider the effect of “awards for damages,” or payoffs that agents receive for filing reports that result in sanction. In the U.S., victims who win lawsuits for workplace sexual harassment can be awarded compensatory and/or punitive damages under Title VII of the 1964 Civil Rights Act. While such awards motivate our discussion, we do not restrict the interpretation of awards, as agents may also gain utility through a sense of vindication or non-monetary utility when the outside party acts on their reports.

We incorporate damage awards in the model by extending the payoff structure so that an agent receives a proportional award \( \alpha x_i \) if the third party acts on a claim of \( x_i > 0 \). Table 4 summarizes an agent’s final payoffs. Note that the payoffs in Table 4 nest the previous payoff structure with \( \alpha = 0 \). All previous results on existence, uniqueness, and comparative

agents have an incentive to deviate if the closed-form solution of agent’s expected payoffs given \( x \) when other agents play reporting threshold \( x^* \), given by \( \pi(x) \) in Equation A.17, satisfies the single crossing condition. As a sufficient condition, the Appendix shows that this condition is satisfied at the limit of \( \sigma \rightarrow \infty \).

\(^{21}\)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 5 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 5. In our model, the incentive for an agent to file a report that is never released when \( \delta \) is very high can also perversely sustain an equilibrium with \( r_{min} = 1 \).
statics of other parameters in Section 1 continue to apply with straightforward modification, and proofs of all results incorporate the payoff structure allowing for \(\alpha \geq 0\).

In this section, we focus on the role of \(\alpha\) for equilibrium behavior. It is straightforward to show that higher \(\alpha\) leads to more reporting because it mitigates the retaliatory cost \(\beta (\partial x^* / \partial \alpha < 0\) and \(\partial \hat{r}(\theta) / \partial \alpha > 0 \forall \theta)\). Proposition 6 captures our main insight from implementing damage awards.

**Proposition 6 (Required awards).** For every \(\beta \in (0, \omega)\), there exists a unique required award \(\alpha^F > 0\) such that the equilibrium reporting threshold equals the “no-retaliation benchmark,” \(x^F = c / \omega\), with:

1. In the baseline game (Section 1), \(\alpha^F = \beta \left(\frac{2 - \gamma}{\gamma}\right)\).
2. In the constant-\(m\) selection game (Section 2), \(\alpha^F = \beta \left(\frac{2s - m}{m}\right)\).
3. In the holding tank game (Section 3), \(\alpha^F = \frac{2}{\gamma(1 - \gamma)} \left(\beta (1 - \frac{1}{2} \gamma) + \bar{r}(\omega - \delta - \beta (1 - \frac{1}{2} \gamma \bar{r}))\right)\).

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

Proposition 6 suggests that awards for coordination success have the potential to significantly and even fully offset the retaliatory cost of coordination failure. A sufficiently high \(\alpha\) 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 \(x^E \equiv \bar{x}\) in equilibrium.\(^{22}\) Even at such a threshold, however, an agent faces possible retaliation if \(\gamma < 1\). An even higher \(\alpha\) 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 an award 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, \(\alpha\) 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 award, which in this case mitigates the fixed cost of reporting \(c\). With \(\alpha = 0\), changes in \(\beta\) or \(\gamma\) cannot generate an equilibrium threshold less than \(x^E\), let alone \(x^F\).

\(^{22}\)The threshold \(x^E\) is also the greatest lower bound of all possible Pareto-optimal thresholds \(\bar{x}\) from Proposition 2 and is plausibly relevant for a policymaker who does not know \(\theta\).
We highlight three other observations from Proposition 6. Part 1 suggests that there is a potential substitution across different policy tools: The required award $α^F$ is lower when $γ$ is high or $β$ is low. Part 2 suggests that $α^F$ is higher when managers can opt out of mentorship since $S > m/γ$. Part 3 suggests that the required award decreases in $δ$ since agents derive more utility from reports even when they remain in the holding tank.

5 Robustness and Further Discussion

As Section 1 noted, the assumptions of unbounded support on $ε_i$, improper uniform prior over $θ$, and linear $Γ(r)$ can be relaxed. Here, we discuss further issues.

5.1 Social welfare

Social welfare analysis 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, the manager’s intent and utility, the utility function $M(θ)$ of mentorship for managers, the benefits of mentorship for agents, and whether managerial sanctions are dead-weight loss. Such an analysis also requires distributional assumptions on $θ$ and welfare weights for agents and managers.

5.2 The outside party and $Γ(r)$

5.2.1 What are the beliefs of the outside party?

The outside party learns the average agent experience $θ$ ex-post. This is because the ex-post aggregate reporting $\hat{r}(θ) = Φ\left(\frac{θ - x^*}{σ}\right)$ 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 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 $\bar{x} < x^*$, $r$ would be higher, and the outside party would be more likely
to act. Note that the outside party learns $\theta$ irrespective of whether agents play $\bar{x}$ or $x^*$ as long as it correctly anticipates agents’ reporting thresholds. Realistically, however, the higher $r$ associated with $\bar{x}$ may increase the likelihood that the outside party acts because it 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 and counters agents’ reports, which the model allows without loss of generality.

5.2.2 What if reported $x_i$’s directly affect the probability of sanction?

The sanction function $\Gamma(r)$ incorporates information about reported $x_i$ through the outside party’s inference about $\theta$ because $r$ is a sufficient statistic for $\theta$. However, the outside party may care disproportionately about some values of reported $x_i$ beyond their inference of $\theta$.

The Online Appendix 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 $\theta$, 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 $\Xi(\{n(x), x\}) \equiv B(\int_x \varphi(x)n(x)dx)$, 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.

5.3 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.

6 Conclusion

This paper shows that strategic uncertainty leads agents to under-report sexual misconduct and highlights several effects of #MeToo relevant for policy and future empirical research.

On the one hand, the model predicts that #MeToo has several intended effects. Making sanctions more responsive to reports, heightened public awareness of misconduct, properly-designed holding tanks, and appropriately-calibrated damage awards can encourage agents to report misconduct.

On the other hand, the model predicts that #MeToo may have unintended effects. Managers may opt out of mentorship, and tools intended to encourage reporting may be less effective than supposed due to strategic spillovers from this opt-out. A holding tank may discourage reporting if it excessively raises the bar for outside parties to access reports.

Future research should further explore the implications of strategic uncertainty for the reporting of misconduct. As noted in the introduction, our insights likely apply to other forms of misconduct, not just sexual misconduct. Further research should also explore how policies that encourage communication among victims of misconduct can encourage reporting. In general, the effects of communication on behavior are rich and complex.\(^{24}\) In our setting, on the one hand, knowing that other agents have experienced misconduct may plausibly encourage agents to report. On the other hand, other communication may be noisy and also may not result in common knowledge about the incidence of misconduct, leaving some strategic uncertainty intact (Morris and Shin, 2003).

\(^{23}\)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 \( \theta \) could improve total agent payoffs without transferring information across agents by directing all agents to play threshold strategies around \( x \). Our definition of under-reporting is related but distinct: it conditions on \( \theta \) 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.

\(^{24}\)Examples of recent theoretical work include Calvó-Armengol et al. (2015) and Sethi and Yildiz (2016). Cooper et al. (1992) provide results from a seminal experiment on communication in coordination games.
References


Appendix A  Proofs of Main Propositions

We prove statements in the text including the possibility that payoffs include \( \alpha \), so that:

\[
\pi(r, x) = \begin{cases} 
  x(\omega + \gamma r \alpha - (1 - \gamma r) \beta) - c & \text{if } x > 0 \\
  x\omega - c & \text{if } x \leq 0
\end{cases}
\]

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 (2004), 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 B2, which is not a direct application. We refer to the reader to Morris and Shin (2003) and Morris and Shin (2004) for their proofs but also provide step-by-step proofs with the appropriate technical modifications to our setting in the Online Appendix.

Note that here we do not require \( \epsilon \) 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 \( \epsilon \), 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 \( x^* > 0 \) solving \( \int_0^1 \pi(r, x^*) \, dr = 0 \).
- **P4. Limit dominance.** There exist \( x \in \mathbb{R} \) and \( x' \in \mathbb{R} \) such that:
  1. \( \pi(r, x) < 0 \) for all \( r \in [0, 1] \) and \( x \leq x' - \varepsilon \); and
  2. \( \pi(r, x) > 0 \) for all \( r \in [0, 1] \) and \( x \geq x' + \varepsilon \), 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:

\[
\frac{\partial \pi}{\partial x}(r, x) = \begin{cases} 
  \omega - \beta + \gamma r (\alpha + \beta) & \text{if } x > 0 \\
  \omega & \text{if } x \leq 0
\end{cases}
\]

While \( \omega - \beta + \gamma r (\alpha + \beta) > 0 \) if \( x > 0 \), \( \omega > 0 \) if \( x \leq 0 \).

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

\[
\int_0^1 x\omega - c \, dr = x\omega - c,
\]

so \( x^- = \frac{x}{\omega} \). But then \( x > 0 \), a contradiction.
irrespective of what other agents play, so playing a cutoff of $x$ is a well-defined decreasing sequence, with $\lim_{x \to 0} \pi(r,x) < 0$ for all $(r,x) \in [0,1] \times (x,\infty)$. Not reporting is payoff dominant if $x \leq \bar{x} - \varepsilon$ where $\bar{x} \equiv \frac{c}{\omega - \beta}$, since $\pi(r,x) < 0$ for all $(r,x) \in [0,1] \times (-\infty, \bar{x})$.

[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 $\lim_{x \to 0} \pi(r,x) = \lim_{x \to 0} \pi(r,x) = \pi(r,0) = -c$. 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 $\theta$ 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 $\theta$:

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

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 $k \geq x$. It suffices to consider $k \geq x$ by iterated deletion of strictly dominated strategies: from the definition of $\pi$, not reporting strictly dominates reporting for any $x < x$ irrespective of what other agents play, so playing a cutoff of $k = x$ strictly dominates playing a cutoff strategy of $k < x$. The one-to-one mapping from $r$ to $\theta$ for $k \geq x > 0$ allows us to write:

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

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

**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 $\{\xi^1, \xi^2, \ldots, \xi^n, \ldots\}$ defined as the solutions to the equations:

$$\pi^*(\xi^1, 0) = 0$$
$$\pi^*(\xi^2, \xi^1) = 0 \ldots$$
$$\pi^*(\xi^n, \xi^{n-1}) = 0 \ldots$$

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

$$\pi^*(\xi^1, \infty) = 0$$
$$\pi^*(\xi^2, \xi^1) = 0 \ldots$$
$$\pi^*(\xi^n, \xi^{n-1}) = 0 \ldots$$

is a well-defined decreasing sequence, with $\lim_{n \to \infty} \xi^n = \bar{\xi}$, 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 $\bar{\xi}$. 

36
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 $s(\xi) = \begin{cases} 0 \text{ [do not report]} & \text{if } \xi < \xi^n, \\ 1 \text{ [report]} & \text{if } \xi > \xi^n. \end{cases}$

Proof. This Lemma follows from Lemma A.2 of Morris and Shin (2004) 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. Let $\xi$ 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, $x^* = \xi$. Thus, $x^*$ is the unique symmetric threshold strategy equilibrium.

Furthermore, by Lemma A2, $x^* = \xi = \lim_{n \to \infty} \xi^n = \lim_{n \to \infty} \xi^\prime$, 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.

Finally, the proof of Lemma A2 Part 3 in Appendix B shows that the belief of the marginal agent over $r$ is uniformly distributed over $[0, 1]$.

A.2 Proof of Proposition 2

A.2.1 Main proof

Proof. Let $H(x) \equiv x(\omega - \beta + \gamma\Phi(\frac{\theta - x}{\sigma})(\alpha + \beta)) - c$. Then $H(x)$ is the expected payoff of the marginal agent given that all agents use reporting threshold $x$ and given the true $\theta$. 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,

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

Since $\theta > x^*$, then $\frac{\partial^2 H}{\partial x^2} < 0$ for all $x \in (x, x^*]$. Since $H(x) < 0$ and $H(x^*) > 0$, then this implies that $\frac{\partial H}{\partial x} > 0$ for all $x \in (x, x^*]$. Let $\tilde{x}$ be the threshold such that $H(\tilde{x}) = 0$. Given that $H(x) < 0$ and $H(x^*) > 0$, such a $\tilde{x} < x^*$ exists. Since $\frac{\partial H}{\partial x} > 0$ for all $x \in (x, x^*]$, 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 $x': W(\theta|x') \equiv \int_{x'}^{\infty} [x(\omega - \beta + \gamma r(\theta|x')(\alpha + \beta)) - c] f(\theta) d\theta$. Since $\frac{\partial W(\theta|x')}{\partial x} < 0$ and
\[ \frac{\partial}{\partial x} \left[ x(\omega - \beta + \gamma r(\theta|x') (\alpha + \beta)) - c \right] > 0, \] then for any \( x' \in [\tilde{x}, x^*], \)

\[ W(\theta|x') = \int_{x'}^{\infty} \left[ x(\omega - \beta + \gamma r(\theta|x') (\alpha + \beta)) - c \right] f(x|\theta)dx \]

\[ = \int_{x'}^{x^*} \left[ x(\omega - \beta + \gamma r(\theta|x') (\alpha + \beta)) - c \right] f(x|\theta)dx + \int_{x^*}^{\infty} \left[ x(\omega - \beta + \gamma r(\theta|x') (\alpha + \beta)) - c \right] f(x|\theta)dx \]

\[ > \int_{x'}^{x^*} \left[ x(\omega - \beta + \gamma r(\theta|x') (\alpha + \beta)) - c \right] f(x|\theta)dx + \int_{x^*}^{\infty} \left[ x(\omega - \beta + \gamma r(\theta|x') (\alpha + \beta)) - c \right] f(x|\theta)dx \]

\[ = \int_{x'}^{x^*} \left[ x(\omega - \beta + \gamma r(\theta|x') (\alpha + \beta)) - c \right] f(x|\theta)dx + W(\theta|x^*) \]

\[ > W(\theta|x^*), \]

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 \( \frac{\partial H}{\partial \sigma} > 0 \) for all \( x \in (x, x^*], \) 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, \( \lim_{\sigma \to 0} \tilde{x} = \bar{\tilde{x}}. \)

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

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

Also, \( \lim_{\sigma \to \infty} \frac{\partial H}{\partial \sigma} = \omega - \beta + \gamma(\alpha + \beta). \) By direct differentiation, \( \frac{\partial H}{\partial \sigma} > 0 \) for all \( x > \theta. \) When \( \theta \leq x^*, \) then \( H(x^*) \leq 0 \) and \( H(\theta) \leq 0 \) 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 \( [\tilde{x}, \theta] \) and satisfy the following conditions:

\[ \dot{\tilde{x}} \left( \omega - \beta + \gamma \Phi(\frac{\theta - \tilde{x}}{\sigma})(\alpha + \beta) \right) - c = 0 \]

\[ \omega - \beta + \gamma(\alpha + \beta) \Phi(\frac{\theta - \tilde{x}}{\sigma}) - \frac{\dot{\tilde{x}} \gamma(\alpha + \beta) \phi(\frac{\theta - \tilde{x}}{\sigma})}{\sigma} \geq 0. \]

Since \( H(\tilde{x}) < 0, H(x^*) \leq 0, \) \( \frac{\partial H}{\partial \sigma} < 0 \) when \( x < \theta, \) and \( \frac{\partial H^2}{\partial \sigma^2} < 0 \) for all \( x \in (x, \theta) \) when \( \sigma > 0, \) then such a \( \tilde{x} < x^* \) exists and is unique when \( \sigma \leq \sigma_0, \) where \( \sigma_0 \) is determined by the value of \( \sigma \) at which \( \max H(x) = 0 \) for \( x \in (\tilde{x}, \theta). \) This condition is satisfied for the pair \( (\tilde{x}, \sigma_0) \) that satisfy Equations A.4 and A.5:

\[ \dot{\tilde{x}}(\omega - \beta + \gamma(\alpha + \beta) \Phi(\frac{\theta - \tilde{x}}{\sigma})) - c = 0 \]

\[ \omega - \beta + \gamma(\alpha + \beta) \Phi(\frac{\theta - \tilde{x}}{\sigma}) - \frac{\dot{\tilde{x}} \gamma(\alpha + \beta) \phi(\frac{\theta - \tilde{x}}{\sigma})}{\sigma} = 0, \]

where Equation A.4 is the requirement that \( H(x^*, \sigma_0) = 0 \) and Equation A.5 is the requirement that \( H(x^*, \sigma) \) is a local maximum. We know that a solution to \( (\tilde{x}, \sigma) \) exists for any \( \theta \in (x^*, x^*) \) due to the following. First, \( \lim_{\sigma \to 0} H(x) = 0 \) and \( \lim_{\sigma \to 0} \frac{\partial H}{\partial \sigma} > 0 \) for all \( x \in (x, \theta), \) implying that \( \lim_{\sigma \to 0} H(x) > 0 \) for all \( x \in (x, \theta). \) Second, \( \lim_{\sigma \to \infty} H(x) < 0 \) for all \( x \in (\tilde{x}, \theta). \) Since \( \frac{\partial H}{\partial \sigma} < 0 \) and \( \frac{\partial H^2}{\partial \sigma^2} < 0 \) when \( x < \theta, \) then by continuity of \( H \) such a \( \sigma \) exists and is unique.
Note that if $\theta = x^*$, the explicit solution for $\sigma$ is $\sigma = \frac{c^\gamma (\alpha + \beta)}{2\pi (\omega - \beta + \frac{1}{2} \gamma (\alpha + \beta))}.$

Let $\theta \leq x^*$ and $\sigma < x$. Since $H(\tilde{x}) < 0$, $H(\tilde{x}) = 0$, and $\frac{\partial H^2}{\partial x} < 0$ when $x < \theta$, then $H(x) < 0$ for $x \in (\tilde{x}, 0)$ 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 $\frac{\partial \tilde{x}}{\partial \theta} > 0$. By direct calculation, $\lim_{\sigma \to a} \tilde{x} = \tilde{x}$.

\[\Box\]

### A.2.2 Application to any threshold equilibrium, including Proposition B2

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 2, not just the unique equilibrium identified by Proposition 1. Section B.6 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

$$E(\pi|x) = E(p^*(\omega x - c) + (1 - p^*)(\omega x - \beta x - c))$$

$$= x(\omega - (1 - E(p^*|x))\beta) - c.$$ 

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. 

\[\Box\]

### 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 $\theta$ 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 
\[
\Psi(r; x, k, \theta < \tilde{\theta}) = \int_{\theta=-\infty}^{\theta=0} f(\theta|x, \theta < \tilde{\theta}) d\theta
\]
\[
= \int_{\theta=-\infty}^{\theta=0} \left( \int_{-\infty}^{\theta} f(x|\theta) \right) \frac{f(\theta|x)}{f(x|\theta)} d\theta
\]
\[
= \int_{z=x-k+F^{-1}(1-r)}^{z=x-k+F^{-1}(1-r)} \frac{f(z)dz}{f(x|\theta)} = \int_{z=x-k+F^{-1}(1-r)}^{z=x-k+F^{-1}(1-r)} \frac{1-F(z)}{1-F(F^{-1}(1-r))}
\]
where we define \( \tilde{r} \) to be the reporting that corresponds to our upper bound \( \tilde{\theta} = k - \sigma F^{-1}(1 - \tilde{r}) \) and we perform a change of variables \( z = \frac{x-k}{\sigma} \) and \( dz = -d\theta \).

For the marginal agent who has \( x = k \), we have
\[
\Psi(r; x = k, k, \theta < \tilde{\theta}) = \frac{1-F(F^{-1}(1-r))}{1-F(F^{-1}(1-r))} = \frac{r}{\tilde{r}}.
\]
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 \( \theta \) (because \( r(\theta; k) = 1 - F(k/\sigma) \)), then the agent’s reporting threshold satisfies:
\[
\pi^*(x, k, \theta < \tilde{\theta}) = \int_{\theta=-\infty}^{\theta=\tilde{\theta}} 1 - \frac{F(k/\sigma)}{\sigma} f\left(\frac{x-\theta}{\sigma}\right) \pi\left(1 - F\left(\frac{k-\theta}{\sigma}\right), x\right) d\theta
\]
\[
= \int_{0}^{\tilde{r}} \psi^*_\sigma(r; x, k, \theta < \tilde{\theta}) \pi(r, x) dr
\]
For the marginal agent, this is 
\[
\pi^*(x, k, \theta < \tilde{\theta}) = \int_{0}^{\tilde{r}} \frac{1}{\tilde{r}} \pi(r, x) dr = 0.
\]
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:
\[
\pi^*(x, k, \theta < \tilde{\theta}) = 0 \quad \int_{0}^{\tilde{r}} \frac{1}{\tilde{r}} (x_S^*[\omega + \gamma r \alpha - (1 - \gamma r) \beta] - c) dr = 0
\]
\[
x_S^* = \frac{c}{\omega - \beta + \frac{1}{\gamma} (\alpha + \beta) \tilde{r}}.
\] (A.6)
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 \( \theta \) with \( \lim_{\theta \to -\infty} M(\theta) > 0 \) and \( S > \frac{M}{\gamma} \). For any given threshold \( x_S^* \), \( -\gamma S\tilde{r}(\theta) \) strictly decreases in \( \theta \). Since \( M(\theta) \) is weakly decreasing in \( \theta \) and \( S > \frac{M}{\gamma} \), type \( \theta \to \infty \) receives a strictly negative payoff from mentoring because he is sure that his \( \tilde{r}(\theta) = 1 \). 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 \( \tilde{r}(\theta) = 0 \). Since \( M(\theta) \) weakly decreases in \( \theta \) and \( -\gamma S\tilde{r}(\theta) \) strictly decreases in \( \theta \), then there exists a unique \( \tilde{\theta} \) such that \( M(\theta) - \gamma S\tilde{r}(\theta) = 0 \). 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 \( M(\tilde{\theta}) - \gamma S \tilde{r}(\tilde{\theta}) = 0 \).

[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 \( \theta \) 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 \( f(\theta | x, \theta > \tilde{\theta}) = \frac{f(x | \theta)}{1 - f(x | \theta) d\theta} \).

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

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

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

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

\[
\pi^*(x, k, \theta > \tilde{\theta}) = 0 \quad \int_{\tilde{r}}^{1} \left( \frac{1}{1 - r} \right) (x^*_S \omega + \gamma r \alpha - (1 - \gamma r) \beta - c) dr = 0 \\
\quad \quad x^*_S = \frac{c}{\omega - \beta + \frac{1}{2} \gamma (\alpha + \beta)(1 + \tilde{r})}. \tag{A.7}
\]

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 \( \theta \). For any given threshold \( x^*_S = -\gamma S \tilde{r}(\theta) \) strictly decreases in \( \theta \). Since \( g(\theta) \) is weakly increasing in \( \theta \) 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 \( \theta \), any type \( \theta \geq \theta' \) receives a strictly positive payoff from mentoring, even if \( \tilde{r}(\theta) = 1 \). 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 \( \theta \).

Since the probability of investigation is \( \gamma r \), the maximum expected sanction that a manager can pay is \( \gamma S \). 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.

\([\alpha]\) No pooling on \( a(\theta) = 0 \) for all \( \theta \): Suppose that all types pool on \( a = 0 \), so each type’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 \( \theta' \) deviates to \( a = 1 \). Then the highest probability of sanction is \( \gamma r' = \gamma(1 - \Phi(\frac{\tilde{\theta} - k'}{\sigma})) = \gamma \Phi(\frac{-k'}{\sigma}) \). So he will deviate if \( m - \gamma \Phi(\frac{-k'}{\sigma})(S) > 0 \), where \( r' \) is increasing in \( \theta' \) and \( \lim_{\theta' \to -\infty} r' = 0 \). Thus, there exists some type \( \theta' \) 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. 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.

\([b]\) No pooling on \( a(\theta) = 1 \) for all \( \theta \): Suppose all types pool on \( a = 1 \), which implies that agents use threshold \( x^* \) as in Proposition 1. Since any type’s payoff is \( m - \gamma \tilde{r}(\theta)S \) and \( \lim_{\theta \to -\infty} \tilde{r}(\theta) = 1 \), then there must exist some type \( \theta' \) sufficiently large that he prefers to deviate to \( a = 0 \) rather than pool on \( a = 1 \) if \( S < \frac{m}{\gamma S} \). 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:

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

\[
m - \gamma \Phi(\frac{\tilde{\theta} - x^*_S}{\sigma})S = 0. \tag{A.9}
\]

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 \( x^*_S = \frac{c}{\omega - \beta + \gamma(\alpha + \beta)} \) and \( \tilde{\theta} = \frac{c}{\omega - \beta + \gamma(\alpha + \beta)} + \sigma \Phi^{-1}(\frac{m}{\gamma S}) \). 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, \( \frac{\partial \tilde{r}}{\partial S} = -\frac{m}{\gamma S} < 0 \) and \( \frac{\partial x^*_S}{\partial S} = \frac{-c\gamma(\alpha + \beta)\frac{m}{\gamma S}}{2(\omega - \beta + \frac{1}{2}r^* \gamma(\alpha + \beta))^2} > 0 \).

[2] What is \( \frac{\partial \tilde{\theta}}{\partial S} \)? From \( \tilde{\theta} = x^*_S + \sigma \Phi^{-1}(\tilde{r}) \), we have \( \frac{\partial \tilde{\theta}}{\partial S} = \frac{\partial x^*_S}{\partial S} + \sigma \frac{\partial \tilde{r}}{\partial S} \Phi^{-1}(\tilde{r}) \). Combining:

\[
\frac{\partial \tilde{\theta}}{\partial S} = \left( \frac{1}{\phi(\Phi^{-1}(\tilde{r}))} - \frac{c\gamma(\alpha + \beta)}{2(\omega - \beta + \frac{1}{2}r^* \gamma(\alpha + \beta))^2} \right) \frac{\partial \tilde{r}}{\partial S}. \]

Since \( \frac{\partial \tilde{r}}{\partial S} < 0 \), the necessary and sufficient condition for the sign of \( \frac{\partial \tilde{\theta}}{\partial S} \) is:

\[
\frac{\partial \tilde{\theta}}{\partial S} < 0 \Leftrightarrow \sigma > \phi(\Phi^{-1}(\tilde{r})) \frac{c\gamma(\alpha + \beta)}{2(\omega - \beta + \frac{1}{2}r^* \gamma(\alpha + \beta))^2}. \tag{A.10}
\]
Recall that \( \hat{r} = \frac{m}{\gamma S} \) in equilibrium. Sufficient conditions for \( \frac{\partial \hat{u}}{\partial S} < 0 \) are then:

\[
S \to \infty \quad \text{since} \quad \hat{r} \to 0 \implies \phi \left( \Phi^{-1} (\hat{r}) \right) \to 0, \\
S \to \frac{m}{\gamma} \quad \text{since} \quad \hat{r} \to 1 \implies \phi \left( \Phi^{-1} (\hat{r}) \right) \to 0, \\
\sigma > \frac{1}{\sqrt{2\pi}} \frac{c\gamma (\alpha + \beta)}{(\omega - \beta + \frac{3}{4} (\alpha + \beta))^2} \quad \text{since} \quad \phi (\cdot) < \frac{1}{\sqrt{2\pi}}.
\]

From Equation A.10, a sufficient condition for \( \frac{\partial \hat{u}}{\partial S} > 0 \) is:

\[
\sigma < \phi \left( \Phi^{-1} \left( \frac{m}{\gamma S} \right) \right) \frac{c\gamma (\alpha + \beta)}{2 (\omega - \beta + \frac{3}{4} (\alpha + \beta))^2} \quad \text{since} \quad \hat{r} \leq 1.
\]

As another example, when \( \hat{r} = 1/2 \), Equation A.10 implies that \( \frac{\partial \hat{u}}{\partial S} > 0 \) if:

\[
\sigma < \frac{1}{\sqrt{2\pi}} \frac{c\gamma (\alpha + \beta)}{(\omega - \beta + \frac{3}{4} (\alpha + \beta))^2}.
\]

### A.5 Proof of Proposition 5

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 \( \tau \in [0, 1] \) only if \( \beta \delta - c < x (\omega - (1 - \gamma (0)) \beta) - c \), which implies \( \delta \in [0, \omega - \beta) \). The bound on the lower dominance region is still \( \underline{x} = \frac{c}{(1 - \gamma (1 - \gamma \beta))} \). When \( \delta \in [0, \omega - \beta) \), the bound on the upper dominance region is \( \overline{x} = \frac{c}{\beta (1 - \gamma \beta)} \). Thus, all previous results on existence, uniqueness, and comparative statics of other parameters in Sections 1-B.7 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 6 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 \( \tau \) yields

\[
\frac{\partial x_H^*}{\partial \tau} = \frac{-c \delta - \omega + \beta (1 - \gamma \tau)}{(\delta \tau + \omega (1 - \tau) - \beta ([1 - \tau] - \frac{1}{2} \gamma - \frac{1}{2} \gamma \tau^2))^2} \tag{A.11}
\]

\[
> 0 \quad \text{if} \quad \delta < \omega - \beta (1 - \gamma \tau) \\
= 0 \quad \text{if} \quad \delta = \omega - \beta (1 - \gamma \tau) \\
< 0 \quad \text{if} \quad \delta > \omega - \beta (1 - \gamma \tau).
\]

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

2. If \( \delta \in [\omega - \beta , \omega] \), then \( \pi_H (r, x) \) fails action monotonicity (P1) for some \( \tau \in [0, 1] \), 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 6. For all \( \delta \in [0, \omega] \), the bound on the lower dominance region becomes \( \underline{x} = \min \left\{ \frac{c}{(1 - \gamma (1 - \gamma \beta))}, \frac{c}{\beta (1 - \gamma \beta)} \right\} \). When \( \delta > \omega - \beta (1 - \gamma \tau) \), the bound on the upper dominance region satisfies \( \pi (\pi_H^*, \tau) = 0 \), which is \( \pi_H^* = \frac{c}{\omega - \beta (1 - \gamma \tau)} \). Thus for all \( \delta \in [0, \omega] \), then the bound on the upper dominance region becomes \( \overline{x} = \max \left\{ \frac{c}{(1 - \gamma (1 - \gamma \beta))}, \frac{c}{\beta (1 - \gamma \beta)} \right\} \).

We can verify that \( x_H^* \in [\underline{x}, \overline{x}] \) for all \( \tau \in [0, 1] \).
Therefore we can write the expected payoffs as:

\[ \pi^* (x, k) = x \delta F \left( \frac{k-x}{\sigma} - F^{-1} (1 - \tau) \right) \]

\[ + x (\omega - \beta) \left( 1 - F \left( \frac{k-x}{\sigma} - F^{-1} (1 - \tau) \right) \right) \]

\[ + x \beta \gamma \int_{k-\sigma F^{-1}(1-\tau)}^{\infty} \frac{1}{\sigma} f \left( \frac{x - \theta}{\sigma} \right) \left( 1 - F \left( \frac{k-\theta}{\sigma} \right) \right) d\theta \]

\[ - c. \]

Given Equation A.11 above, it is straightforward to show that \( \frac{\partial x_H^*}{\partial \tau} |_{\tau=0} < 0 \) for all \( \delta \in (\omega - \beta, \omega] \) and that there is a unique \( \tau_{min} \in (0, 1] \) that minimizes \( x_H^* \) for all \( \delta \in (\omega - \beta, \omega] \). Since \( \frac{\partial x_H^*}{\partial \tau} = 0 \) when \( \delta = \omega - \beta(1 - \gamma) \), then \( \tau_{min} = \min \{ \delta - (\omega - \beta) \} \) and \( x_H^*(\tau_{min}) < x_H^*(0) \). In particular, \( \tau_{min} \in (0, 1) \) when \( \delta \in (\omega - \beta, \omega - \beta(1 - \gamma)) \) and \( \tau_{min} = 1 \) when \( \delta \in [\omega - \beta(1 - \gamma), \omega] \). Moreover, note that \( \frac{\partial \tau_{min}}{\partial \beta} > 0 \) if \( \tau_{min} \in (0, 1) \).

From Equation A.11, it follows that \( x_H^*(\tau) \leq x_H^*(\tau = 0) \) for all \( \tau \in (0, 1] \) if and only if \( x_H^*(\tau = 1) \leq x_H^*(\tau = 0) \). This holds if and only if \( \delta \geq \omega - \beta(1 - \gamma/2) \). When \( \delta \in [\omega - \beta, \omega - \beta(1 - \gamma/2)) \), \( x_H^*(\tau) > x_H^*(\tau = 0) \) for all \( \tau > r'' \) where \( r'' \) satisfies \( x_H^*(r'') = x_H^*(\tau = 0) \). By direct calculation, \( r'' = 2(\delta - (\omega - \beta)) \). Thus implementing a holding tank with sufficiently high \( \tau \) whenever \( \delta \in [\omega - \beta, \omega - \beta(1 - \gamma/2)) \) results in \( x_H^*(\tau) > x_H^*(\tau = 0) \).

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 \( \theta \), 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, \( f (\theta | x) = \frac{1}{\sigma} f \left( \frac{x - \theta}{\sigma} \right) \). For simplicity, assume Gaussian noise at this point. Observe that:

\[ \int_{-\infty}^{\infty} \frac{1}{\sigma} f \left( \frac{x - \theta}{\sigma} \right) d\theta = \int_{-\infty}^{\infty} \frac{1}{\sigma} f \left( \frac{\theta - x}{\sigma} \right) d\theta \]

\[ = F \left( \frac{z - x}{\sigma} \right). \tag{A.12} \]

Therefore we can write the expected payoffs as:

\[ \pi^* (x, k) = x \delta F \left( \frac{k-x}{\sigma} - F^{-1} (1 - \tau) \right) \]

\[ + x (\omega - \beta) \left( 1 - F \left( \frac{k-x}{\sigma} - F^{-1} (1 - \tau) \right) \right) \]

\[ + x \beta \gamma \int_{k-\sigma F^{-1}(1-\tau)}^{\infty} \frac{1}{\sigma} f \left( \frac{x - \theta}{\sigma} \right) \left( 1 - F \left( \frac{k-\theta}{\sigma} \right) \right) d\theta \]

\[ - c. \]

\[ = x \delta F \left( \frac{k-x}{\sigma} - F^{-1} (1 - \tau) \right) \]

\[ + x (\omega - \beta) \left( 1 - F \left( \frac{k-x}{\sigma} - F^{-1} (1 - \tau) \right) \right) \]

\[ + x \beta \gamma \int_{k-\sigma F^{-1}(1-\tau)}^{\infty} \frac{1}{\sigma} f \left( \frac{x - \theta}{\sigma} \right) \left( 1 - F \left( \frac{k-\theta}{\sigma} \right) \right) d\theta \]

\[ - c. \]

Consider the “little integral”:

\[ \text{little integral} = \int_{k-\sigma F^{-1}(1-\tau)}^{\infty} \frac{1}{\sigma} f \left( \frac{x - \theta}{\sigma} \right) d\theta. \]
Use a u-substitution:

\[ z = \frac{x - \theta}{\sigma}, \quad dz = -\frac{1}{\sigma}d\theta \]

\[ k - \frac{\theta}{\sigma} = \frac{x - \theta}{\sigma} - \frac{x - k}{\sigma} = z - \frac{x - k}{\sigma} = \frac{k - x}{\sigma} + z \]

\[ \theta = k - \sigma F^{-1}(1 - \tau) \]

\[ \zeta = \frac{x - \theta}{\sigma} = \frac{x - k + \sigma F^{-1}(1 - \tau)}{\sigma} = \frac{x - k}{\sigma} + F^{-1}(1 - \tau) \]

\[ \bar{\theta} = \infty \]

\[ \bar{\zeta} = \frac{x - \theta}{\sigma} = -\infty, \]

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

\[ \text{little integral} = 1 - F \left( \frac{k - x}{\sigma} - F^{-1}(1 - \tau) \right). \]

Now consider the “big integral”:

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

Use the same u-substitution to get:

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

From Owen (1980),

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

where \( BvN \) is the bivariate normal cumulative distribution function. We can apply the formula to the “big integral” with:

\[ \rho = -\frac{1}{\sqrt{2}} \]

\[ l = \frac{1}{\sqrt{2}} \frac{k - x}{\sigma}, \]

\[ h = \frac{x - k}{\sigma} + F^{-1}(1 - \tau), \]

so:

\[ \text{big integral} = BvN \left( \frac{x - k}{\sigma} + F^{-1}(1 - \tau), \frac{1}{\sqrt{2}} \frac{k - x}{\sigma}; -1 \right). \]
The expected payoff when agents play cutoff $k$ then becomes:

$$\pi^* (x, k) = x (\omega - \beta (1 - \gamma)) - x (\omega - \beta (1 - \gamma) - \delta) F \left( \frac{k - x}{\sigma} - F^{-1} (1 - \tau) \right)$$

$$- x \beta \gamma \left[ BvN \left( \frac{x - k}{\sigma} + F^{-1} (1 - \tau), \frac{1}{\sqrt{2}} \frac{k - x}{\sigma} ; \frac{-1}{\sqrt{2}} \right) \right] - c.$$  \hspace{1cm} (A.16)

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

$$x^*_H \equiv \frac{c}{\tau \delta + (1 - \tau) (\omega - \beta) + \frac{1}{2} \beta \gamma (1 - \tau^2)}.$$  

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

$$\pi^* (x^*_H, x^*_H) = x^*_H (\omega - \beta (1 - \gamma)) - x^*_H (\omega - \beta (1 - \gamma) - \delta) \tau \left[ \text{using } F^{-1} (1 - r) = -F^{-1} (r) \right]$$

$$- x^*_H \beta \gamma \left[ BvN \left( F^{-1} (1 - \tau), 0; \frac{-1}{\sqrt{2}} \right) \right] - c$$

$$= x^*_H (\omega - \beta (1 - \gamma)) - x^*_H (\omega - \beta (1 - \gamma) - \delta) \tau - x^*_H \beta \gamma \frac{1}{2} (1 - \tau^2) - c$$

$$= x^*_H \left[ \delta \tau + (1 - \tau) (\omega - \beta) + \frac{1}{2} \beta \gamma (1 - \tau^2) \right] - c$$

$$= 0.$$  \hspace{1cm}

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

$$BvN \left( F^{-1} (1 - \tau), 0; \frac{-1}{\sqrt{2}} \right) = T \left( F^{-1} (1 - \tau), 0 \right) + T \left( 0, \infty \right) - T \left( F^{-1} (1 - \tau), \frac{1}{\sqrt{2}} F^{-1} (1 - \tau) \right)$$

$$- T \left( 0, \infty \right) + F \left( F^{-1} (1 - \tau) \right) F (0), \text{ using Owen (1980) Equation 3.2}$$

$$= - T \left( F^{-1} (1 - \tau), 1 \right) + \frac{1}{2} (1 - \tau), \text{ using Owen (1980) Equation 2.1}$$

$$= - F \left( F^{-1} (1 - \tau) \right) \left[ 1 - F \left( F^{-1} (1 - \tau) \right) \right] /2 + \frac{1}{2} (1 - \tau),$$

using Owen (1980) Equation 2.3

$$= \frac{1}{2} (1 - \tau)^2.$$  \hspace{1cm}

From Equation A.16, define the function $\pi(x)$ as:

$$\pi (x) \equiv \pi^* (x, x^*_H) \hspace{1cm} (A.17)$$

A symmetric threshold equilibrium exists at $x^*_H$ when $\pi(x) < 0$ for all $x < x^*_H$ and $\pi(x) > 0$ 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 A.17 satisfies the single-crossing condition. To see this, note from
Figure A1: Expected payoff gain function $\pi^*(x, k = x^*_H)$. This figure plots $\pi^*(x, k = x^*_H)$ for $x \in (\underline{x}, \bar{x})$ with $\{\beta, c, \omega, \gamma, \delta, \tau\} = \{0.85, 0.3, 1, 0.5, 0.5, 0.2\}$. The vertical line denotes $x^*_H$. In Panel (a), $\sigma=0.005$ and $x^*_H$ is not an equilibrium. In Panel (b), $\sigma = 0.5$ and $x^*_H$ is the unique symmetric threshold equilibrium.

Equation A.16 that $\pi^*(x, k)$ is a continuous function. Consider $\sigma$ large. Then for any $x$ and $k$:

$$\pi^*(x, k) = x (\omega - \beta (1 - \gamma)) - x (\omega - \beta (1 - \gamma) - \delta) F \left( 0 - F^{-1} (1 - \tau) \right)$$

$$- x \beta \gamma \left[ \text{BvN} \left( 0 + F^{-1} (1 - \tau), \frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}} \right) \right] - c$$

$$= x (\omega - \beta (1 - \gamma)) - x (\omega - \beta (1 - \gamma) - \delta) \tau$$

$$- x \beta \gamma \left[ \text{BvN} \left( F^{-1} (1 - \tau), 0, -\frac{1}{\sqrt{2}} \right) \right] - c$$

$$= x \left[ \delta \tau + (1 - \tau) (\omega - \beta) + \frac{1}{2} \beta \gamma (1 - \tau^2) \right] - c$$

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 A1 illustrates when the equilibrium does and does not exist for low and high $\sigma$, consistent with the limit result.

A.6 Proof of Proposition 6

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

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

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

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

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

The $\alpha^E$ such that $x^* = x^E$ must satisfy $\delta \tau + \omega (1 - \tau) - \beta (1 - \tau - \frac{1}{2} \gamma (1 - \tau^2)) = \omega - (1 - \gamma) \beta$. Thus $\alpha^E = \max \left\{ \frac{2 \gamma}{\gamma - 1} \left( \frac{1}{2} \beta \gamma + \tau (\omega - \delta - \beta (1 - \frac{1}{2} \gamma) \tau) \right), 0 \right\}$.

Clearly, $\alpha^F$ and $\alpha^E$ are unique and positive for any given $\beta$. 

47
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 \(f_2(z) = \frac{1}{\sigma} f\left(\frac{z + \xi}{\sigma}\right)\), and \(F_2(z; x_1) = \int_{-\infty}^{z} \frac{1}{\sigma} f\left(\frac{s + \xi}{\sigma}\right) \, ds < \int_{-\infty}^{z} \frac{1}{\sigma} f\left(\frac{x_1 + s}{\sigma}\right) \, ds = F_2(z; x_2)\) 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:

\[
\pi^*(x, k) = \int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x - \theta}{\sigma}\right) \pi\left(1 - F\left(\frac{k - \theta}{\sigma}\right), x\right) \, d\theta \\
= \int_{z=-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x + z}{\sigma}\right) \pi\left(1 - F\left(\frac{k + z}{\sigma}\right), x\right) \, dz
\]

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 \(x_1 < x_2\),

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

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 \(\{\xi^n\}\) 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 \(x < \bar{x}\), so \(\pi^*(x, 0) < 0\) for all \(x < \bar{x}\). But we also know that \(\pi^*(x, 0) > 0\) for all \(x > \bar{x}\). Define \(\xi^0 = 0\) and \(\bar{\xi}^0 = \infty\). By continuity in \(x\), there exists at least one solution \(x\) with \(\pi^*(x, \xi^0) = 0\), where \(x \in [\bar{x}, \bar{x}]\). Call \(\xi^1\) the smallest such solution. Define \(\xi^1 \in [\bar{x}, \bar{x}]\) analogously to be the largest such solution with \(\pi^*(x, \bar{\xi}^0) = 0\). Note that \(\xi^0 < \xi^1 < \bar{\xi}^1 < \bar{\xi}^0\); if the inside inequality did not hold, then \(0 = \pi^*(\xi^1, \bar{\xi}^0) \geq \pi^*(\xi^1, \xi^0) > \pi^*(\xi^1, \bar{\xi}^0) = 0\), a contradiction.

Our starting point for the induction is as follows. Given \(\xi^1\) and \(\bar{\xi}^1\) with \(\xi^0 < \xi^1 < \bar{\xi}^1 < \bar{\xi}^0\), \(\pi^*(\xi^1, \bar{\xi}^0) = 0\), and \(\pi^*(\xi^1, \xi^0) = 0\), we claim there exists a smallest solution \(\xi^2\), \(\xi^2\) of \(\pi\left(\xi^2, \xi^1\right) = 0\) and a largest solution \(\bar{\xi}^2\), \(\pi\left(\bar{\xi}^2, \xi^1\right) = 0\)
of $\pi (\hat{\xi}_2, \hat{\xi}_1) = 0$, and that $\xi_1 < \xi_2 < \xi_0 < \xi_1$. We know $\pi^* (\hat{\xi}_1, \hat{\xi}_0) = 0 > \pi^* (\hat{\xi}_1, \xi_1)$, and $\pi^* (\hat{\xi}_1, \xi_0) = 0 < \pi^* (\hat{\xi}_1, \xi_1)$. Note for the latter inequality that $\hat{\xi}_0 > \hat{\xi}_1$. By continuity, there exists a smallest solution $\xi_2 \in (\hat{\xi}_1, \xi_1)$ with $\pi^* (\xi_2, \hat{\xi}_1) = 0$. Analogously, we know $\pi^* (\xi_1, \xi_0) = 0 > \pi^* (\xi_1, \hat{\xi}_1)$, and $\pi^* (\xi_1, \xi_0) = 0 < \pi^* (\xi_1, \xi_1)$; by continuity there exists a largest solution $\xi_0 \in (\hat{\xi}_1, \xi_1)$. Note that $\xi_1 < \xi_2 < \xi_0 < \xi_1$; if the inside inequality did not hold, then $0 = \pi^* (\xi_2, \xi_1) > \pi^* (\xi_0, \xi_1) = 0$, a contradiction.

The inductive hypothesis is that, given $\xi_n$ and $\hat{\xi}_n$ with $\hat{\xi}_n \in (\hat{\xi}_1, \xi_1)$, there exists a smallest solution $\xi_{n+1}$ of $\pi (\xi_{n+1}, \hat{\xi}_n) = 0$ and a largest solution $\hat{\xi}_{n+1}$ of $\pi (\xi_{n+1}, \xi_n)$, and $\hat{\xi}_{n+1} < \hat{\xi}_n$. By continuity, there exists a smallest solution $\xi_{n+1} \in (\hat{\xi}_n, \xi_n)$ with $\pi^* (\xi_{n+1}, \hat{\xi}_n) = 0$. Similarly, we know $\pi^* (\xi_{n+1}, \xi_n) = 0 > \pi^* (\xi_{n+1}, \xi_n)$. By continuity there exists a largest solution $\xi_{n+1} \in (\xi_{n+1}, \xi_n)$ with $\pi^* (\xi_{n+1}, \xi_n) = 0$. Note that $\xi_n < \xi_{n+1} < \xi_{n+1} < \xi_n$; if the inside inequality did not hold, then $0 = \pi^* (\xi_{n+1}, \xi_n) > \pi^* (\xi_{n+1}, \xi_n) = 0$, a contradiction.

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

Note that we can write:

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

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

$$\Psi (r; x, k) = \int_{-\infty}^{k - \sigma F^{-1} (1 - r)} f (\theta \mid x) \, d\theta$$

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

$$= \int_{z = \frac{x - k}{\sigma} + F^{-1} (1 - r)}^{\infty} f (z) \, dz \text{ for } z = \frac{x - \theta}{\sigma}, \, dz = -\frac{1}{\sigma} \, d\theta$$

$$= 1 - F \left( \frac{x - k}{\sigma} + F^{-1} (1 - r) \right).$$

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 $\pi^* (x, r) = \int_{0}^{1} \pi (r, x) \, dr$. By Property P3, there is exactly one such solution $\xi$. From [2], it must be that $\xi = \xi = \xi$ and that this is the unique threshold equilibrium.
B.1.2 Proof of Lemma A3

Proof. Let $\Sigma$ 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 $\xi$ when other players play $\Sigma$. 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 $\xi^1$ and monotonicity of $\pi^* (x, k)$ in $x$,

$$\xi < \xi^1 \Rightarrow \text{for any } \Sigma, \tilde{\pi}^i (\xi, \Sigma) \leq \pi^* (\xi, 0) < \pi^* (\xi^1, 0) = 0.$$

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

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

In words, reporting strictly dominates not reporting ($\tilde{\pi}^i (\xi, \Sigma) > 0$) 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:

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

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

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

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 $\Sigma^n$ 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:

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

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

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

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

$$\xi > \xi^{n+1} \Rightarrow \text{for any } \Sigma, \tilde{\pi}^i (\xi, \Sigma^n) \geq \pi^* (\xi, \xi^n) > \pi^* (\xi^{n+1}, \xi^n) = 0,$$

so that reporting strictly dominates not-reporting ($\tilde{\pi}^i (\xi, \Sigma^n) > 0$) whenever $\xi > \xi^{n+1}$, 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 $\epsilon$ has bounded support and is possibly asymmetric.
Suppose $\epsilon$ has bounded support with cumulative distribution function (CDF) $F$ and probability density function (PDF) $f$. In particular, suppose $\epsilon$ has CDF representation:

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

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

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

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

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

$$
f_{x_i}(x \mid \theta) = \begin{cases} 
0 & x < \theta - \sigma \bar{l} \\
\tilde{f}_{x_i}(x \mid \theta) & x \in [\theta - \sigma \bar{l}, \theta + \sigma \bar{l}] \\
0 & x > \theta + \sigma \bar{l}
\end{cases},
$$

for:

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

The CDF for $x_i$ is:

$$
F_{x_i}(x \mid \theta) = \begin{cases} 
0 & x < \theta - \sigma \bar{l} \\
\tilde{F}_{x_i}(x \mid \theta) & x \in [\theta - \sigma \bar{l}, \theta + \sigma \bar{l}] \\
1 & x > \theta + \sigma \bar{l}
\end{cases},
$$

for:

$$
\tilde{F}_{x_i}(x \mid \theta) = \int_{\theta - \sigma \bar{l}}^{x} \frac{1}{\sigma} \tilde{f}_{\epsilon}\left(\frac{s - \theta}{\sigma}\right) ds
= \tilde{F}_{\epsilon}\left(\frac{x - \theta}{\sigma}\right).
$$
The posterior in $\theta$ conditional on $x$ equals:

$$f_\theta(\theta \mid x) = \frac{f_x(x \mid \theta) f_\theta(\theta)}{\int f_x(x \mid \theta) f_\theta(\theta) d\theta} = f_x(x \mid \theta)$$

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

$$= \begin{cases} 
0 & \theta < x - \sigma \bar{l} \\
\frac{1}{2} \tilde{f}_x\left(\frac{x - \theta}{\sigma}\right) & \theta \in [x - \sigma \bar{l}, x + \sigma \bar{l}] \\
0 & \theta > x + \sigma \bar{l} 
\end{cases},$$

under the improper prior assumption. Thus:

$$F_\theta(\theta \mid x) = \begin{cases} 
0 & \theta < x - \sigma \bar{l} \\
\int_0^\theta f_x(x \mid \theta) d\theta & \theta \in [x - \sigma \bar{l}, x + \sigma \bar{l}] \\
1 & \theta > x + \sigma \bar{l} 
\end{cases} = 1 - \tilde{F}_c\left(\frac{x - \theta}{\sigma}\right) \theta \in [x - \sigma \bar{l}, x + \sigma \bar{l}].$$

**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:

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

It follows that:

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

is a bijection for any $r \in (0, 1)$ to $\theta \in (k - \sigma \bar{l}, k + \sigma \bar{l})$, where the open intervals are important. The 1-1 map fails if $r = 1$, since then $\theta \geq k + \sigma \bar{l}$ 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 “$\pi^*$ weakly decreases in $k$ (strictly decreasing for $x > 0$ and $k > 0$ for $k \in (x - \sigma(l + \bar{l}), x + \sigma(l + \bar{l}))$.”

**Proof.** The payoff gain $\pi(r, x)$ continues to satisfy Properties P1-P5 in Lemma A1.
\[ \pi^* (x, k) \text{ increases in } x: \text{ As before, implement a change in variables } z = -\theta. \text{ Then:} \]

\[
\pi^* (x, k) = \int_{\theta=x-\sigma}^{\theta=x+\sigma} \frac{1}{\sigma} f_\theta \left( \frac{x-\theta}{\sigma} \right) \pi \left( 1 - F_\varepsilon \left( \frac{k-\theta}{\sigma} \right), x \right) d\theta \\
= \int_{z=x-\sigma}^{z=x+\sigma} \frac{1}{\sigma} f_\theta \left( \frac{x+z}{\sigma} \right) \pi \left( 1 - F_\varepsilon \left( \frac{k+z}{\sigma} \right), x \right) dz
\]

It is trivial to show that, for \( x_1 < x_2, F_\varepsilon (x_1 + \epsilon) \leq F_\varepsilon (x_2 + \epsilon) \forall \epsilon \). 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 \( k \in (x - \sigma (1 + l), x + \sigma (1 + l)) \). Observe:

\[
\pi^* (x, k) = \int_{\theta=x-\sigma}^{\theta=x+\sigma} f_\theta (\theta | x) \pi \left( 1 - F_\varepsilon \left( \frac{k-\theta}{\sigma} \right), x \right) d\theta
\]

\[
= \begin{cases}
  f_{\theta=x+\sigma} f_\theta (\theta | x) \pi (0, x) d\theta = \pi (0, x) & x < k - \sigma (1 + l) \\
  f_{\theta=x-\sigma} f_\theta (\theta | x) \pi (0, x) d\theta & x \in (k - \sigma (1 + l), k) \\
  f_{\theta=x+\sigma} f_\theta (\theta | x) \pi (1 - F_\varepsilon (k+\theta), x) d\theta & x \in (k, k + \sigma (1 + l)) \\
  f_{\theta=x-\sigma} f_\theta (\theta | x) \pi (1, x) d\theta & x > k + \sigma (1 + l)
\end{cases}
\]

Let \( k_1 < k_2 \) be given. Re-write as:

\[
\pi^* (x, k) = \int_{\theta=x-\sigma}^{\theta=x+\sigma} f_\theta (\theta | x) \pi \left( 1 - F_\varepsilon \left( \frac{k-\theta}{\sigma} \right), x \right) d\theta
\]

\[
= \begin{cases}
  F_1 (x, k) = \int_{\theta=x-\sigma}^{\theta=x+\sigma} f_\theta (\theta | x) \pi (0, x) d\theta = \pi (0, x) & k > x + \sigma (1 + l) \\
  F_2 (x, k) = \int_{\theta=x-\sigma}^{\theta=x+\sigma} f_\theta (\theta | x) \pi (0, x) d\theta & k \in (x, x + \sigma (1 + l)) \\
  F_3 (x, k) = \int_{\theta=x-\sigma}^{\theta=x+\sigma} f_\theta (\theta | x) \pi (1, x) d\theta & k \in (x - \sigma (1 + l), x) \\
  F_4 (x, k) = \int_{\theta=x-\sigma}^{\theta=x+\sigma} f_\theta (\theta | x) \pi (1, x) d\theta = \pi (1, x) & k < x - \sigma (1 + l)
\end{cases}
\]

Notice that \( F_1 (x, k^1) < F_2 (x, k^2) < F_3 (x, k^3) < F_4 (x, k^4) \) 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:

\[
F_2(x, k_2) - F_2(x, k_1) = \int_{\theta = x - \sigma l}^{\theta = k_2 - \sigma l} f_\theta(\theta \mid x) \pi(0, x) \, d\theta \\
+ \int_{\theta = k_2 - \sigma l}^{\theta = x + \sigma l} f_\theta(\theta \mid x) \pi(1 - F_\epsilon\left(\frac{k_2 - \theta}{\sigma}\right), x) \, d\theta \\
- \int_{\theta = x - \sigma l}^{\theta = k_1 - \sigma l} f_\theta(\theta \mid x) \pi(0, x) \, d\theta \\
- \int_{\theta = k_1 - \sigma l}^{\theta = x + \sigma l} f_\theta(\theta \mid x) \pi(1 - F_\epsilon\left(\frac{k_1 - \theta}{\sigma}\right), x) \, d\theta \\
= \int_{\theta = k_1 - \sigma l}^{\theta = k_2 - \sigma l} f_\theta(\theta \mid x) \pi(0, x) \, d\theta \\
+ \int_{\theta = k_2 - \sigma l}^{\theta = x + \sigma l} f_\theta(\theta \mid x) \pi(1 - F_\epsilon\left(\frac{k_2 - \theta}{\sigma}\right), x) \, d\theta \\
- \int_{\theta = k_1 - \sigma l}^{\theta = x + \sigma l} f_\theta(\theta \mid x) \pi(1 - F_\epsilon\left(\frac{k_1 - \theta}{\sigma}\right), x) \, d\theta \\
\leq \int_{\theta = k_1 - \sigma l}^{\theta = k_2 - \sigma l} f_\theta(\theta \mid x) \pi(0, x) \, d\theta \\
+ \int_{\theta = k_2 - \sigma l}^{\theta = x + \sigma l} f_\theta(\theta \mid x) \pi(1 - F_\epsilon\left(\frac{k_1 - \theta}{\sigma}\right), x) \, d\theta \\
- \int_{\theta = k_1 - \sigma l}^{\theta = x + \sigma l} f_\theta(\theta \mid x) \pi(1 - F_\epsilon\left(\frac{k_1 - \theta}{\sigma}\right), x) \, d\theta \\
= \int_{\theta = k_1 - \sigma l}^{\theta = k_2 - \sigma l} f_\theta(\theta \mid x) \pi(0, x) \, d\theta \\
- \int_{\theta = k_2 + \sigma l}^{\theta = k_1 - \sigma l} f_\theta(\theta \mid x) \pi(1 - F_\epsilon\left(\frac{k_1 - \theta}{\sigma}\right), x) \, d\theta \\
< 0,
\]
F_3(x,k_2) - F_3(x,k_1) = \int_{\theta=x-\sigma L}^{\theta=k_1+\sigma L} f_{\theta}(\theta | x, \pi) \left( 1 - F_\varepsilon \left( \frac{k_2 - \theta}{\sigma} \right), x \right) d\theta \\
+ \int_{\theta=k_2+\sigma L}^{\theta=x+\sigma L} f_{\theta}(\theta | x, \pi) (1, x) d\theta \\
- \int_{\theta=x+\sigma L}^{\theta=k_1+\sigma L} f_{\theta}(\theta | x, \pi) \left( 1 - F_\varepsilon \left( k_1 - \theta \sigma \right), x \right) d\theta \\
- \int_{\theta=x-\sigma L}^{\theta=k_2+\sigma L} f_{\theta}(\theta | x, \pi) (1, x) d\theta \\
= - \int_{\theta=x-\sigma L}^{\theta=k_2+\sigma L} f_{\theta}(\theta | x, \pi) (1, x) d\theta \\
+ \int_{\theta=x-\sigma L}^{\theta=k_1+\sigma L} f_{\theta}(\theta | x, \pi) \left( 1 - F_\varepsilon \left( k_2 - \theta \sigma \right), x \right) d\theta \\
- \int_{\theta=x+\sigma L}^{\theta=k_1+\sigma L} f_{\theta}(\theta | x, \pi) \left( 1 - F_\varepsilon \left( k_1 - \theta \sigma \right), x \right) d\theta \\
= \int_{\theta=k_1+\sigma L}^{\theta=k_2+\sigma L} f_{\theta}(\theta | x, \pi) \left( 1 - F_\varepsilon \left( k_2 - \theta \sigma \right), x \right) d\theta \\
- \int_{\theta=k_1+\sigma L}^{\theta=x+\sigma L} f_{\theta}(\theta | x, \pi) (1, x) d\theta \\
< 0,

since \pi is an increasing function of \( r \) and \( 1 - F_\varepsilon \left( \frac{k_2 - \theta}{\sigma} \right) > 0 \) for \( \theta > k_1 - \sigma L \). In Region 3:

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

Our starting point for the induction is as follows. Given \( \xi^1 \) and \( \xi^1 \) with \( \xi^0 < \xi^1 < \xi^1 < \xi^0 \), and \( \pi^* (\xi^1, \xi^0) = 0 \), we claim there exists a smallest solution \( \xi^2 \) of \( \pi^* (\xi^2, \xi^1) = 0 \) and a largest solution \( \xi^2 \) of \( \pi (\xi^2, \xi^1) = 0 \), and that \( \xi^1 < \xi^2 < \xi^2 < \xi^1 \). We know \( \pi^* (\xi^2, \xi^1) = 0 \) > \( \pi^* (\xi^2, \xi^1) \), and \( \pi^* (\xi^1, \xi^0) = 0 < \pi^* (\xi^2, \xi^1) \leq \pi^* (\xi^2, \xi^1) \), where the first inequality is strict because \( \pi^* (\xi^2, \xi^1) \) is strictly decreasing in
k. By continuity, there exists a smallest solution \( \xi^*_n \in (\xi^1_n, \xi^1_l) \) with \( \pi^* (\xi^2_n, \xi^1_l) = 0 \). Analogously, we know \( \pi^* (\xi^2_n, \xi^1_l) = 0 > \pi^* (\xi^2_n, \xi^1_l) \geq \pi^* (\xi^1_n, \xi^1_l) \), and \( \pi^* (\xi^2_n, \xi^1_l) = 0 < \pi^* (\xi^1_n, \xi^1_l) \); by continuity there exists a largest solution \( \xi^2_n \in (\xi^1_n, \xi^1_l) \) with \( \pi^* (\xi^2_n, \xi^1_l) = 0 \). Note that \( \xi^2_n < \xi^2_n < \xi^2_n < \xi^2_l \); if the inside inequality did not hold and \( \xi^2_n \geq \xi^2_n \), then \( 0 = \pi^* (\xi^2_n, \xi^1_l) > \pi^* (\xi^2_n, \xi^1_l) \geq \pi^* (\xi^2_n, \xi^2_l) \geq \pi^* (\xi^2_n, \xi^1_l) = 0 \), a contradiction. The inequalities are because \( \pi^* (\xi^2_n, \xi^1_l) \) strictly decreases in \( k \) locally, the contradiction assumption with \( \pi^* \) strictly increasing in \( x \), and that \( \pi^* \) is globally weakly decreasing with \( \xi^2_n \in (\xi^1_n, \xi^1_l) \), respectively.

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

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

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

Importantly, conditional on \( x \), we must have \( \theta \in [x - \sigma \hat{l}, x + \sigma \hat{l}] \).

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

\[
\Psi (r; x, k) = \int_{x - \sigma \hat{l}}^{k - \sigma \hat{l}} f_\theta (\theta | x) \, d\theta
\]

\[
= \int_{x - \sigma \hat{l}}^{k - \sigma \hat{l}} \frac{1}{\sigma} f_\xi \left( \frac{x - \theta}{\sigma} \right) \, d\theta
\]

\[
= \int_{z = \frac{x - \hat{l}}{\sigma}}^{z = \hat{l}} f_\xi (z) \, dz \text{ for } z = \frac{x - \theta}{\sigma}, \, dz = -\frac{1}{\sigma} \, d\theta
\]

noting that \( z (k - \sigma \hat{l}) = \frac{x - (k - \sigma \hat{l})}{\sigma} = \frac{x - k}{\sigma} + \hat{l} \)

and that \( z (x - \sigma \hat{l}) = \frac{x - (x - \sigma \hat{l})}{\sigma} = \hat{l} \)

\[
= 1 - F_\xi \left( \frac{x - k}{\sigma} + \hat{l} \right).
\]

If \( k - \sigma \hat{l} < x - \sigma \hat{l} \) then \( x > k \) and the probability is zero by the definition of \( f_\theta (\theta) \). If \( k - \sigma \hat{l} > x + \sigma \hat{l} \) or
equivalently if $\frac{k-x}{\sigma} > \bar{l} + \frac{1}{\bar{l}}$ then the probability is 1. So:

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

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 $\theta < k + \sigma \bar{l}$ plus the probability that $r = 1$, which is the probability that $\theta \geq k + \sigma \bar{l}$. 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 $\theta(r; k) = k - \sigma \bar{F}_\epsilon^{-1}(1 - r) \in (k - \sigma \bar{l}, k + \bar{l})$. We also must have $\theta \in [x - \sigma \bar{l}, x + \sigma \bar{l}]$ in the posterior distribution of $\theta$. Some useful facts to reference later:

$$
k - \sigma \bar{F}_\epsilon^{-1}(1 - r) > x - \sigma \bar{l} \iff r > 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right)
$$

$$
x \geq k + \sigma (\bar{l} + \frac{1}{\bar{l}}) \Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) = 1
$$

$$
x < k + \sigma (\bar{l} + \frac{1}{\bar{l}}) \Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) < 1
$$

$$
x < k \Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right) = 0,
$$

$$
k - \sigma \bar{F}_\epsilon^{-1}(1 - r) < x + \sigma \bar{l} \iff r < 1 - \tilde{F}_\epsilon^{-1}\left(\frac{k-x}{\sigma} - \bar{l}\right)
$$

$$
x \leq k - \sigma (\bar{l} + \frac{1}{\bar{l}}) \Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} - \bar{l}\right) = 0
$$

$$
x > k - \sigma (\bar{l} + \frac{1}{\bar{l}}) \Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} - \bar{l}\right) > 0
$$

$$
x > k \Rightarrow 1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} - \bar{l}\right) = 1.
$$

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

Several cases can occur:

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

$$
k - \sigma \bar{F}_\epsilon^{-1}(1 - r) \in (x - \sigma \bar{l}, x + \sigma \bar{l}) \Rightarrow r \in \left(1 - \tilde{F}_\epsilon\left(\frac{k-x}{\sigma} + \bar{l}\right), 1 - \tilde{F}_\epsilon^{-1}\left(\frac{k-x}{\sigma} - \bar{l}\right)\right)
$$

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

10
If $k - \sigma \bar{F}^{-1}_e (1 - r) > x - \sigma \bar{l}$, then $r \in \left( 1 - \tilde{F}_e \left( \frac{k-x}{\sigma} + \bar{l} \right), 1 \right)$, and furthermore:

$$
\Psi \left( r; x, k \right) = \int_{x - \sigma \bar{l}}^{k - \sigma \bar{F}^{-1}_e (1 - r)} f_\theta \left( \theta \mid x \right) \, d\theta \\
= \int_{x - \sigma \bar{l}}^{k - \sigma \bar{F}^{-1}_e (1 - r)} \frac{1}{\sigma} f_e \left( \frac{x - \theta}{\sigma} \right) \, d\theta \\
= \int_{z = \bar{l}}^{\frac{x}{\sigma} \bar{F}^{-1}_e (1 - r)} f_z (z) \, dz \text{ for } z = \frac{x - \theta}{\sigma}, \, dz = -\frac{1}{\sigma} \, d\theta
$$

noting that $z \left( k - \sigma \bar{F}^{-1}_e (1 - r) \right) = x - \left( k - \sigma \bar{F}^{-1}_e (1 - \bar{r}) \right) = \frac{x - k}{\sigma} + \bar{F}^{-1}_e (1 - r)$

and that $z \left( x - \sigma \bar{l} \right) = \frac{x - (x - \sigma \bar{l})}{\sigma} = \bar{l}$

$$
= 1 - \bar{F}_e \left( \frac{x - k}{\sigma} + \bar{F}^{-1}_e (1 - r) \right)
$$

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

$$
\Psi \left( r; x, k \right) = -\int_{x - \sigma \bar{l}}^{x - \sigma \bar{l}} f_\theta \left( \theta \mid x \right) \, d\theta \\
= 0.
$$

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

$$
\Psi \left( r; x, k \right) = -\int_{k - \sigma \bar{F}^{-1}_e (1 - r)}^{x - \sigma \bar{l}} f_\theta \left( \theta \mid x \right) \, d\theta \\
= 0.
$$

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

$$
\Psi \left( r; x, k \right) = \int_{x - \sigma \bar{l}}^{x + \sigma \bar{l}} f_\theta \left( \theta \mid x \right) \, d\theta \\
= 1.
To summarize:

\[
\Psi (r; x) = \begin{cases} 
0 & r = 0 \\
0 & r \in (0, 1) \\
1 & r = 1,
\end{cases}
\]

\[
\Psi (r; x) = \begin{cases} 
0 & r = 0 \\
0 & r \in \left(0, 1 - \hat{F}_{\sigma} \left(k - x + \bar{l} \right) \right) \\
1 - \hat{F}_{\sigma} \left(\frac{x - k}{\sigma} + \bar{l} \right) & r \in \left(1 - \hat{F}_{\sigma} \left(k - x + \bar{l} \right), 1 \right) \\
1 & r = 1,
\end{cases}
\]

\[
\Psi (r; x) = \begin{cases} 
1 - \hat{F}_{\sigma} \left(\frac{x - k}{\sigma} + \bar{l} \right) & r = 0 \\
1 - \hat{F}_{\sigma} \left(\frac{x - k}{\sigma} + \bar{l} \right) \left(1 - r \right) & r \in (0, 1) \\
1 & r = 1,
\end{cases}
\]

\[
\Psi (r; x) = \begin{cases} 
1 & r = 0 \\
1 & r \in (0, 1) \\
1 & r = 1,
\end{cases}
\]

\[
\Psi (r; x = k) = \begin{cases} 
0 & r = 0 \\
r & r \in (0, 1) \\
1 & r = 1.
\end{cases}
\]

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:

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

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

For the marginal agent with \( x = k \), this payoff equals:

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

Given the marginal agent’s posterior belief over \( \theta \) translates into a uniform belief over \( r \), we can write:

\[
\pi^*(x, x) = \int_{0}^{1} \pi (r, x) dr.
\]

\[\blacksquare\]

**B.3 Finite number of agents**

We show that the equilibrium solution and overall conclusions are virtually identical to that of a finite \( N \)-agent version of the game.

The manager employs \( N \) agents. This is common knowledge. Each agent who reports increments the probability of sanction by \( \frac{1}{N} \gamma \) for \( \gamma \leq 1 \). If \( n \) agents report, probability of sanction is \( \frac{n}{N} \gamma \).
If \( n \) other agents report, the payoff gain for any individual agent from reporting equals:

\[
E (\text{Report} - \text{NoReport} | x_i, n \text{ other agents report}) = \begin{cases} 
    x (\omega - (1 - \frac{m+1}{N}) \beta) - c & \text{if } x \geq 0 \\
    x \omega - c & \text{if } x < 0. 
\end{cases}
\]

(B.4)

This payoff function implies that there are dominance regions in \( x_i \). Agent \( i \) reports as a dominant strategy if \( x_i > \overline{x} = \frac{c}{\omega - (1 + \frac{1}{N}) \beta} \). Similarly, agent \( i \) does not report as a dominant strategy if \( x_i < \overline{x} = \frac{c}{\omega - (1 - \frac{1}{N}) \beta} \).

The payoff function is also monotonic in \( x \) and \( n \), since \( \beta < \omega \) and \( \gamma > 0 \).

Conditional on experience \( x_i \), agent \( i \)'s posterior beliefs about other agents' experiences are determined by Bayes' Rule. Specifically, her posterior belief about any agent \( j \)’s experience \( x_j = \theta + \epsilon_j \) is normal with:

1. \( E [x_j | x_i] = E [\theta | x_i] = x_i. \)
2. \( \text{Var} [x_j | x_i] = \text{Var} [\theta | x_i] + \sigma^2 = 2\sigma^2 \Rightarrow \text{SD} [x_j | x_i] = \sqrt{2}\sigma. \)

This derivation uses the fact that \( x_j \) and \( x_i \) are i.i.d. jointly normal. Given joint normality, \( 1/\text{Var}[\theta | x_i] = 1/\text{Var}[\theta] + 1/\sigma^2 \); given the agent’s improper prior over \( \theta \), \( \text{Var}[\theta | x_i] = \sigma^2 \).

Conjecture a symmetric strategy equilibrium where agents report if and only if \( x > k \). In agent \( i \)'s beliefs,

1. The probability agent \( j \) does not report equals: \( p(x_j < k | x_i) = \Phi \left( \frac{k-x_i}{\sqrt{2}\sigma} \right) \).
2. The probability agent \( j \) reports equals: \( p(x_j > k | x_i) = 1 - \Phi \left( \frac{k-x_i}{\sqrt{2}\sigma} \right) \).
3. The probability of \( m \) other agents reporting equals:

\[
Pr (n = m | x_i; k) = \left( \frac{N-1}{m} \right) p(x,k)^m (1 - p(x,k))^{N-1-m}. \quad (B.5)
\]

The expected payoff gain for any agent conditional on experience \( x \) equals:

\[
\pi^* (x, k) = \sum_{m=0}^{N-1} \left( \frac{N-1}{m} \right) p(x,k)^m (1 - p(x,k))^{N-1-m} \left[ x \left( \omega - \left(1 - \frac{m+1}{N} \gamma \right) \beta \right) - c \right] \quad (B.6)
\]

The indifference condition and cutoff threshold \( x^* \) satisfies \( \pi^* (x^*, x^*) = 0 \). Note that \( p(x^*, x^*) = 1/2 \). An agent who draws \( x = x^* \) assigns \( 1/2 \) probability to each other agent when agents play threshold strategies around \( x^* \). Thus, the indifference condition satisfies:

\[
0 = \sum_{m=0}^{N-1} \left( \frac{N-1}{m} \right) \left( \frac{1}{2} \right)^{N-1} \left[ x \left( \omega - \left(1 - \frac{m+1}{N} \gamma \right) \beta \right) - c \right] \quad (B.7)
\]

\[
= \left( \frac{1}{2} \right)^{N-1} x \left( 2^{N-1} (\omega - \beta) + \frac{1}{N} \sum_{m=0}^{N-1} \left( \frac{N-1}{m} \left( m+1 \right) \gamma \beta \right) - \left( \frac{1}{2} \right)^{N-1} 2^{N-1} c \right) \quad (B.8)
\]

\[
= x \left( (\omega - \beta) + \left( \frac{1}{2} \right)^{N-1} \frac{1}{N} (N-1) 2^{N-2} + 2^{N-1} \gamma \beta \right) - c \quad (B.9)
\]

\[
= x \left( (\omega - \beta) + \frac{N+1}{N} 2^{N-1} \gamma \beta \right) - c, \quad \text{if } x > 0 
\]

(B.10)

where we use the fact that \( \sum_{m=0}^{n} \left( \frac{n}{m} \right) = 2^n \) and \( \sum_{m=0}^{n} m \left( \frac{n}{m} \right) = n2^{n-1} \).
The above analysis implies that a candidate equilibrium cutoff is given by:

\[ x^* = \frac{c}{(\omega - \beta) + \frac{N+1}{2} \gamma \beta}. \]  

(B.11)

Existence and uniqueness of the equilibrium follow from the fact that \( x^* \) increases in \( x \) and decreases in \( k \). Since \( x^* \) increases in \( x \), \( \pi^*(x, x^*) > 0 \) for \( x > x^* \) and \( \pi^*(x, x^*) < 0 \) for \( x < x^* \). Thus, \( x^* \) is an equilibrium. Observing that \( \pi^*(x, k) \) also decreases in \( k \), the usual global games argument shows that this is the unique equilibrium due to the iterated deletion of strictly dominated strategies (Morris and Shin, 2003).

The claim that \( \pi^*(x, k) \) increases in \( x \) follows because \( p(x, k) \) increases in \( x \). Stochastic dominance arguments then show that \( \pi^*(x, k) \) increases in \( x \), just as in the proof of Lemma A2. To see this, note that \( p(x, k) \) increases in \( x \). This implies that, for any \( x_1 < x_2 \), the binomial distribution of the number of reports out of \( N - 1 \) agents under \( p(x_2, k) \) stochastically dominates the distribution under \( p(x_1, k) \). Since \( x (\omega - (1 - \frac{m+1}{N} \gamma) \beta) - c \) is itself an increasing function of \( x \), then stochastic dominance also implies that \( \pi^*(x, k) \) increases under \( x \).

The claim that \( \pi^*(x, k) \) decreases in \( k \) follows because \( x (\omega - (1 - \frac{m+1}{N} \gamma) \beta) - c \) is invariant to \( k \). By the same stochastic dominance argument, we have \( \pi^*(x, k) \) decreases in \( k \).

We summarize with:

**Proposition B1** (Baseline Equilibrium with \( N \) agents). Suppose agents have improper uniform priors over \( \theta \). There exists a globally unique symmetric threshold strategy equilibrium where all agents play a threshold strategy and report \( s(x_i) = 1 \) if and only if \( x_i \geq x^* \), where

\[ x^*_N = \frac{c}{(\omega - \beta) + \frac{N+1}{2} \gamma \beta}. \]

(B.12)

and \( x^* > \frac{c}{\omega} > 0 \). The threshold strategy is the unique strategy that survives the iterated deletion of strictly dominated strategies. In equilibrium, the belief of the marginal agent who draws \( x_i = x^* \) over the number of agents reporting \( r \) is uniformly distributed over \([0, 1]\).

Proposition B1 is a close analog of Proposition 1 in the text. Notice that for \( N \) large, \( x^*_N \) from Proposition B1 converges to \( x^* = \frac{c}{(\omega - \beta) + \frac{1}{2} \gamma \beta} \) from Proposition 1. The slight difference in the equilibrium cutoff occurs because agents are atomistic in the continuum agent game but not in the finite agent game. In the finite agent game, each agent’s report increments the probability of sanction by a discrete amount, \( 1/N \). All else equal, greater \( N \) results in a greater equilibrium reporting threshold because the effect of each agent’s own report on the probability of sanction diminishes. As the number of agents approaches infinity, the effect of each agent’s own report on the probability of sanction becomes vanishingly small, and the cutoff converges to the solution with a continuum of agents.

**B.4 Endogenous Sanction Functions**

We provide a few stylized models that describe channels through which the outside party’s objective function can generate a coordination problem among agents and show that the results still hold qualitatively when \( \Gamma(r) \) is endogenous.

In each of these models, the agents receive their \( x_i \)’s and decide whether or not to report. Then the outside party (the firm) observes \( r \) and decides whether or not to sanction the manager. For simplicity, there is no uncertainty in the sanction process.

1. **Model 1**: Suppose the firm values the manager’s productivity but that its payoff decreases in \( r \).

Here, the firm doesn’t mind how the manager treats his subordinates, but does find it costly to deal with reports of misconduct. This payoff function generates an endogenous \( \Gamma(r) \) where the firm
sanctions the manager if reports exceed an equilibrium sanction threshold \( r' \) that depends only on the manager’s productivity parameters. Because the endogenous \( \Gamma(r) \) does not depend on agents' payoff parameters, all results from the exogenous \( \Gamma(r) \) model described in the paper apply.

The manager’s value (productivity) to the firm is \( v > 0 \). If the firm does not sanction the manager, then it keeps the manager but there is a \( c_I(r) \) (due to the possibility of employee turnover, violation of federal law, reputational risks, etc). Let \( c_I(r) \) strictly increase in \( r \) with \( c_I(0) = 0 \). If the firm sanctions the manager, he is replaced by another manager with productivity \( v' \).

Given \( r \), the firm’s payoff is \( v - c_I(r) \) if it does not sanction, and \( v' \) if it sanctions. Note that if \( v - c_I(1) > v' \), then the firm would never sanction. If \( v < v' \), then the firm always wants to sanction/fire the manager regardless. Thus we focus on the case of interest, \( v - c_I(1) < v' < v \).

Thus, the firm would sanction for any \( r \geq r' \) where \( v - c_I(r') = v' \) and not sanction for any \( r < r' \). Since \( c_I(r) \) is strictly increasing in \( r \) and \( v - c_I(1) < v' < v \), then there exists a unique \( r' \in (0,1) \) such that \( v - c_I(r') = v' \). Note that \( r' \) decreases in \( v' \) and increases in \( v \).

This implies that when deciding whether or not to report, agents take the sanction function \( \Gamma(r) \) as

\[
\Gamma(r) = \begin{cases} 
0 & \text{if } r < r' \\
1 & \text{if } r \geq r'. 
\end{cases}
\]

Since we have shown in the manuscript that \( x^* = \frac{c}{\omega - \beta + \beta \int_0^1 \Gamma(r) dr} \), then the solution to the equilibrium is

\[
x^* = \frac{c}{\omega - \beta r'} = v - c_I(r') = v'.
\]

Because the threshold \( r' \) does not depend on agents’ payoff parameters, all results from the exogenous \( \Gamma(r) \) model described in the manuscript follow.

2. Model 2: Suppose the firm values the manager’s productivity but that its payoff decreases in \( \theta \).

Here, the firm does not want to employ a manager that treats his subordinates poorly, irrespective of the number of reports. This generates an endogenous \( \Gamma(r) \) with an equilibrium sanction threshold \( r' \) that depends on agents’ payoff parameters through their equilibrium reporting threshold \( x^* \). As such, changing agents’ payoff parameters affects \( x^* \) both directly and indirectly through its effect on \( r' \). Nevertheless, we show that the comparative statics of this game for \( x^* \) qualitatively match the comparative statics of the paper where \( \Gamma(r) \) is exogenous.

The manager’s value to the firm (productivity) is \( v > 0 \). The firm dislikes bad apples (i.e., high-\( \theta \) type). If the firm sanctions the manager, he is replaced by another manager with productivity \( v' \). The firm’s payoff is \( v - f(\theta) \) if it does not sanction the manager, where \( f(\theta) \) is non-negative and weakly increasing in \( \theta \). The firm’s payoff is \( v' \) if it sanctions the manager. Assume that \( v > v' \) so the firm does not want to fire the manager ex ante.

Conjecture an equilibrium in which agents use threshold strategies. Given a realized \( r \), in equilibrium the firm learns \( \theta \). Since \( f(\theta) \) is non-negative and weakly increasing in \( \theta \), then the firm will sanction for any \( \theta \geq \tilde{\theta} \) where \( v - v' - f(\tilde{\theta}) = 0 \) and not sanction for any \( \theta < \tilde{\theta} \).

Since the use of threshold strategies generates a one-to-one mapping between \( r \) and \( \theta \) where \( r = 1 - \Phi \left( \frac{v' - \theta}{\sigma} \right) = \Phi \left( \frac{\theta - v'}{\sigma} \right) \), then in equilibrium the firm sanctions for any \( r \geq r' \) where \( r' = \Phi \left( \frac{v - v'}{\sigma} \right) \) and does not sanction if \( r < r' \).
This implies that when deciding whether or not to report, agents take the sanction function $\Gamma(r)$ as

$$\Gamma(r) = \begin{cases} 0 & \text{if } r < r' \\ 1 & \text{if } r \geq r'. \end{cases}$$

Thus the solution to the equilibrium is $(\tilde{\theta}, r', x^*)$ such that

$$x^* = \frac{c}{\omega - \beta v'}$$

(B.13)

$$r' = \Phi \left( \frac{\tilde{\theta} - x^*}{\sigma} \right)$$

(B.14)

$$v - v' - f(\tilde{\theta}) = 0.$$ 

(B.15)

Note that $r'$ is the solution to

$$r' - \Phi \left( \frac{f^{-1}(v - v') - \frac{c}{\omega - \beta v'}}{\sigma} \right) = 0,$$ 

(B.16)

and

$$\tilde{\theta} = f^{-1}(v - v').$$ 

(B.17)

Let $G(r) = r - \Phi \left( \frac{f^{-1}(v - v') - \frac{c}{\omega - \beta v'}}{\sigma} \right)$. Since $G(r = 1) > 0$ and $G(r = 0)$, then there exists an $r' \in (0, 1)$ that satisfies Equation B.16. Moreover,

$$\frac{\partial G}{\partial r} = 1 - \phi \left( \frac{f^{-1}(v - v') - \frac{c}{\omega - \beta v'}}{\sigma} \right) \left( \frac{\beta c}{\sigma} \right) (\omega - \beta r)^{-2} > 0.$$ 

Thus this $r'$ is unique.

Here, note that the threshold $r'$ depends on agents’ payoff parameters. Intuitively, as agents become more reluctant to report ($x^*$ increases), then a lower $r$ reveals $\theta$ in equilibrium so a lower threshold $r'$ is required to generate sanction for a given $\theta$.

Consider the total effect on $x^*$ when $\beta$ increases ($\frac{\partial x^*}{\partial \beta}$).

Applying the implicit function theorem, we have

$$\frac{\partial r'}{\partial \beta} = -\frac{\partial G}{\partial x'}.$$

(B.18)

We have already shown that $\frac{\partial G}{\partial r'} > 0$.

$$\frac{\partial G}{\partial \beta} = -\phi \left( \frac{f^{-1}(v - v') - \frac{c}{\omega - \beta v'}}{\sigma} \right) \left( \frac{r c}{\sigma} (\omega - \beta r)^{-2} \right) > 0.$$ 

(B.19)
Thus, consistent with the above intuition $\frac{\partial \Omega}{\partial \beta} < 0$.

Since $x^* = \frac{c}{\omega - \beta r}$, then the effect of $\frac{\partial x^*}{\partial \beta}$ is determined by $\frac{\partial}{\partial \beta}(\beta r')$:

$$
\frac{\partial}{\partial \beta}(\beta r') = r' + \beta \left( \frac{\partial r'}{\partial \beta} \right) = r' - \beta \left( \frac{\phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{c}{\sigma} \right)(\omega - \beta r)^{-2}}{1 + \phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{c}{\sigma} \right)(\omega - \beta r)^{-2}} \right).
$$

Note that

$$
\frac{\phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{c}{\sigma} \right)(\omega - \beta r)^{-2}}{1 + \phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{c}{\sigma} \right)(\omega - \beta r)^{-2}} < \frac{\phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{c}{\sigma} \right)(\omega - \beta r)^{-2}}{\phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{\beta c}{\sigma} \right)(\omega - \beta r)^{-2}} = \frac{r}{\beta}.
$$

Thus,

$$
\frac{\partial}{\partial \beta}(\beta r') = r' + \beta \left( \frac{\partial r'}{\partial \beta} \right) = r' - \beta \left( \frac{\phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{c}{\sigma} \right)(\omega - \beta r)^{-2}}{1 + \phi \left( f^{-1}(v-v') - \frac{\beta c}{\sigma} \right) \left( \frac{c}{\sigma} \right)(\omega - \beta r)^{-2}} \right).
$$

Thus, $\frac{\partial}{\partial \beta}(\beta r') > 0$, so $\frac{\partial x^*}{\partial \beta} > 0$. That is, the indirect effect of increasing $\beta$ through lowering $r'$ is weaker than the direct effect of increasing $\beta$.

A similar exercise shows that $\frac{\partial x^*}{\partial c} > 0$.

Finally, note that, for example, if $v$ increases (the manager becomes more valuable to the firm), this clearly increases $r'$ and therefore increases $x^*$. Increasing $v$ is akin to increasing $\gamma$ in the model described in the manuscript.

3. Model 3: Suppose the firm values the manager’s productivity, that its payoff decreases $\theta$, and faces a constraint whereby it must have $r \geq \gamma$ reports to sanction the manager.

Here, the firm may not want to employ a manager that treats his subordinates poorly but is required to provide “just cause” in the form of a sufficient number of reports to sanction the manager. Depending on the parameters, this case generates endogenous $\Gamma(r)$ that is akin to either Model 1 or Model 2. Thus, the results still hold qualitatively.

The manager’s value to the firm (productivity) is $v > 0$. The firm dislikes high-$\theta$ types. However, it is constrained to provide “just cause” for sanctioning the manager, which means that it must also provide $r \geq \gamma$ to sanction (take $\gamma$ as an exogenous constraint, like a legal requirement). If the firm sanctions the manager, he is replaced by another manager with productivity $v'$. The firm’s payoff is $v - f(\theta)$ if it does not sanction the manager, where $f(\theta)$ is non-negative and weakly increasing in $\theta$. The firm’s payoff is $v'$ if it sanctions the manager. Assume that $v > v'$ so the firm does not want to fire the manager ex ante.
To derive the solution, there are two relevant cases based on whether the “just cause” constraint binds in equilibrium:

(a) The “just cause” constraint does not bind in equilibrium. In this case, the solution must satisfy Equations B.13, B.16, and B.17. Thus, the solution is the same as in Model 2.

When does this case occur? The requirement that \( r \geq r \) does not bind if the solution \( r' \) to Equations B.13, B.16, and B.17 satisfies \( r' \geq r \). Intuitively, this must be where \( \overline{\theta} \) is sufficiently high (we solve for it explicitly after analyzing the second case).

(b) The “just cause” constraint binds in equilibrium. That is, the solution \( r' \) to Equations B.13, B.16, and B.17 fails \( r' \geq r \).

In this case, agents know that \( r \) binds as the minimal reporting at which the firm sanctions. Thus, agents take \( \Gamma(r) \) as

\[
\Gamma(r) = \begin{cases} 
0 & \text{if } r < r \\
1 & \text{if } r \geq r, 
\end{cases}
\]

Thus, the solution is analogous to that of Model 1 and the equilibrium reporting threshold is

\[
x^* = \frac{c}{\omega - \beta r}. \quad (B.19)
\]

Let \( \theta \) be the minimal manager type such that the firm sanctions managers if and only if \( \theta \geq \overline{\theta} \). Then we must have:

\[
\begin{align*}
\overline{r} &= \Phi\left(\frac{\theta - x^*}{\sigma}\right) \\
\theta &= x^* + \sigma \Phi^{-1}(\overline{r}) \\
\overline{\theta} &= \frac{c}{\omega - \beta \overline{r} + \sigma \Phi^{-1}(\overline{r})}. \quad (B.20)
\end{align*}
\]

Thus, the solution must satisfy Equations B.19 and B.20.

Note that it must be that \( v - v' - f(\theta) < 0 \), i.e., \( \theta > \overline{\theta} \). That is, if the “just cause” constraint binds, then the firm would prefer to sanction managers with \( \theta \in [\overline{\theta}, \overline{\theta}] \), but cannot do so because there is not enough reporting for the firm to legally do so.

Thus \( \overline{\theta} \) is pinned down by Equation B.17 and \( \overline{\theta} \) is pinned down by Equation B.20. When \( \overline{\theta} \geq \overline{\theta} \), the solution satisfies Equations B.13, B.16, and B.17 (akin to Model 2) because the “just cause” constraint does not bind in equilibrium. When \( \overline{\theta} < \overline{\theta} \), the solution satisfies Equations B.19 and B.20 (akin to Model 1) because the “just cause” constraint binds in equilibrium.

### B.5 Alternative payoffs from not reporting

We derive a variation of the model that is identical to that of Section 1 in the manuscript except that the agent’s payoff from not reporting equals \( ax_i \), where \( a \geq 0 \), instead of zero when the manager is sanctioned. The baseline model in Section 1 corresponds to \( a = 0 \). We show that all qualitative insights remain as long as \( a < \beta \) (i.e., the payoff function is monotone in \( r \)).

Table B1 describes agent payoffs.

The payoff gain \( \pi(r, x_i) \) is monotone in \( r \) and \( x_i \) if \( a < \beta \):

\[
\pi(r, x_i) = \begin{cases} 
x_i(\omega - \beta + \gamma r(\beta - a)) - c & \text{if } x_i > 0, \\
x_i(\omega - \gamma r a) - c & \text{if } x_i \leq 0.
\end{cases}
\]

18
<table>
<thead>
<tr>
<th>Sanction</th>
<th>No Sanction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Report</td>
<td>$x_i\omega - c$</td>
</tr>
<tr>
<td>No Report</td>
<td>$ax_i$</td>
</tr>
</tbody>
</table>

**Table B1**: Agent payoffs when she earns a payoff in (No Report, Sanction)

The new dominance regions are delimited by $\bar{x}$ and $\bar{x}$:

$$\bar{x} = \frac{c}{\omega - \beta + \gamma(\beta - a)} \quad (B.21)$$

$$\bar{x} = \frac{c}{\omega - \beta} \quad (B.22)$$

As expected, $\frac{\partial \bar{x}}{\partial a} > 0$ and $\frac{\partial \bar{x}^*}{\partial a} = 0$.

Finally, $x^* = \frac{c}{\omega - \beta(1 - \frac{1}{2}\gamma) - \frac{1}{2}\gamma a}$ where $\frac{\partial x^*}{\partial a} > 0$:

$$\int_0^1 \pi(r, x^*) dr = 0 \quad (B.23)$$

$$x^*(\omega - \frac{1}{2}\gamma a - (1 - \frac{1}{2}\gamma)\beta) - c = 0 \quad (B.24)$$

$$x^* = \frac{c}{\omega - \beta(1 - \frac{1}{2}\gamma) - \frac{1}{2}\gamma a} \quad (B.25)$$

Thus, the main differences are that $\bar{x}$ and $x^*$ are higher when $a > 0$ than when $a = 0$. That is, the reporting threshold is higher when agents can free-ride off other agents’ reports. Proposition 2 (and Corollary 1.1) still applies, using in the proof the new $\bar{x}$, $x^*$, and $H(x) \equiv x(\omega - \beta + \gamma \Phi(\frac{\theta - x}{a})(\beta - a)) - c$, which is the new expected payoff of the marginal agent given that all agents use reporting threshold $x$ and given the true $\theta$.

Note that if instead the agent receives $ax_i$ only if $x_i > 0$ and zero otherwise, then we would have the same results above because $a < \beta$ implies that reporting negative $x_i$ is a dominated strategy.

### B.6 General sanction function

The outside party observes $X \equiv \{(n(x), x) | \forall x \in \mathbb{R}\}$, the frequency and value of each reported $x$. Let $\Xi(X|X) : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, 1]$ represent the sanction probability given the set of all reports and the value of those reports. Consider a class of sanction functions $\Xi(X) = B(\int \varphi(x)n(x)dx)$, 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 $\Xi(X) = B(r) = \Gamma(r)$. 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 \( n(x) = 1_{[x>x^*]} \frac{1}{\sigma} \phi \left( \frac{x-\theta}{\sigma} \right) \), giving

\[
\Xi(X) = B \left( \int_{x^*}^{\infty} \frac{1}{\sigma} \phi \left( \frac{x-\theta}{\sigma} \right) dx \right)
\]

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

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

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

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

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

2. For more general \( \varphi(x) \), the sanction function can be reformulated as \( \Xi(X) = \Gamma(x^*, r) \) (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

\[
\Xi(X) = B \left( \int_{x^*}^{\infty} \frac{\varphi(x)}{\sigma} \phi \left( \frac{x-\theta}{\sigma} \right) dx \right) = B \left( \int_{x^*}^{\infty} \frac{\varphi(x)}{\sigma} \phi \left( \frac{x-x^*}{\sigma} - \Phi^{-1}(r) \right) dx \right) = \Gamma(x^*, r),
\]

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

\[
x^* = \frac{c}{\omega - \beta + \left( \int_0^1 \Gamma(x^*, r) dr \right) (\alpha + \beta)}.
\]
To show the existence of such an $x^*$, note that we can still define the dominance regions in which the agent never reports if $x \leq \bar{x}$ (i.e., even if everyone reports and the manager is definitely sanctioned) and always reports if $x \geq \bar{x}$ (i.e., even if no one reports so the manager is definitely not sanctioned), where

$$\bar{x} = \frac{c}{\omega + \alpha},$$

$$\bar{x} = \frac{c}{\omega - \beta}.$$

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 $G(\bar{x}) > 0$ 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 $B(s) : [0, \infty) \rightarrow [0, 1]$ is weakly monotone increasing (and strictly increasing at some $s$), then $\int_0^1 \Gamma(x^*, r)dr > 0$. Note that $G(\bar{x}) < 0$ 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 $B(s) : [0, \infty) \rightarrow [0, 1]$ is weakly monotone increasing, and strictly increasing at some $s$. Thus, such an $x^*$ exists, and $x^* \in (\underline{x}, \bar{x})$. Given such an $x^*$, it is straightforward to verify that agents with $x_i < x^*$ do not report and agents with $x_i \geq x^*$, since $x_i \omega - \beta + \left(\int_0^1 \Gamma(x^*, r)dr\right)(\alpha + \beta)$ 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.26, 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 $\theta \in (\bar{x}, x^*)$ we still have $\pi(1, x) > 0 > \pi(0, x)$. Since $x^* \in (\underline{x}, \bar{x})$, 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 $\theta$ is sufficiently high, there always exists some $\hat{x} < x^*$ such that there is a Pareto improvement in agent payoffs. When $\theta$ is intermediate, there exists some $\hat{x} < x^*$ such that there is a Pareto improvement in agent payoffs if $\sigma$ is sufficiently small.

**Lemma B2.** Let $g$ and $h$ be density functions, where $g \geq h$ if $\int_{-\infty}^{\xi} g(s)ds \leq \int_{-\infty}^{\xi} h(s)ds$. If $g \geq h$ 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 $g \geq h$, then $G(z) \leq H(z)$.

Using integration by parts,

$$\int_a^\infty u(z)g(z)dz = \int_a^\infty u(z)G'(z)dz$$

$$= u(z)G(z)|_a^\infty - \int_a^\infty G(z)u'(z)dz$$

$$\int_a^\infty u(z)h(z)dz = \int_a^\infty u(z)H'(z)dz$$

$$= u(z)H(z)|_a^\infty - \int_a^\infty H(z)u'(z)dz.$$
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 \( \theta' \) be the minimum value of \( \theta \) such that the marginal agent’s expected probability of sanction is less than or equal to the realized probability of sanction:
\[
\theta' = \min\{ \theta : \int_{0}^{1} B \left( \int_{x_{\ast}}^{\infty} \frac{\varphi(x)}{\sigma} \phi \left( \frac{x-x_{\ast}}{\sigma} - \Phi^{-1}(r) \right) dx \right) dr \leq B \left( \int_{x_{\ast}}^{\infty} \frac{\varphi(x)}{\sigma} \phi \left( \frac{x-\theta'}{\sigma} \right) \right) \}
\]

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

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

Recall that \( \Xi(\theta, x^\ast) = B \left( \int_{x_{\ast}}^{\infty} \frac{1}{\varphi(x)} \phi \left( \frac{x-\theta}{\sigma} \right) dx \right) \). For any \( \hat{x} < x^\ast \), total agent payoffs equal:
\[
W(\theta | \hat{x}) = \int_{a}^{\infty} \left[ x(\omega - \beta + \Xi(\theta, \hat{x}) (\alpha + \beta)) - c \right] f(x | \theta) dx
+ \int_{\hat{x}}^{x^\ast} \left[ x(\omega - \beta + \Xi(\theta, \hat{x}) (\alpha + \beta)) - c \right] f(x | \theta) dx.
\]

We know that equilibrium total payoffs are such that:
\[
W(\theta | x^\ast) = \int_{a}^{\infty} \left[ x(\omega - \beta + \Xi(\theta, x^\ast) (\alpha + \beta)) - c \right] f(x | \theta) dx
< \int_{a}^{\infty} \left[ x(\omega - \beta + \Xi(\theta, \hat{x}) (\alpha + \beta)) - c \right] f(x | \theta) dx,
\]
for any \( \hat{x} < x^\ast \).

The remaining claim is that there exists a \( \hat{x} < x^\ast \) such that:
\[
K \equiv \int_{\hat{x}}^{x^\ast} \left[ x(\omega - \beta + \Xi(\theta) (\alpha + \beta)) - c \right] f(x | \theta) dx > 0.
\]

By Lemma B2, we know that \( \Xi(\theta, \hat{x}) \) is weakly increasing in \( \theta \), which implies that when \( \theta > \theta' \):
\[
0 = x^\ast \left( \omega - \beta + \int_{0}^{1} \Gamma(x^\ast, r)dr (\alpha + \beta) \right) - c
< x^\ast \left( \omega - \beta + \Xi(\theta, x^\ast) (\alpha + \beta) \right) - c.
\] (B.27)

Let \( H(x) \equiv x(\omega - \beta + \Xi(\theta, x) (\alpha + \beta)) - c \). Note that \( H(x^\ast) > 0 \) from Equation B.27.

Suppose \( H \) is continuous at \( x^\ast \). If \( H'(x^\ast) < 0 \), then there exists a \( \hat{x} < x^\ast \) such that:
\[
0 < x^\ast \left( \omega - \beta + \Xi(\theta, x^\ast) (\alpha + \beta) \right) - c
< \hat{x} \left( \omega - \beta + \Xi(\theta, \hat{x}) (\alpha + \beta) \right) - c.
\]
If \( H'(x^*) > 0 \), there exists some \( \varepsilon > 0 \) such that, for \( \hat{x} = x^* - \varepsilon \),
\[
0 < \hat{x} (\omega - \beta + \Xi(x|\theta, \hat{x})(\alpha + \beta)) - c < x^* (\omega - \beta + \Xi(x|\theta, x^*)(\alpha + \beta)) - c.
\]

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,
\[
\frac{\partial}{\partial x} \int_\hat{x}^\infty \frac{1}{\sigma} \varphi(x) \phi \left( \frac{x-\theta}{\sigma} \right) dx < 0.
\]
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:
\[
K = \int_\hat{x}^{x^*} \left[ (\omega - \beta + \Xi(x|\theta, \hat{x})(\alpha + \beta)) - c \right] f(x | \theta) dx
\]
\[
> \int_\hat{x}^{x^*} \left[ \hat{x} (\omega - \beta + \Xi(x|\theta, \hat{x})(\alpha + \beta)) - c \right] f(x | \theta) dx
\]
\[
= \left[ \hat{x} (\omega - \beta + \Xi(x|\theta, \hat{x})(\alpha + \beta)) - c \right] \times \left[ \Phi \left( \frac{x^* - \theta}{{\sigma}} \right) - \Phi \left( \frac{\hat{x} - \theta}{{\sigma}} \right) \right]
\]
\[
> 0.
\]

Note that this is a Pareto improvement because for any \( x \in (\hat{x}, x^*) \),
\[
x (\omega - \beta + \Xi(x|\theta, \hat{x})(\alpha + \beta)) - c > H(\hat{x}) > 0,
\]
whereas these agents all receive 0 when playing a threshold strategy around \( x^* \).

(b) Suppose \( \underline{x} < \theta \leq \theta' \), where \( \underline{x} \) is defined below. If \( \sigma \) is sufficiently low, there exists some \( \hat{x} \in (\underline{x}, \theta) \) 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 \( \hat{x} < \theta \), \( B \left( \int_\hat{x}^{\infty} \frac{\varphi(x)}{\sigma} \phi \left( \frac{x-\theta'}{\sigma} \right) dx \right) \)
increases as \( \sigma \) decreases. By direct computation, \( \frac{\partial}{\partial x} \int_\hat{x}^{\infty} \frac{1}{2} \varphi(x) \phi \left( \frac{x-\theta}{\sigma} \right) dx < 0 \). Define \( \underline{x} \) as the threshold such that
\[
\underline{x} \left( \omega - \beta + \lim_{\sigma \to 0} B \left( \int_\underline{x}^{\infty} \frac{\varphi(x)}{\sigma} \phi \left( \frac{x-\theta'}{\sigma} \right) \right) (\alpha + \beta) \right) - c = 0.
\]
By construction, \( \lim_{\sigma \to 0} H(x) = 0 \) and \( x^* > \underline{x} \geq \hat{x} \).

Suppose \( H(\cdot) \) is continuous on the interval \([\underline{x}, \theta]\) for all \( \sigma \). By direct differentiation, \( \lim_{\sigma \to 0} \frac{\partial H}{\partial \sigma} > 0 \)
for all \( \hat{x} < \theta \). Thus, \( \lim_{\sigma \to 0} H(\hat{x}) > 0 \) for all \( \hat{x} \in (\underline{x}, \theta) \). By Lemma B2, \( H(\hat{x}) \) is decreasing as \( \sigma \) increases. But by continuity of \( H \), for any \( \hat{x} \in (\underline{x}, \theta) \), there exists some \( \sigma > 0 \) sufficiently small that \( H(\hat{x}) > 0 \).

Suppose \( H(x) \) is not continuous on the interval \([\underline{x}, \theta]\) for all \( \sigma \). Since \( \underline{x} \) 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 \( Z = \{ z : \lim_{\sigma \to 0^+} B(x) \neq \lim_{\sigma \to 0^-} B(x) \} \). 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:
\[
B \left( \int_{\underline{x}}^{\infty} \frac{\varphi(x)}{\sigma} \phi \left( \frac{x-\theta}{\sigma} \right) dx \right) = y_0.
\]
Since for all \( \hat{x} < \theta \),

\[
B \left( \int_{\hat{x}}^{\infty} \frac{\phi(x)}{\sigma} \left( \frac{x - \theta}{\sigma} \right) \right)
\]

increases as \( \sigma \) decreases, and \( B(s) \) is increasing in \( s \),

then there exists some \( \epsilon > 0 \) such that \( H(x) \) is continuous on the interval \([x, x + \epsilon)\) for all \( \sigma < \sigma \) and we can apply the preceding argument. Thus for any \( \hat{x} \in (x, x + \epsilon) \), there exists some \( \sigma > 0 \) sufficiently small that \( H(\hat{x}) > 0 \).

### B.7 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 a specific manager or information about misconduct in the broader population of managers. Proposition B2 characterizes behavior in this environment. The proof strategy follows Morris and Shin (2004). We defer the proof of the Proposition to the end of this section.

**Proposition B2** (Equilibrium with proper priors). Suppose agents have common proper priors that \( \theta \) is distributed normally with mean \( y \) and standard deviation \( \tau \). Let \( p(\theta) \) be.

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

\[
x_i^* = \frac{c}{\omega - \beta \left( 1 - \gamma \Phi \left( \frac{y - x_i^*}{\kappa} \right) \right)}, \tag{B.28}
\]

for \( \kappa \equiv \frac{\sigma^2 + \tau^2}{\sigma^2 + \tau^2} \sqrt{\frac{\sigma^2 + 2\tau^2}{\sigma^2 + \tau^2}} \) and where \( x_i^* > x > 0 \).

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 \( \kappa > \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{(\omega - \beta (1 - \frac{x_i^*}{\omega}))^2} \).

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

   (c) For \( y \to \infty \), there is a unique threshold equilibrium \( x_i^* \) with \( x_i^* \to \pi \). For \( y \to -\infty \), there is a unique threshold equilibrium \( x_i^* \) with \( x_i^* \to \pi \).

   (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 \( \Psi_I(\cdot) \):

\[
\Psi_I(r) \equiv \Phi \left( \frac{\sigma}{\tau \sqrt{\sigma^2 + \tau^2}} (x_i^* - y) + \frac{1}{\tau} \sqrt{\sigma^2 + \tau^2} \Phi^{-1}(r) \right), \tag{B.29}
\]

and has expectation \( E_I^* [r] = \Phi \left( \frac{y - x_i^*}{\kappa} \right) \).
Table B2: Comparative statics with proper priors. Comparative statics for $\omega$, $c$, $\gamma$, and $\beta$ are identical to those in Section 1.6. Note that $y > x_I^*$ if and only if $y > \frac{c}{\omega - \beta(1 - \gamma^2)}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>(1) Reporting Threshold $x^*$</th>
<th>(2) Aggregate Reporting $\hat{r}(\theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$: Public belief of average type</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$\tau$: Public belief of type dispersion</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>(for $y &gt; x_I^*$; flip signs if $&gt; 0$ if “=”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B.7.1 Equilibrium uniqueness and under-reporting

There is a unique equilibrium when priors are sufficiently diffuse or when private experiences are informative for $\theta$ relative to priors. Specifically, Part 2(a) of Proposition B2 is satisfied for high $\tau$ or when $\sigma$ is low relative to $\tau$. The reason is that $\kappa$ increases in $\tau$; furthermore, $\kappa$ decreases in $\sigma$ 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, 2004). Part 2(b) illustrates equilibrium multiplicity. Suppose $y = c\omega - \beta(1 - \gamma^2)$, so that $y = x_I^*$ and $E^*_I[r] = 1/2$. Multiplicity occurs if $\tau \approx 0$ and $\sigma$ is not too large. 25 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 B2 (Appendix A.2.2). Condition (1) holds so long as $\theta > x_I^* + \frac{\sigma}{\kappa} (y - x_I^*)$. Condition (2) holds for $\theta \in \left( x + \frac{\sigma}{\kappa} (y - x_I^*) , x_I^* + \frac{\sigma}{\kappa} (y - x_I^*) \right]$ and $\sigma$ 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 assumption 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.

B.7.2 Implications for #MeToo

Table B2 summarizes how changes in the mean and standard deviation of the prior belief, $y$ and $\tau$, affect equilibrium reporting. Corollary B2.1 highlights the key implication relevant for the #MeToo movement. 26

---

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

26 The comparative statics for $\omega$, $c$, $\gamma$, and $\beta$ are identical to those in Section 1.6. For $\sigma$, distinct from the improper prior case in Section 1.6, an increase in $\sigma$ can either increase or decrease the reporting threshold $x_I^*$, because the change in $\sigma$ has two effects on beliefs over $r$. First, it affects the marginal agent’s assessment
Figure B2: 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 B2.

Corollary B2.1 (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 $\tau$ 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 aggregate reporting $\hat{r}(\theta)$ rises for every $\theta$. 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 $y = \frac{c}{\omega - \beta(1 - \gamma)}$ so that $x^*_I = y$. The marginal agent believes $E_I^*[r] = \frac{1}{2}$, and Figure B2 plots the density of her belief over $r$. Note that, from Equation B.29, 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 B2 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 $\tau$ 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 $\tau$, 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 $\tau$ 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 B2.1 is a robust first-order prediction irrespective of underlying parameters: Part 2(c) from Proposition B2 guarantees that the equilibrium is unique when $y \to \infty$ and that $x^*_I \to \frac{c}{2}$, irrespective 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 $\sigma^2/\tau^2$ and whether $y$ is low or high. For brevity, we omit details about $\sigma$ as it seems less directly relevant to the #MeToo movement.
other parameters. Part 2(c) also shows that Corollary B2.1 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 \( (x, \bar{x}) \).

Overall, Corollary B2.1 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 \( \tau \)) 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-\( \tau \)) environment purely from a change in beliefs about whether other agents are reporting.

B.7.3 Proof of Proposition B2

Suppose the prior belief of \( \theta \) with density \( p(\theta) \) is normally distributed with mean \( y \) and standard deviation \( \tau \), 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 \( f(\theta | x) \) is a normal density with:

\[
\text{mean } \lambda = \frac{\sigma^2 y + \tau^2 x}{\sigma^2 + \tau^2} = \frac{ty + ux}{t + u},
\]

\[
\text{standard deviation } v = \sqrt{\frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}} = \frac{1}{\sqrt{t + u}},
\]

\[
\text{precision } h = t + u = \frac{\sigma^2 + \tau^2}{\sigma^2 \tau^2}.
\]

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 \( \theta \): \( r(\theta; k) = 1 - \Phi \left( \frac{k - \theta}{\sigma} \right) \) so \( \theta(\tau; k) = k - \sigma \Phi^{-1}(1 - r) \), 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:

\[
\pi^*(x, k) = \int_{-\infty}^{\infty} f(\theta | x) \pi(r(\theta; k), x) \ d\theta
= x \left( \omega - \beta + \gamma (\alpha + \beta) \int_{-\infty}^{\infty} f(\theta | x) \left( 1 - \Phi \left( \frac{k - \theta}{\sigma} \right) \right) \ d\theta \right) - c. \tag{B.30}
\]

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

B.7.4 Intermediate results

Lemma B3. Any solution \( x^* \) to \( \pi^*(x, x) = 0 \) satisfies the implicit equation:

\[
x^* = \frac{c}{\omega - \beta + \gamma (\alpha + \beta) \Phi \left( \frac{k - \theta}{\sigma} \right)},
\]

where we drop the \( I \) subscript on \( x^*_I \) for notational brevity.

Proof. We have \( f(\theta | x) = \frac{1}{v} \phi \left( \frac{\theta - \lambda}{v} \right) \). Then Equation B.30 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 \( z = \frac{\theta - \lambda}{\sigma} \) and \( a + bz = \frac{k - \lambda - \frac{v}{\sigma} z}{\sigma - \frac{\nu}{\sigma}} \), we have:

\[
\int_{-\infty}^{\infty} \frac{1}{v} \Phi\left(\frac{\theta - \lambda}{\sigma}\right) \left(1 - \Phi\left(\frac{k - \theta}{\sigma} - \frac{v}{\sigma} z\right)\right) \, d\theta = 1 - \int_{-\infty}^{\infty} \phi(z) \Phi\left(\frac{k - \lambda - \frac{v}{\sigma} z}{\sqrt{1 + \frac{v^2}{\sigma^2}}}\right) \, dz
\]

\[
= 1 - \Phi\left(\frac{k - \lambda}{\sqrt{\sigma^2 + \nu^2}}\right)
\]

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

\[
\pi^*(x, k) = x \left(\omega - \beta + \gamma (\alpha + \beta) \Phi\left(\frac{\lambda - k}{\sqrt{\sigma^2 + \nu^2}}\right)\right) - c \text{ if } x > 0,
\]

\[
= x\omega - c \text{ if } x \leq 0.
\]

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

\[
\Phi\left(\frac{\lambda - x}{\sqrt{\sigma^2 + \nu^2}}\right) = \Phi\left(\frac{\frac{y - x}{\tau^2}}{\sqrt{\sigma^2 + \nu^2}}\right)
\]

\[
= \Phi\left(\frac{y - x}{\kappa}\right) \text{ for } \kappa = \frac{\tau^2}{\nu^2} \sqrt{\sigma^2 + \nu^2}.
\]

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 B4.** There exists at least one symmetric threshold strategy equilibrium \( x^*_I \in (x, \pi) \) where agents report for \( x \geq x^*_I \) and do not report for \( x < x^*_I \).

**Proof.** From Lemma B3, any potential equilibria must solve \( G(x^*) = 0 \), where:

\[
G(x) = \pi^*(x, x) = x \left(\omega - \beta + \gamma E(x) (\alpha + \beta)\right) - c,
\]

\[
E(x) = \Phi\left(\frac{y - x}{\kappa}\right).
\]

For notational brevity, we drop the \( I \) subscript in \( x^*_I \).

First, we claim there exists a \( x^* \in (x, \pi) \) 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 \( G\left(\frac{\omega - \beta}{\omega + \gamma \alpha (1 - \gamma)\beta}\right) > 0 \), and \( \frac{\omega - \beta}{\omega + \gamma \alpha (1 - \gamma)\beta} < 0 \), and \( \frac{\omega - \beta}{\omega + \gamma \alpha (1 - \gamma)\beta} > \frac{\omega - \beta}{\omega + \gamma \alpha (1 - \gamma)\beta} \). 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 B.31 in Lemma B3 that the payoff from reporting when an agent’s signal is \( x > 0 \), conditional on
other players playing threshold strategies around \( k \), equals:
\[
\pi^* (x, k) = x \left( \omega - \beta + \gamma (\alpha + \beta) \Phi \left( \frac{x^2 y + \gamma x^2 - k}{\sqrt{\sigma^2 + \nu^2}} \right) \right) - c.
\]
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 \( x < x^* \), \( \pi^* (x, x^*) < 0 \).

Lemmas B5 and B6 substitute for Lemma A2 in the informed prior case.

**Lemma B5.** 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 \( \xi \) solve \( \pi^* (\xi, \xi) = 0 \). The sequence \( \{\xi^1, \xi^2, \ldots, \xi^n \ldots\} \) defined as the solutions to the equations:

\[
\begin{align*}
\pi^* (\xi^1, 0) &= 0 \\
\pi^* (\xi^2, \xi^1) &= 0 \\
\vdots \\
\pi^* (\xi^{n+1}, \xi^n) &= 0 \\
\end{align*}
\]

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

\[
\begin{align*}
\pi^* (\xi^1, \infty) &= 0 \\
\pi^* (\xi^2, \xi^1) &= 0 \\
\vdots \\
\pi^* (\xi^{n+1}, \xi^n) &= 0 \\
\end{align*}
\]

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

**Proof.**

1. These properties are evident from Equation B.31 in Lemma B3.
2. Property [1] implies that all \( \xi^n \) and \( \xi^i \) are well-defined. We know that not reporting is dominant for \( x < \bar{x} \), so \( \pi^* (x, 0) < 0 \) for all \( x < \bar{x} \). But we also know that \( \pi^* (x, 0) > 0 \) for all \( x > \bar{x} \). Define \( \xi^0 \equiv \bar{x} \). By continuity in \( x \), there exists at least one solution \( x \) with \( \pi^* (x, x^0) = 0 \), where \( x \in [\bar{x}, \bar{x}] \). Call \( \xi^1 \) the smallest such solution. Note that \( \xi^1 > 0 \). Furthermore, note that \( \xi^1 < \xi \in (\bar{x}, \bar{x}] \): if not, then \( 0 = \pi^* (\xi^1, 0) > \pi^* (\xi, 0) \), a contradiction.

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

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

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

An analogous argument works identically to show that \( \{\xi^n\} \) is a bounded decreasing sequence, that there exists a \( \xi \) with \( \lim_{n \to \infty} \xi^n = \xi \), and that \( \xi \) is the largest such solution to \( \pi^* (\xi, \xi) = 0 \).
Lemma B6. Uniqueness of equilibrium (allowing for either strategy to be played at $x^*$):

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

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

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

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

[c] For $y \to \infty$, there is a unique threshold equilibrium $x^*$ with $x^* \to \bar{x}$. 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 $G(x) < G(\bar{x})$ with at least one solution in between where $G$ is increasing from Lemma B4, and because $G$ is differentiable for all $x$, a necessary and sufficient condition for uniqueness is that $G'(x^*) > 0$ for all $x^* \in (\bar{x}, \bar{x})$ such that $G(x^*) = 0$. We have

$$G'(x) = \gamma (\alpha + \beta) x E'(x) + (\omega - \beta + \gamma (\alpha + \beta) E(x))$$  \hspace{1cm} (B.32)

and also:

$$G'(x) = \gamma (\alpha + \beta) x E'(x) + \frac{c}{x}$$  \hspace{1cm} (B.33)

Substituting in $E'(x) = -\frac{1}{\gamma} \phi \left( \frac{y-x}{\kappa} \right)$ into Equation B.34, a necessary and sufficient condition for uniqueness is, for all solutions $x^*$:

$$\gamma (\alpha + \beta) (x^*)^2 \phi \left( \frac{y-x^*}{\kappa} \right) \frac{1}{\kappa} < c.$$  \hspace{1cm} (B.35)

Recall from Lemmas B3 and B4 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 $\phi(z) < \frac{1}{\sqrt{2\pi}}$ and $x^* < \bar{x} = \frac{c}{\omega - \beta}$, a sufficient condition for uniqueness from Equation B.35 is:

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

[b] For $y = \frac{c}{\omega - \beta + \frac{1}{2}\gamma (\alpha + \beta)}$, note that $x^* = y$, $E(y) = 1/2$, $E'(y) = -\frac{1}{\gamma \sqrt{2\pi}}$, and we have the following expression for $G'(y)$ from Equation B.32:

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

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

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 $\bar{x}$.
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 \( \bar{x} \).

To show uniqueness in both cases, observe from Equation B.33 that:

\[
\lim_{y \to \infty} G'(x) = \omega - \beta + \gamma (\alpha + \beta) > 0,
\]
\[
\lim_{y \to -\infty} G'(x) = \omega - \beta > 0.
\]

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 \( x^* \).

\[ [d] \text{ If } G(x) \text{ has a unique solution } x^*, \text{ then from Lemma B5 we have } x^* = \xi = \xi = \xi. \text{ By Lemma A3, the only strategy which survives the iterated deletion of dominated strategies is the } x^*\text{-threshold strategy. This implies that the } x^*\text{-threshold equilibrium is the globally unique equilibrium.} \]

**Lemma B7.** The marginal agent has beliefs over \( r \) summarized by the cumulative distribution function \( \Phi \left( \frac{1}{\sqrt{r+t}} (x-y) + \frac{\sqrt{r+t}}{\sqrt{u}} \Phi^{-1}(r) \right) \), with expectation \( E_1^* \{ r \} = \Phi \left( \frac{\sqrt{x^*}}{\kappa} \right) \).

**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 \( \Psi (\tilde{r}; x, k) = \Pr (r < \tilde{r} | x) \) that the true proportion of players reporting is less than \( \tilde{r} \) equals the probability that the true \( \theta \) satisfies \( r(\theta; k) = 1 - F \left( \frac{k - \theta}{\sigma} \right) < \tilde{r} \), or equivalently that \( \theta \) is such that fewer than \( \tilde{r} \) players observe a signal greater than \( k \); in turn, this equals the probability that the true \( \theta \) is less than \( k - \sigma F^{-1}(1 - \tilde{r}) \), integrated against the conditional density \( f(\theta | x) \).

With some abuse of notation,

\[
\Psi (r; x, k) = \int_{-\infty}^{k - \sigma \Phi^{-1}(1 - r)} f(\theta | x) \, d\theta
\]
\[
= \int_{-\infty}^{k - \sigma \Phi^{-1}(1 - r)} \frac{1}{\sqrt{v}} \phi \left( \frac{\theta - \lambda}{\sqrt{v}} \right) \, d\theta
\]
\[
= \int_{z = \frac{k - \lambda}{\sqrt{v}} - \frac{\sigma \Phi^{-1}(1 - r)}{\sqrt{v}}}^{\infty} \phi (z) \, dz \text{ for } z = \frac{\theta - \lambda}{\sqrt{v}}, \, dz = \frac{1}{\sqrt{v}} \, d\theta
\]
\[
= \Phi \left( \frac{k - \lambda}{\sqrt{v}} - \frac{\sigma}{\sqrt{v}} \Phi^{-1}(1 - r) \right).
\]

For the marginal agent with \( x = k \),

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

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

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

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 \( \theta \), 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):

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

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

where $E^*_I [r] = \int_0^1 r \psi (r; x, x) \, dr$. But then by Lemma B3, $E^*_I [r] = \Phi \left( \frac{y - x^*}{\kappa} \right)$.

\[\Box\]

B.7.5 Proposition B2

Proof. Part 1 follows from Lemma B4. Part 2 follows from Lemma B6. Part 3 follows from Lemma B7. \[\Box\]