Reporting Sexual Misconduct in the #MeToo Era*

Ing-Haw Cheng† Alice Hsiaw‡

October 2019

Abstract

We show that, due to a coordination problem, individuals under-report sexual misconduct if a manager's misconduct is widespread. The coordination problem rationalizes under-reporting within a unique equilibrium and occurs because reporting misconduct is costly when other individuals do not corroborate a pattern of behavior. We then study three unintended and intended consequences of #MeToo. First, managers may opt out of mentoring individuals due to possible reporting, spilling over into individuals' willingness to report. Second, heightened public awareness of misconduct encourages individuals to report otherwise-hidden misconduct. Finally, whistleblower rewards mitigate this coordination problem when appropriately calibrated.

Keywords: Misconduct, Reporting, Retaliation, Coordination

*The authors thank Andrew Little, Lisa Camner McKay, Ashley Miller, Stephen Morris, Matthew Rabin, Kathy Spier, Courtney Stoddard, Dustin Tingley, Wei Xiong, Kwok Yu, Nina Zipser and seminar participants at Brandeis University for helpful comments and discussions.

†Dartmouth College, Tuck School of Business. Email: ing-haw.cheng@tuck.dartmouth.edu.

‡Brandeis University, International Business School. Email: ahsiaw@brandeis.edu.

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

This paper shows that a *coordination problem* leads economic agents to under-report sexual misconduct if misconduct is widespread and examines its implications for #MeToo and broader policy changes. The coordination problem occurs because an agent's cost of reporting misconduct is higher if other agents fail to report and provide corroboration of a pattern of behavior. Such corroboration provides "safety in numbers" by increasing the chances that any relevant outside parties will believe the agent's report. Corroboration thus reduces the chances that a reporting agent will face retaliation, stigma, or reprisal.

This coordination problem lies at the heart of an agent's reporting decision. In practice, corroboration is critical because misconduct occurs in private and because disputes over misconduct often come down to "he-said/she-said" situations (Fitzgerald et al., 1995; Ayres and Unkovic, 2012). Worry about reprisal and retaliation is the top-cited motive for why individuals do not report sexual assault and rape (U.S. Department of Justice, 2013), a worry supported by academic and anecdotal evidence (Cortina and Magley, 2003; McDonald, 2012; for press accounts of the Harvey Weinstein case and others, see Engel, 2018; Farrow, 2017). "What if I'm not believed?" is a common refrain in press accounts of individuals who weigh whether to report misconduct (see, e.g., Forde, 2018; Willingham and Maxouris, 2018).

Simple economic analysis has difficulty characterizing agents' reporting decisions in the presence of a coordination problem. The reason is that beliefs about other agents' actions and equilibria are self-fulfilling. On the one hand, if agents believe others will not report a manager's misconduct, they will not report misconduct themselves. This situation precipitates a self-fulfilling "no-report" equilibrium where no agents report, vindicating their initial beliefs. On the other hand, this is not a concrete prediction since a self-fulfilling "all-

report" equilibrium where all agents report a manager's misconduct also exists. Indeed, if a manager engaged in widespread sexual misconduct, one might conjecture that the all-report equilibrium prevails over the no-report equilibrium, as it features higher payoffs for agents.

Our theory models agents' actions when they have *strategic uncertainty* about whether other agents will report misconduct. Strategic uncertainty exists due to an information friction: agents have imperfect knowledge of other agents' private experiences with a manager. The distribution of experiences across agents is determined by the manager's type, which equals the average experience and is unknown to agents. An agent's experience need not be bad, and an agent and manager may perceive the same interaction differently. If an agent receives a bad experience, she has an intrinsic motive to report it to an outside party (e.g., human resources, law enforcement, or other authorities).[1] However, if the outside party resolves the case in the manager's favor, the agent incurs a retaliatory cost. The coordination problem exists because the outside party is more likely to resolve a case in favor of a reporting agent and sanction the manager if other agents also report.

We derive agents' endogenous beliefs about how many other agents will report and provide conditions for a unique equilibrium. In the unique equilibrium, agents employ "switching strategies" and report if and only if their experience is worse than a unique reporting threshold. Our technique draws from work on global games and coordination problems in finance and macroeconomics (for early work, see Carlsson and van Damme, 1993a,b and Morris and Shin, 1998; for reviews, see Morris and Shin, 2003 and Angeletos and Lian, 2016).

Our first observation is that the coordination problem leads to under-reporting when misconduct is most widespread. In a stark example, 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 because the threshold to report is too high. The reason is that, despite knowing that almost all experiences are identically bad, agents are effectively uninformed over how many other agents will report. This "reverberant doubt" (Morris and Shin, 2002) over other agents' actions depresses an agent's motive to report. More broadly, whenever agents experience significant misconduct on average, a lower reporting threshold Pareto-improves agents' payoffs over

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

the equilibrium threshold, due to strategic uncertainty. Information frictions thus provide a theoretical basis for empirical evidence on under-reporting of misconduct.

However, the consequences of encouraging reporting are not well understood, in practice and theory. Our subsequent analysis considers the implications of encouraging reporting surrounding the #MeToo movement.

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

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

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

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

Finally, we consider the effect of rewarding reports that result in sanction, commonly known as rewards for whistle-blowers. Such a reward could be monetary, but could also be a sense of vindication or public recognition. In our model, whistle-blower rewards mitigate the coordination problem induced by retaliation, which imposes a cost on reports that do not result in a sanction. Such rewards can be calibrated to a benchmark level where agents report as if this coordination problem did not exist. Rewards greater than this calibrated level incentivize agents to report a broader range of behaviors than in this benchmark.

Overall, our theory provides a rationale for why individuals do not report misconduct and a study of both unintended and intended consequences of #MeToo for reporting. Our contribution is to understand and organize thinking about agents' reporting decisions and the effects of #MeToo around the coordination problem. Existing work on misconduct often focuses on its ramifications without addressing the role of coordination (e.g., Basu, 2003; Chen and Sethi, 2018; Hersch, 2018).

Our focus on the coordination problem draws on work from global games and enables us to "capture the idea that economic agents may be pushed into taking a particular action because of a belief that others are taking such actions" (Morris and Shin, 2000, 2003). The literature has yet to take this approach to study the reporting of sexual misconduct and thus offers little understanding of the #MeToo movement, whose premise is to raise awareness of misconduct and change individuals' beliefs about whether others might report. 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 a stronger incentive to file a lawsuit if they corroborate previous lawsuits. While frictions in timing can generate coordination problems, neither paper investigates how agents endogenously form beliefs about other agents' actions. Our approach fully endogenizes agents' beliefs about whether other agents will report misconduct, yielding concrete implications about the effects of #MeToo and policies intended to encourage reporting. We conclude with questions for future research.

1 The Reporting Problem

1.1 Model setup

Agents in our model receive private experiences from a manager of an unobserved type. Specifically, a continuum of risk-neutral agents with mass 1 indexed by $i \in (0,1)$ work for a manager at a firm. The manager has a type $\theta \in \mathbb{R}$, a fixed characteristic that generates heterogeneous *private experiences* $\{x_i \in \mathbb{R}\}$, for which θ is the average experience: $x_i = \theta + \sigma\epsilon_i$ for $\epsilon_i \sim N(0,1)$, where ϵ_i is i.i.d. across agents. In the reporting game, the manager exogenously generates experiences and is not a player in the game.[2]

Each experience x_i reflects agent i's perception of an interaction with the manager. We do not model the manager's experience, but the manager may perceive the same interaction differently, and agent i may truthfully disagree.[3] For example, a large body of evidence suggests there are gender differences in the perception of sexual harassment (McDonald, 2012), with women perceiving a wider range of behaviors as harassing (Rotundo et al., 2001); Bénabou, Falk and Tirole (2019) propose a model of why different groups, such as men and women, may interpret the same behavior according to different moral standards.

The key information friction in our model is that agents do not know θ but form beliefs about it through their private heterogeneous ($\sigma > 0$) experiences x_i, where a larger σ reflects a wider range of heterogeneity. Many interactions are private, especially in the context of sexual misconduct. For simplicity, we begin by assuming agents are uninformed about θ but generalize this in Section 3; for now, we note that this case is a useful starting point that reflects the beliefs of an uninformed agent who believes that true manager characteristics are ex-ante unknown.

[2] In reality, managers have agency over their behavior and the experiences they generate. Given our focus on the reporting problem, however, we abstract away from a managerial strategic motive for experiences. In plain words, our theory does not seek to explain why some managers commit sexual misconduct. Pei and Strulovici (2019) studies how a criminal strategically decides the frequency of criminal activity.

[3] For example, one could write agent i's experience as $x_i = \theta + u_i + q_i$ and the manager's experience with agent i as $x_i' = \theta + u_i + q_i'$, where u_i, q_i, and q_i' are jointly normally distributed, i.i.d. across agents, and u_i and q_i have mean zero. The sum $\theta + u_i$ represents the common component of the interaction between agent i and the manager, while q_i and q_i' represent agent i's and the manager's differing interpretation of that interaction, respectively. Then $x_i \neq x_i'$ if $q_i \neq q_i'$. As a concrete example, $\theta + u_i$ may reflect whether there was some physical contact, and $q_i \neq q_i'$ could reflect different views as to whether such contact constituted harassment or whether there was consent. In our notation, $\sigma\epsilon_i = u_i + q_i$.

1.1.1 Misconduct and the coordination problem in reporting

Before turning to the agent's payoffs to reporting misconduct, we review what we mean by misconduct and discuss the coordination problem in reporting misconduct.

Misconduct. While legal definitions and lay perceptions differ across countries and cultures, sexual misconduct generally refers to unwelcome and unreasonable sex-related conduct, encompassing a wide range of behaviors from verbal remarks to physical gestures to violent acts (Hersch, 2010; Cortina and Berdahl, 2008). In the United States, sexual harassment as defined by federal law includes acts such as unwelcome sexual advances and verbal or physical conduct that creates a hostile work environment. Sexual assault, a category that includes rape, is a crime that the U.S. Department of Justice defines as "any nonconsensual sexual act proscribed by Federal, tribal, or State law, including when the victim lacks capacity to consent" (U.S. Department of Justice, 2019b).[4]

Organizations often implement policies that deal with sexual misconduct, a broad category of behavior that includes but is not limited to legal definitions of harassment and assault (examples include Brandeis University, 2019, Dartmouth College, 2019). Misconduct is generally an umbrella term that describes unwelcome behavior of a sexual nature that is committed without consent and at the expense of another.

As a matter of terminology, we say an agent in the model experiences *sexual misconduct* if his or her experience x_i weakly exceeds a normalized value of zero, with higher values of x_i corresponding with progressively worse acts ranging from harassment to assault and even rape. Negative values of x_i correspond to good experiences. We interpret the model to encompass a wide range of $x_i > 0$ behaviors under the umbrella term of misconduct. We emphasize that we do so not because all forms of misconduct are similar, but because this paper's central focus is on the reporting decisions faced by victims of such behavior, which share a common central feature: the coordination problem.

A given manager may generate both positive and negative x_i experiences across agents. For example, perpetrators of harassment may generate $x_i > 0$ for victims due to a relatively high θ, but also several $x_i < 0$ for other agents. Managers who never commit misconduct generate $x_i < 0$ for an arbitrarily large number of agents due to θ sufficiently less than zero.

The coordination problem in reporting misconduct. Individuals face a coordina-

[4]The definitions and legal consequences of sexual assault and rape vary by state; for a detailed overview of the complex U.S. legal landscape, see Tracy et al. (2012).

tion problem when deciding whether to report misconduct: Agents would like to report if other agents are reporting. The reason is that the cost of reporting misconduct is higher if others do not provide corroborating reports of misconduct, and lower if others do. Because sexual misconduct usually happens in private, reports of misconduct are often unlikely to be believed unless other victims also come forward.

Examples of the importance of corroboration in individuals' decisions to report are abundant. Victims of Harvey Weinstein repeatedly weighed the possibility that their claims would not be taken seriously in the absence of corroboration (Farrow, 2017). Bill Cosby's initial trial included the testimony from only one other accuser and resulted in a mistrial due to jury deadlock. He was convicted upon retrial, when five additional accusers provided similar accounts of sexual assault. Commentary from legal experts indicates that the corroboration from other accusers likely significantly bolstered the primary accuser's credibility (Bowley, 2018; Weiser et al., 2012). These concerns are not limited to high-profile cases, and press accounts suggest that they may be even more warranted when there is no "spotlight" on their case (Weiser et al., 2012).[5]

A key reason that victims face skepticism is that accusations often come down to "he-said/she-said" situations around ambiguous behaviors. For example, women often have different views as to what constitutes harassment and consent (in assault and rape cases) than men have (Jozkowski et al., 2014; Rotundo et al., 2001). Such issues occur even when authorities think all parties are reporting what they believe to be the truth and even when there is physical evidence of a sexual act (Anderson, 2004; Tracy et al., 2012). Accusers also face skepticism because of suspicion that they may be fabricating outright lies, although the incidence of deceitful fabrication appears to be relatively low in practice, and lower than common conception. The rate of rape cases deemed "unfounded" in the FBI Uniform Crime Report (UCR) statistics was 6% in 2006, but likely exceeds the true fabrication rate, as it includes cases that authorities deem truthful yet do not meet the criteria of assault (Lisak et al., 2010; Lonsway, 2010; for similar evidence in Europe, see Kelly, Lovett and Regan, 2005; Lovett and Kelly, 2009). Accusers also face skepticism over why they did not report immediately or physically resist (Lonsway, 2010; Bohner et al., 2013).

The effects of skepticism potentially manifest in the relatively low rates of reporting

[5]Indeed, the importance of corroboration was codified in law, as until recently a number of U.S. states required more corroboration for accusations of sexual assault than for other crimes (Anderson, 2004; Tracy et al., 2012). Most of these laws were repealed in the 1980s.

and arrest for misconduct. The rate of reporting for rape and sexual assault equaled 25% compared to 45% for other types of violent crime in 2018 (U.S. Department of Justice, 2019a), while the clearance rate (rate of cases solved by police, including the arrest rate) for rape equaled 35% compared to 62% for murder in 2017 (U.S. Department of Justice, 2017), patterns that are common (for a review of related statistics, see Kruttschnitt et al., 2014). Daly and Bouhours (2010) survey evidence across the United States, Australia, Canada, and United Kingdom, and estimate that the overall conviction rate for rape from 1970-2005 is 15%. For sexual harassment, the EEOC reports that only 23% of charges resulted in resolutions favorable to accusers (U.S. Equal Employment Opportunity Commission, 2019a).

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

Overall, skepticism underscores the importance of corroboration in an individual's decision to report misconduct. Corroboration provides "safety in numbers" by increasing the chances that relevant outside parties believe a report of sexual assault and mitigates the chances of costly retaliation.

1.1.2 Agent strategies and payoffs

In the model, an agent can report her experience x_i to an outside party, who can potentially sanction the manager. To keep the model flexible, we do not take a strong stand on the identity of the outside party, other than that their purpose is to sanction managers for bad behavior. The outside party could be law enforcement, an internal committee charged with evaluating claims of harassment, or even the court of public opinion. This flexibility allows us to capture a variety of outside parties who may differ in their objective functions, frictions in decision-making (e.g., the outside party need not know θ or any parameters), and belief formation processes, while still allowing us to analyze how the outside party's willingness to act affects agents' willingness to report.

Motivated by our discussion of the coordination problem in the real world, we focus on the central reduced form feature we wish to capture: outside parties are more likely to act if more agents report. Formally, a strategy for an agent is $s(x_i) = 1$ if an agent reports their experience and reveals their x_i to the outside party, and $s(x_i) = 0$ if an agent does not report. If r agents report misconduct (play $s(x_i) = 1$ for $x_i \geq 0$; $r \equiv \int_0^\infty s(x_i)\, di$), the outside party sanctions the manager with probability $\Gamma(r)$, where $\Gamma(r)$ is an increasing function and where the probability is independent of other uncertainty in the model. In our simplest formulation, we assume $\Gamma(r) = \gamma r$, where $\gamma \in [0, 1]$ is the sensitivity of sanction to reporting, but we also analyze more general $\Gamma(r)$.

$\Gamma(r)$ is a reduced form representation of how the outside party's objective function and beliefs translates into action. Although in principle $\Gamma(r)$ is endogenously determined, assuming a reduced form allows our model to focus on the agents' coordination problem without requiring further assumptions about an outside party's objective function and beliefs, which may vary widely across different institutional settings.[6] For example, the legal system may commit to a high standard of proof to assess criminal liability, while a human resources department may commit to a different standard due to concerns about civil liability for a firm. Even after a standard has been adopted, there is residual uncertainty about whether sanctions will follow from a given number of reports due to potential disputes and idiosyncrasies in jury trials and committee proceedings. In both cases, however, once the outside party commits to a standard ex-ante, agents take the probabilistic function $\Gamma(r)$ as given.[7]

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. This would leave the outside party in precisely the "he-said/she-said" situation that motivates the need for corroboration in the first place.[8] For simplicity, the model does not include deceitful fabrication, or for

[6] Chassang and Miquel (2019) study the optimal mechanism to elicit reporting from a single whistleblower. In their setting, a criminal can only potentially retaliate against the whistleblower once the outside party intervenes against the criminal.

[7] An alternative formulation for $\Gamma(r)$ would be a known standard of proof $\overline{r}$ such that sanction occurs if and only if $r \geq \overline{r}$. We view this formulation as less realistic as it features no residual uncertainty. Nonetheless, all of our intuitions and results hold qualitatively with such a function.

[8] For similar reasons, we also assume that the probability the outside party sanctions the manager is based on the number of agents reporting, not the specific x_i reported, which is subject to "he-said/she-said"

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

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

agents to know x_i but strategically represent a different x to the outside party. The objective incidence of outright fabrication appears to be low (as Section 1.1.1 notes), and thus would not meaningfully unravel the coordination problem.

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

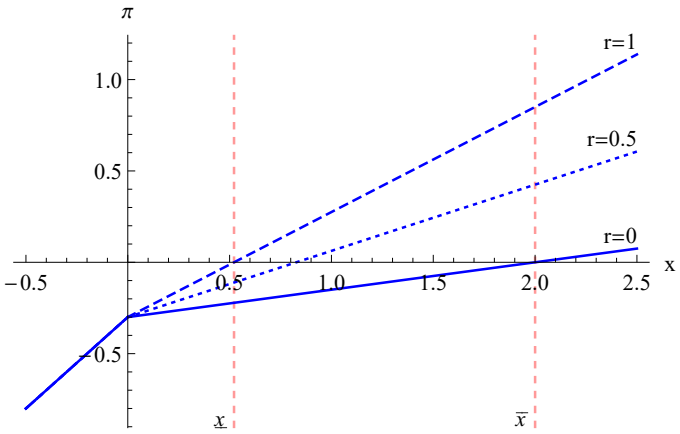
Our payoff specification appeals to the following intuitions. The intrinsic motivation $\omega > 0$ can be driven by moral obligation, a sense of justice, or civic duty.[9] The retaliatory cost $\beta > 0$ can be driven by outright retaliation by the manager, public humiliation, or other costs (Section 1.1.1). We assume that $\omega > \beta$: even if an agent knew an outside party would not sanction the manager, her incentive to report increases as experiences become worse and x_i increases. We also assume that the an agent incurs a retaliatory cost only for reports of misconduct ($x_i > 0$). We assume that both benefit ω and cost β are proportional to x_i: Agents are intrinsically more motivated to report progressively worse experiences, but the retaliation is larger when agents report worse experiences that do not result in sanction.

A key quantity we consider is the *payoff gain* $\pi(r, x_i)$ from reporting, defined as the

disagreement. In the real world, even severe allegations of rape or assault may not lead to sanction or even investigation (Lonsway, 2010; for a recent popular account, see Safronova and Halleck, 2019).

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

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



expected utility from reporting experience x_i given r agents reporting, minus the expected utility from not reporting: $\pi(r,x_i) = Eu(Report,r,x_i) - Eu(NoReport,r,x_i)$, where the expectation is taken over the possibility of sanction given r. Because sanction occurs with probability $\Gamma(r) = \gamma r$, based on Table 1 we have:

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

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

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

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

Finally, there is a *coordination problem*. For any $x_i \in (\underline{x}, \overline{x})$, the payoff gain can be either positive or negative, depending on the number of reporting agents r, due to the existence of

the retaliatory cost $\beta > 0$. If no agents report, then $\pi(0, x) < 0$; if all agents report, then $\pi(1, x) > 0$. A higher retaliatory cost β widens the region of experiences $(\underline{x}, \overline{x})$ where there is a coordination problem, and also worsens the experiences by increasing both $\underline{x}$ and $\overline{x}$. The coordination problem distinguishes this game from games such as the prisoner's dilemma that feature only dominant strategies, and motivates the analysis in the rest of the paper.

1.2 Self-fulfilling equilibria in the absence of information frictions

In the presence of a coordination problem, simple economic theory has difficulty characterizing agents' reporting decisions because equilibria are *self-fulfilling*. To illustrate, consider the following thought exercise that removes the information friction in our model. Suppose that the manager engages in misconduct that is so widespread that every agent experiences identical misconduct: $\sigma = 0$ so that $x_i = \theta \in (\underline{x}, \overline{x})$ $\forall i$ and all agents learn θ and every other agents' experience for certain. The best outcome for agents is for all agents to report because the outside party is most likely to act. These facts are common knowledge among all agents. What does standard analysis say will happen?

No concrete prediction emerges. If an agent expects all other agents to report ($r = 1$), it is optimal to report because the payoff gain of reporting is positive: $\pi(1, x) > 0$ for $x \in (\underline{x}, \overline{x})$, as Figure 1 illustrates. If all agents share this belief, then all agents report, vindicating agents' initial beliefs in an "all-report" equilibrium. Conversely, if an agent expects no other agents to report ($r = 0$), then it is optimal to not report because $\pi(0, x) < 0$, precipitating a self-fulfilling "no-report" equilibrium.[10]

In general, in the absence of information frictions, all-reporting is the payoff-dominant equilibrium for any $\theta > x^E \equiv \underline{x}$, and no-reporting is the payoff-dominated equilibrium. The reason all-reporting is payoff-dominant is that $\pi(1, x) > \pi(0, x)$ for any such θ.

Despite this ordering of equilibrium payoffs, the model without information frictions cannot tell us which equilibrium emerges. Whether reporting occurs is due to a "sunspot" that determines whether agents believe $r = 1$ or $r = 0$. The model is thus unable to answer questions such as whether under-reporting is likely in the real world, and how policies or the #MeToo movement affect reporting. The fundamental reason is that, whenever $\sigma = 0$, all

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

experiences $x_i = \theta$ are common knowledge among all agents, and beliefs about other agents' actions are self-fulfilling.

1.3 Under-reporting with information frictions

Our model focuses on the more realistic situation where $\sigma > 0$ so that agents are uncertain about θ and do not know other agents' experiences x_i. This approach provides structure for how expectations of other agents' actions are determined in equilibrium, which ultimately determines agents' actions. We start by stating the unique equilibrium in the baseline game where agents are ex-ante uninformed about θ.

Proposition 1 (Baseline Equilibrium)**.** *Suppose agents have an improper prior over θ.*

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

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

where:

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

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

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

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

The role of *strategic uncertainty* and beliefs about how many other agents will report play a central role in the intuition. Each agent must ask herself: Given my x, what is the probability that a proportion less than r of other agents have experiences worse than mine? If agents are playing switching strategies around x, then the proportion of agents receiving experiences worse than x is given by $1-\Phi\left((x-\theta)/\sigma\right)$ given the true θ, where $\Phi(.)$ is the normal cumulative distribution function. Re-arranging, the proportion is less than r if $\theta < x-\sigma\Phi^{-1}(1-r)$. If the agent has no idea what θ equals ex-ante, her posterior belief over θ after experiencing x is normally distributed with mean x and standard deviation σ. Then, the probability the proportion falls below r equals $\Phi\left(\left(x-\sigma\Phi^{-1}(1-r)-x\right)/\sigma\right)=r$.[11]

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

1.3.1 Under-reporting

We say there is *under-reporting* in equilibrium if a lower reporting threshold (and thus a higher level of aggregate reporting) would Pareto-improve agent payoffs, given the true distribution of experiences among agents as determined by θ. The following Corollary 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 1.2.

Corollary 1.1 (Under-reporting with widespread misconduct: $\sigma \to 0$)**.** *Let $\sigma \to 0$. There is under-reporting for any manager with $\theta \in (\underline{x}, x^*)$: $r \to 0$ in the unique equilibrium, even though $\pi(1,x) > 0 > \pi(0,x)$ for all but an arbitrarily small mass of agents.*

Corollary 1.1 describes what we call an *open secret equilibrium*. All agents have nearly identically bad experiences (σ is vanishingly small and $\theta > x^E = \underline{x}$). Furthermore, all agents

[11] Proposition 1 does not require normally distributed ϵ; an improper prior is sufficient. As we show in Appendix A, the proof admits any continuous density f with support over the real line, and can admit densities with bounded support with technical modification.

know that all other agents have nearly identical experiences (agents know σ is small), know that all other agents share nearly identical beliefs about θ (the manager's type is an "open secret"), *and* know that all-reporting is the best outcome for all agents ($\pi(1, x) > 0 > \pi(0, x)$ for almost all agents). Yet virtually no agents report because almost all $x_i < x^*$, leading to under-reporting and $r \approx 0$ as the unique outcome. In contrast, every agent would be weakly better off, with some agents strictly better off, if they had played a switching strategy around x^E, as this would achieve $r \approx 1$.

The reason for under-reporting in equilibrium is that strategic uncertainty persists whenever $\sigma > 0$, however small. If $\sigma = 0$ as in Section 1.2, every agent learns θ for certain: the manager's type is not an open secret, it is simply "out in the open." In that case, all agents are also certain of other agent's actions. In contrast, in the presence of information frictions, Part 3 of Proposition 1 says that the marginal agent is uninformed about how many other agents are reporting, even when all experiences are nearly identical, and agents know that all experiences are nearly identical. The key issue is that the experiences themselves are not common knowledge, creating "reverberant doubt" (Morris and Shin, 2002) and uncertainty over whether other agents' actions, even though there is little uncertainty over θ and agents anticipate all other agents' strategy functions $s(\cdot)$ in equilibrium.

Away from the limiting case of $\sigma \to 0$, the following Corollary shows that there may be under-reporting even when $\theta > x^*$.

Corollary 1.2 (Under-reporting with widespread misconduct: High θ)**.** *Let $\sigma > 0$. For $\theta > x^*$, there exists a $\hat{x} < x^*$ such that every agent would be weakly better off, with some agents strictly better off, playing a switching strategy around $\hat{x}$.*

Intuitively, if the average experience θ is greater than x^*, then the true r will be higher than the marginal agent's expectation of r, which is $1/2$. Some agents with x_i below x^* would then be better off by reporting. 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 $\hat{x} < x^*$ thus results in a Pareto improvement. Agents fail to achieve this higher set of payoffs in equilibrium due to their strategic uncertainty, which stems from uncertainty about θ.

Together, Corollaries 1.1 and 1.2 show that under-reporting is likely precisely when misconduct is most widespread, either because misconduct is so wholesale that it is an "open secret" (small σ) or because agents experience significant misconduct on average (any $\theta > x^*$). Information frictions thus provide a rationale for why under-reporting of misconduct can

persist even when large payoff gains for agents are possible from more reporting.[12]

How important is the role of the improper prior in this analysis? The statement of Corollary 1.1 is unchanged even when agents hold proper priors; Corollary 1.2 holds over a different range of θ related to the average type in the prior belief. We defer the details to Section 3, where we discuss the publicity effects of the #MeToo movement.

1.3.2 Implications

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

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

Lowering β also lowers x^*, but in practice retaliation has proven difficult for policymakers to reduce. Although federal law prohibits workplace retaliation, there is ample evidence that various forms of retaliation are prevalent (Cortina and Magley, 2003; McDonald, 2012; U.S. Equal Employment Opportunity Commission, 2016). Increasing ω or lowering c also lower x^*. However, changing these values may prove difficult since these parameters more closely reflect agents' preferences rather than the outside party (γ) or the manager and firm (β).[13]

To see how effects on x^* translate into the amount of equilibrium reporting, the second column of Table 2 considers the effect of changing parameters on *aggregate reporting* $\hat{r}(\theta)$, which equals the total number of agents that report type θ in equilibrium:

$$\hat{r}(\theta) \equiv \int_0^\infty s(x) f(x \mid \theta)\, dx = \int_{x^*}^\infty \phi\left(\frac{x-\theta}{\sigma}\right) dx = \Phi\left(\frac{\theta - x^*}{\sigma}\right),$$

[12] Corollary 1.1 also supports the notion that equilibrium outcomes are constrained-inefficient (Angeletos and Pavan, 2007; Angeletos and Lian, 2016). The reason is that, for $\sigma \to 0$, a planner could improve total agent payoffs without knowledge of θ and without transferring information across agents by directing all agents to play switching strategies around x^E. Our definition of under-reporting is related but distinct: it conditions on θ and thus the ex-post distribution of experiences, and requires that more reporting create a Pareto-improvement in agent payoffs, not just an improvement in the sum of payoffs.

[13] With improper priors, experience dispersion σ does not affect x^*. With proper priors, the comparative static for σ changes. Except for σ, all other comparative statics are invariant to assuming a proper or improper prior. For brevity, we omit details about σ.

Table 2: Comparative statics.

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

where $\phi(.)$ and $\Phi(.)$ are the normal density and cumulative distribution functions.

To continue our discussion of the sanction sensitivity γ, note that increasing γ increases the aggregate number of agents reporting $\hat{r}$ for every type θ, precisely because it lowers the reporting threshold x^*. This insight generalizes to broader sanction functions $\Gamma(r)$, so long as $\Gamma(r)$ is an increasing function. One would replace $\gamma/2$ in Equation 2 with $\int_0^1 \Gamma(r)\, dr$.

Policymakers who encourage reporting through #MeToo should then not be surprised if the marginal reported behavior becomes less reprehensible (x^* falls) and the reporting rate for all manager types rises ($\hat{r}(\theta)$ increases).[14] These effects increase payoffs among agents within the model, although it is important to note that we have not modeled the manager's utility. We next consider a setting where managers can avoid sanction by avoiding agents.

2 The Effects of Endogenous Mentorship

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

We show a manager may avoid mentoring agents in order to avoid sanction, and that

[14]Even if one thought in principle that lowering the reporting threshold might not be warranted because some low-x behaviors are relatively innocuous, our model points out that such a lowering leads to more reporting of high-θ types who commit many high-x acts of misconduct.

the mentoring decision has strategic spillovers onto agents' reporting strategies. Suppose a manager knows his own type θ, and chooses whether or not to mentor agents, $a \in \{0, 1\}$.[15] If $a = 1$, he mentors all agents, after which all agents realize experiences x_i and then make their reporting decisions. An agent who is mentored plays strategy $s(x_i, a)$, where $s(x_i, 1) = 1$ if the agent reports and $s(x_i, 1) = 0$ if the agent does not report. If $a = 0$, the manager mentors no agents, and no agents report: $s(x_i, 0) = 0$.[16] As in Section 1, the manager's type θ determines the average agent experience. We maintain the assumption that agents have improper priors over θ.

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

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

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

[15] 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.

[16] 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.

2.1 Strategic spillovers

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

Proposition 2 formalizes this observation and shows that this strategic spillover creates a trade-off between selection and reporting. If high-misconduct (high-θ) types opt out of mentorship, agents become less willing to report misconduct because corroboration is less likely when most other agents are not experiencing misconduct; if low-misconduct types opt out of mentorship, then agents become more willing to report misconduct. The proposition also shows that which effect prevails depends on whether $M(\theta)$ increases or decreases in θ.

Proposition 2 (Selection and strategic spillovers)**.** *Let $s(x_i, a)$ denote an agent's reporting strategy:*

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

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

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

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

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

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

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

at a finite θ' as follows:

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

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

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

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

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

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

Part 2(b) shows that low-misconduct types are likely to opt out when $M(\theta)$ is increasing, which could happen because high-misconduct types are predatory and derive significant utility from misconduct. The sufficient condition starkly illustrates with an $M(\theta)$ that discontinuously jumps at θ'. In this case, low-misconduct types opt out as they do not receive positive mentoring utility ($M(\theta) < 0$ for $\theta < \theta'$), while high-misconduct types mentor as their

$M(\theta)$ is so high that it exceeds expected penalties even if all agents report ($M(\theta) > \gamma S$ for $\theta > \theta'$). More broadly, if $M(\theta)$ is continuous and increasing in θ, but there exists some θ' such that $M(\theta) < 0$ for $\theta < \theta'$, a set of intermediate-θ types may also opt out in addition to the low-θ types.

Proposition 2 presents two starkly contrasting cases that illustrate the robust prediction that a manager's decision to mentor always generates a strategic spillover on reporting. In reality, the true shape of the $M(\theta)$ function is an open empirical question beyond the scope of this paper, but the strategic spillover is always present.[17] The Proposition also provides a theoretical basis for the popular concern, supported by survey evidence (Atwater et al., 2018), that #MeToo could have the unintended effect of dampening the career prospects of women through fewer networking and mentoring opportunities.

2.2 Unintended effects of tools intended to encourage reporting

We now examine how tools that are intended to encourage reporting or sanction misconduct have unintended consequences on agents' reporting decisions when managers can endogenously choose whether to mentor. We study two interventions: increasing the sanction penalty S, and increasing the sensitivity of sanction to reporting γ. To fix ideas, we analyze the simplest case where mentorship utility does not change with manager types: $M(\theta) = m$, where $m > 0$ is constant. Proposition 3 characterizes the equilibrium of this environment.

Proposition 3 (Equilibrium when $M(\theta) = m$)**.** *Suppose agents have an improper prior over θ. If $S > m/\gamma$, there is a unique symmetric perfect Bayesian equilibrium in which the manager mentors ($a = 1$) if $\theta \leq \tilde{\theta}$ and does not mentor ($a = 0$) if $\theta > \tilde{\theta}$. The marginal type $\tilde{\theta}$ that mentors agents is given by:*

$$\tilde{\theta} = \frac{c}{\omega - \beta\left(1 - \frac{m}{2S}\right)} + \sigma\Phi^{-1}\left(\frac{m}{\gamma S}\right). \tag{3}$$

Agent's reporting strategies equal:

$$s(x_i, 1) = \begin{cases} 1 & \text{if } x_i \geq x_S^* \\ 0 & \text{if } x_i < x_S^*, \end{cases} \tag{4}$$

$$x_S^* = \frac{c}{\omega - \beta\left(1 - \frac{m}{2S}\right)}. \tag{5}$$

[17] Proposition 2 holds even if sanction S depends on the accusations x. The sanction function would affect which types opt out, which affects reporting in the same way as Proposition 2.

Table 3: Comparative statics given manager selection.

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

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

In this unique equilibrium, high-misconduct types are reluctant to mentor agents, so Part 1 of Proposition 2 applies. Because agents realize they are mentored by a $\theta < \tilde{\theta}$ type, they are more reluctant to report any misconduct because they are more pessimistic about how many others will report due to the right-truncation in beliefs over r: $x_S^* > x^*$. Table 3 summarizes comparative statics within the game considered by Proposition 3.

2.2.1 The effect of sanctions

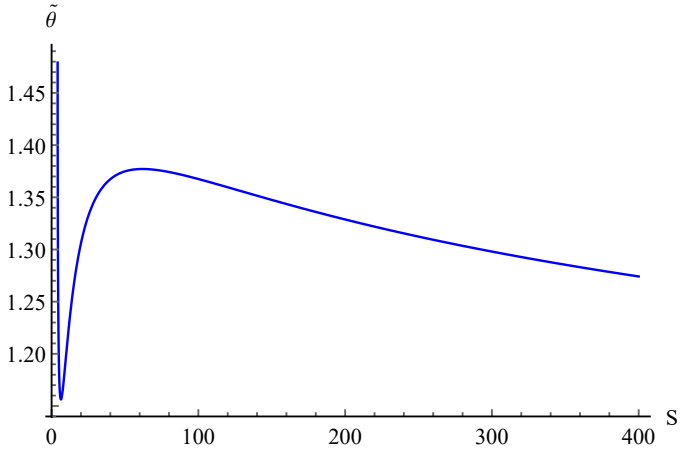
Corollary 3.1 shows that increasing the sanction S depresses reporting and can have potentially adverse effects for agents through several channels.

Corollary 3.1. *The following hold with respect to S:*

1. *Increasing sanction S raises the reporting threshold and lowers the reporting of the marginal type: $\frac{\partial x_S^*}{\partial S} > 0$, and $\frac{\partial \tilde{r}}{\partial S} < 0$.*
2. *Increasing sanction S can lower $\tilde{\theta}$: $\frac{\partial \tilde{\theta}}{\partial S} < 0$ when $S = \frac{m}{\gamma}$ or as $S \to \infty$. A sufficient condition for $\frac{\partial \tilde{\theta}}{\partial S} < 0$ is:*

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

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



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

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

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

Parts 2 and 3 together show that increasing sanctions have ambiguous effects on the identity of the marginal type $\tilde{\theta}$ that chooses to mentor. The intuition for the ambiguity is because $\tilde{r}$ can fall in two ways when S rises: 1) directly because x_S^* rises, and 2) indirectly because the marginal type $\tilde{\theta}$ changes. In Part 2, the higher x_S^* has a small effect on $\tilde{r}$, so $\tilde{\theta}$ must fall to induce the required fall in $\tilde{r}$ that compensates for an increase in S. In Part 3, the higher x_S^* may have such a large effect on $\tilde{r}$ that $\tilde{\theta}$ can increase. Figure 2 illustrates a case where $\tilde{\theta}$ is non-monotone with respect to S.[18]

This leads to the following striking conclusion: there can exist different levels of S that result in the same amount of mentoring and the same level of expected sanction for the

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

marginal type, yet the *lowest* S among these leads to the lowest reporting threshold for agents and the highest aggregate reporting of any given manager. One can see this by drawing a horizontal line in Figure 2, as it can intersect the $\tilde{\theta}$ curve at up to three points. All three points correspond with identical amounts of mentoring. Because the expected sanction is also the same for all three points, the point with the lowest S must correspond with the highest probability of sanction and highest $\tilde{r}$.[19]

2.2.2 The effect of sanction sensitivity

Corollary 3.2 highlights that selection effects blunt the effectiveness of tools intended to encourage reporting through an increase in γ, the sensitivity of sanction to reporting.

Corollary 3.2. *In the presence of selection, agents' reporting strategies are less responsive to increases in γ compared to the case without selection: $\frac{\partial x_S^*}{\partial \gamma} = 0 < \frac{\partial x^*}{\partial \gamma}$.*

Intuitively, without selection, an increase in γ increases an agent's incentive to report because it increases the probability of sanction for any given level of reporting r. However, in the presence of selection with constant $M(\theta)$, agents expect a lower r because they expect fewer agents will report in equilibrium (Part 1 of Proposition 2), weakening this effect. When $\Gamma(r) = \gamma r$, the selection effect is strong enough to reduce the effect of γ on reporting to zero. One can show analogous results for a class of non-linear functions, $\Gamma(r) = \gamma r^{\frac{1}{n}}$ where $n > 0$.

2.3 Takeaways

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

[19] 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.

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

3 The Publicity Effect of #MeToo

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

We show that heightened public awareness of misconduct increases reporting of otherwise-hidden misconduct by coordinating beliefs over strategic uncertainty. We associate changes in public information with changes in agents' common priors (Angeletos and Lian, 2016). In Section 1.3, we assumed that agents had a common improper uniform prior over θ. Now suppose that agents share an informed prior $p(\theta)$ that is normally distributed with mean y and standard deviation τ. Such a prior could be shaped by public information about a specific manager or information about misconduct in the broader population of managers. Proposition 4 characterizes behavior in this environment. The proof strategy follows Morris and Shin (2004a); Appendix A provides all details.

Proposition 4 (Equilibrium with informed priors)**.** *Suppose agents have a common informed prior that θ is distributed normally with mean y and standard deviation τ.*

1. *Existence: There exists at least one symmetric switching strategy equilibrium $x_I^* \in$*

$(\underline{x}, \overline{x})$ *where agents report if and only if* $x_i \geq x_I^*$*, where* x_I^* *is implicitly defined by:*

$$x_I^* = \frac{c}{\omega - \beta\left(1 - \gamma\Phi\left(\frac{y - x_I^*}{\kappa}\right)\right)}, \tag{6}$$

for $\kappa \equiv \frac{\sigma^2+\tau^2}{\sigma}\sqrt{\frac{\sigma^2+2\tau^2}{\sigma^2+\tau^2}}$.

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

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

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

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

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

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

$$\Psi_I(r) \equiv \Phi\left(\frac{\sigma}{\tau}\frac{1}{\sqrt{\sigma^2+\tau^2}}(x_I^* - y) + \frac{1}{\tau}\sqrt{\sigma^2+\tau^2}\,\Phi^{-1}(r)\right), \tag{7}$$

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

3.1 Equilibrium uniqueness and under-reporting

There is a unique equilibrium when priors are sufficiently diffuse or when private experiences are informative for θ relative to priors. Specifically, Part 2(a) of Proposition 4 is satisfied for high τ or when σ is low relative to to τ. The reason is that κ increases in τ; furthermore, κ decreases in σ when $\sigma^2/\tau^2 < \sqrt{2}$. Indeed, as $\tau \to \infty$ or $\sigma \to 0$, we converge to the unique equilibrium in Proposition 1, because $\kappa \to \infty$ and $\Psi_I(r) \to r$.

This observation comports with well-known results in the literature about global games: precise public information induces coordination and multiple equilibria, while private information hinders coordination because signals are not common knowledge (Angeletos and

Table 4: Comparative statics with informed priors. Comparative statics for ω, c, γ, and β are identical to those in Section 1.3.2. Note that $y > x_I^*$ if and only if $y > \frac{c}{\omega - \beta\left(1-\frac{\gamma}{2}\right)}$.

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

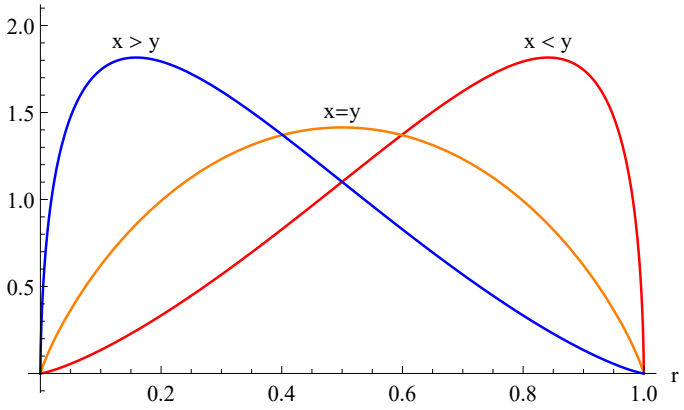
Lian, 2016; Morris and Shin, 2003, 2004a). Part 2(b) illustrates equilibrium multiplicity. Suppose $y = \frac{c}{\omega - \beta\left(1-\frac{\gamma}{2}\right)}$, so that $y = x_I^*$ and $E_I^*[r] = 1/2$. Multiplicity occurs if $\tau \approx 0$ and σ is not too large.[20] We discuss the relevance of Part 2(c) below.

From here on, we assume the environment satisfies the condition for a unique switching 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.

In the unique equilibrium, under-reporting occurs if misconduct is widespread. Because we converge to the equilibrium of Proposition 1 when $\sigma \to 0$, the statement of Corollary 1.1 is unchanged even when priors are informed. Corollary 1.2 also holds with a slight modification: when priors are informed, there is a Pareto-dominating threshold $\hat{x} < x_I^*$ so long as $\theta > x_I^* + \frac{\sigma}{\kappa}(y - x_I^*)$. If the prior mean y is low, Corollary 1.2 can hold even for $\theta < x_I^*$, implying under-reporting for an even larger range of θ.

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

Figure 3: Probability density function $\psi_I(r)$. This figure plots the probability density function of the marginal agent's belief over the number of agents reporting for three different cases denoted in the figure. The cumulative distribution function for this belief is $\Psi_I(r)$ given in Proposition 4.



3.2 Implications for #MeToo

Table 4 summarizes how changes in the mean and standard deviation of the prior belief, y and τ, affect equilibrium reporting. Corollary 4.1 highlights the key implication relevant for the #MeToo movement.[21]

Corollary 4.1 (Publicity effect of #MeToo)**.** *Agents become more willing to report (x_I^* falls) when y increases. When y is high, agents also become more willing to report when τ falls.*

If y increases so that the public believes that managers as a whole are engaging in worse misconduct on average than previously believed, the reporting threshold x_I^* falls and aggregate reporting $\hat{r}(\theta)$ rises for every θ. The reason is that an increase in y makes agents believe the average experience is worse and hence more agents are likely to report.

The intuition is easiest to see when we are starting from an equilibrium where $y = \frac{c}{\omega - \beta\left(1-\frac{\gamma}{2}\right)}$ so that $x_I^* = y$. The marginal agent believes $E_I^*[r] = \frac{1}{2}$, and Figure 3 plots the density of her belief over r. Note that, from Equation 7, 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 the marginal agent will think

[21] The comparative statics for ω, c, γ, and β are identical to those in Section 1.3.2. For σ, distinct from the uninformed prior case in Section 1.3.2, an increase in σ can either increase or decrease the reporting threshold x_I^*, because the change in σ has two effects on beliefs over r. First, it affects the marginal agent's assessment of how likely her experience was relative to what other agents might be experiencing based on her priors, but it also affects how much she revises her belief about what experience she expects others to have when forming her posteriors. Which effect dominates depends on σ^2/τ^2 and whether y is low or high. For brevity, we omit details about σ as it seems less directly relevant to the #MeToo movement.

that more agents will report because they are having worse experiences, making her willing to report and no longer indifferent.

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

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

Corollary 4.1 is a robust first-order prediction irrespective of underlying parameters: Part 2(c) from Proposition 4 guarantees that the equilibrium is unique when $y \to \infty$ and that $x_I^* \to \underline{x}$, irrespective of other parameters. Part 2(c) also shows that Corollary 4.1 has first-order effects on the magnitudes of x_I^*, as changes in y can move the equilibrium threshold x_I^* across the entire range of $(\underline{x}, \overline{x})$.

Overall, Corollary 4.1 is consistent with more agents coming forward with accusations in the wake of #MeToo, which publicized several major incidents of misconduct and arguably raised y (and perhaps decreased τ) by raising public awareness of sexual misconduct. Our model suggests that heightened awareness led directly to more reporting even though: 1) experiences may have remained unchanged, and 2) there was no direct impact on agent payoffs $\pi(r, x)$. In particular, agents with "hidden x_i" who were previously not reporting due to a low-y environment may come forward in a higher-y (possibly lower-τ) environment purely from a change in beliefs about whether other agents are reporting.

Table 5: Agent's Payoffs with Whistleblower Rewards. The parameter α is the whistleblower reward. $\mathbf{1}_{[x_i \geq 0]}$ is an indicator function that equals one if $x_i \geq 0$.

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

4 Rewarding Whistleblowers

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

Given these difficulties, we next consider the effect of rewarding reports that result in subsequent action. Such a "whistleblower reward" for coordination success is a natural counterweight to a retaliatory cost for coordination failure. For example, in practice, victims of workplace sexual harassment can win emotional distress damages in Title VII lawsuits. More broadly, aside from monetary rewards, agents may also receive utility when the outside party acts, such as social recognition or a sense of vindication.

We incorporate whistleblower rewards in the model by extending the payoff structure so that an agent receives a proportional reward αx_i if the third party acts on a claim of $x_i \geq 0$. Table 5 summarizes an agent's final payoffs. Given these payoffs, an agent's payoff gain function π_R equals:

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

Note that the payoffs in Table 5 and the payoff gain function $\pi_R(r, x)$ nest the previous payoff structure with $\alpha = 0$. All previous results on existence, uniqueness, and comparative statics of other parameters in Sections 1-3 continue to apply with straightforward modification, and proofs of all results incorporate the extended payoff structure allowing for $\alpha \geq 0$. For brevity, we provide the formulas for the reporting and selection thresholds when $\alpha > 0$ in Appendix A and omit them in the main text.

In this section, we focus on the role of α for equilibrium behavior. It is straightforward

to show that α mitigates the retaliatory cost β because agents are more willing to report for higher α ($\partial x^*/\partial\alpha < 0$ and $\partial\hat{r}(\theta)/\partial\alpha > 0\,\forall\theta$). The following Proposition captures our main insight from implementing whistleblower rewards.

Proposition 5 (Required rewards)**.** *For every $\beta \in (0,\omega)$, there exists a unique* required reward $\alpha^F > 0$ *such that the equilibrium reporting threshold equals the "no-retaliation benchmark," $x^F = c/\omega$. If $\alpha < \alpha^F$, the equilibrium threshold exceeds x_F, so that there is less reporting than in the no-retaliation benchmark. Conversely, if $\alpha > \alpha^F$, the equilibrium threshold is less than x_F, so that there is more reporting.*

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

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

Proposition 5 suggests that rewards for coordination success have the potential to significantly and even fully offset the retaliatory cost of coordination failure. A sufficiently high α can induce each agent to report as though she were sure that others were all reporting: that is, so that agents play the reporting threshold $x^E = \underline{x} = c/(\omega - \beta(1-\gamma))$ in equilibrium. Recall from the discussion of Corollary 1.1 that x^E is the threshold that achieves the highest payoff for agents when $\sigma \to 0$ and $\alpha = 0$, given all other parameters. Even at such a threshold, however, an agent faces possible retaliation if $\gamma < 1$. An even higher α can induce each agent to report as though she were sure she would not face retaliation: that is, so that agents play the threshold $x^F = c/\omega$ in equilibrium, where $x^F < x^E$.

The Proposition also suggests that a reward of $\alpha > \alpha^F$ leads to an equilibrium reporting threshold that is lower than x^F and hence to more reporting than in the no-retaliation benchmark. In this case, α itself creates a coordination motive: agents report behavior that they otherwise would not report even in a world where retaliation did not exist. To see this, consider the extreme case where $\beta = 0$ so that $\alpha^F = 0$, but where $\alpha > 0$. Agents in this case are motivated to report when other agents report only due to the possibility of the reward, which in this case mitigates the fixed cost of reporting c.[22] With $\alpha = 0$, changes in β or γ cannot generate an equilibrium threshold less than x^E, let alone x^F.

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

We highlight three other observations from Proposition 5. Part 1 suggests that there is a potential substitution across different policy tools: The required reward α^F is lower when γ is high or β is low. Part 2 suggests that a higher reward is needed to counteract the strategic spillover effect of manager selection on reporting when $M(\theta)$ is constant. With the same improper prior, α^F is higher in the game with selection than without selection, since $S > m/\gamma$. Likewise, the required reward increases with the size of the sanction and decreases with the size of mentorship utility.

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

5 Conclusion

This paper shows that coordination and strategic uncertainty are fundamental to understanding economic agents' decisions to report sexual misconduct. The coordination problem leads agents to under-report sexual misconduct if misconduct is widespread, and provides an explanation for unintended and intended effects of #MeToo.

Future research should expand on the implications of the coordination problem. Potential theoretical avenues include considering specific legal contexts that allow for a closer micro-foundation of the outside party and further endogenizing managers' decisions to allow for social welfare calculations. Empirical avenues include quantifying the causal relationship of retaliation with the frequency and severity of reports of misconduct, studying whether encouraging reporting leads to fewer soft opportunities for agents, and estimating required inputs for social welfare analysis. Given that the #MeToo movement is leading both business leaders and policymakers to re-think how they deal with sexual misconduct, these are first-order questions for future research.

References

American Economic Association, "AEA Professional Climate Survey: Main Findings," September 15 2019. https://www.aeaweb.org/resources/member-docs/

final-climate-survey-results-sept-2019 (accessed September 19, 2019).

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

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

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

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

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

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

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

Ayres, Ian and Cait Unkovic, "Information Escrows," *Michigan Law Review*, 2012, *111* (2), 145.

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

Bénabou, Rolad, Armin Falk, and Jean Tirole, "Narratives, Imperatives, and Moral Persuasion," September 2019.

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

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

Bohner, Gerd, Friederike Eyssel, Afroditi Pina, Frank Siebler, and G Tendayi Viki, "Rape Myth Acceptance: Cognitive, Affective and Behavioural Effects of Beliefs that Blame the Victim and Exonerate the Perpetrator," in "Rape," Willan, 2013, pp. 40–68.

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

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

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

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

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

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

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

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

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

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

Daly, Kathleen and Brigitte Bouhours, "Rape and Attrition in the Legal Process: A Comparative Analysis of Five Countries," *Crime and Justice*, 2010, *39* (1), 565–650.

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

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

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

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

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

Forde, Kaelyn, "What Experts Say about 'Corroboration,' a Key Word From the Kavanaugh Hearings," *ABC News*, October 5 2018. https://abcnews.go.com/US/kavanaugh-hearings-wrong-corroboration-hurts-survivors/story?id=58254664 (accessed July 25, 2019).

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

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

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

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

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

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

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

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

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

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

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

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

Kruttschnitt, Candace, William D. Kalsbeek, and Carol C. House, *Estimating the Incidence of Rape and Assault*, Washington, D.C.: National Academies Press, 2014.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Safronova, Valeriya and Rebecca Halleck, "These Rape Victims Had to Sue to Get the Police to Investigate," *The New York Times*, May 23 2019. https://www.nytimes.com/2019/05/23/us/rape-victims-kits-police-departments.html (accessed May 23, 2019).

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

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

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

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

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

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

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

__ , "Crime in the United States, 2017," 2017. https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topic-pages/clearance-browse-by/national-data (accessed September 24, 2019).

__ , "Criminal Victimization, 2018," September 2019. NCJ #253043.

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

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

__ , "Charges Alleging Sexual Harassment, FY 2010 - FY 2018," 2019. https://www.eeoc.gov/eeoc/statistics/enforcement/sexual_harassment_new.cfm. (accessed September 14, 2019).

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

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

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

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

Appendix A Proofs of Main Propositions

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

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

A.1 Proof of Proposition 1

Lemmas A1, A2, and A3 set the stage from which Proposition 1 follows. The latter two lemmas are a direct application of Proposition 2.1 of Morris and Shin (2003) and Lemma A.2 of Morris and Shin (2004a), with technical modification to account for the fact that r only includes the number of agents reporting $x_i > 0$. We spell out the Lemmas for completeness and because we will use Lemma A3 in proving Proposition 4, which is not a direct application. We refer to the reader to Morris and Shin (2003) and Morris and Shin (2004a) for their proofs but also provide step-by-step proofs with the appropriate technical modifications to our setting in Appendix B.

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

A.1.1 Intermediate results

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Let $r(\theta; k)$ equal the proportion of other agents drawing $x > k$ for type θ for any $k \geq 0$, 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 $k \geq 0$, there is a one-to-one mapping from r into θ:

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

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

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

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

Lemma A2. *The following properties hold:*

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

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

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

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

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

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

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

Proof. We know $\pi(r, x)$ satisfies Properties P1-P5 in Lemma A1. By Proposition 2.1 from Morris and Shin (2003), the Lemma follows. Appendix B.1 contains the detailed step-by-step proof appropriately modified for our setting. ■

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

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

A.1.2 Proposition 1

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

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

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

A.1.3 Corollary 1.2

Proof. For any $\hat{x} < x^*$, total agent payoffs equal:

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

We know that equilibrium total payoffs are such that:

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

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

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

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

Observe that for $\theta > x^*$,

$$\begin{aligned} 0 &= x^* \left(\omega - \beta + \gamma \frac{1}{2}(\alpha + \beta)\right) - c \\ &< x^* \left(\omega - \beta + \gamma \Phi\left(\frac{\theta - x^*}{\sigma}\right)(\alpha + \beta)\right) - c. \end{aligned} \tag{A.2}$$

Let $H(x) \equiv x\left(\omega - \beta + \gamma\Phi\left(\frac{\theta - x}{\sigma}\right)(\alpha + \beta)\right) - c$. Note that $H(x^*) > 0$ from Equation A.2. If $H'(x^*) < 0$, then there exists a $\hat{x} < x^*$ such that:

$$\begin{aligned} 0 &< x^*\left(\omega - \beta + \gamma\Phi\left(\frac{\theta - x^*}{\sigma}\right)(\alpha + \beta)\right) - c \\ &< \hat{x}\left(\omega - \beta + \gamma\Phi\left(\frac{\theta - \hat{x}}{\sigma}\right)(\alpha + \beta)\right) - c. \end{aligned}$$

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

$$\begin{aligned} 0 &< \hat{x}\left(\omega - \beta + \gamma\Phi\left(\frac{\theta - \hat{x}}{\sigma}\right)(\alpha + \beta)\right) - c \\ &< x^*\left(\omega - \beta + \gamma\Phi\left(\frac{\theta - x^*}{\sigma}\right)(\alpha + \beta)\right) - c. \end{aligned}$$

If $H'(x^*) = 0$, observe that because $H^{(n)}(x^*) \neq 0$ for some $n > 0$, we can apply a similar argument to find an $\hat{x} < x^*$ such that $H(\hat{x}) > 0$. Either way, there exists some $\hat{x} < x^*$ such that $H(\hat{x}) > 0$.

But then:

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

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

$$x\left(\omega - \beta + \gamma\hat{r}\left(\theta \mid \hat{x}\right)(\alpha + \beta)\right) - c \;>\; H(\hat{x}) > 0,$$

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

A.2 Proof of Proposition 2

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

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

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

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

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

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

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

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

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

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

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

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

For the marginal agent, this is $\pi^*(x, k, \theta < \tilde{\theta}) = \int_0^{\tilde{r}} (\frac{1}{\tilde{r}}) \pi(r, x) dr = 0$. To find the solution - denote the cutoff $x_i = k \equiv x_S^*$. Given the marginal agent's beliefs, x_S^* is given by Equation A.3:

$$\begin{aligned}\pi^*(x, k, \theta < \tilde{\theta}) &= 0 \\ \int_{r=0}^{\tilde{r}} (\frac{1}{\tilde{r}}) \left(x_S^* [\omega + \gamma r \alpha - (1 - \gamma r)\beta] - c \right) dr &= 0 \\ x_S^* &= \frac{c}{\omega - \beta + \frac{1}{2}\gamma(\alpha + \beta)\tilde{r}}. \qquad \text{(A.3)}\end{aligned}$$

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

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

unique. As shown above, the equilibrium $\tilde{\theta}$ and x_S^* are thus given by Equations A.3 and $M(\tilde{\theta}) - \gamma S\hat{r}(\tilde{\theta}) = 0$.

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

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

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

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

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

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

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

Because there is a one-to-one mapping of r and θ, then the agent's reporting threshold satisfies: $\pi^*(x, k, \theta > \tilde{\theta}) = \int_{\tilde{r}}^{1} \psi_\sigma^*(r; x, k, \theta > \tilde{\theta})\pi(r, x)dr$. For the marginal agent, this is $\pi^*(x, k, \theta > \tilde{\theta}) = \int_{\tilde{r}}^{1} (\frac{1}{1-\tilde{r}})\pi(r, x)dr$. To find the solution - denote the cutoff $x_i = k \equiv x_S^*$. Given the marginal agent's beliefs, x_S^* is given by Equation A.4:

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

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

$$x_S^* = \frac{c}{\omega - \beta + \frac{1}{2}\gamma(\alpha+\beta)(1+\tilde{r})}. \tag{A.4}$$

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

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

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

where $\theta' \in (-\infty, \infty)$, $g(\theta') < 0$, $h(\theta') > \gamma S$, and $g(\theta) : \mathbb{R} \to \mathbb{R}$ and $h(\theta) : \mathbb{R} \to \mathbb{R}$ are weakly increasing functions of θ. For any given threshold x_S^*, $-\gamma S\hat{r}(\theta)$ strictly decreases in θ. Since $g(\theta)$ is weakly increasing in θ and $g(\theta') < 0$, any $\theta < \theta'$ receives a strictly negative payoff from mentoring. Thus any type $\theta < \theta'$ would choose $a = 0$ in any equilibrium. Since $h(\theta') > \gamma S$ and $h(\theta)$ is weakly increasing in θ, any type $\theta \geq \theta'$ receives a strictly positive payoff from mentoring, even if $\hat{r}(\theta) = 1$. Thus types $\theta \geq \theta'$ would choose $a = 1$ in any equilibrium. Therefore the unique separating equilibrium is that equilibrium such that types choose $a = 1$ if and only if $\theta \geq \theta'$. As shown above, the equilibrium reporting threshold x_S^* is thus given by Equation A.4 where $\tilde{r} = \Phi(\frac{\theta' - x_S^*}{\sigma})$ and $\tilde{\theta} = \theta'$.

A.3 Proof of Proposition 3

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

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

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

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

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

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

This leaves a unique equilibrium in which types $\theta \leq \tilde{\theta}$ choose $a = 1$ and types $\theta > \tilde{\theta}$ all choose $a = 0$. Given this selection, the proof of Part 1 of Proposition 2 applies. The equilibrium is the solution to Equations A.5 and A.6, which are the indifference conditions for the marginal agent and manager, respectively:

$$\int_{r=0}^{\tilde{r}} (\frac{1}{\tilde{r}}) \left(x_S^*[\omega + \gamma r\alpha - (1 - \gamma r)\beta] - c\right) dr = 0 \tag{A.5}$$

$$m - \gamma\Phi(\frac{\tilde{\theta} - x_S^*}{\sigma})S = 0. \tag{A.6}$$

Note that Equation A.6 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.5, we obtain $x_S^* = \frac{c}{\omega - \beta + (\frac{m}{2S})(\alpha+\beta)}$ and $\tilde{\theta} = \frac{c}{\omega - \beta + (\frac{m}{2S})(\alpha+\beta)} + \sigma\Phi^{-1}(\frac{m}{\gamma S})$. Clearly, there is no incentive for any type to deviate given x_S^* and $\tilde{\theta}$.

A.3.1 Corollary 3.1

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

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

$$\frac{\partial \tilde{\theta}}{\partial S} = \left(\sigma\frac{1}{\phi\left(\frac{\tilde{\theta} - x_S^*}{\sigma}\right)} - \frac{c\gamma\left(\alpha + \beta\right)\frac{\partial \tilde{r}}{\partial S}}{2\left(\omega - \beta + \frac{1}{2}\tilde{r}\gamma\left(\alpha + \beta\right)\right)^2}\right)\frac{\partial \tilde{r}}{\partial S}. \tag{A.7}$$

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

$$\begin{aligned}
\sigma \frac{1}{\phi\left(\frac{\tilde{\theta}-x_S^*}{\sigma}\right)} &> \frac{c\gamma\left(\alpha+\beta\right)}{2\left(\omega-\beta+\frac{1}{2}\tilde{r}\gamma\left(\alpha+\beta\right)\right)^2} \\
\sigma &> \phi\left(\frac{\tilde{\theta}-x_S^*}{\sigma}\right)\frac{c\gamma\left(\alpha+\beta\right)}{2\left(\omega-\beta+\frac{1}{2}\tilde{r}\gamma\left(\alpha+\beta\right)\right)^2} \\
\sigma &> \phi\left(\Phi^{-1}\left(\tilde{r}\right)\right)\frac{c\frac{1}{2}\gamma\left(\alpha+\beta\right)}{\left(\omega-\beta+\frac{1}{2}\tilde{r}\gamma\left(\alpha+\beta\right)\right)^2}. \qquad \text{(A.8)}
\end{aligned}$$

Thus a sufficient condition for $\frac{\partial \tilde{\theta}}{\partial S} < 0$ is $\sigma > \frac{1}{\sqrt{2\pi}}\frac{c\frac{1}{2}\gamma(\alpha+\beta)}{(\omega-\beta)^2}$. When $S = \frac{m}{\gamma}$, $\tilde{r} = 1$ so from Equation A.7 we have $\frac{\partial \tilde{\theta}}{\partial S} < 0$ when $S = \frac{m}{\gamma}$. As $S \to \infty$, $\tilde{r} \to 0$ so $x_S^* \to \underline{x}$ and $\tilde{\theta} \to -\infty$. From Equation A.8, we have $\frac{\partial \tilde{\theta}}{\partial S} < 0$ as $S \to \infty$.
[3] A sufficient condition for $\frac{\partial \tilde{\theta}}{\partial S} > 0$ when $\tilde{r} = 1/2$ is $\sigma < \frac{1}{\sqrt{2\pi}}\frac{c\gamma(\alpha+\beta)}{2\left(\omega-\beta+\frac{1}{4}\gamma(\alpha+\beta)\right)^2}$.

A.4 Proof of Proposition 4

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

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

Define $r\left(\theta,k\right)$ as in the proof of Proposition 1. Recall that we can consider $k \geq 0$ and that for such k there is a one-to-one map of r into θ: $r\left(\theta;k\right) = 1 - \Phi\left(\frac{k-\theta}{\sigma}\right)$ so $\theta\left(r;k\right) = k - \sigma\Phi^{-1}\left(1-r\right)$, 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 switching strategies around $k \geq 0$ when $x > 0$ equals:

$$\begin{aligned}
\pi^*\left(x,k\right) &= \int_{-\infty}^{\infty} f\left(\theta \mid x\right)\pi\left(r\left(\theta;k\right),x\right)\,d\theta \\
&= x\left(\omega-\beta+\gamma\left(\alpha+\beta\right)\int_{-\infty}^{\infty} f\left(\theta \mid x\right)\left(1-\Phi\left(\frac{k-\theta}{\sigma}\right)\right)\,d\theta\right) - c. \qquad \text{(A.9)}
\end{aligned}$$

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

A.4.1 Intermediate results

Lemma A4. *Any solution x^* to $\pi^*\left(x,x\right) = 0$ satisfies the implicit equation:*

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

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

Proof. We have $f(\theta \mid x) = \frac{1}{v}\phi\left(\frac{\theta-\lambda}{v}\right)$. Then Equation A.9 becomes:

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

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

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

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

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

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

$$\begin{aligned}\pi^*(x,k) &= x\left(\omega-\beta+\gamma(\alpha+\beta)\,\Phi\left(\frac{\lambda-k}{\sqrt{\sigma^2+v^2}}\right)\right)-c \text{ if } x>0,\\ &= x\omega - c \text{ if } x\le 0.\end{aligned} \tag{A.10}$$

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

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

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

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

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

Lemma A5. *There exists at least one symmetric switching strategy equilibrium $x_I^* \in (\underline{x}, \overline{x})$ where agents report for $x \ge x_I^*$ and do not report for $x < x_I^*$.*

Proof. From Lemma A4, any potential equilibria must solve $G(x^*) = 0$, where:

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

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

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

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

Second, we show that any such solution x^* constitutes a symmetric switching equilibrium. Recall from Equation A.10 in Lemma A4 that the payoff from reporting when an agent's signal is $x > 0$, conditional on other players playing switching strategies around k, equals:

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

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

Lemmas A6 and A7 substitute for Lemma A2 in the informed prior case.

Lemma A6. *The following properties hold:*

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

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

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

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

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

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

Proof. [1] These properties are evident from Equation A.10 in Lemma A4.

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

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

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

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

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

Lemma A7. *Uniqueness of equilibrium (allowing for either strategy to be played at x^*):*

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

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

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

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

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

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

Proof. Because $G\left(\underline{x}\right) < G\left(\bar{x}\right)$ with at least one solution in between where G is increasing from Lemma A5, and because G is differentiable for all x, a necessary and sufficient condition for uniqueness is that $G'\left(x^*\right) > 0$ for all $x^*\in\left(\underline{x},\bar{x}\right)$ such that $G\left(x^*\right) = 0$. We have

$$\begin{aligned} G'\left(x\right) &= \gamma\left(\alpha+\beta\right)xE'\left(x\right) + \left(\omega-\beta+\gamma\left(\alpha+\beta\right)E\left(x\right)\right) && \text{(A.11)}\\ &= \omega-\beta+\gamma\left(\alpha+\beta\right)\left(xE'\left(x\right)+E\left(x\right)\right), && \text{(A.12)}\end{aligned}$$

and also:

$$G'\left(x\right) = \gamma\left(\alpha+\beta\right)xE'\left(x\right)+\frac{c}{x} \quad \text{(A.13)}$$

Substituting in $E'\left(x\right) = -\frac{1}{\kappa}\phi\left(\frac{y-x}{\kappa}\right)$ into Equation A.13, a necessary and sufficient condition for uniqueness is, for all solutions x^*:

$$\gamma\left(\alpha+\beta\right)\left(x^*\right)^2\phi\left(\frac{y-x^*}{\kappa}\right)\frac{1}{\kappa} < c. \quad \text{(A.14)}$$

Recall from Lemmas A4 and A5 that any solution to $G(x)$ constitutes a symmetric switching equilibrium and that there exists at least one such equilibrium. We now provide conditions under which such an equilibrium is unique or not unique.

[a] Using the fact that $\phi\left(z\right) < \frac{1}{\sqrt{2\pi}}$ and $x^* < \bar{x} = \frac{c}{\omega-\beta}$, a sufficient condition for uniqueness from Equation A.14 is:

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

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

expression for $G'(y)$ from Equation A.11:

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

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

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

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

To show uniqueness in both cases, observe from Equation A.12 that:

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

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

[d] If $G(x)$ has a unique solution x^*, then from Lemma A6 we have $x^* = \xi = \underline{\xi} = \overline{\xi}$. By Lemma A3, the only strategy which survives the iterated deletion of dominated strategies is the x^*-switching strategy. This implies that the x^*-switching equilibrium is the globally unique equilibrium. ■

Lemma A8. *The marginal agent has beliefs over r summarized by the cumulative distribution function $\Phi\left(\frac{t}{\sqrt{t+u}}(x-y) + \frac{\sqrt{t+u}}{\sqrt{u}}\Phi^{-1}(r)\right)$, with expectation $E_I^*[r] = \Phi\left(\frac{y-x^*}{\kappa}\right)$.*

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

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

For the marginal agent with $x = k$,

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

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

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

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

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

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

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

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

where $E_I^*[r] = \int_0^1 r\, \psi(r; x, x)\, dr$. But then by Lemma A4, $E_I^*[r] = \Phi\left(\frac{y - x^*}{\kappa}\right)$. ■

A.4.2 Proposition 4

Proof. Part 1 follows from Lemma A5. Part 2 follows from Lemma A7. Part 3 follows from Lemma A8. ■

A.5 Proof of Proposition 5

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

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

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

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

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

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

To obtain the comparative statics:

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

[*b*] By direct differentiation:

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

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

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

Appendix B Supplemental Detail for Proofs (Internet Appendix)

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 Proof of Lemma A2

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

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

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

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

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

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

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

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

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

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

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

[3] Note that we can write:

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

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

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

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

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

B.2 Proof of Lemma A3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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