Monitoring Harassment in Organizations

Laura Boudreau
Columbia University

Sylvain Chassang
Princeton University

Ada González-Torres
Ben Gurion University

Rachel Heath*
University of Washington

July 24, 2022

Abstract

We study the value of garbled survey methods as a tool to monitor harassment. Theory predicts that randomly switching reports that no harassment took place to reports that harassment did take place can improve information transmission by guaranteeing participants plausible deniability in the event they file an incriminating report. We evaluate this prediction in a phone-based survey of workers at apparel manufacturing plants in Bangladesh. We vary the survey method (direct or garbled), the degree of personally identifiable information (team id) associated with the report, as well as the degree of rapport built with respondents. We find that garbling increases reporting of sexual harassment by about 306%, physical harassment by 295%, and threatening behavior by 56%. We also find a negative effect of attaching team id to the report. We use the improved data to assess policy-relevant aspects of harassment: How prevalent is it? What share of managers is responsible for the misbehavior? How isolated are victims? How do harassment rates compare for men and women? Based on the answers to these questions, we draw implications for decision-makers.

KEYWORDS: Harassment, garbling, secure survey design, gender, garments, Bangladesh

*This project is funded by the Private Enterprise Development in Low-income Countries (PEDL) Initiative and by Columbia University's Provost's Diversity Grants Program for Junior Faculty. We are grateful to Ferdausi Sumana, Raied Arman, and Krishna Kamepalli for their excellent research assistance.

1 Introduction

Organizations' ability to take action against harassment is limited by their ability to elicit information from relevant parties. Reporting harassment, especially if one's report is associated with personally identifying information (PII), is a difficult step for individuals who have been victimized and for witnesses concerned with possible retaliation and reputational costs. This prevents organizations from responding to individual issues, but also from assessing the scope and nature of their harassment problem. We study the impact of garbled survey methods on information transmission. Theory predicts that randomly switching reports that no harassment took place to reports that harassment did take place can improve information transmission by guaranteeing participants plausible deniability in the event they file an incriminating report (Warner, 1965, Chassang and Padró i Miquel, 2018, Chassang and Zehnder, 2019). We evaluate this prediction in a phone-based survey experiment about workplace harassment with employees of a large apparel manufacturer in Bangladesh.

This paper's first goal is to evaluate the impact of different aspects of survey design – hard garbling (HG), rapport building (RB), and removing team-level information – on respondents' propensity to report misbehavior. In HG, the surveyor exogenously imposes information garbling, for example, by programming a computer-based survey to randomly flip a share of "no" responses to "yeses." In RB, the surveyor allocates survey time to build "rapport," or trust, with the respondent, chatting about family and hobbies in a natural but pre-specified manner, beyond the minimum small talk in a typical social science survey. Finally, the surveyor can limit the amount of PII requested from the respondent; we eliminate questions about the identity of the respondent's direct supervisor and their production team.

HG aims to increase information transmission by reducing the expected retaliation and/or reputational cost of reporting harassment, while the latter two approaches aim to reduce respondents' perception of the risk of survey data leakage. In all three approaches, the possible benefit of increased willingness to report comes at a cost: HG provides a noisy signal of misbehavior, which constrains the severity of organizational responses to reports; RB requires

careful planning of the RB process, additional training of survey enumerators, and more time to conduct the survey; and removing team-level information precludes computation of manager-level statistics that are important to characterize the nature of an organization's harassment problem.

This paper's second goal, if we find that one or more of our experimental design aspects increases reporting of harassment, is to use our improved data to assess several policy-relevant aspects of harassment: How prevalent is it? Is a small share of managers responsible for the bulk of the misbehavior? How isolated are its victims? How do harassment rates compare for men and women? The answers to these questions are crucial inputs to determining the policies that can be used to address harassment. For example, if a small share of managers is responsible for the harassment, the organization could simply fire them. In contrast, if most managers are involved, firing them all is likely impossible, and other remedial actions need to be taken.

We collaborated with the apparel producer to conduct phone-based surveys with workers at two of its plants. We surveyed 2,197 workers and had a response rate of 61%.[1] We randomly assigned survey respondents to 9 different combinations of the treatment conditions: HG, RB, and low PII. The status quo or baseline treatment arm entailed direct elicitation (DE) of respondents' experience of harassment, no RB, and elicitation of PII. We examine the effects of our survey design interventions on three pre-specified outcomes: reporting of threatening behavior, physical harassment, and sexual harassment by respondents' direct supervisors.

We find that reporting rates in the survey's control group are low, especially for physical and sexual harassment: 9.47% of respondents report threatening behavior, 1.46% report physical harassment, and 1.70% report sexual harassment. HG increased reporting of threatening behavior by 56%, sexual harassment by about 306%, and physical harassment by 295%. We also find that reducing the amount of PII increased reporting of threatening behavior by 39% and of physical harassment by 279%, but it had no effect on reporting of sexual

[1]Nearly all non-response was due to our inability to reach workers by phone.

harassment. Finally, we find that RB had positive but weak effects. There is evidence of complementarity between treatment arms for threatening behavior and physical harassment. That is, combining hard garbling with rapport-building and limiting PII significantly increases reporting compared to the sum of the effects of implementing each feature alone.

We find a surprising pattern of heterogeneous treatment effects (HTEs) by respondents' sex. Compared to women, men's baseline reporting rates were higher for threatening behavior and physical harassment and slightly lower for sexual harassment. The effects of HG were substantially larger for men compared to women for both threatening behavior and sexual harassment, although we lack power to detect statistical differences between the effects for men and women.

Next, we aim to use our improved reporting data to estimate several policy-relevant statistics of harassment. Doing so requires using garbled data to construct estimators of statistics that depend on respondents' *intended* reports. Warner (1965) derives a consistent estimator for the mean intended reporting rate using garbled data. We extend this result to the team case. We derive consistent estimators of team-level statistics under varying HG schemes, including independent and identically distributed (i.i.d.) HG and what we refer to as blocked HG. With blocked HG, the surveyor ensures that a target number of possible "no" reports are flipped to "yeses," either in the overall sample or per team. Blocked HG, in particular at the team-level, substantially reduces the variances of estimators.[2]

Using the garbled data, we estimate that 17.3% of workers reported experiencing some form of harassment by their supervisor in the past year. 14% reported threatening behavior, 5.5% reported physical harassment, and 8% reported sexual harassment. On average, there are 9 workers per production team in the HG arm; considering teams of this size, we find that over 80% had at least one worker who had experienced some form of harassment, nearly 80% had at least one who had been threatened, just over 50% had at least one who had been sexually harassed, and just under 50% had at least one who had been physically harassed.

[2]It also affords workers with less protection in case of a data leakage, which could be an important consideration in some contexts.

These statistics indicate that harassment is widespread in this organization, and a policy of firing all misbehaving supervisors is unlikely to be feasible. Among teams that had at least one worker report some form of harassment, 57% had at least two workers report. This suggests that escrow mechanisms along the lines of Ayres and Unkovic (2012), which seek to help coordinate the reports of multiple victims, may be relevant in the majority of harassment cases. That said, victims' isolation varies significantly by harassment type. Physical harassment victims were most isolated, with the likelihood of a second worker reporting conditional on the first falling to 13.5%. If an escrow system were to operate at the team-by-issue type-level, or if victims only felt protected by an escrow when reporting the same type of harassment, an escrow system would miss a large share of victims.

Our experimental results and statistics of harassment elucidate that lack of plausible deniability caused severe under-reporting of harassment in this organizational setting. They shed light on the nature of harassment: It was widespread among teams. The extent to which victims were isolated across teams varied substantially by harassment type. This is the first field evidence on HG outside of a lab and in a real-world organizational setting; our findings demonstrate that HG can dramatically increase reporting of harassment and improve the measurement of key statistics. Reducing PII also helps, at the cost of being unable to calculate policy-relevant manager-level statistics.

This research contributes to an emerging literature in economics on workplace harassment, in particular sexual harassment, and its implications for labor markets. Cheng and Hsiaw (2020) consider reasons for underreporting of sexual harassment; they develop a model in which harassment is underreported if there are multiple victimized individuals because of coordination problems. Dahl and Knepper (2021) also examine causes of underreporting, providing evidence that U.S. employers use the threat of retaliatory firing to coerce workers not to report sexual harassment. Folke and Rickne (2022) study sexual harassment and gender ratios in Sweden's labor market; they find that women and men, respectively, are more likely to be harassed in occupations and workplaces that are dominated by the other sex. Wages do not appear to compensate for sexual harassment risk, and individuals who report

sexual harassment have lower job satisfaction. We contribute evidence that lack of plausible deniability causally negatively effects reporting of workplace harassment. Our findings indicate that estimates of labor supply and other responses to harassment may be severely biased when harassment is measured using formal complaints, as it may be in workplaces where harassment is most problematic that workers face the highest retaliation/reputation costs associated with reporting, and reporting is most suppressed.

In the context of developing countries, sexual harassment in the workplace and in public spaces is considered to be a key barrier to women's labor market participation (Jayachandran, 2021).[3] There is a dearth of evidence, however, on the effects of sexual harassment and violence in the workplace on workers' labor supply and well-being.[4] Further, in light of workers' lack of access to secure internal reporting channels (Boudreau, 2022) and to recourse through criminal justice systems, as well as relatively stronger gender norms, we expect underreporting to be even more of a concern in many developing countries. We contribute to our understanding of the prevalence and nature of harassment in a low-skill manufacturing sector that is common to many developing countries. Our evidence confirms that harassment against women by managers who are men is common, and it shows that harassment by men against subordinate men is also substantial. In the context of the garments sector, the large majority of workers are women, so research and policymaking that focuses on reducing harassment against women is of paramount concern, but harassment against men in garments and similar sectors needs more attention.

This research also contributes to the literature on the detection and deterrence of collusion, corruption, and other forms of misbehavior in organizational settings. A large body of contract theory literature with principal-agent-monitor set-ups considers the possibility of

[3]One stream of literature establishes that harassment is prevalent in public spaces and transit systems in cities ranging from Rio de Janeiro to Delhi and that it reduces women's educational investments and labor supply (Aguilar et al., 2021, Kondylis et al., 2020, Borker, 2018, Chakraborty et al., 2018, Siddique, forthcoming).

[4]The poor working conditions (Boudreau et al., 2022) and extreme gender imbalances between managers and workers (Macchiavello et al., 2020) documented in the literature on Bangladesh's garments sector are suggestive of possible harassment concerns.

bribes in collusive relationships between monitors and agents to limit information transmission to the principal (Tirole, 1986, Laffont and Martimort, 1997, 2000, Prendergast, 2000, Faure-Grimaud et al., 2003, Chassang and Ortner, 2019). More recently, a smaller strand of literature considers that collusion may come in the form of punishments against informants, or whistleblowers (Heyes and Kapur, 2009, Bac, 2009, Makowsky and Wang, 2018). Chassang and Padró i Miquel (2018) develop a model in which misbehaving agents can commit to effective retaliation and show that only garbled intervention policies are effective at disciplining behavior. They clarify how to evaluate such policies even in the hypothetical presence of malicious workers wrongfully reporting well-behaved managers. We contribute by bringing HG into a real-world organizational setting to test its effects on information transmission in a context that shares many features of these frameworks. Our finding of very large effects of HG on information transmission presents a promising direction for future theoretical and empirical work to consider HG in contexts where credible threats or reputation costs limit information transmission.

Finally, this research contributes to a literature on garbled survey designs and on inference from garbled surveys dating back to Warner (1965) and Greenberg et al. (1969). Warner (1965) proposed randomized response (RR), a "soft garbling" technique in which the surveyor instructs a respondent only to provide sensitive information on a probability basis, such as privately rolling a die to determine whether to answer the sensitive question or an unrelated question.[5] The empirical literature on survey design for sensitive questions has found that RR performs better than DE (Rosenfeld et al., 2016), but that survey respondents often do not comply with the protocol to garble, and they provide the least sensitive responses (Chuang et al., 2020). In organizational settings, theory suggests that HG can increase information transmission relative to RR (Chassang and Zehnder, 2019), which we discuss in Section 3.4. Our contribution is twofold. First, we clarify theoretically that blocked HG designs deliver more precise estimates than i.i.d. garbling, which is the only option available

[5]For example, a respondent may be instructed to answer the question "Is the sky blue?" if the die lands on 1 or 2, and to answer to question "Have you experienced harassment?" if the die lands on 3-6. The surveyor does not observe the respondent's die roll.

under RR. This is especially valuable when baseline reporting rates are low and sampling error can dwarf the statistic of interest. Second, we derive consistent estimators for team- or group-level statistics of intended responses for inference using garbled data.

The remainder of this paper is organized as follows. Section 2 provides background on Bangladesh's garments sector and the anonymous apparel producer whom we partner with. Section 3 details the policy-relevant statistics that we are interested in, which must be estimated using workers' reports. It develops a theoretical framework for a worker's decision to report harassment and how HG, RB, and removing team-level identifiers might influence this decision. Finally, it derives consistent estimators for team-level statistics of harassment and discusses statistical inference. Section 4 presents the research design. Section 5 presents the results of the reporting experiment. Section 6 uses the garbled survey data to characterize the apparel producer's harassment problem. Finally, Section 7 discusses our findings and concludes.

2 Context

2.1 Garments Production in Bangladesh

Garments production is organized into cutting, sewing, and finishing sections; some factories also have wet and dry washing sections, which adds texture and/or fading to sewn garments (e.g., denim jeans). Within these sections, workers are organized into production teams or lines, with team assignments that are largely stable over time. The organizational structure is very hierarchical, above teams of workers are supervisors, followed by line chiefs or team incharges, floor-supervisors and/or assistant production managers, production manager(s), and finally, the managing director. Production sections vary dramatically in their sex composition: Cutting and wet washing sections typically exclusively employ men, sewing and finishing sections mostly employ women, and dry washing sections are often closer to parity. In contrast, more than 90% of managers in all sections are men.

Ethnographic evidence and evidence from community-based surveys suggests that harassment is a long-running problem in Bangladesh's garments sector (Siddiqi, 2003, Sumon et al., 2018, Kabeer et al., 2020). Workers' precarious livelihoods and lack of legal recourse, as well as conservative societal norms around gender and sex, contribute to an enabling environment (Siddiqi, 2003). While there is reason to believe that harassment is widespread, measuring and constructing informative statistics of harassment is extremely challenging, even in social science research conducted outside of the workplace. For example, using data from Kabeer et al. 2020's community-based survey of garment workers, we find that while 20% (11%) of workers report witnessing physical (sexual) harassment, only 1% (0%) report experiencing it themselves.

2.2 Collaboration with a Bangladeshi Apparel Producer

We conducted this research in collaboration with one of the largest apparel producers in Bangladesh.[6] The producer owns 7 plants that collectively employ around 26,000 workers. The producer's senior leadership team sought a collaboration with our research team because it wished to improve relations with its workers and to improve workers' well-being. To achieve this, it aimed to directly collect feedback from workers' on their experiences in the workplace and relationships with their managers. It then aimed to use this feedback to inform its HR policies. In the short-term, we agreed to collaborate on a survey of workers at 2 of its plants. In the longer-term, the senior management team's goal was to set-up a reporting system for workers to provide continuous feedback in real-time.

Turning to production, the firm specializes in manufacturing denim jeans, sweaters, suits, and other apparel items. Between 34-42% of workers in the 2 plants are employed on sewing lines, 16-18% are employed in finishing, and between 10-14% are employed in washing, with remaining workers employed in smaller, supporting production sections. The organizational hierarchy and gender composition of production sections are similar to those described above

[6]We have a confidentiality agreement with the apparel manufacturer.

for the sector. 93% of managers are men.

3 Framework

We consider an organization consisting of $m \in \mathbb{N}$ teams. Each team $a \in M \equiv \{1, \cdots, m\}$ consists of a manager (also denoted by a) and L workers indexed by $i \in I \equiv \{1, \cdots, L\}$. Altogether, the organization consists of $n \equiv m \times L$ workers and m managers.

We assume for simplicity that all harassment is performed by managers against workers under their span of control. For any manager a and worker i, we denote by $h_{i,a} = 1$ the event that manager a harassed worker i, and by $h_{i,a} = 0$ the event that they did not. We denote by $h_a \in \{0,1\}^L$ the profile of harassment choices made by manager a.

3.1 Policy-relevant statistics.

Throughout the paper, we take as given the behavior of managers, and seek to elicit information about patterns of harassment $(h_a)_{a \in M}$. Specifically, we are interested in identifying a number of policy-relevant statistics helpful in assessing policy options. We emphasize that these statistics are not directly computable since they depend on harassment patterns not directly observed by the decision maker. We discuss the reporting of harassment below.

Patterns of harassment. We are primarily interested in computing the following statistics:

$$\begin{aligned} \mathsf{S}_V &\equiv \frac{1}{n} \sum_{a,i \in M \times I} h_{i,a}, \\ \mathsf{S}_{PM} &\equiv \frac{1}{m} \sum_{a \in M} \max_{i \in I} h_{i,a}, \\ \forall k \in \{1, \cdots, L\}, \quad \mathsf{S}_{TV \geq k} &\equiv \frac{1}{m} \sum_{a \in M} \mathbf{1}_{\sum_{i \in I} h_{i,a} \geq k}, \\ \mathsf{E}_{2V|1V} &\equiv \frac{\mathsf{S}_{TV \geq 2}}{\mathsf{S}_{TV \geq 1}}. \end{aligned}$$

Statistic S_V measures the share of victimized workers. This allows us to gauge the magnitude of the harassment problem in the organization, allowing stakeholders to correctly prioritize the issue, as well as allocate suitable resources.

Statistic S_{PM} measures the share of problem managers, i.e. managers that have harassed at least one person. It is a special case of statistic $\mathsf{S}_{TV\geq k}$ for $k = 1$. This last statistic measures the share of managers that have harassed at least k workers. The behavior of $\mathsf{S}_{TV\geq k}$ as k increases clarifies policy options. If there exists k large such that $\mathsf{S}_{TV\geq k}$ is small, but $k \times \mathsf{S}_{TV\geq k}$ is large, then this means that a relatively small share of managers is responsible for a large amount of the damage. This means that simply firing repeated offenders may be a viable policy option for the organization. If instead S_{PM} is large but $k\mathsf{S}_{TV\geq k}$ is small for k large, then this means that many managers are involved in harassment, and that it's not possible to address a significant number of cases by firing a small amount of managers. Since firing many managers is likely impossible for the organization, this means that other remedial action will have to be taken, such as improved training or better monitoring.

Finally, $\mathsf{E}_{2V|1V}$ measures the likelihood that a manager has at least 2 victims given that they have at least one. This allows us to assess how isolated victims are. If $\mathsf{E}_{2V|1V}$ is small, then victims are isolated. This implies that escrow mechanisms along the lines of Ayres and Unkovic (2012), which seek to help coordinate the reports of multiple victims, are unlikely to be helpful. In contrast, if $\mathsf{E}_{2V|1V}$ is close to 1, then victims are rarely isolated. This means that escrow mechanisms could be helpful, and that once someone complains, it may be possible to cross-validate reports of misbehavior, permitting more effective action.

Sensitive statistics. These four statistics differ in the sensitivity of information required to compute them. Specifically, it is not necessary to know a particular worker's team to compute S_V. Anonymous reports, provided they are unique, are sufficient. In contrast, S_{PM}, $\mathsf{S}_{TV\geq k}$, and $\mathsf{E}_{2V|1V}$ all require the respondent to associate some team identifier to their report. Otherwise, it is not possible to match the reports of different workers on the same team. For this reason, these statistics are intrinsically more sensitive than S_V: surveys needed to

compute these sensitive statistics will need to include both team ids and harassment reports.

Third-party witnesses. In principle, harassment may be observed by workers other than the victim, and decision makers may be interested in statistics of harassment calculated using information furnished by witnesses. In this paper, our focus is on reporting of one's own harassment status. We leave the question of witnesses' role in detecting and counter-acting harassment to future research.

3.2 Reporting and the value of garbled survey designs

Understanding and addressing harassment is made difficult by the fact that victims are often unwilling to come forward. This may be because of explicit or implicit threats of retaliation, concerns over one's own reputation, or negative impacts on one's career and private life, even if the organization takes action against the perpetrator.

We consider binary surveys in which worker i in team a can submit intended responses $r_{i,a} \in \{0, 1\}$. Rates of complaints in our application are low, and the implicit stigma associated with reports of harassment (especially of a sexual nature) is high. For this reason, we assume there are no false positives: $r_{i,a} \in \{0, h_{i,a}\}$. We discuss the possibility of subversion in the longer-run in Appendix B.4.

Following Chassang and Padró i Miquel (2018) and Chassang and Zehnder (2019), we allow for garbled survey methods that add noise to the report sent by a worker. An intended report $r \in \{0, 1\}$ is associated with potentially random recorded report $\widetilde{r} = \phi(r) \in \Delta(\{0, 1\})$.

Concretely, we are interested in the following survey designs:

- *Direct Elicitation*, in which $\phi(r) = r$: the recorded report is equal to the intended report.

- *Hard Garbling*, in which $\phi(1) = 1$, but

$$\phi(0) = \begin{cases} 0 & \text{with probability } 1 - \pi \\ 1 & \text{with probability } \pi \end{cases}$$

  where $\pi \in (0, 1)$. In words, reports of harassment are systematically recorded, but reports indicating no harassment are switched to reports of harassment with an interior probability π.

For the remainder of this section, unless otherwise noted, we refer to Hard Garbling as "garbling." The rationale for garbling surveys is to guarantee the worker plausible deniability in the event that their record is leaked. In particular, we assume that the worker assigns subjective probability $p \in [0, 1]$ on their recorded report $\widetilde{r}_i^a$ being leaked. We do not take a stance on whether leaks actually occur or not. In our experimental application, leaks exist only in the mind of respondents: other than our analysis, no data has or will be released. However, we are interested in the use of reporting systems for ongoing monitoring in organizations. In such a context, "leaks" may simply correspond to the fact that some action is taken by the organization on the basis of the recorded report.[7] Such leaks are inevitable even under ideal governance.

Worker i's utility U_i associated with an intended report r consists of direct benefits from reporting, as well as potential reputational costs:

$$U_i(r|h_{i,a}) = \mathsf{PB}(\mathrm{r}|\mathrm{h_{i,a}}) + \mathsf{SB}(\widetilde{\mathrm{r}}|\mathrm{h_{i,a}}) + \mathsf{p} \times \mathsf{RC}(\widetilde{\mathrm{r}})$$

where:

- PB is a psychological benefit from taking action such that $\mathsf{PB}(1|1) > 0$ and for simplicity $\mathsf{PB}(1|0) = \mathsf{PB}(0|1) = \mathsf{PB}(0|0) = 0$.

- SB is a social benefit from realized report $\widetilde{r}$ as perceived by the worker, either because

[7] For instance, the manager is sent to a training seminar.

it triggers an investigation, or because it helps the organization design better policies. For simplicity, we assume that $\mathsf{SB}(1|1) > 0$, $\mathsf{SB}(1|0) < 0$ and $\mathsf{SB}(0|1) = \mathsf{SB}(0|0) = 0$.[8]

- $\mathsf{RC}(\widetilde{\mathrm{r}})$ is a reputational cost in case recorded report $\widetilde{r}$ is leaked; we assume it takes the form $\mathsf{RC}(\widetilde{r}_{i,a}) = -\mathsf{K}(\text{prob}(r_{i,a} = 1|\widetilde{r}_{i,a}))$ where K is a positive continuous strictly increasing function, and $\text{prob}(r_{i,a} = 1|\widetilde{r}_{i,a})$ is the posterior belief about $r_{i,a}$ conditional on report $\widetilde{r}_{i,a}$.[9]

The value of garbling reports. In equilibrium, a non-harassed worker always finds it optimal to submit intended report $r_{i,a} = 0$. The expected payoffs from sending reports $r_{i,a} = 1$ and $r_{i,a} = 0$ are

$$U_i(1|0) = \mathsf{SB}(1|0) - \mathsf{pK}(\text{prob}(\mathrm{r_{i,a}} = 1|\widetilde{\mathrm{r}}_{\mathrm{i,a}} = 1)) < 0$$
$$U_i(0|0) = \pi \times (\mathsf{SB}(1|0) - \mathsf{pK}(\text{prob}(\mathrm{r_{i,a}} = 1|\widetilde{\mathrm{r}}_{\mathrm{i,a}} = 1))) < 0.$$

In turn, a harassed worker's payoffs are

$$U_i(1|1) = \mathsf{PB}(1|1) + \mathsf{SB}(1|1) - \mathsf{pK}(\text{prob}(\mathrm{r_{i,a}} = 1|\widetilde{\mathrm{r}}_{\mathrm{i,a}} = 1))$$
$$U_i(0|1) = \pi \times (\mathsf{SB}(1|1) - \mathsf{pK}(\text{prob}(\mathrm{r_{i,a}} = 1|\widetilde{\mathrm{r}}_{\mathrm{i,a}} = 1))).$$

Hence a harassed worker is willing to send intended report $r = 1$ if and only if

$$\mathsf{PB}(1|1) + (1 - \pi)[\mathsf{SB}(1|1) - \mathsf{pK}(\text{prob}(\mathrm{r_{i,a}} = 1|\widetilde{\mathrm{r}}_{\mathrm{i,a}} = 1))] \geq 0. \tag{1}$$

[8] The assumption that $\mathsf{SB}(1|0) < 0$ implies that arbitrarily high garbling rates π are not a priori desirable.

[9] This functional form is well suited to capture concerns over ex post retaliation by managers and related career concerns. We could also include a term dependent on $\text{prob}(h_{i,a} = 1|\widetilde{r}_{i,a})$. It would not change the analysis.

The posterior belief $\text{prob}(r_{i,a} = 1|\widetilde{r}_{i,a} = 1)$ takes the form

$$\text{prob}(r_{i,a} = 1|\widetilde{r}_{i,a} = 1) = \frac{1}{1 + \pi \frac{\text{prob}(r=0)}{1-\text{prob}(r=0)}}.$$

Taking as given the share of null reports $\text{prob}(r = 0)$, increasing garbling rate π increases the worker's incentives to send report $r_{i,a} = 1$ for two reasons: it shrinks the reputational cost savings associated with sending a null report $r_{i,a} = 0$; it reduces the reputational impact of a positive recorded report $\widetilde{r}_{i,1}$.

Altogether, this set of results can be summarized as follows.

Proposition 1 (the value of garbling)**.** *Taking as given the behavior of managers,*

(i) intended reports underreport true harassment: $r_{i,a} \leq h_{i,a}$*;*

(ii) equilibrium reporting weakly increases with the garbling rate π*;*

(iii) equilibrium reporting weakly decreases with the perceived leakage probability p*.*

A corollary of Proposition 1 is that both garbling and reducing the perceived leakage probability increase the accuracy of intended reports.

Let S^r_V, S^r_{PM}, and $\mathsf{S}^r_{TV\geq k}$ denote analogues of S_V, S_{PM}, and $\mathsf{S}_{TV\geq k}$ computed using intended reports $r_{i,a}$ instead of actual harassment status $h_{i,a}$.

Corollary 1. *Measurement errors* $|\mathsf{S}_V - \mathsf{S}^r_V|$, $|\mathsf{S}_{PM} - \mathsf{S}^r_{PM}|$, *and* $|\mathsf{S}_{TV\geq k} - \mathsf{S}^r_{TV\geq k}|$ *are decreasing in garbling rate* π *and increasing in the perceived leakage probability* p*.*

3.3 Measurement

Corollary 1 argues that garbling intended messages reduces the bias of policy-relevant statistics. However these policy-relevant statistics are computed on the basis of intended reports which need not be made available to the analyst.

We now discuss the measurement of S^r_V, S^r_{PM}, and $\mathsf{S}^r_{TV\geq k}$ from garbled data under different garbling schemes. Appendix B briefly discusses the pros and cons of adopting a "trusted analyst" approach in which the econometrician uses intended responses to estimate key relationships, but only releases garbled data and a small set of estimated regression coefficients.

Inference from garbled reports. A key insight of Warner (1965) is that statistic S^r_V can be consistently estimated from garbled data, even though it depends on intended reports. The following estimator is consistent

$$\mathsf{S}^{\widetilde{r}}_V \equiv \frac{\frac{1}{n}\sum_{a,i\in M\times I}\widetilde{r}_{i,a} - \pi}{1-\pi}. \tag{2}$$

It turns out the same is true for other statistics of intended reports, such at $\mathsf{S}^{\widetilde{r}}_{\mathsf{TV}\geq\mathsf{k}}$.

Let $\mu \in \Delta(\{0,1\}^L)$ and $\widetilde{\mu} \in \Delta(\{0,1\}^L)$ respectively denote the sample distribution of profiles of intended and recorded reports, $(r_a)_{a\in M}$ and $(\widetilde{r}_a)_{a\in M}$ across teams:

$$\forall r \in \{0,1\}^I, \quad \mu(r) \equiv \frac{1}{m}\sum_{a\in M}\mathbf{1}_{r_a=r} \quad \text{and} \quad \widetilde{\mu}(r) \equiv \frac{1}{m}\sum_{a\in M}\mathbf{1}_{\widetilde{r}_a=r}.$$

Proposition 2 (identification under i.i.d. garbling)**.** *Assume that garbling is independent across workers. As m grows large, the sample distribution of intended reports μ is identified from the sample distribution of recorded reports $\widetilde{\mu}$.*

A consistent estimator as the number of teams m grows large is provided in the proof (Appendix A). This generalization of Warner (1965) allows us to compute consistent estimates of statistics S^r_V, S^r_{PM}, and $\mathsf{S}^r_{TV\geq k}$, which are all functions of distribution μ.

Balanced garbling. A difficulty with estimators constructed under i.i.d. garbling, such as estimator $\mathsf{S}^{\widetilde{r}}_V$ defined in equation 2, is that sampling error due garbling can dwarf the mean reporting rate $\overline{r} \equiv \frac{1}{n}\sum_{a,i\in M\times I} r_{i,a}$. If this is the case, it may be useful to use correlated garbling to ensure that the realized potential garbling rate is equal to the expected garbling

rate. We clarify the efficiency gain from balanced, or blocked, garbling in the context of Warner's estimator (eqn. 2).

Garbled reports can be expressed as

$$\widetilde{r}_i = r_i + (1 - r_i)\eta_i \tag{3}$$

where $\eta_i \in \{0, 1\}$ is a Bernoulli random variable equal to 1 with probability π.

For concision, we index players by $j \in \{1, \cdots, n\}$ rather than $a, i \in M \times I$. The sum of garbled reports can be expressed as

$$\sum_{j=1}^{n} \widetilde{r}_j = \sum_{j=1}^{n} r_j + \underbrace{\sum_{j=1}^{n} \eta_j}_{A} - \underbrace{\sum_{j=1}^{n} r_j \eta_j}_{B}.$$

When garbling terms η_i are i.i.d. across workers, then

$$\mathsf{Var}\left(\sum_{j=1}^{n} \widetilde{r}_j \middle| \mathbf{r}\right) = (1 - \overline{r})\pi(1 - \pi)n.$$

When the average reporting rate $\overline{r}$ is small most of the variance is due to sampling error in aggregate garbling term A.

For this reason, whenever the mean reporting rate $\overline{r}$ is small, it is beneficial to consider garbling terms η_j that are exchangeable but ex ante balanced, i.e. that $\sum_{j=1}^{n} \eta_j = n\pi$. This ensures that A is a constant so that the only remaining uncertainty is assigned to term B.

Since terms $(\eta_j)_{j \in \{1, \cdots, n\}}$ are ex post balanced,[10] $\mathsf{Cov}(\eta_j, \eta_{j'}) = -\frac{\pi(1-\pi)}{n-1}$ and

$$\mathsf{Var}\left(\sum_{j=1}^{n} \widetilde{r}_j \middle| \mathbf{r}\right) = \mathsf{Var}\left(\sum_{j=1}^{n} r_j \eta_j \middle| \mathbf{r}\right) = \overline{r}\left(1 - \frac{\overline{r}n - 1}{n - 1}\right)\pi(1 - \pi)n. \tag{4}$$

A proof of the second equality in eqn. 4 is provided in Appendix A. Whenever $\overline{r}$ is small,

[10] By ex post balanced, we mean that the joint distribution of $(\eta_1, \cdots, \eta_n)$ is unchanged when labels are switched, but the variables are weakly correlated, so that the realized sum is equal to the ex ante sum.

as is the case in our application, this induces a significant reduction in the variance of the estimator $\mathsf{S}_V^{\widetilde{r}}$.

The idea can be further exploited to reduce errors in our estimate of not just $\overline{r}$ but other moments of the team-level distribution of complaints μ. However, balance must be ensured at the team level. Specifically, for any team a, we ensure that parameters $\eta_{i,a}$ are exchangeable, and such that $\sum_{i \in I} \eta_{i,a} = g$, where g is a target number of garbled potential null reports. In our application $g = 2$ for random samples of mean size 9 from teams. We refer to this garbling scheme as *blocked garbling*.

We show in Appendix B that Proposition 2 extends to the case of blocked garbling under some conditions. In addition, we provide simulations showing significant standard error reduction for our sample size.

A small-dimensional data-generating process (DGP). In principle, Proposition 2 allows us to recover any arbitrary distribution μ of team-level report profiles. However, in practice, for a mean of 9 workers surveyed in the HG condition per team, μ exhibits 9 degrees of freedom, and we must make inferences from a relatively small number of teams (112). For this reason, it is worth it to consider estimating more constrained sets of distributions μ. One possibility is to set $\mu(k) = 0$ for $k > 4$, thereby reducing the search space to 4 degrees of freedom.

An alternative, which perhaps provides a better intuitive understanding of the underlying phenomena, is to estimate through maximum likelihood a parametric class of data generating processes. Specifically, we consider the following class of environments: A manager $a \in M$ can be one of three types $\theta \in \{L, M, H\}$, with respective probabilities q_L, q_M and q_H. Conditional on a type θ, the manager harasses each worker i under their span of control independently with fixed probability ρ_θ where we assume that $\rho_L = 0$ and $\rho_M \leq \rho_H$. The data generating process is entirely specified by the 4 dimensional vector $\gamma = (q_M, q_H, \rho_M, \rho_H)$.

Given γ, the distribution of observable data $\widetilde{\mathbf{r}}$ depends on the garbling scheme used (no garbling, i.i.d. garbling, blocked garbling). Appendix B derives the associated likelihood

functions.

3.4 Discussion

Capturing other interventions. While our theoretical framework focuses on the impact of garbling on reporting, it also provides structure to other aspects of our experiment, specifically varying the elicitation of team identifiers (needed to compute team statistics), and improving rapport building with enumerators. Intuitively, we believe that both can be captured through the worker's subjective probability of a leak p.

Removing team information intuitively reduces the likelihood that a leaked report may be linked to a specific worker, thereby reducing the worker's perceived expected reputational or retaliation cost. Similarly, building rapport may increase the caller's trust that survey enumerators are trust-worthy, and in particular, unlikely to leak any information. If rapport affects workers through trust, there may be complementarities between rapport and HG: HG is only effective if workers trust that it is implemented as described; increasing trust may therefore increase the impact of the HG treatment.

HG compared to RR. RR is one of a limited number of survey methods considered to be well-suited to elicit sensitive information (Blair et al., 2015, Rosenfeld et al., 2016). Why then did we opt for HG instead of RR? The key distinction between HG and RR is the nature of the garbling: In HG, it is hard, or exogenous. In other words, the surveyor casts the die. In RR, it is soft. In other words, the respondent rolls the die. This distinction conveys three types of benefits to HG relative to RR.

First, as we showed above, HG allows for blocked HG designs that deliver more precise estimates than i.i.d. garbling, which is the only option under RR. This is especially valuable when baseline reporting rates are low and sampling error can dwarf the statistic of interest. Second, implementing RR typically relies on the availability of a randomization aid such as a die, which is not required for HG. This consideration becomes relevant when surveys are being conducted remotely, and respondents may not be relied upon to have an aid on hand.

Further, the widespread use of computer-assisted surveys means that HG can be programmed into a survey's design, which reduces the amount of time spent garbling responses during the survey's implementation.[11]

Finally, because HG imposes exogenous garbling, it does not rely on the respondent's compliance with the injunction to garble their response. In an organizational setting, when actions may be taken on the basis of reports, Chassang and Zehnder (2019) clarify theoretically that RR is potentially problematic because of its reliance on respondents' compliance; when reputational or retaliation costs are high, respondents may maximize their payoffs under RR by strategically providing the least sensitive response. In contrast, HG can improve information transmission in equilibrium. This means that it is more durable in an organizational setting compared to RR.

4 Experiment Design

We collaborated with the apparel producer to conduct surveys with workers at 2 plants. Prior to the survey's launch, the factories' HR departments made an announcement on the PA system that workers may be invited to participate in a survey that it was running in collaboration with independent researchers. The BRAC Institute for Governance and Development (BIGD), a well-respected arm of BRAC University in Bangladesh, conducted all data collection. The research team prepared a pre-analysis plan (PAP) for the experiment's design and registered it on the AEA's RCT registry. We adhere to our PAP in the analysis.

The survey process entailed 3 phone calls conducted outside of working hours. The first phone call entailed introducing the survey, establishing a baseline level of trust, and recruiting the prospective respondent; the second entailed the main survey; and the third entailed a follow-up survey conducted two weeks later. During the first call, workers who consented to participate were requested to suggest a time for the main survey when they could find a private place where they felt comfortable talking about difficult workplace issues. We in-

[11] HG can also be implemented using paper-based survey formats.

formed participants that aggregated results would be shared with senior management and would inform HR policy. All survey enumerators for the study were women.[12]

Table 1: Harassment definitions

Type of harassment	Actions read aloud to respondent
Threatening behavior	Threatened you; Told you that they will harm you if you do not agree to or fulfill their demands.
Physical	Hit, slapped, or punched you; Cut or stabbed you; Tripped you; Otherwise intentionally caused you physical harm.
Sexual	Made remarks about you in a sexual manner; Asked you to enter into a love or sexual relationship; Asked or forced you to perform sexual favors; Asked or forced you to meet outside of the factory or meet them alone in a way that made you feel uncomfortable; Touched you in a sexual manner or in a way that made you feel uncomfortable or scared; Shown you pictures of sexual activities.

Notes: For each type of harassment, respondents were asked, "In the past year, has your line supervisor taken any of the following actions toward you against your will?"

4.1 Harassment Outcomes

The research team was interested in measuring workers' experience of three types of harassment: Threatening behavior, physical harassment, and sexual harassment. For each type of harassment, we asked workers, "In the past year, has your line supervisor taken any of the following actions toward you against your will?" We then listed, for each respective type of harassment, the actions in the second column of Table 1. Ex ante, we hypothesized that

[12]Budget constraints prohibited the research team from randomly assigning, after stratifying respondents by their sex, the sex of the survey enumerator. Based on its knowledge of the context and guidance from local survey staff, the research team expected that it would be more socially acceptable for enumerators who are women to survey respondents who are men.

threatening behavior would be the least sensitive to report and that sexual harassment would be the most sensitive to report.

4.2 Treatment Conditions

We randomly assigned survey participants to different combinations of treatment conditions. First, we varied whether the survey method garbled respondents' intended reports. Second, we varied the level of identifiability of a workers' team and manager. Third, we varied the extent to which the survey enumerator built rapport with the surveyed individual. As discussed in Section 3.4, the latter two conditions aim to reduce the worker's subjective probability of a leak p. The experimental conditions are:

Survey method for harassment-related questions:

1.a) Direct elicitation (DE): Directly ask the survey respondent about sensitive information.

1.b) Hard garbling (HG): For a yes or no question, where *yes* is the more sensitive answer, exogenously flip *no* answers to *yes* with $\pi = 0.2$.

DE is the status quo survey method and the control condition. HG is the treatment condition; in contrast to DE, HG provides respondents with plausible deniability if they submit a sensitive answer. This is because if a respondent's answer is recorded as *yes*, it is impossible to know from the record whether they actually answered yes or no. We can estimate the statistics of harassment using the garbled data and the consistent estimator derived in Section 3. To reduce noise in our estimates, we fix the flipping rate at 20% in groups of 10 HG surveys, which means that the flipping rate is typically fixed across strata of workers that are from the same team or are on teams located nearby to each other in the factory. Section 4.3 provides more information on sampling and on assignment to treatment. The paper's Supplementary Materials reports the script used to explain HG to respondents.

Personally-identifying information (PII):

2.a) Status quo approach: Ask survey respondents to answer questions that reveal relatively more PII; questions include production section or line number and direct supervisor.

2.b) Low PII approach: Limit the amount of PII requested from the survey respondent; no questions asked about production section or line number or direct supervisor.

Asking questions that reveal relatively more PII is the status quo approach because surveys in organizational settings often explicitly or de facto reveal respondents' identities, hence, this is the control condition.

Rapport-building (RB):

3.a) Status quo approach: Survey enumerators follow a typical social science research introduction script before beginning the survey and then ask the survey questions.

3.b) RB approach: Survey enumerators allocate survey time to build rapport, or trust, with the participant. RB entails chatting about family and hobbies in a natural but pre-specified manner, beyond the minimum small talk typical in the standard social science approach.[13] We developed our RB treatment modules by combining insights from practitioners and policy-makers conducting surveys on sensitive issues, such as sexual abuse and gender-based violence (e.g. United Nations Human Rights Office, 2011, United Nations Statistical Office, 2014, Muraglia et al., 2020) and from research focused on protocols for criminal investigations of sexual abuse allegations (e.g. Cowles, 1988, Vallano and Compo, 2011, Hershkowitz et al., 2014). For details on the development of our RB approach and our RB modules, see the paper's Supplementary Materials.

[13]During training, survey enumerators developed and practiced the RB approach using role plays. The senior research associate running this training module had to approve each survey enumerator on their RB approach before the survey was launched.

– RB1: In the baseline rapport-building section, the enumerator signals that they care about the worker, getting to know the respondent, using emotional mirroring and acknowledging them.

– RB2: In this extended rapport-building section, the enumerator becomes personable with the worker, who has the chance to ask them questions. The enumerator also shares a related experience.

The status quo approach (condition 3.a) is the control condition. RB is the treatment condition. We conduct a shorter and a longer version of RB to test for the possibility that the marginal returns of building rapport decrease quickly.

Table 2: Treatment Arms & Surveyed (Planned) sample sizes

		No Rapport	Rapport 1	Rapport 2	TOTAL
Direct elicitation	PI	Arm 1 412(476)	Arm 2a 190(225)	Arm 2b 188(229)	790(930)
	No PI	Arm 3 197(226)	Arm 4 189(220)		386(446)
Hard garbling	PI	Arm 5 397(487)	Arm 6a 178(225)	Arm 6b 189(227)	764(939)
	No PI		Arm 7 257(305)		257(305)
	Total	1006(1189)	814(975)	377 (456)	2197(2620)

Table 2 summarizes the combinations of the experimental treatment arms that we tested. Treatment arm 1 is the benchmark, as it represents the status quo survey approach. Ex ante, we identified treatment arm 7 as the most protective. This may not be the case, however, if RB, which entails asking the respondent for more information about themselves that is not recorded in the survey, erodes the benefit of not asking for respondents' PII. We shed light on this possibility by comparing Arms 3 and 4. The experimental conditions were introduced

after respondents completed all non-harassment related survey modules.[14] Appendix Figure A.1 displays the survey modules and the treatment interventions' locations in the survey.

4.3 Sampling & Assignment to Treatment Arms

Sampling. We conducted a stratified random selection of workers to participate in the survey. Using the entire list of employees in the two plants, we sampled workers from four types of production teams: Sewing production lines; finishing teams; dry washing teams; and wet washing teams. Among these teams, we chose teams with a sufficiently large number of workers (approximately above 15), because we aimed to stratify the treatment assignment by team and gender. We were left with 112 eligible teams and a total of 5,948 eligible workers out of a workforce of 7,727 workers (77% of workers).

We stratified workers on eligible teams by their sex, which we identified based on name (male, female, uncertain).[15] In some cases, there are teams with very small numbers of one sex; in these cases, we aggregated small groups of workers to the smallest level that yielded a sufficiently large group sizes (e.g., production section-floor). Next, we selected 9 workers per stratum, which aimed to ensure a minimum of one per stratum assigned to each treatment arm. We then sampled larger strata in proportion to their share of the overall eligible worker population.

We targeted a sample size of 2,620 workers. Because we had access to the complete population of workers at the 2 plants, we were able to replace workers who were unreachable or who declined to participate. We attempted to recruit a total of 3,581 workers by phone, and we achieved a final sample size of 2,197 workers (61% response rate). The main reason for non-response was that we were not able to reach workers by phone (85% of cases); of workers whom we reached, 92% agreed to participate.[16] We did not achieve our target sample

[14] This is with the exception of one survey module on COVID-prevention behavior that was included to enable measurement of possible surveyor demand effects of rapport building. See Section 5.3 for a discussion.

[15] Names in Bangladesh are highly gendered. As such, we were able to categorize names as male or female for 99.7% of eligible workers.

[16] Survey enumerators were allowed to call workers a total of 9 times to recruit them. We obtained workers' phone numbers from the apparel manufacturer's HR department, so it is possible that the phone numbers

size despite our ability to replace workers because we stratified our selection by team and gender, and for some strata, we ran out of candidate replacement workers.

Assignment to Treatment Arms. The unit of randomization is a worker, stratified by plant-production team and sex. As detailed under sampling above, in cases where there were too few men or women on a production team, we aggregated to the next highest level that yielded a sufficiently large stratum size. We implemented the randomization in Stata. We first randomly assigned one worker per stratum to each treatment arm because we wanted to ensure that all strata were represented in all treatment arms. For larger strata, we then randomly assigned workers to each treatment arm with probabilities of assignment that corresponded to the treatment arm's target share of the overall sample size. We used the *randtreat* package by Carril (2017) to address misfits across strata. We conducted 10 randomizations and selected the one that performed best in terms of balance on two covariates available to the research team (tenure and skill group) (Banerjee et al., 2020).[17]

4.4 Data & Sample Characteristics

Table 3 presents summary statistics of our sample. Appendix Table A.1 presents team-level summary statistics for the teams represented in the survey.

4.5 Internal Validity

Appendix Table A.2 shows balance tests for workers' characteristics across the main treatment conditions. Appendix Table A.3 presents balance tests for workers' characteristics separately across no rapport, short rapport, and long rapport treatment arms. Across 48 tests, one is statistically significant at 10% level, which is no more than we would expect by chance.

listed for some workers were outdated.

[17]We created two skill groups: Low-skill workers in helper positions and higher-skill workers.

Table 3: Summary Statistics

	Mean	SD	Min	p25	p50	p75	Max
Female	0.81	0.39	0	1	1	1	1
Currently Working	0.96	0.20	0	1	1	1	1
Age	26.8	5.14	17	23	26	30	55
Experience (yrs)	5.20	3.57	0	2.83	4.42	7.17	28.8
Tenure (yrs)	2.91	2.45	0.0027	0.61	2.87	4.21	16.9
Tenure in Team (yrs)† [N=1554]	2.58	2.52	0	0.50	1.92	3.92	14.5
Years of Education	6.71	3.38	0	5	6.50	9	16
Marital Status (1=Yes)	0.82	0.38	0	1	1	1	1
Children (1=Yes)	0.74	0.44	0	0	1	1	1
Sewing Section	0.49	0.50	0	0	0	1	1
Finishing Section	0.34	0.47	0	0	0	1	1
Washing Section	0.17	0.38	0	0	0	0	1
Position: Helper	0.18	0.39	0	0	0	0	1
Position: Ironing/Folding	0.086	0.28	0	0	0	0	1
Position: Operator	0.59	0.49	0	0	1	1	1
Position: Packer	0.046	0.21	0	0	0	0	1
Position: Quality	0.096	0.29	0	0	0	0	1

Notes: This table reports summary statistics on workers' characteristics. Unless otherwise noted, the sample includes 2,197 workers who participated in our survey. †This variable is available for the 1554 respondents who were assigned to status quo PII collection treatment arms, in which we collected respondents' team id, manager id, and tenure on their team.

5 The Impact of Survey Design

In this section, we report the results of the survey experiment. First, we present the results for the main treatment conditions. Next, we assess HTEs by gender. We then examine whether HG, RB, and removing team-level identifying information are substitutes versus complements for each other. Finally, we conduct robustness checks for our results.

5.1 Specifications

We aim to estimate the following regression specification:

$$r_{is} = \alpha\, HG_i + \gamma\, NoPII_i + \beta\, Rapport_i + \mu_s + \theta X_i + \epsilon_{is} \tag{5}$$

where r_{is} is the intended reporting outcome of interest for individual i in stratum s. HG_i, $Rapport_i$ and $NoPII_i$ are hard-garbling, rapport, and not asking for personally identifying information, respectively. μ_s are stratum fixed-effects. We present results without and with controls for individuals' characteristics X_i, which are selected using the post double selection lasso (Belloni et al., 2014).

Identification of intended responses. We do not observe r_i for individuals in the HG arms, only their $\widetilde{r}_i$. Consequently, following Blair et al. (2015), we transform each respondent's $\widetilde{r}_i$ by applying eqn. 3 from Section 3, which relates recorded and intended reports:

$$\widetilde{r}_i = r_i + (1 - r_i)(\pi + \varepsilon_i)$$

where we've rewritten η_i from eqn. 3 as π plus an error term ε_i that equals $(1 - \pi)$ with probability π and equals $-\pi$ with probability $(1 - \pi)$.

This equation can be expressed as

$$\frac{\widetilde{r}_i - \pi}{1 - \pi} = \underbrace{r_i + \frac{1 - r_i}{1 - \pi}\varepsilon_i}_{\hat{r}_i}.$$

We transform reporting outcomes by applying the equation on the lefthandside of this equality with $\pi = 0.2$ for the HG group and $\pi = 0$ for the DE group. $\hat{r}_i$ is the transformed outcome that we use in our analysis. The second term on the righthandside of the equation indicates that intended responses are measured with a heteroskedastic error term. We assume that r_i satisfies eqn. 5. This implies that $\hat{r}_i$ satisfies an equation (5b) that has heteroskedastic errors with an expected value of 0 conditional on covariates. Consequently, OLS is consistent, and we get correct standard errors (SEs) for parameter estimates by using robust SEs. We estimate the following equation, in which ξ_{is} is now the residual, and report robust SEs:

$$\hat{r}_{is} = \alpha\, HG_i + \gamma\, NoPII_i + \beta\, Rapport_i + \mu_s + \theta X_i + \xi_{is} \tag{5b}$$

HTE analysis by respondents' sex. We also estimate treatment effects separately for women and for men:

$$\hat{r}_{is} = \alpha_f \, HG_i * Female_i + \alpha_m \, HG_i * Male_i + \beta_f \, Rapport_i * Female_i + \beta_m \, Rapport_i * Male_i + \gamma_f \, NoPII_i * Female_i + \gamma_m \, NoPII_i * Male_i + \lambda \, Female_i + \mu_s + \theta X_i + \xi_{is} \quad (6)$$

Testing for effects by treatment arm. Finally, we test for complementarity and/or substitutability among treatments by estimating the treatment effects separately for each treatment arm. In this regression, the omitted category is $\mathbb{1}(\text{DE} \times \text{PII} \times \text{No RB})_i = 1$, which is treatment arm 1, the control condition.

$$\begin{aligned}\hat{r}_{is} = {} & \alpha_1 \, \mathbb{1}(\text{DE} \times \text{PII} \times \text{RB 1})_i + \alpha_2 \, \mathbb{1}(\text{DE} \times \text{PII} \times \text{RB 2})_i + \alpha_3 \, \mathbb{1}(\text{DE} \times \text{No PII} \times \text{RB 1})_i \\ & + \alpha_4 \, \mathbb{1}(\text{DE} \times \text{No PII} \times \text{No RB})_i + \beta_1 \, \mathbb{1}(\text{HG} \times \text{PII} \times \text{No RB})_i \\ & + \beta_2 \, \mathbb{1}(\text{HG} \times \text{PII} \times \text{RB 1})_i + \beta_3 \, \mathbb{1}(\text{HG} \times \text{PII} \times \text{RB 2})_i + \beta_4 \, \mathbb{1}(\text{HG} \times \text{No PII} \times \text{RB 1})_i \\ & + \mu_s + \theta X_i + \xi_{is} \end{aligned} \quad (7)$$

5.2 Results

Main effects of survey design on reporting. Table 4 reports the main results.[18] Before examining the treatment effects, it is important to discuss the reporting rates in the baseline arm, which is (DE × PII × No RB). 9.47% of workers in this arm report experiencing threatening behavior, 1.46% report being physically harassed, and 1.70% report being sexually harassed by their supervisor. Among all workers who report being harassed under DE, meaning respondents in arms 1-5, 42% who experienced threatening behavior reported it through one of their factory's internal channels, 55% who were physically harassed reported it, and 67% of those who were sexually harassed did. This means that from

[18] Appendix Table A.4 reports the main results with separate indicator variables for short- and long-RB conditions.

the producer's perspective, it would have detected that 4%, 0.81%, and 1.4% of workers, respectively, experienced threatening behavior, physical harassment, and sexual harassment by their supervisor in the past year.

Turning to the effects of survey design, in Table 4 and other regression tables, odd-numbered columns display the results from the baseline specification, while even-numbered columns display the results with PDS lasso-selected controls. Columns (1)-(2) show that HG increases reporting of threatening behavior by 5.3-5.4 percentage points (ppts), an increase of 56% (p <0.01). Removing questions about respondents' supervisor and team increases reporting by 3.6-3.7 ppts, an increase of 38% ($p < 0.10$). Finally, building rapport also appears to have a positive effect, a 2.7-2.8 ppts or 29% increase, but it is not statistically significant ($p = 0.129$).

Table 4: Effects of Survey Design on Reporting of Harassment

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
HG Treatment	0.0535***	0.0532***	0.0417**	0.0430**	0.0501***	0.0520***
	(0.0203)	(0.0196)	(0.0179)	(0.0173)	(0.0184)	(0.0178)
No PII Treatment	0.0358*	0.0370*	0.0405**	0.0408**	0.0100	0.0118
	(0.0210)	(0.0203)	(0.0177)	(0.0171)	(0.0181)	(0.0176)
Rapport Treatment	0.0277	0.0271	-0.0094	-0.0096	0.0123	0.0111
	(0.0188)	(0.0183)	(0.0161)	(0.0156)	(0.0168)	(0.0162)
Control Group Mean	.0947	.0947	.0146	.0146	.017	.017
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Yes	No	Yes	No	Yes
Observations	2194	2194	2194	2194	2194	2194

Notes: This table reports OLS estimates of treatment effects on workers' reporting. Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the treatment indicator and stratification variables. Even-numbered columns also include controls selected using the PDS lasso. Robust standard errors are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Columns (3)-(4) present the results for reporting of physical harassment. HG increases

reporting by 4.2-4.3 ppts, an increase of 295% (p <0.01). Removing questions about respondents' supervisor and team also increases reporting by 4.08 ppts, an increase of 279% (p <0.05). Finally, building rapport has no effect. The effects of HG on reporting of sexual harassment are similarly large, more than a 300% increase in reporting. In contrast to threatening behavior and physical harassment, however, there is no main effect of removing PII. There is a positive but imprecisely estimated effect of RB on sexual harassment, a 1.1 ppt or 65% increase in reporting (not statistically significant).

Together, these results paint a clear, compelling case for the importance of plausible deniability in the design of transmission systems for sensitive information in organizational settings. They also elucidate that from this organization's perspective, the harassment reported through formal channels is only a fraction of the harassment experienced by workers. If we assume that workers who reported harassment under HG would report through the producer's channels at the same rate as workers who reported under DE, then our estimates suggest that the producer would have detected 27%, 14%, 20% of cases of threatening behavior, physical harassment, and sexual harassment, respectively.[19] Finally, as expected based on our framework, the relative effects of plausible deniability are larger for physical and sexual harassment, for which the retaliation and/or reputation costs are plausibly larger.

Removing team-level identifying questions also has large effects for physical harassment, and to a lesser extent, threatening behavior. But the results are not consistent across all outcomes: It has no effect on reporting of sexual harassment. It also comes at the cost of not being able to calculate manager-level statistics that may be valuable to decision-makers. Finally, based on RB's main effects, we cannot reject that it has no effect on reporting. It is possible, though, that the imprecision of our estimates for RB indicate heterogeneity in its effects across respondents or that RB's effects depend on other survey design features. Next, we explore the possibility of HTEs by respondents' sex. We then test for effects separately

[19]These are likely upper bounds, as many workers who report under HG would be unwilling to report under DE. This suggests that they would likely also less willing to report through the producer's internal systems. We only asked DE respondents who responded "yes" to a harassment question about reporting through the producer's reporting channels. We did not ask HG respondents because doing so would reveal information about their true harassment status in the record.

by treatment arms to further examine how the effects of removing team-level identifying information and RB may depend on their combination with other features of survey design.

Effects of survey design on reporting by men and women. Motivated by the possibility that the experience of harassment and the utility generated by reporting harassment may be different for men and women, we estimate the main effects separately by sex in Table 5. Again, it is important to begin by examining baseline reporting rates. 18.06% of men report experiencing threatening behavior, 4.17% report experiencing physical harassment, and 1.39% report experiencing sexual harassment. Reporting rates among women are very different: 7.65% report threatening behavior, 0.88% report physical harassment, and 1.76% report sexual harassment. We cannot disentangle whether these differences are due to differential incidences of harassment or differential reporting. Among respondents who report being harassed under DE, across all forms of harassment, women are more likely to say that they reported their experience through an internal channel.

Beginning with threatening behavior, columns (1)-(2) show that women are almost 50% more likely to report threatening behavior under HG compared to DE (p <0.10). The effect is even larger for men, despite a higher baseline reporting rate, a 67% increase (p <0.05). The difference in these effects is marginally statistically significant ($p = 0.112$). For removing team-level identifying information, the magnitude of the effect in ppts is similar, and there is no statistical difference between the effects for both groups, but the relative effect for women is larger. Finally, an interesting pattern emerges for RB: For women, the effect is positive and statistically significant, while for men, the estimate is negative, although not statistically significant. We are unable to reject that the effects are the same ($p = 0.189$). We discuss this pattern of opposite effect signs for RB below.

Turning to physical harassment, columns (3)-(4) show that HG increases women's reporting by 3.81 ppts, an increase of 433% ($p < 0.05$). For men, the increase is 6.45 ppts or 155%. While large, the effect on men is imprecisely estimated and is not statistically significant. We also cannot reject that the effects are the same for both sexes. For removing team-level

identifying information, again, the magnitude of the effect in terms of a ppt increase is similar for men and women, although only statistically significant for women. Finally, RB has no effect on reporting by women or men.

Table 5: Effects of Survey Design on Reporting of Harassment,Heterogeneity by Sex

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
HG Treatment × Female	0.0382*	0.0376*	0.0361*	0.0381**	0.0415**	0.0442**
	(0.0223)	(0.0215)	(0.0197)	(0.0190)	(0.0204)	(0.0198)
HG Treatment × Male	0.1226**	0.1212**	0.0673	0.0645	0.0859**	0.0861**
	(0.0493)	(0.0478)	(0.0438)	(0.0423)	(0.0430)	(0.0419)
Rapport × Female	0.0402*	0.0393*	-0.0125	-0.0125	0.0256	0.0248
	(0.0209)	(0.0202)	(0.0178)	(0.0171)	(0.0187)	(0.0179)
Rapport × Male	-0.0240	-0.0271	0.0049	0.0025	-0.0513	-0.0502
	(0.0442)	(0.0426)	(0.0393)	(0.0380)	(0.0390)	(0.0380)
No PII Treatment × Female	0.0350	0.0363	0.0428**	0.0419**	0.0143	0.0147
	(0.0233)	(0.0224)	(0.0195)	(0.0187)	(0.0204)	(0.0199)
No PII Treatment × Male	0.0318	0.0358	0.0309	0.0380	-0.0107	-0.0058
	(0.0496)	(0.0482)	(0.0424)	(0.0414)	(0.0401)	(0.0389)
Female	-0.0813	-0.0872	-0.0263	-0.0134	0.0440	0.0578
	(0.0836)	(0.0806)	(0.0661)	(0.0642)	(0.0648)	(0.0629)
Control Mean - Female	.0765	.0765	.0088	.0088	.0176	.0176
Control Mean - Male	.1806	.1806	.0417	.0417	.0139	.0139
p(HGxFemale - HGxMale)	[0.120]	[0.112]	[0.516]	[0.570]	[0.352]	[0.368]
p(RapportxFemale - RapportxMale)	[0.189]	[0.159]	[0.687]	[0.718]	[0.076]	[0.073]
p(NoPIIxFemale - NoPIIxMale)	[0.954]	[0.993]	[0.799]	[0.932]	[0.579]	[0.642]
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Lasso	No	Lasso	No	Lasso
Observations	2194	2194	2194	2194	2194	2194

Notes: This table reports OLS estimates of treatment effects by gender heterogeneity on workers' reporting. Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the gender interactions of treatment variables and stratification variables. Even-numbered columns also include controls selected using the PDS lasso. Robust standard errors are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Finally, for sexual harassment, columns (5)-(6) show a surprising pattern of effects for HG. The effect for women is a 4.46 ppt or 253% increase in reporting. The effect for men

is more than twice as large, a 8.60 ppt or 619% increase in reporting. While qualitatively these results appear to be different, due to our smaller sample size for men, we are unable to reject that the effects for men and women are the same ($p = 0.373$). For removing team-level identifying information, the effect is positive but not statistically significant for women. There is no effect for men. Finally, for RB, the effect for women is positive and large, a 2.48 ppt or 141% increase, but not statistically significant. For men, it is negative and large in magnitude but is not statistically significant. While we cannot reject that the effects of RB for men and women are different from zero, we can reject that they are equal, indicating more positive effects on women compared to men ($p = 0.075$).

The fact that the estimated effects of RB on men are either negative or zero suggests that RB may actually have backfired with men. In our study's context, this may be related to the fact that the survey enumerators were all women,[20] and for Bangladeshi men, small talk with a woman who is not a family member may not increase their feelings of trust toward the surveyor and may even backfire. It may also be related to the fact that the RB script included questions about respondents' children that may have been more engaging for women. Whether RB approaches explicitly designed with respondents who are men in mind are more effective is a question for future research.

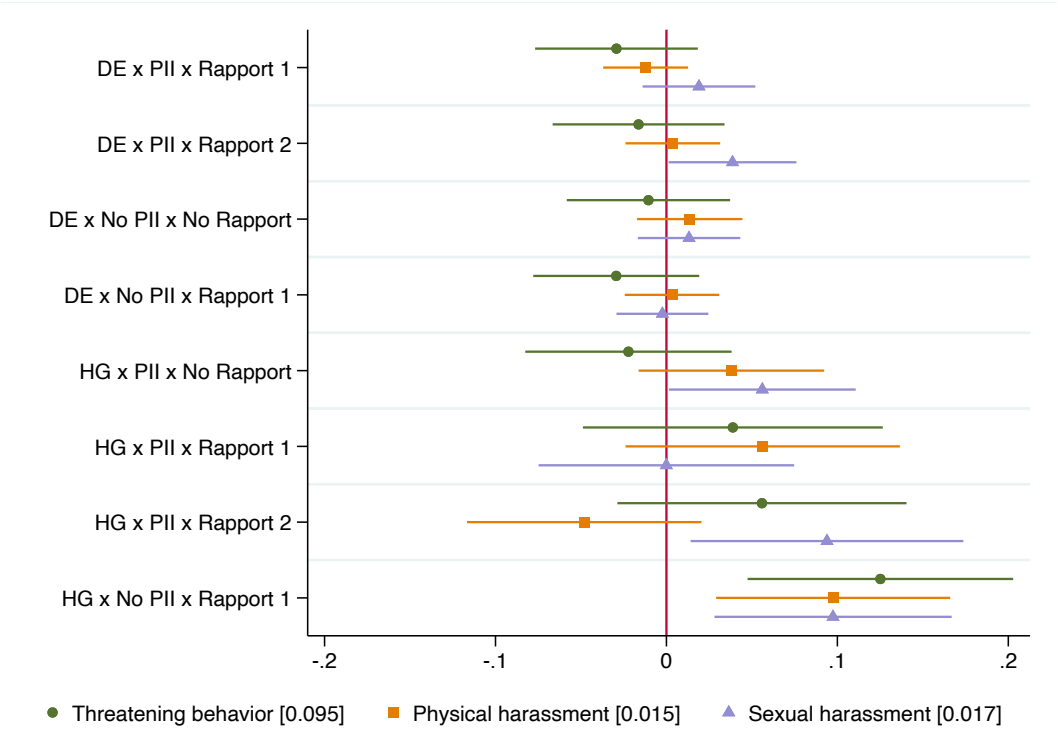
Interactions among treatment conditions. Finally, we examine the possibility that the treatment conditions may substitute or complement each other when implemented together. We test this possibility using regression equation 7. Figure 1 summarizes the results, which are presented in Appendix Table A.5. Recall that in this regression, the omitted category is treatment arm 1, $\mathbb{1}(\text{DE} \times \text{PII} \times \text{No RB})_i = 1$, which is the control condition.

Beginning with the effect of (DE × PII × RB 1), Figure 1 shows that adding the short RB treatment when using DE and eliciting PII does not significantly affect reporting of any type of harassment. The estimated effect for threatening behavior is negative, while

[20] This design choice was made as a result of budget constraints, as it was prohibitively costly to increase the sample size enough to randomize the match between respondents and enumerators of different sexes.

that for sexual harassment is positive. If we increase the amount of RB, which is arm (DE × PII × RB 2), then the effect on reporting of sexual harassment becomes statistically significant. In both cases, as shown in Appendix Figure A.2, the negative effect of RB on reporting of threatening behavior is driven by men, while its positive effect on reporting of sexual harassment is driven by women. In sum, there is suggestive evidence that RB may increase reporting by women in contexts with surveys that use DE and elicit more PII.

Figure 1: Treatment effects by survey arm



Notes: This figure reports coefficients from separate regressions of the outcome variable on the treatment arm indicators, strata fixed effects, and controls selected using the PDS lasso. The regression specification is eqn. 7. The whiskers are 95% confidence intervals estimated using robust standard errors. The omitted category is treatment arm 1, $\mathbb{1}(\text{DE} \times \text{PII} \times \text{No RB})_i = 1$, which is the control condition. The number in square brackets is the reporting rate for this group.

Turning to (DE × No PII × No RB), removing team-level identifying information, on its own, does not significantly increase reporting. The estimated coefficients for physical harassment and sexual harassment are fairly large and positive, but they are not statistically

significant. Combining removal of team-level identifying information and a short amount of RB, in (DE × No PII × RB 1), also does not significantly impact reporting of any outcome. In short, when directly eliciting survey respondents' experience with harassment, the longer RB appears to have a positive effect on reporting of sexual harassment, but the effects of short RB and of removing team-level identifying information are too small to significantly increase reporting.

Next, examining the effects when using HG, beginning with (HG × PII × No RB), simply changing the survey method from DE to HG significantly increases reporting of sexual harassment and appears to increase reporting of physical harassment. The effects on both outcomes are large: A 334% increase for sexual harassment and a 259% increase for physical harassment. Adding RB in (HG × PII × RB 1) does not have conclusive effects: The point estimates for threatening behavior and physical harassment become more positive, while that for sexual harassment actually gets closer to zero. With the longer RB treatment, (HG × PII × RB 2), the effects for threatening behavior and for sexual harassment get more positive while that for physical harassment flips to negative. These noisier results reflect the smaller cell sizes for these treatment arms, as shown in Table 2.

Finally, turning to the ex ante most protective arm, (HG × No PII × RB 1), the estimated effects are large and positive across all three types of harassment. With PDS lasso controls, we reject that the effects of this arm are different from the effects of each of the DE arms, respectively, with $p < 0.05$ except for the effect of (DE × PII × RB 2), which has $p = 0.124$. Comparing this arm to the other HG arms, we cannot reject the equality of effects for certain outcomes and arms, but with the exception of the effect of (HG × PII × RB 2) on sexual harassment, the effects of the ex ante most protective arm are qualitatively larger than those of other treatment arms.

Motivated by the fact that the point estimate for the effect of (HG × No PII × RB 1) is larger than the sum of the point estimates for (DE × PII × RB 1)+(DE × No PII × No RB)+ (HG × PII × No RB) for all three outcomes, we test the null hypothesis of no complementarity among HG, removing team-level identifying information, and RB in complementarity

test 1 at the bottom of Appendix Table A.6. Focusing on even-numbered columns, which include PDS-lasso-selected controls, this test is rejected for threatening behavior (p=0.001), marginally rejected for physical harassment (p=0.120), and fails to reject for sexual harassment (p=0.428). We interpret this as suggestive evidence of complementarity among the design features.

We can also test for complementarity or substitutability in the effects of HG and RB in the case that PII is collected. Again, the point estimate for the combined effect is larger for 2 of 3 outcomes, so we report the test for complementarity: (HG × PII × RB 1) ≤ (DE × PII × RB 1) + (HG × PII × No RB). We do not find strong evidence in favor of complementarity in this case; we reject this test with p=0.052 for threatening behavior but do not reject it for physical harassment (complementarity test 2). For sexual harassment, we find evidence of substitutability between the two arms, as the p-value for this test is simply $1 - p_{\text{complementarity}} = 1 - 0.939 = 0.061$. We interpret the evidence from these tests as inconclusive.

5.3 Robustness Checks

Confusion in the HG condition. One concern with HG is that it is a more complicated mechanism compared to DE. This means that it takes more time to explain HG, which increases survey duration by 5% on average (Appendix Table A.7). More importantly, it also means that respondents may be confused by HG and that confusion may be more likely under HG compared to DE. This concern is especially relevant in our context, in which the average survey respondent has 6.71 years of schooling (Table 3), or a little less than a seventh grade education. We were concerned about this possibility, so we included two comprehension questions in the HG module.[21] Respondents answered these prior to being asked the questions about harassment, and survey enumerators explained the answers to the

[21]The questions were, "Can you please tell me whether the following statements are true or false: (a) If I respond 'Yes,' no one can ever know this for sure. (b) The system will record at least one out of every five workers' responses as 'Yes.' " The script explaining HG, including the comprehension questions, is included in the paper's Supplementary Materials.

comprehension questions after asking them.

We find that 8.8% of HG respondents answer at least 1 comprehension question incorrectly, while 4.8% answer 2 incorrectly. Women and men answer incorrectly at somewhat similar rates: 9.6% of men and 8.5% of women in HG answer at least 1 question incorrectly ($p = 0.637$), while 7.0% of men and 4.3% of women answer 2 questions incorrectly ($p = 0.131$). We also test robustness to confusion separately by gender.

While the surveyor would desire for respondents who are confused by HG to respond by answering "no" to avoid false positives, in practice, reporting rates are weakly higher among confused respondents. Consequently, we must evaluate whether asymmetric confusion among respondents in the HG versus the DE arm could explain our HG results. We adopt a very conservative approach to this test, which is to re-estimate our main results, considering all respondents who answer at least 1 comprehension question incorrectly as confused and setting all confused respondents' answers to harassment questions equal to "no."

Panel A of Appendix Table A.8 reports the main results; focusing on the HG effects and comparing them to the estimates in Table 4, for threatening behavior, column (2) shows that the effect is now a 4.2 ppt increase in reporting ($p < 0.05$) compared to a 5.3 ppt increase ($p < 0.01$). For physical harassment, the effect is a 3.2 ppt increase ($p < 0.10$) compared to a 4.3 ppt increase ($p < 0.05$). For sexual harassment, the effect is a 3.9 ppt increase ($p < 0.05$) compared to a 5.2 ppt increase ($p < 0.01$). Evidently, even under the extreme assumption that all confused respondents intended to report "no" to all questions, the effects of HG are positive, large, and statistically significant. Turning to Panel B, while the point estimates for both sexes are attenuated by similar magnitudes as in Panel A, there is not a differential pattern of attenuation by sex, and the patterns of heterogeneity are unchanged.

Strategic misreporting by workers. In our conceptual framework, we assume that there are no false positives in reporting; workers either report their true harassment status or they report that they have not been harassed. As discussed in Section 3, we think that

this is an appropriate assumption for our setting, at least in the short-run.[22] One may still be concerned, though, that this is a strong assumption. For example, workers who are motivated by career concerns may take advantage of the plausible deniability provided by garbling to try to take down innocent supervisors. This may especially be true for men, who are much more likely to be promoted into supervisor positions. If so, it provides an alternative explanation for the patterns of HTEs that we find.

We think that this possibility is a priori unlikely. In addition, we can provide empirical evidence consistent with this view. To do so, we split our sample by sex and by whether the respondent has at least 8 years of schooling, an informal cutoff used by factories to determine workers' eligibility to become a supervisor.[23] If workers are strategically misreporting, we expect that our effects will be be driven by workers with at least 8 years of schooling, who are differentially more eligible to become supervisors, especially among men. Appendix Table A.9 presents the results. It shows that there is no consistent pattern of HTEs for men or women with more or less than 8 years of schooling. Sometimes the effects are larger for the group with less schooling, sometimes smaller, and sometimes the same. This evidence goes against the possibility that strategic misreporting due to career concerns is driving our results.

6 Understanding Harassment

In this section, we use our improved survey data to assess the scope and nature of the harassment problem in the apparel producer's organization. Given the large effects of HG on reporting, we use the garbled data to characterize harassment. We pool across all HG treatment arms, including the RB arms and the arms in which we do not collect team-level identifying information. We begin by describing the patterns of harassment in the

[22] See Appendix B.4 for a discussion of the possibility of subversion in the longer run.

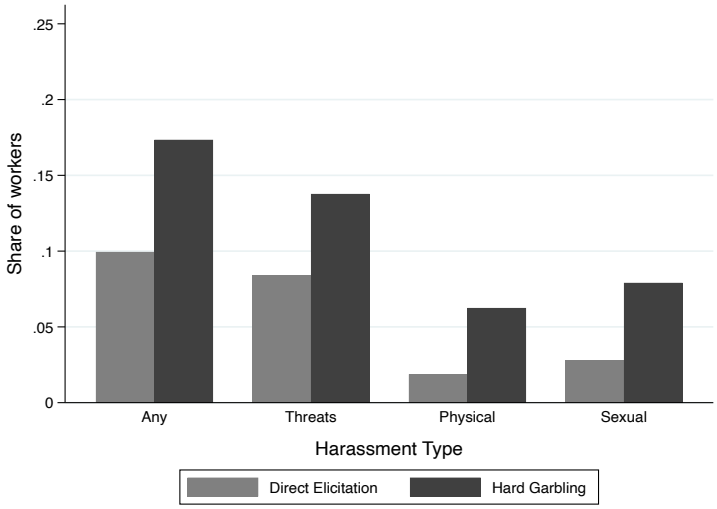
[23] In a survey conducted with supervisors and other lower-level managers employed by the apparel producer, 87% of supervisors had at least 8 years of schooling. 22% had exactly 8 years of schooling, a large jump up from the 8% of managers reporting having the next lower category, "some middle school education."

organization, and then we discuss the policy implications for the producer.[24]

6.1 Patterns of harassment

We use the formal non-parametric inference result from the extension of Proposition 2 for the case of blocked garbling (see Appendix B) to estimate the statistics. As discussed in Section 3, on average, there are 9 workers per production team in the HG arm. Consequently, we apply our result for teams of size 9 to identify the statistics.

Figure 2: Share of workers who have been victimized (S_V)



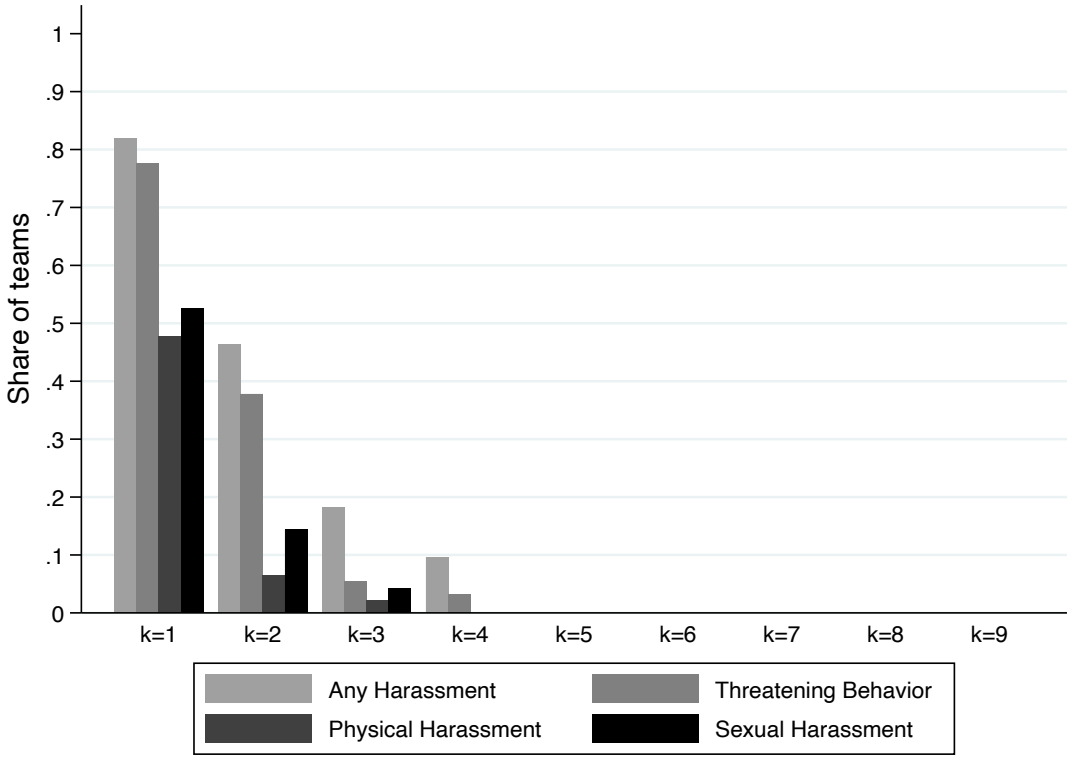
Notes: This figure reports harassment rates estimated using reporting with DE and HG, respectively. For both DE and HG, we pool across all treatment arms, including the RB arms and the arms in which we do not collect team-level identifying information. "Any" harassment indicates that a respondent reported at least one of the three types of harassment.

Figure 2 displays S_V, the share of workers who have been victimized by their supervisors. The Figure compares the rate estimated using HG to that using DE (pooling across all arms).

[24]One may be concerned that supervisors who engage in more harassment may have pressured workers not to participate the survey. In the context of the descriptive analysis, this would mean that our statistics of harassment are downward biased. We examine this possibility in Appendix Table A.10, which reports the correlation between the team-level reporting rates for harassment with DE and HG, respectively, and the team-level response rate to the survey. In all cases, the correlations are weak or are zero. This indicates that our statistics are not downward biased due to selective non-response by team's harassment levels.

It shows that using HG, 17.3% of workers report experiencing any form of harassment by their supervisor in the past year, while using DE, the reporting rate for any harassment is 9.9%. By type of harassment, 13.8% reported threatening behavior with HG compared to 8.4% with DE, 6.3% reported physical harassment with HG compared to 1.9% with DE, and 7.9% reported sexual harassment with HG compared to 2.8% with DE. Evidently, reporting rates using HG are uniformly higher than those using DE. They are also higher than the rates that would have been detected by the apparel producer through its reporting channels: 4.8% for any form of harassment, 0.81% for physical harassment, and 1.4% for sexual harassment.

Figure 3: Share of teams of size 9 with at least k workers who are victims ($\mathbf{S}_{TV \geq k}$)



Next, we turn to the question of how many managers are responsible for the harassment. We conduct our analysis at the production-team-level.[25] Figure 3 presents the results. Begin-

[25]Production teams typically had 2 supervisors overseeing distinct subgroups of workers. In the analysis,

ning with S_{PM}, which is the share of teams with at least one worker who has been victimized, the figure shows that over 80% of teams have at least one worker who has experienced any form of harassment, nearly 80% have at least one who has been threatened, just over 50% have at least one who has been sexually harassed, and just under 50% have at least one who has been physically harassed.

Examining $\mathsf{S}_{TV\geq k}$ for $k \in \{1, \cdots, 9\}$ in Figure 3, the figure shows that nearly 50% of teams of size 9 have at least 2 workers who have experienced any form of harassment, nearly 20% have at least 3, and nearly 10% have at least 4. Nearly 40% of teams of size 9 have at least 2 who have been threatened, 5% have at least 3, and 3% have at least 4. For physical harassment, 6% have at least two workers who have been victimized and 2% have at least three. For sexual harassment, 14% of teams have at least 2 workers who have been victimized and 4% have at least 3. These distributions illustrate that supervisors vary in how problematic their behavior is: Some supervisors harass many more workers than others.

The distributions also elucidate that the extent to which victims of harassment are isolated across teams varies with the type of harassment. Table 6 reports $\mathsf{E}_{2V|1V}$, the likelihood that a supervisor has harassed at least 2 workers conditional on harassing at least 1. This likelihood is the highest for threatening behavior (48.7%), then for sexual harassment (27.4%), and finally physical harassment (13.5%).

Table 6: Likelihood that a supervisor has at least 2 victims given that they have at least 1 ($\mathsf{E}_{2V|1V}$)

Statistic	Any Harassment	Threatening Behavior	Physical Harassment	Sexual Harassment
$S_{tv\geq 1}$	0.819	0.776	0.477	0.525
$S_{tv\geq 2}$	0.464	0.378	0.064	0.144
$E_{2V\|1V}$	0.566	0.487	0.135	0.274

Finally, as discussed in Section 3, a more intuitive way of describing supervisors' en-

we treat teams as if they had 1 supervisor because we can only link workers to their supervisors if they are in an arm in which we collected team-level identifying information. We lack this information for 643 respondents (29% of the sample).

gagement in harassment is to categorize them into three types $\theta \in \{L, M, H\}$, with the low harassment type not harassing any workers on their team, and then to estimate via ML the data-generating process specified by $\gamma = (q_M, q_H, \rho_M, \rho_H)$. Supervisors' types have respective probabilities q_θ, and conditional on their type, supervisors harass a given worker on their team with probability ρ_θ.

Table 7: ML estimates of supervisor types, shares, and harassment rates

Parameter	Any Harassment (1)	Threatening Behavior (2)	Physical Harassment (3)	Sexual Harassment (4)
ρ_L	0	0	0	0
	()	()	()	()
ρ_M	0.151	0.123	0.060	0.075
	(0.041)	(0.031)	(0.020)	(0.026)
ρ_H	0.263	0.223	0.168	0.180
	(0.125)	(0.144)	(0.176)	(0.154)
q_L	0.062	0.051	0.125	0.140
	(0.046)	(0.042)	(0.095)	(0.096)
q_M	0.528	0.547	0.547	0.519
	(0.304)	(0.319)	(0.293)	(0.283)
q_H	0.410	0.402	0.329	0.340
	(0.304)	(0.319)	(0.288)	(0.278)

Table 7 presents the results. There are several observations worth noting. First, the estimated share of supervisors who do not harass any workers (q_L) is small, between 5-14% depending on the type of harassment. Depending on the harassment type, between 52-55% of supervisors are estimated to be the intermediate type. These supervisors harass workers with 6% probability for physical harassment, 7.5% for sexual harassment, and 12% for threatening behavior. Finally, 41% of supervisors are estimated to be the high type when considering workers experiencing any harassment. The estimated shares for the different types of harassment range from 33% (physical harassment) to 40% (threatening behavior). High type supervisors harass workers with 17% probability for physical harassment, 18% probability for sexual harassment, and 22% probability for threatening behavior, respectively. Comparing

the the intermediate and high harassment types, for any harassment, high types (41% of supervisors) are estimated to be responsible for 57.5% of the harassment. For threatening behavior, high types (40% of supervisors) are responsible for 57% of the harassment. For physical harassment, high types (33%) are responsible for 63% of the harassment, and for sexual harassment, high types (34%) are responsible for 61% of the harassment.

6.2 Policy Implications for the Producer

These statistics shed light on the scope and nature of the producer's harassment problem, and they allow us to draw several policy implications for reducing harassment in this organization. First, the estimated reporting rates under HG are substantially higher than the reporting rates through the producer's internal channels. Our estimates suggest that the producer would have detected, at most, 35%, 13%, 18% of cases of threatening behavior, physical harassment, and sexual harassment, respectively. This suggests that lack of information about the scale of harassment taking place in the organization may have prevented decision-makers from acting.

The team-level statistics indicate that harassment by supervisors is widespread among teams, with greater than 80% of teams of size 9 having at least 1 worker who has experienced some type of harassment by their supervisor. This high rate of harassment across teams implies that firing all problematic managers is unlikely to be a viable policy option. The ML estimates from table 7 indicate that low-grade harassers, who comprise 53% of supervisors, are responsible for 42.5% of the harassment, while the worst harassers, who comprise 41% of supervisors, are responsible for 63% of it. Consequently, targeting the worst offenders, for example, by firing them, could address up to 63% of the harassment. For each of the worst offenders that the producer fired, our estimates imply that it would reduce harassment by about 2.4 workers.[26] Evidently, though, even if the producer fired all of the worst offenders, it would not eradicate the harassment problem. The producer would still need a way to

[26]They also imply that targeting firing at the low-grade harassers would reduce harassment by 1.4 workers per fired supervisor.

ameliorate the low-grade harassers' behavior, for example, through sensitivity training or increased monitoring.

The estimated distributions of harassment across teams also shed light on the implications of setting different burdens of proof for harassment. Our estimates of $\mathsf{E}_{2V|1V}$ with teams of size 9 suggest that a burden of proof that requires multiple victims of harassment to come forward, for example, to avoid "he said, she said" situations, would miss 86.5% of physical harassment cases and 73% of sexual harassment cases in which harassment is perpetrated against a single worker. Further, while not the focus of this paper, we collected data about witnessing harassment, and witnessed harassment at the team-level is not strongly correlated with reported harassment at the team-level (see Appendix Table A.11). Consequently, the producer cannot rely on witnesses to bolster a victim's story. Eradicating harassment requires having actions available that the producer can take in cases when only one victim comes forward.

The estimated distributions also shed light on the usefulness of policy options for reporting systems such as information escrows, which have been proposed to facilitate reporting in groups (Ayres and Unkovic, 2012). In the producer's case, if a reporting escrow were to operate at the team-by-harassment type level, we might expect it to be most successful at bringing threatening behavior to light, less successful for sexual harassment, and least successful for physical harassment. If it were to pool over issue types, say by operating at the team-level for all types of harassment, it would be most likely to trigger. In this case, though, it may provide victims with less protection, for example, if managers target retaliation based on the type of harassment reported.

7 Discussion

In this paper, we evaluate the impact of different aspects of survey design – HG, RB, and removing team level information – on respondents' propensity to report harassment in a survey experiment with employees of a large Bangladeshi apparel producer. We then use our

improved reporting data to assess several policy-relevant aspects of harassment.

Our experimental results and more informative statistics of harassment elucidate that lack of plausible deniability causes severe under-reporting of harassment in this organizational setting. They shed light on the nature of harassment, which is widespread among teams. They also shed light on the isolation of victimized workers; workers who experience physical and sexual harassment are typically isolated within teams. This is the first field evidence on HG outside of a lab and in a real-world organizational setting; our findings demonstrate that HG can dramatically increase reporting of harassment and reduce measurement of key statistics. Reducing PII also helps, at the cost of being able to calculate, in our case, manager-level statistics of interest. Our findings presents a promising direction for organizations to adopt HG in contexts in which credible threats or reputation costs limit information transmission via reporting systems. We hope that researchers will explore whether our findings replicate in other settings.

We focus on introducing HG with an aim to increase workers' willingness to report. As Chassang and Padró i Miquel (2018) note, however, the effectiveness of monitoring mechanisms often differs over the short-run versus the long-run. In the long-run, the relevant players figure out ways to game the system for their own benefit. Two are of specific concern in our setting: Managers may increase the magnitude of retaliation in response to the increased anonymity provided by HG; and workers may start strategically misreporting innocent managers out of idiosyncratic spite or to further their own careers. Chassang and Padró i Miquel (2018) offers some guidance on how to approach sustainable policy-making in such a context, which we discuss in Appendix B.4. An implication of their theory is that increasing the reporting rate is a key first step. In this regard, one can interpret the findings of our paper as establishing such a starting point. A next question raised by our findings and by the guidance from Chassang and Padró i Miquel (2018) is how to scale up enforcement actions taken as a function of reports. This strikes us as a particularly valuable direction for research.

Finally, our results hold promise for improving survey methodology for sensitive ques-

tions, even outside of organizational settings. Relative to the RR method, we show that HG allows for blocked HG designs that deliver more precise estimates than i.i.d. garbling under RR. This is especially valuable when baseline reporting rates are low and sampling error can dwarf the statistic of interest. Further, HG is easier to implement in the sense that it does not require randomization aids. Testing the HG approach with different populations and survey modalities would be valuable.

References

Aguilar, A., E. Gutiérrez, and P. S. Villagrán (2021): "Benefits and Unintended Consequences of Gender Segregation in Public Transportation: Evidence from Mexico City's Subway System," *Economic Development and Cultural Change*, 69, 1379–1410.

Ayres, I. and C. Unkovic (2012): "Information escrows," *Mich. L. Rev.*, 111, 145.

Bac, M. (2009): "An economic rationale for firing whistleblowers," *European Journal of Law and Economics*, 27, 233–256.

Banerjee, A. V., S. Chassang, S. Montero, and E. Snowberg (2020): "A Theory of Experimenters: Robustness, Randomization, and Balance," *American Economic Review*, 110, 1206–30.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects after Selection among High-Dimensional Contr," *Review of Economic Studies*, 81, 608–650.

Blair, G., K. Imai, and Y.-Y. Zhou (2015): "Design and analysis of the randomized response technique," *Journal of the American Statistical Association*, 110, 1304–1319.

Borker, G. (2018): "Safety First: Perceived Risk of Street Harassment and Educational Choices of Women," Tech. rep., mimeo.

Boudreau, L. (2022): "Multinational enforcement of labor law: Experimental evidence on strengthening occupational safety and health (OSH) committees," Tech. rep., mimeo.

Boudreau, L., R. Heath, and T. H. McCormick (2022): "Migrants, Experience, and Working Conditions in Bangladeshi Garment Factories," Tech. rep., mimeo.

Carril, A. (2017): "Dealing with misfits in random treatment assignment," *Stata Journal*, 17, 652–667.

Chakraborty, T., A. Mukherjee, S. R. Rachapalli, and S. Saha (2018): "Stigma of sexual violence and women's decision to work," *World Development*, 103, 226–238.

Chassang, S. and J. Ortner (2019): "Collusion in auctions with constrained bids: Theory and evidence from public procurement," *Journal of Political Economy*, 127, 2269–2300.

Chassang, S. and G. Padró i Miquel (2018): "Crime, Intimidation, and Whistleblowing: A Theory of Inference from Unverifiable Reports," *Review of Economic Studies*, 86, 2530–2553.

Chassang, S. and C. Zehnder (2019): "Secure Survey Design in Organizations: Theory and Experiments," .

Cheng, I.-H. and A. Hsiaw (2020): "Reporting Sexual Misconduct in the MeToo Era," Tech. rep., mimeo.

Chuang, E., P. Dupas, E. Huillery, and J. Seban (2020): "Sex, Lies, and Measurement: Do Indirect Response survey methods work?" .

Cowles, K. V. (1988): "Issues in qualitative research on sensitive topics," *Western Journal of Nursing Research*, 10, 163–179.

Dahl, G. B. and M. Knepper (2021): "Why is Workplace Sexual Harassment Underreported? The Value of Outside Options Amid the Threat of Retaliation," Tech. rep., mimeo.

Davison, A. C. and D. V. Hinkley (1997): *Bootstrap methods and their application*, 1, Cambridge university press.

Efron, B. (1987): "Better bootstrap confidence intervals," *Journal of the American statistical Association*, 82, 171–185.

Faure-Grimaud, A., J.-J. Laffont, and D. Martimort (2003): "Collusion, delegation and supervision with soft information," *The Review of Economic Studies*, 70, 253–279.

Folke, O. and J. Rickne (2022): "Sexual Harassment and Gender Inequality in the Labor Market," *Quarterly Journal of Economics*, 1–50.

Greenberg, B. G., A.-L. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz (1969): "The unrelated question randomized response model: Theoretical framework," *Journal of the American Statistical Association*, 64, 520–539.

Hershkowitz, I., M. E. Lamb, and C. Katz (2014): "Allegation rates in forensic child abuse investigations: Comparing the revised and standard NICHD protocols." *Psychology, Public Policy, and Law*, 20, 336.

Heyes, A. and S. Kapur (2009): "An economic model of whistle-blower policy," *The Journal of Law, Economics, & Organization*, 25, 157–182.

Jayachandran, S. (2021): "Social Norms as a Barrier to Women's Employment in Developing Countries," *IMF Economic Review*, 69, 576–595.

Kabeer, N., L. Huq, and M. Sulaiman (2020): "Paradigm Shift or Business as Usual? Workers' Views on Multi-stakeholder Initiatives in Bangladesh," *Development and Change*, 0, 1–39.

Kondylis, F., A. Legovini, K. Vyborny, A. Zwager, and L. Andrade (2020): "Demand for "Safe Space": Avoiding Harassment and Stigma," Tech. rep., mimeo.

Laffont, J.-J. and D. Martimort (1997): "Collusion under asymmetric information," *Econometrica*, 65, 875–911.

——— (2000): "Mechanism design with collusion and correlation," *Econometrica*, 68, 309–342.

Macchiavello, R., A. Menzel, A. Rabbani, and C. Woodruff (2020): "Challenges of Change: An Experiment Promoting Women to Managerial Roles in the Bangladeshi Garment Sector," Tech. rep., mimeo.

Makowsky, M. D. and S. Wang (2018): "Embezzlement, whistleblowing, and organizational architecture: An experimental investigation," *Journal of Economic Behavior & Organization*, 147, 58–75.

Muraglia, S., A. Vasquez, and J. Reichert (2020): "Conducting research interviews on sensitive topics," *Illinois Criminal Justice Information Authority (ICJIA)*.

Poi, B. P. (2004): "From the help desk: Some bootstrapping techniques," *The Stata Journal*, 4, 312–328.

Prendergast, C. (2000): "Investigating Corruption," Tech. rep., Working Paper, World Bank Development Group.

Rosenfeld, B., K. Imai, and J. N. Shapiro (2016): "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions," *American Journal of Political Science*, 60, 783–802.

Siddiqi, D. M. (2003): "The Sexual Harassment of Industrial Workers: Strategies for Intervention in the Workplace and Beyond," Tech. rep., Center for Policy Dialogue, Dhaka, Bangladesh.

Siddique, Z. (forthcoming): "Media reported violence and Female Labor Supply," *Economic Development and Cultural Change*.

Sumon, M. H., A. Borhan, and N. Shifa (2018): "Garment Workers' Rights: Situation analysis in Dhaka, Gazipur, Narayanganj, and Chittagong," Tech. rep., Manusher Jonno Foundation, Dhaka, Bangladesh.

Tirole, J. (1986): "Hierarchies and bureaucracies: On the role of collusion in organizations," *Journal of Law, Economics, & Organizations*, 2, 181–214.

United Nations Human Rights Office (2011): "Manual on human rights monitoring," *OHCHR UN Publications*.

United Nations Statistical Office (2014): "Guidelines for producing statistics on violence against women," *United Nations, Department of Economic and Social Affairs Statistics*.

Vallano, J. P. and N. S. Compo (2011): "A comfortable witness is a good witness: Rapport-building and susceptibility to misinformation in an investigative mock-crime interview," *Applied cognitive psychology*, 25, 960–970.

Warner, S. L. (1965): "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, 60, 63–69.

Appendix

A Proofs

Proof of Proposition 2. Since workers are exchangeable, the distributions μ and $\widetilde{\mu}$ are entirely described by the associated distribution of the number of positive reports: $\forall k \in \{1, \cdots, L\}$

$$p_k \equiv \text{prob}_{\mu}\left(\sum_{i \in I} r_i = k\right) \quad \text{and} \quad \widetilde{p}_k \equiv \text{prob}_{\widetilde{\mu}}\left(\sum_{i \in I} \widetilde{r}_i = k\right).$$

Under i.i.d. garbling with garbling rate π, distribution parameters $(p_k)_{k \in \{1,\cdots,L\}}$ and $(\widetilde{p}_k)_{k \in \{1,\cdots,L\}}$ are related as follows:

$$\begin{aligned}
&\widetilde{p}_0 = p_0(1-\pi)^L \\
&\widetilde{p}_1 = p_0\binom{L}{1}\pi(1-\pi)^{L-1} + p_1(1-\pi)^{L-1} \\
&\widetilde{p}_2 = p_0\binom{L}{2}\pi^2(1-\pi)^{L-2} + p_1\binom{L-1}{1}\pi(1-\pi)^{L-2} + p_2(1-\pi)^{L-2} \\
\forall k \in \{1, \cdots, L\}, \quad &\widetilde{p}_k = \sum_{n=0}^{k} p_n \binom{L-n}{k-n}\pi^{k-n}(1-\pi)^{L-k}.
\end{aligned}$$

This is a triangular system of linear equation which means we can infer p_ks using observed $\widetilde{p}_k$s using the following recursion:

$$\begin{aligned} p_0 &= \frac{1}{(1-\pi)^L}\,\widetilde{p}_0 \\ p_1 &= \frac{1}{(1-\pi)^{L-1}}\,\widetilde{p}_1 - p_0\binom{L}{1}\pi \\ \forall k \in \{2,\cdots,L\}, \quad p_k &= \frac{1}{(1-\pi)^{L-k}}\,\widetilde{p}_k - \sum_{n=0}^{k-1} p_n \binom{L-n}{k-n}\pi^{k-n}. \end{aligned}$$

This concludes the proof that μ is identified given $\widetilde{\mu}$. ■

Proof of Eqn 4.

$$\begin{aligned} \mathsf{Var}\left(\sum_{j=1}^{n} r_j\eta_j \middle| \mathbf{r}\right) &= \sum_j r_j \mathsf{Var}(\eta_j) + \sum_{j\neq j'} r_j r_{j'} \mathsf{Cov}(\eta_j, \eta_{j'}) \\ &= \overline{r}n\pi(1-\pi) - \sum_{j,j'} r_j r'_j \frac{\pi(1-\pi)}{n-1} + \sum_j r_j \frac{\pi(1-\pi)}{n-1} \\ &= \overline{r}n\pi(1-\pi) - [\overline{r}n]^2\pi(1-\pi) + \overline{r}n\frac{\pi(1-\pi)}{n-1} \\ &= \overline{r}\left(1 - \frac{\overline{r}n-1}{n-1}\right)\pi(1-\pi)n. \end{aligned}$$

■

B Extensions

B.1 Measurement under alternative frameworks

B.1.1 Blocked garbling

Proposition B.1 (identification under blocked garbling)**.** *Whenever $\widetilde{\mu}(L) = 0$, the true distribution μ of team-level intended reports is identified from $\widetilde{\mu}$.*

Proof of Proposition B.1. As in the case of Proposition 2, since workers are exchangeable, the distributions μ and $\widetilde{\mu}$ are entirely described by the associated distribution of the

number of positive reports: $\forall k \in \{1, \cdots, L\}$

$$p_k \equiv \text{prob}_\mu \left(\sum_{i \in I} r_i = k \right) \quad \text{and} \quad \widetilde{p}_k \equiv \text{prob}_{\widetilde{\mu}} \left(\sum_{i \in I} \widetilde{r}_i = k \right).$$

Under blocked garbling with 2 null responses garbled ex ante per team, distribution parameters $(p_k)_{k \in \{1, \cdots, L\}}$ and $(\widetilde{p}_k)_{k \in \{1, \cdots, L\}}$ are related as follows:

$$\begin{aligned}
&\widetilde{p}_0 = 0 \\
&\widetilde{p}_1 = 0 \\
&\widetilde{p}_2 = \left[p_0 \binom{L}{2} + p_1 \binom{L-1}{1} + p_2 \right] \Big/ \binom{L}{2} \\
&\widetilde{p}_3 = \left[p_1 \binom{L-1}{2} + p_2 \binom{2}{1} \binom{L-2}{1} + p_3 \binom{3}{2} \right] \Big/ \binom{L}{2} \\
\forall k \in 2, \cdots, L, \quad &\widetilde{p}_k = \left[p_{k-2} \binom{L-k+2}{2} + p_{k-1} \binom{k-1}{1} \binom{L-k+1}{1} + p_k \binom{k}{2} \right] \Big/ \binom{L}{2}.
\end{aligned}$$

In general this system of equations is not invertible. However it is invertible whenever $\widetilde{p}_L = 0$. This implies that $p_L = p_{L-1} = p_{L-2} = 0$. In turn, p_ks for $k < L - 2$ can be recovered using the backward recursion

$$p_k = \left[\widetilde{p}_{k+2} \binom{L}{2} - p_{k+2} \binom{k+2}{2} - p_{k+1} \binom{k+1}{1} \binom{L-k-1}{1} \right] \Big/ \binom{L-k}{2}.$$

■

B.2 Likelihoods for the 3 types, conditionally independent harassment model

This section provides likelihood functions for the small dimensional model of harassment described in Section 3.

A manager $a \in M$ can be one of three types $\theta \in \{L, M, H\}$, with respective probabilities q_L, q_M and q_H. Conditional on a type θ, the manager harasses each worker i under their span of control independently with fixed probability ρ_θ where we assume that $\rho_L = 0$ and $\rho_M \leq \rho_H$. The data generating process is entirely specified by the 4 dimensional vector $\gamma = (q_M, q_H, \rho_M, \rho_H)$.

Given γ, the likelihood of observable data $\widetilde{\mathbf{r}}$ associated with different garbling schemes is as follows.

To reflect real data, we allow the size of each team to vary with the manager. We denote by L_a the size of the team reporting to manager a.

Intended responses. The likelihood $p_{k,a}$ of observing k intended reports equal to 1 in team a takes the form

$$p_{k,a} = \sum_{\theta \in \{L,M,H\}} q_\theta \rho_\theta^k (1-\rho_\theta)^{L_a - k} \binom{L_a}{k}.$$

I.i.d. garbling. Under i.i.d. garbling the likelihood of observing k garbled reports $\widetilde{r}_{i,a} = 1$ in team a is

$$\widetilde{p}_{k,a} = \sum_{n=0}^{k} p_{n,a} \pi^{k-n} (1-\pi)^{L_a - k} \binom{L_a - n}{k-n}.$$

Blocked garbling. In turn, under blocked garbling, if a number g_a of potential null reports are set to 1 in team a, the likelihood of observing k garbled reports $\widetilde{r}_{i,a} = 1$ in team a is

$$\widetilde{p}_{k,a} = \sum_{n=k-g_a}^{k} p_{n,a} \binom{L_a - n}{k-n} \binom{g_a - k + n}{n} \Big/ \binom{L_a}{g_a}.$$

Log likelihood. Let us denote by $\widetilde{k}_a \equiv \sum_{i \in I} \widetilde{r}_{i,a}$. Altogether the log likelihood associated with observing data $\widetilde{\mathbf{r}}$ is

$$L(\widetilde{\mathbf{r}}) = \sum_{a \in M} \log(\widetilde{p}_{\widetilde{k}_a, a}).$$

B.3 Trusted analyst

Under the trusted analyst paradigm, we assume that two pieces of information are associated with a recorded report: the recorded report $\widetilde{r}_i$, potentially made available to the principal, on the basis of which action may be taken; and an encoded version of the intended report which the analyst can use to recover intended reports.

Importantly, releasing the distribution of intended reports $\widehat{\mu}$ has a vanishing impact on inferences that can be made from realized reports as the number of teams gets large.

Proposition B.2. *Take as given an equilibrium behavior. With probability approaching 1 as the number of team m gets large,*

$$|prob(r_{i,a} = 1|\widetilde{r}_{i,a} = 1) - prob(r_{i,a} = 1|\widetilde{r}_{i,a} = 1, \widehat{\mu})| \to 0.$$

B.4 HG & subversion over the long-run

We focus on introducing HG with an aim to increase respondents' willingness to report. As Chassang and Padró i Miquel (2018) note, however, the effectiveness of monitoring mechanisms often differs over the short-run versus the long run. In the long run, relevant players will figure out ways to game the system for their own benefit. Two are of specific concern: the first one is that managers may increase the magnitude of retaliation in response to the increased anonymity provided by hard garbling; the second is that workers may start strategically misreporting well behaving managers out of idiosyncratic spite, or to further their own careers. In principle, such subversive behavior can break the connection between reports of misbehavior and actual misbehavior: a drop in incriminating reports may be driven by threats of greater retaliation rather than a reduction in misbehavior, and an increase in incriminating reports may be driven by malicious reporting rather than underlying crime.

Chassang and Padró i Miquel (2018) offers some guidance on how to approach sustainable policy-making in such a context. They explicitly allow for subversion in a model of whistleblowing that includes endogenous threats by managers as well as malicious workers benefiting from reporting good managers. In general it is difficult to identify optimal garbling policy on the basis of what are ultimately unverifiable reports. However, Chassang and Padró i Miquel (2018) show that it's possible to pick whistleblowing policies offering robust guarantees by comparing outcomes under various sets of experiments. The key policy dimensions to vary are: (1) the level of enforcement, i.e. the impact of a positive report on a manager's outcome; (2) the information content of a positive realized report, captured by the likelihood ratio $\text{prob}(r = 1|\widetilde{r} = 1)/\text{prob}(r = 0|\widetilde{r} = 1)$.

Chassang and Padró i Miquel (2018) show that one can reach whistleblowing policies than deliver robust guarantees on the underlying level of misbehavior using the following steps:

1. starting from a low level of enforcement, reduce the information content of reports up to a point where workers are willing to complain;

2. keeping the information content of positive reports the same, scale up enforcement.

Chassang and Padró i Miquel (2018) show that the reduction in reported misbehavior along a policy path with reports of constant informativeness corresponds to actual reductions in misbehavior from managers. However, reductions in reported crime are only achievable if one can find an initial policy configuration where there are some reports of misbehavior. Therefore finding a garbled low-enforcement policy configuration that yields high reporting is a key preliminary requirement for robust long-term policy elaboration. One can interpret the findings of this paper as establishing such a starting point.

C Figures & Tables

Figure A.1: Survey Modules & Treatment Interventions

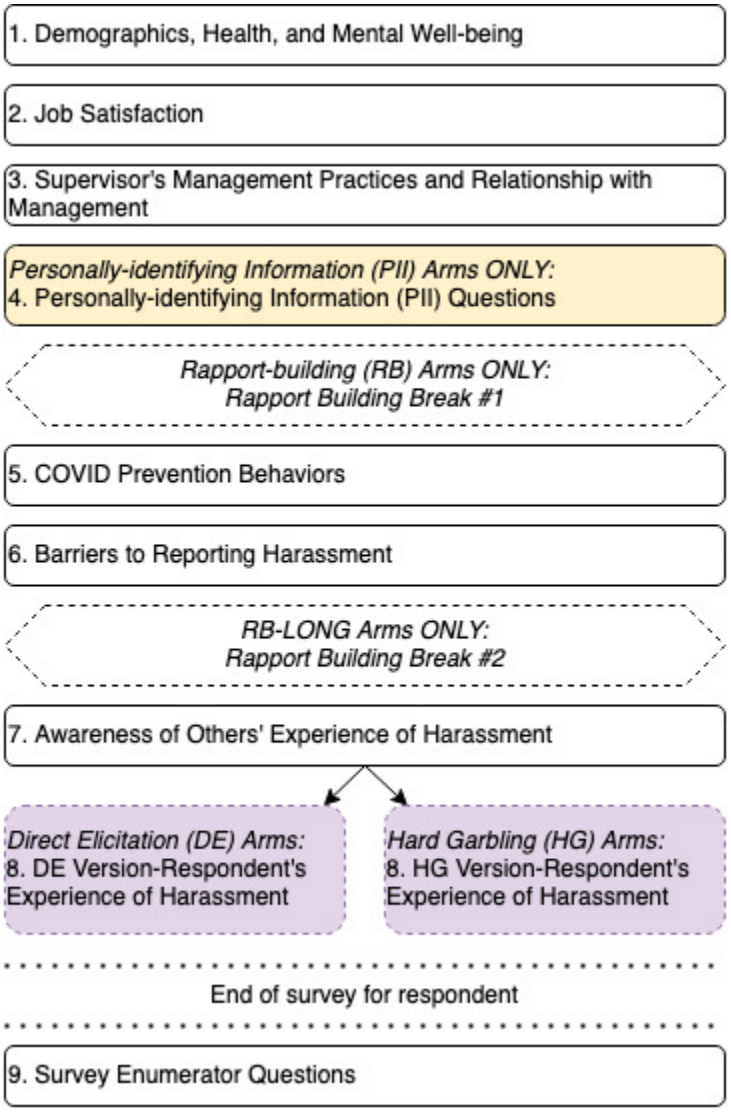
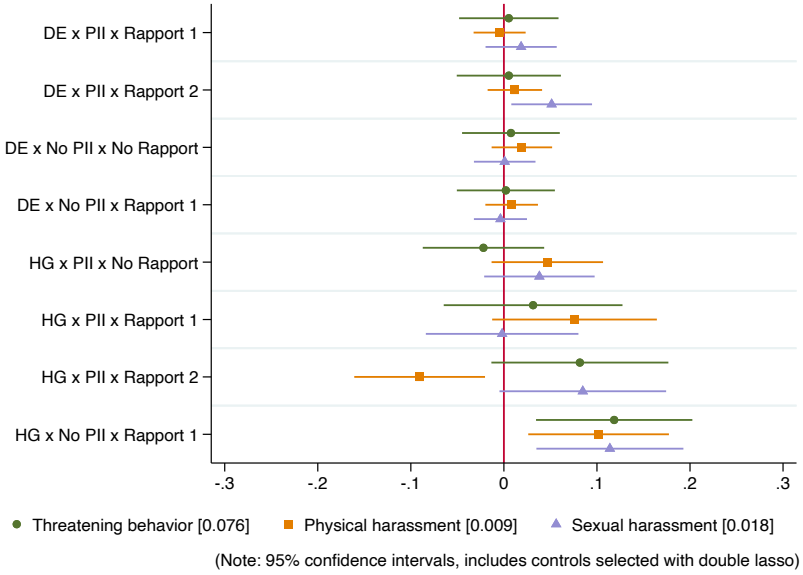


Figure A.2: Treatment effects by survey arm, separately by sex

(a) For Women



(b) For Men

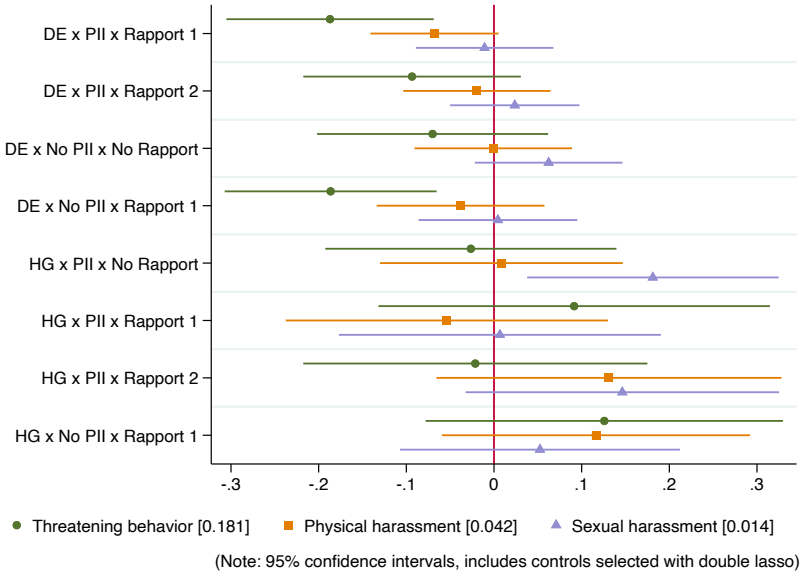


Table A.1: Summary Statistics (Team)

	Mean	SD	Min	p25	p50	p75	Max	N
Panel A: Number of workers in a team								
Team Size: Overall	53.1	20.8	17	35	54	72	98	112
Team Size: Factory 1	54.9	23.1	19	32	55.5	74.5	98	60
Team Size: Factory 2	51	17.7	17	37	47.5	69	74	52
Team Size: Sewing Section	70.9	7.75	49	67.5	72	74.5	90	48
Team Size: Finishing Section	35.8	8.98	20	30	35.5	39	65	46
Team Size: Washing Section	49.8	27.0	17	26	47	65	98	18
Panel B: Share of Female workers in a team								
Team's Female Share: Overall	0.82	0.26	0	0.84	0.92	0.96	1	112
Team's Female Share: Factory 1	0.85	0.26	0	0.88	0.94	0.97	1	60
Team's Female Share: Factory 2	0.79	0.25	0	0.81	0.88	0.93	1	52
Team's Female Share: Sewing Section	0.95	0.033	0.86	0.93	0.96	0.98	1	48
Team's Female Share: Finishing Section	0.89	0.062	0.72	0.85	0.89	0.93	1	46
Team's Female Share: Washing Section	0.30	0.28	0	0.063	0.19	0.58	0.82	18

Table A.2: Balance tests: Main treatment conditions

	Mean / (SE)						Difference in means / [p-value]		
Variable	DE	HG	No Rapport	Rapport	PII	No PII	HG-DE	Diff Rapport	Diff PII
Female	0.810	0.816	0.813	0.812	0.815	0.806	0.007	0.009	-0.006
	(0.393)	(0.388)	(0.390)	(0.391)	(0.388)	(0.396)	[0.185]	[0.102]	[0.314]
Currently Working	0.957	0.961	0.955	0.962	0.959	0.958	0.003	0.005	-0.003
	(0.202)	(0.194)	(0.207)	(0.191)	(0.197)	(0.201)	[0.746]	[0.529]	[0.781]
Age	26.681	26.881	26.661	26.870	26.813	26.680	0.200	0.103	-0.102
	(5.030)	(5.254)	(5.064)	(5.195)	(5.205)	(4.964)	[0.351]	[0.631]	[0.657]
Experience (yrs)	5.188	5.204	5.141	5.241	5.191	5.206	-0.022	0.064	0.039
	(3.628)	(3.510)	(3.554)	(3.589)	(3.590)	(3.533)	[0.878]	[0.662]	[0.809]
Tenure (yrs)	2.923	2.895	2.888	2.929	2.890	2.959	-0.012	-0.073	0.079
	(2.458)	(2.441)	(2.486)	(2.419)	(2.388)	(2.594)	[0.885]	[0.398]	[0.419]
Years of Education	6.773	6.640	6.716	6.706	6.728	6.668	-0.115	0.024	-0.070
	(3.381)	(3.386)	(3.362)	(3.402)	(3.350)	(3.465)	[0.407]	[0.863]	[0.641]
Marital Status (1=Yes)	0.837	0.811	0.827	0.823	0.822	0.830	-0.028*	-0.008	0.006
	(0.370)	(0.392)	(0.378)	(0.382)	(0.382)	(0.376)	[0.087]	[0.602]	[0.737]
Children (1=Yes)	0.736	0.744	0.742	0.739	0.739	0.742	0.007	-0.004	0.004
	(0.441)	(0.436)	(0.438)	(0.439)	(0.439)	(0.438)	[0.703]	[0.825]	[0.827]
Observations	1,176	1,021	1,006	1,191	1,554	643	2,197	2,197	2,197
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table summarizes workers' characteristics. The table reports the mean values of each variable for each treatment condition. Robust standard errors are reported. The final three columns report mean differences between each treatment condition. In column (4), Rapport pools the short and long rapport conditions. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.3: Balance tests: No rapport, short rapport, and long rapport arms

	Mean / (SE)			**Difference in means / [p-value]**		
Variable	No Rapport	Short Rapport	Long Rapport	(1) - (0)	(2) - (0)	(2) - (1)
	(0)	(1)	(2)			
Female	0.813	0.819	0.796	0.009	0.008	-0.003
	(0.390)	(0.385)	(0.404)	[0.118]	[0.251]	[0.646]
Currently Working	0.955	0.964	0.958	0.008	0.000	-0.008
	(0.207)	(0.185)	(0.202)	[0.402]	[0.989]	[0.526]
Age	26.661	26.861	26.889	0.124	0.090	-0.020
	(5.064)	(5.101)	(5.398)	[0.595]	[0.775]	[0.951]
Experience (yrs)	5.141	5.338	5.033	0.167	-0.178	-0.357
	(3.554)	(3.574)	(3.617)	[0.296]	[0.400]	[0.101]
Tenure (yrs)	2.888	2.894	3.002	-0.092	-0.062	0.091
	(2.486)	(2.410)	(2.442)	[0.336]	[0.610]	[0.464]
Years of Education	6.716	6.690	6.740	0.019	0.051	0.025
	(3.362)	(3.439)	(3.326)	[0.897]	[0.797]	[0.906]
Marital Status (1=Yes)	0.827	0.823	0.822	-0.009	-0.006	0.000
	(0.378)	(0.382)	(0.383)	[0.624]	[0.799]	[0.985]
Children (1=Yes)	0.742	0.744	0.727	0.003	-0.017	-0.019
	(0.438)	(0.436)	(0.446)	[0.892]	[0.515]	[0.476]
Observations	1,006	814	377	1,820	1,383	1,191
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table summarizes workers' characteristics. The table reports the mean values of each variable for each treatment arm. Robust standard errors are reported. The final three columns report mean differences between each treatment arm. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A.4: Effects of Survey Design on Reporting of Harassment

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
HG Treatment	0.0539***	0.0537***	0.0394**	0.0407**	0.0524***	0.0544***
	(0.0203)	(0.0196)	(0.0178)	(0.0172)	(0.0184)	(0.0178)
No PII Treatment	0.0379*	0.0401*	0.0246	0.0249	0.0259	0.0280
	(0.0221)	(0.0213)	(0.0187)	(0.0180)	(0.0186)	(0.0181)
Rapport Treatment (Short)	0.0252	0.0235	0.0091	0.0089	-0.0061	-0.0077
	(0.0212)	(0.0206)	(0.0185)	(0.0178)	(0.0184)	(0.0178)
Rapport Treatment (Long)	0.0316	0.0329	-0.0393*	-0.0394*	0.0423	0.0413
	(0.0280)	(0.0270)	(0.0231)	(0.0223)	(0.0262)	(0.0252)
Control Group Mean	.0947	.0947	.0146	.0146	.017	.017
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Yes	No	Yes	No	Yes
Observations	2194	2194	2194	2194	2194	2194

Notes: This table reports OLS estimates of treatment effects on workers' reporting. Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the treatment indicator and stratification variables. Even-numbered columns also include controls selected using the PDS lasso. Robust standard errors are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Table A.5: Effects of Survey Design on Reporting of Harassment, Recorded HG Responses (Full Interactions)

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
DE × PII × Rapport 1	-0.0276	-0.0292	-0.0091	-0.0121	0.0229	0.0190
	(0.0246)	(0.0243)	(0.0126)	(0.0127)	(0.0169)	(0.0168)
DE × PII × Rapport 2	-0.0132	-0.0163	0.0062	0.0037	0.0387*	0.0387**
	(0.0262)	(0.0257)	(0.0142)	(0.0141)	(0.0200)	(0.0191)
DE × No PII × No Rapport	-0.0098	-0.0105	0.0184	0.0136	0.0124	0.0132
	(0.0253)	(0.0244)	(0.0156)	(0.0157)	(0.0152)	(0.0153)
DE × No PII × Rapport 1	-0.0296	-0.0294	0.0087	0.0033	-0.0004	-0.0024
	(0.0258)	(0.0248)	(0.0146)	(0.0141)	(0.0140)	(0.0137)
HG × PII × No Rapport	-0.0205	-0.0222	0.0406	0.0380	0.0557*	0.0561**
	(0.0319)	(0.0308)	(0.0287)	(0.0277)	(0.0289)	(0.0279)
HG × PII × Rapport 1	0.0457	0.0389	0.0597	0.0564	-0.0010	-0.0000
	(0.0461)	(0.0448)	(0.0427)	(0.0409)	(0.0396)	(0.0381)
HG × PII × Rapport 2	0.0529	0.0559	-0.0468	-0.0481	0.0955**	0.0939**
	(0.0449)	(0.0432)	(0.0361)	(0.0350)	(0.0423)	(0.0407)
HG × No PII × Rapport 1	0.1253***	0.1252***	0.0946***	0.0975***	0.0944***	0.0975***
	(0.0412)	(0.0396)	(0.0363)	(0.0349)	(0.0366)	(0.0354)
Control Group Mean	.0947	.0947	.0146	.0146	.017	.017
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Lasso	No	Lasso	No	Lasso
Observations	2194	2194	2194	2194	2194	2194

Notes: This table reports OLS estimates of treatment effects on workers' reporting. Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the full interactions of treatment variables and stratification variables. Even-numbered columns also include controls selected using the PDS lasso. Robust standard errors are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

See Table A.6 on next page for p-values.

Table A.6: Effects of Survey Design on Reporting of Harassment, Recorded HG Responses (Full Interactions, p-values of differences between coefficients)

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
DExPIIxRB1 − DExPIIxRB2	[0.617]	[0.646]	[0.331]	[0.314]	[0.495]	[0.381]
DExPIIxRB1 − DExNoPIIxNoRB	[0.528]	[0.494]	[0.103]	[0.128]	[0.586]	[0.764]
DExPIIxRB1 − DExNoPIIxRB1	[0.945]	[0.996]	[0.265]	[0.325]	[0.206]	[0.244]
DExPIIxRB1 − HGxPIIxNoRB	[0.835]	[0.833]	[0.088]	[0.075]	[0.291]	[0.219]
DExPIIxRB1 − HGxPIIxRB1	[0.125]	[0.143]	[0.113]	[0.103]	[0.564]	[0.633]
DExPIIxRB1 − HGxPIIxRB2	[0.084]	[0.059]	[0.312]	[0.317]	[0.097]	[0.077]
DExPIIxRB1 − HGxNoPIIxRB1	[0.000]	[0.000]	[0.005]	[0.002]	[0.062]	[0.035]
DExPIIxRB2 − DExNoPIIxNoRB	[0.910]	[0.842]	[0.511]	[0.595]	[0.231]	[0.231]
DExPIIxRB2 − DExNoPIIxRB1	[0.589]	[0.657]	[0.888]	[0.978]	[0.073]	[0.050]
DExPIIxRB2 − HGxPIIxNoRB	[0.838]	[0.863]	[0.252]	[0.237]	[0.607]	[0.582]
DExPIIxRB2 − HGxPIIxRB1	[0.226]	[0.243]	[0.223]	[0.214]	[0.353]	[0.343]
DExPIIxRB2 − HGxPIIxRB2	[0.164]	[0.116]	[0.158]	[0.155]	[0.210]	[0.204]
DExPIIxRB2 − HGxNoPIIxRB1	[0.002]	[0.001]	[0.018]	[0.009]	[0.163]	[0.125]
DExNoPIIxNoRB − DExNoPIIxRB1	[0.497]	[0.500]	[0.602]	[0.569]	[0.453]	[0.354]
DExNoPIIxNoRB − HGxPIIxNoRB	[0.760]	[0.725]	[0.466]	[0.409]	[0.154]	[0.146]
DExNoPIIxNoRB − HGxPIIxRB1	[0.248]	[0.289]	[0.352]	[0.318]	[0.744]	[0.736]
DExNoPIIxNoRB − HGxPIIxRB2	[0.183]	[0.142]	[0.088]	[0.097]	[0.056]	[0.054]
DExNoPIIxNoRB − HGxNoPIIxRB1	[0.002]	[0.001]	[0.045]	[0.023]	[0.029]	[0.021]
DExNoPIIxRB1 − HGxPIIxNoRB	[0.795]	[0.833]	[0.284]	[0.226]	[0.061]	[0.044]
DExNoPIIxRB1 − HGxPIIxRB1	[0.121]	[0.146]	[0.249]	[0.212]	[0.987]	[0.952]
DExNoPIIxRB1 − HGxPIIxRB2	[0.081]	[0.060]	[0.142]	[0.159]	[0.026]	[0.020]
DExNoPIIxRB1 − HGxNoPIIxRB1	[0.000]	[0.000]	[0.022]	[0.009]	[0.011]	[0.006]
HGxPIIxNoRB − HGxPIIxRB1	[0.202]	[0.223]	[0.705]	[0.706]	[0.234]	[0.219]
HGxPIIxNoRB − HGxPIIxRB2	[0.150]	[0.111]	[0.053]	[0.048]	[0.426]	[0.431]
HGxPIIxNoRB − HGxNoPIIxRB1	[0.002]	[0.001]	[0.228]	[0.166]	[0.390]	[0.339]
HGxPIIxRB1 − HGxPIIxRB2	[0.906]	[0.771]	[0.054]	[0.050]	[0.089]	[0.085]
HGxPIIxRB1 − HGxNoPIIxRB1	[0.171]	[0.124]	[0.526]	[0.436]	[0.071]	[0.054]
HGxPIIxRB2 − HGxNoPIIxRB1	[0.206]	[0.207]	[0.005]	[0.003]	[0.983]	[0.946]
Complementarity Test 1	[0.002]	[0.001]	[0.188]	[0.120]	[0.474]	[0.428]
Complementarity Test 2	[0.051]	[0.052]	[0.295]	[0.275]	[0.942]	[0.939]
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS Lasso Controls	No	Lasso	No	Lasso	No	Lasso

Notes: This table reports p-values of the difference between fully interacted treatment groups from the OLS regression of treatment effects on workers' reporting. Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the full interactions of treatment variables and stratification variables. Even-numbered columns also include controls selected using the PDS lasso.

Complementarity Test 1: HGxNoPIIxRB1 ≤ DExPIIxRB1 + DExNoPIIxNoRB + HGxPIIxNoRB. We test for complementarity because for all outcomes, the point estimate for HGxNoPIIxRB1 is greater than the sum of the point estimates for the other three arms.

Complementarity Test 2: HGxPIIxRB1 ≤ DExPIIxRB1 + HGxPIIxNoRB. We focus on the test for complementarity because the point estimate for HGxPIIxRB1 is greater than the sum of the coefficients for the DExPIIxRB1 and HGxPIIxNoRB arms for 2 of 3 outcomes.

Robust standard errors are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Table A.7: Effects of Survey Design on Survey Duration

	Rapport Treatment (Pooled)		Rapport Treatment	
	(1)	(2)	(3)	(4)
HG Treatment	1.9513***	1.9333***	2.0155***	1.9978***
	(0.5263)	(0.5046)	(0.5275)	(0.5055)
No PII Treatment	-1.6580***	-1.6193***	-1.1822*	-1.1369*
	(0.5782)	(0.5555)	(0.6357)	(0.6114)
Rapport Treatment (Pooled)	6.0851***	6.0393***		
	(0.5325)	(0.5130)		
Rapport Treatment (Short)			5.5365***	5.4826***
			(0.6166)	(0.5941)
Rapport Treatment (Long)			6.9836***	6.9438***
			(0.7746)	(0.7466)
Control Group Mean	41.85	41.85	41.85	41.85
p(Long − Short Rapport)			[0.098]	[0.083]
Strata FE	Yes	Yes	Yes	Yes
PDS lasso controls	No	Yes	No	Yes
Observations	2155	2155	2155	2155

Notes: This table reports OLS estimates of treatment effects on survey duration (in minutes) which is trimmed below and above at 1 and 99 percentiles respectively. Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the treatment indicator and stratification variables. Even-numbered columns also include controls selected using the PDS lasso. Robust standard errors are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Table A.8: Main treatment effects, estimated with response = "no" for confused respondents

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main effects						
HG Treatment	0.0423**	0.0421**	0.0316*	0.0331*	0.0372**	0.0391**
	(0.0201)	(0.0195)	(0.0177)	(0.0171)	(0.0181)	(0.0176)
No PII Treatment	0.0336	0.0350*	0.0419**	0.0424**	0.0105	0.0120
	(0.0209)	(0.0201)	(0.0175)	(0.0168)	(0.0180)	(0.0175)
Rapport Treatment	0.0272	0.0266	-0.0127	-0.0131	0.0140	0.0131
	(0.0187)	(0.0181)	(0.0159)	(0.0154)	(0.0166)	(0.0160)
Control Group Mean	.0947	.0947	.0146	.0146	.017	.017
Panel B: Heterogeneity by sex						
HG Treatment × Female	0.0285	0.0281	0.0310	0.0335*	0.0271	0.0298
	(0.0221)	(0.0214)	(0.0196)	(0.0189)	(0.0201)	(0.0195)
HG Treatment × Male	0.1049**	0.1030**	0.0353	0.0312	0.0793*	0.0797*
	(0.0490)	(0.0474)	(0.0422)	(0.0407)	(0.0426)	(0.0415)
No PII Treatment × Female	0.0323	0.0336	0.0428**	0.0417**	0.0174	0.0178
	(0.0232)	(0.0223)	(0.0193)	(0.0186)	(0.0203)	(0.0197)
No PII Treatment × Male	0.0302	0.0350	0.0375	0.0460	-0.0215	-0.0185
	(0.0491)	(0.0476)	(0.0409)	(0.0399)	(0.0394)	(0.0383)
Rapport × Female	0.0430**	0.0420**	-0.0139	-0.0142	0.0285	0.0279
	(0.0208)	(0.0201)	(0.0176)	(0.0170)	(0.0184)	(0.0176)
Rapport × Male	-0.0382	-0.0413	-0.0063	-0.0080	-0.0558	-0.0542
	(0.0439)	(0.0424)	(0.0379)	(0.0366)	(0.0389)	(0.0378)
Female	-0.0898	-0.0970	-0.0325	-0.0201	0.0383	0.0482
	(0.0833)	(0.0803)	(0.0652)	(0.0634)	(0.0644)	(0.0625)
Control Mean - Female	.0765	.0765	.0088	.0088	.0176	.0176
Control Mean - Male	.1806	.1806	.0417	.0417	.0139	.0139
p(HGxFemale - HGxMale)	[0.156]	[0.151]	[0.927]	[0.960]	[0.270]	[0.278]
p(NoPIIxFemale - NoPIIxMale)	[0.968]	[0.979]	[0.907]	[0.922]	[0.380]	[0.401]
p(RapportxFemale - RapportxMale)	[0.095]	[0.075]	[0.857]	[0.879]	[0.050]	[0.049]
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Lasso	No	Lasso	No	Lasso
Observations	2194	2194	2194	2194	2194	2194

Notes: This table reports OLS estimates of treatment effects on workers' reporting (also heterogeneity by sex). Each column in the table reports the estimated coefficient from a separate regression. The dependent variable in each column is regressed on the gender interactions of treatment variables and stratification variables. Even-numbered columns also include controls selected using the PDS lasso. Robust standard errors are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Table A.9: HTEs by respondents' schooling qualification for supervisor position

	Threatening behavior	Physical harassment	Sexual harassment
	(1)	(2)	(3)
HG Treatment × Female × Min Grade 8	0.0267	0.0379	0.0989***
	(0.0333)	(0.0306)	(0.0323)
HG Treatment × Female × Below Grade 8	0.0489*	0.0336	0.0029
	(0.0293)	(0.0257)	(0.0258)
HG Treatment × Male × Min Grade 8	0.0935	0.1003	0.0617
	(0.0643)	(0.0612)	(0.0556)
HG Treatment × Male × Below Grade 8	0.1389*	0.0358	0.1000
	(0.0711)	(0.0617)	(0.0629)
Rapport Treatment	0.0287	-0.0091	0.0109
	(0.0189)	(0.0161)	(0.0168)
No PII Treatment	0.0352*	0.0398**	0.0116
	(0.0210)	(0.0177)	(0.0181)
Control Mean-Female & Above	.068	.0068	.0136
Control Mean-Female & Below	.0829	.0104	.0207
Control Mean-Male & Above	.2105	.0263	.0263
Control Mean-Male & Below	.1471	.0588	0
p(HGXFemaleXHigh-HGXFemaleXLow)	[0.612]	[0.913]	[0.020]
p(HGXMaleXHigh-HGXMaleXLow)	[0.628]	[0.457]	[0.641]
Strata FE	Yes	Yes	Yes
Observations	2194	2194	2194

Notes: Main effects of gender and schooling included but not displayed. Rapport pools the short and long rapport conditions. Robust standard errors in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Table A.10: Correlation of team-level reporting rates and response rate to the survey

Correlations	DE	HG
ρ(Threat, Survey Response Rate)	-0.036	-0.088
	(0.101)	(0.091)
	[-0.249,0.153]	[-0.252,0.097]
ρ(Physical, Survey Response Rate)	-0.005	0.050
	(0.100)	(0.099)
	[-0.215,0.178]	[-0.123,0.269]
ρ(Sexual, Survey Response Rate)	0.030	-0.025
	(0.125)	(0.090)
	[-0.228,0.260]	[-0.195,0.163]

Notes: This table reports the correlation between the team-level response rate to the survey and the team-level reporting rates of harassment. Standard errors (in parenthesis) are computed from 1000 bootstrap replications, drawing samples of reporting rates at the team-level. Confidence intervals [in brackets] are bias corrected and accelerated (BCa), following (Efron, 1987, Davison and Hinkley, 1997), implemented using STATA package **bootstrap** (Poi, 2004).

Table A.11: Correlation of team-level reporting rates with witnessed harassment

Panel A: Witness Reports: Average number of workers in own team witnessed being harassed

	Witnessed Sexual Harassment		*Witnessed Physical Harassment*	
Correlations	DE	HG	DE	HG
ρ(Threat, Witness Reports)	0.155	0.026	0.035	-0.102
	(0.088)	(0.097)	(0.076)	(0.082)
	[-0.023,0.317]	[-0.165,0.204]	[-0.107,0.189]	[-0.254,0.078]
ρ(Physical, Witness Reports)	0.108	0.075	0.072	-0.082
	(0.123)	(0.129)	(0.100)	(0.087)
	[-0.062,0.448]	[-0.141,0.366]	[-0.078,0.343]	[-0.227,0.160]
ρ(Sexual, Witness Reports)	0.041	-0.009	0.052	0.034
	(0.103)	(0.111)	(0.088)	(0.108)
	[-0.097,0.343]	[-0.173,0.262]	[-0.083,0.283]	[-0.155,0.285]

Panel B: Witness Reports: Frequency with which other workers are witnessed being harassed

	Witnessed Sexual Harassment		*Witnessed Physical Harassment*	
Correlations	DE	HG	DE	HG
ρ(Threat, Witness Reports)	0.085	-0.043	0.065	-0.090
	(0.087)	(0.106)	(0.105)	(0.118)
	[-0.089,0.250]	[-0.253,0.168]	[-0.120,0.293]	[-0.290,0.170]
ρ(Physical, Witness Reports)	-0.041	-0.019	0.148	-0.048
	(0.085)	(0.089)	(0.128)	(0.077)
	[-0.185,0.157]	[-0.187,0.156]	[-0.065,0.454]	[-0.193,0.116]
ρ(Sexual, Witness Reports)	0.101	-0.043	-0.054	-0.018
	(0.091)	(0.089)	(0.076)	(0.098)
	[-0.064,0.302]	[-0.204,0.124]	[-0.171,0.109]	[-0.216,0.163]

Notes: This table reports the correlation between team-level measures of witnessed harassment and team-level reporting rates of harassment. Standard errors (in parenthesis) are computed from 1000 bootstrap replications, drawing samples of reporting rates at the team-level. Confidence intervals [in brackets] are bias corrected and accelerated (BCa), following (Efron, 1987, Davison and Hinkley, 1997), implemented using STATA package **bootstrap** (Poi, 2004).