NBER WORKING PAPER SERIES

MONITORING HARASSMENT IN ORGANIZATIONS

Laura E. Boudreau
Sylvain Chassang
Ada Gonzalez-Torres
Rachel M. Heath

Working Paper 31011
http://www.nber.org/papers/w31011

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2023

This project is funded by the Private Enterprise Development in Low-income Countries (PEDL) Initiative, Columbia University's Provost's Diversity Grants Program for Junior Faculty and the Israeli Science Foundation. We are grateful to Ferdausi Sumana, Raied Arman, and Krishna Kamepalli for their excellent research assistance. We are grateful to Jana Gallen, Rocco Macchiavello, and Benjamin Roth for detailed discussions. We are indebted to Dan Ben-Moshe, Laura Doval, Florian Englmaier, Nathaniel Hendren, Danielle Li, Tomasso Porzio, and Andrea Prat for helpful comments. We are grateful to seminar participants at Bar-Ilan, Berkeley, Chicago-Booth, Columbia, CUNEF, Haifa, Hebrew U, LMU Munich, Reichman, Rochester, UBC, and USC, as well as participants at the Barcelona Summer Forum, the CEPR/CESifo/Imo-ENT Conference, the Economics of Firms and Labor Conference in Munich, the German-Israeli Frontiers of Humanities Symposium, the NBER Organizational Economics Meeting, the NBER Personnel SI, the SIOE Conference, and the Women in Applied Microeconomics Conference for stimulating comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Monitoring Harassment in Organizations
Laura E. Boudreau, Sylvain Chassang, Ada Gonzalez-Torres, and Rachel M. Heath
NBER Working Paper No. 31011
March 2023
JEL No. C42,D82,J70,J71,J81,J83,M54

ABSTRACT

We evaluate secure survey methods designed for the ongoing monitoring of harassment in organizations. We use the resulting data to answer policy relevant questions about the nature of harassment: How prevalent is it? What share of managers is responsible for the misbehavior? How isolated are its victims?To do so, we partner with a large Bangladeshi garment manufacturer to experiment with different designs of phone-based worker surveys. Garbling responses to sensitive questions by automatically recording a random subset as complaints increases reporting of physical harassment by 288%, sexual harassment by 269%, and threatening behavior by 46%. A rapport-building treatment has an insignificant aggregate effect, but may affect men and women differently. Removing team identifiers from survey responses does not significantly increase reporting and prevents the computation of policy-relevant team-level statistics. The resulting data shows that harassment is widespread, that the problem is not restricted to a minority of managers, and that victims are often isolated in teams.

Laura E. Boudreau
Columbia University
l.boudreau@columbia.edu

Sylvain Chassang
Department of Economics
Princeton University
Julis Romo Rabinowitz Building
Princeton, NJ 08544
and NBER
chassang@princeton.edu

Ada Gonzalez-Torres
Ben-Gurion University of the Negev
Department of Economics
P.O. Box 653
Beer Sheba 8410501
Israel
adagt@bgu.ac.il

Rachel M. Heath
Department of Economics
University of Washington
Box 353330
Seattle, WA 98103
rmheath@uw.edu

A data appendix is available at
http://www.nber.org/data-appendix/w31011
A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/7103

1 Introduction

Organizations' ability to take action against harassment is limited by their ability to elicit information from relevant parties. Reporting harassment 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. In this paper, we study the impact of survey methods that seek to offer plausible deniability, increase trust in the survey enumerator, and reduce the perceived likelihood of leaks, on information transmission. We do so in the context of a phone-based survey experiment implemented in partnership with a large Bangladeshi apparel manufacturer. We use the resulting survey data to draw policy relevant inferences about harassment.

Our theoretical framework builds on a principal-agent-monitor model (Chassang and Padró i Miquel, 2018, Chassang and Zehnder, 2019). A monitor, here the victim, is asked to report harassment behavior by the agent to the principal. The difficulty is that the agent can engage in retaliation, and victims may be concerned that reports could be leaked. Leakages may be the result of legitimate steps taken by the principal to investigate or to address the issue, as well as malicious or erroneous revelation by either the principal or survey collectors. The theoretical framework predicts that steps that increase plausible deniability, i.e. that make it harder to infer a respondent's intended message, as well as steps that increase trust in the enumerator, and reduce the perceived likelihood of leaks, can increase reporting by reducing the perceived risk of retaliation.

This motivates three concrete treatments. First, hard garbling (HG) recorded information by automatically setting a random subset of reports as reports that harassment took place, which provides respondents with plausible deniability in the event that they file an incriminating report (Warner, 1965, Chassang and Padró i Miquel, 2018, Chassang and Zehnder, 2019). Second, rapport building (RB) by the survey enumerator, i.e., chatting about family and hobbies in a natural but pre-specified manner beyond the minimum small talk typical in a social science survey, which may increase the respondents' trust in the enumerator, as well as their trust in the fact that protocol will be followed. Third, reducing the amount of personally identifying information collected in the survey (Low PII), including the name of workers' direct supervisor and their production team, which may alleviate the concern that leaked data could be traced back to the respondent.

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; 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 is to use the collected survey data to assess several policy-relevant aspects of harassment: How prevalent is it? What share of managers is responsible for the bulk of the misbehavior?[1] How isolated are 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 investigate and 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 a Bangladeshi apparel producer to conduct phone-based surveys with workers at two of its plants. We surveyed 2,245 workers and had a response rate of 63%.[2] 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 team-level 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: 9.9% of respondents report threatening behavior, 1.52% report physical harassment, and 1.78% report sexual harassment. HG increased reporting of threatening behavior by 46%, sexual harassment by about 269%, and physical harassment by 288%. We also find that low PII and RB had positive but weak effects. There is suggestive evidence of complementarity between treatment arms; that is, combining hard garbling with rapport-building and low PII increases reporting compared to the sum of the effects of implementing each feature alone.

[1] In the context that we study, harassment by managers perpetrated against workers is the primary concern. Section 2 provides more information on the context.

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

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 lower for sexual harassment. The effects of HG were substantially larger for men compared to women for both threatening behavior and sexual harassment, although for sexual harassment, we lack power to detect the statistical differences between the effects for men and women.

Next, we 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 different 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 reports are set to automated "yeses," either in the overall sample or per team. Blocked HG, in particular at the team-level, substantially reduces the variances of estimators.[3]

Using data from treatment arms including both HG and PII, we estimate that 13.6% reported threatening behavior, 5.7% reported physical harassment, and 7.7% reported sexual harassment. On average, there are 7 workers per production team in HG/PII arms. Considering teams of this size, we find that 59% of teams had at least one worker who had been threatened, just over 38% of teams had at least one who had been sexually harassed, and 27% had at least one who had been physically harassed. These statistics indicate that harassment is widespread in this organization, and a policy of firing all misbehaving supervisors is unlikely to be feasible. Conditional on a type of harassment, victims tend to be isolated, and more so for graver types of harassment. The probability of having at least two victims on the team, conditional on having at least one are respectively 37% for threatening behavior, 20% for physical harassment, and 24% for sexual harassment. These results shed light on the implications of setting different burdens of proof for harassment. In contexts where victims are isolated, requiring multiple victims to come forward, for example, to avoid "he said, she said" situations, will miss the majority of cases; eradicating harassment requires organizations to have actions available that can be taken in cases when only one victim comes forward.

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

This paper 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. Adams-Prassl et al. (2022) document that experiencing harassment leads to adverse employment outcomes for victims and perpetrators and Folke and Rickne (2022) show that sexual harassment contributes to gender inequality in the labor market. We contribute evidence that lack of plausible deniability causally negatively affects 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: it may be that reporting is most suppressed in workplaces where harassment is most problematic.

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).[4] 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.[5] 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.

[4]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).

[5]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.

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 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, Ortner and Chassang, 2018). 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 a retaliation strategy. They show that garbled intervention policies are needed to discipline their behavior. They also clarify how to experimentally 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. The large experimental effect of HG on information transmission in our setting suggests that this class of mechanisms deserves further exploration in other environments 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). Warner (1965) proposed randomized response (RR) as a way to offer survey respondents a form of plausible deniability when answering sensitive questions. Under RR, the surveyor instructs respondents to roll a dice, and answer the question truthfully or not depending on the outcome. For instance, a respondent may be instructed to submit the response "Yes" if the dice lands on 1 or 2, and to answer the question "Have you experienced harassment?" truthfully if the dice lands on 3-6. The surveyor does not observe the respondent's dice roll. RR admits several variants, which we discuss later in the paper. Provided that people comply with the instructions of the surveyor, RR offers plausible deniability: a recorded response "Yes" may be due to the fact that the dice landed on 1 or 2. The empirical literature on survey design for sensitive questions has found that RR performs better than DE, at least in single shot, large scale surveys (Rosenfeld et al., 2016).

We argue that RR and related designs, such as list experiments (LE), are poorly suited for ongoing use in organizations. Because the randomization is entirely under the control of the respondent, respondents can freely ignore instructions to randomly respond "Yes" if they are worried about retaliation. In equilibrium, this causes plausible deniability to unravel

altogether. This concern is empirically validated by the work of Chuang et al. (2020): survey respondents often do not comply with the protocol to garble, and systematically provide the least sensitive response. Because the garbling in RR relies on the compliance of respondents, we refer to mechanisms in this class as soft garbling. Instead, in our design, responses are mechanically switched at an exogenous rate. This is why we refer to our design as hard garbling. Chassang and Zehnder (2019) show that in contrast to RR, the value of HG does not unravel in equilibrium. For this reason, we believe it is better suited for ongoing use in organizations. Our analysis makes two additional contributions. First, we derive consistent estimators of team-level statistics of intended responses using garbled data, extending the estimator of population-level reporting rates proposed by Warner (1965). Second, we show that using sequences of garbling errors that satisfy a small law of large numbers – i.e., blocking – considerably improves inference. This is especially important when baseline reporting rates are low so that sampling error can dwarf the statistic of interest.

The remainder of the 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 provides a simple theoretical framework that clarifies incentives for information transmission under various designs. We explicitly discuss the pros and cons of HG vs. RR or LE, and provide estimators for team-level statistics based on garbled reports. 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. Section 7 discusses our findings and concludes.

2 Context

We conducted this research in collaboration with a large apparel producer in Bangladesh, employing upwards of 25,000 workers in roughly half a dozen factories.[6] The manufacturer'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. For the purpose of the experiment, we agreed to survey workers at 2 of its plants. In the longer-term, the senior management team's goal was to set-up a reporting system for

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

workers to provide continuous feedback in real-time.

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 for managers with power over workers to harass them (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.

The manufacturer's operations are representative of garment manufacturing in Bangladesh. 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: teams of workers are typically overseen by 2 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 considerably 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 more mixed. In contrast, more than 90% of managers in all sections are men.

Within the two plants that were surveyed, 34-42% of workers are employed on sewing lines, 16-18% are employed in finishing, and 10-14% are employed in washing. The remaining workers are employed in smaller, supporting production sections. 93% of managers are men.

3 Framework

Our objective is to collect policy-relevant statistics of harassment. Some statistics, such as the share of victimized workers, do not require collecting information about workers' teams (i.e., their team id). In contrast, statistics associated with team-level patterns do: for instance, assessing whether victimized workers are isolated or assessing the number of managers engaging in misbehavior. Using a principal-agent-monitor framework, we show

how the gap between true statistics of harassment and their counterparts based on intended reports can be affected by different survey procedures.

In addition, we show how to infer statistics of intended reports based on garbled reports alone, and how different garbling structures can affect statistical power. Finally, we clarify the pros and cons of using different versions of HG instead of common alternatives, such as RR and LE.

3.1 Policy-relevant statistics of harassment

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. Throughout the paper, we take as given the behavior of managers, and we seek to elicit information about patterns of harassment $(h_a)_{a \in M}$ in the organization.

We are interested in identifying four statistics helpful in assessing policy options. We emphasize that these statistics are not directly computable because they depend on harassment patterns that are not directly observed by the decision-maker. We discuss workers' decisions to report harassment below. 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 decision-makers to gauge the magnitude of the harassment problem in their organization, which allows stake-

holders to correctly prioritize the issue and to allocate suitable resources.

Statistic S_{PM} measures the share of problematic managers, in other words, managers who have harassed at least one person. It is a special case of statistic $\mathsf{S}_{TV\geq k}$, which measures the share of managers who have harassed at least k workers, for $k = 1$. The behavior of $\mathsf{S}_{TV\geq k}$ as k increases clarifies policy options. For example, if there exists a 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 investigating and firing repeat 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 it is not possible to address a significant number of cases by firing a small number 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 decision-makers 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 such cases. In addition, rules limiting investigations to cases where multiple victims come forward would lead the organization to ignore the majority of problem cases. 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.

Sensitivity of statistics. These statistics differ in the sensitivity of information required to compute them. It is not necessary to know a particular worker's team to compute S_V. 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. We will return to this consideration in our discussion of workers' decision to report harassment and the design of our survey experiment.

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.[7]

3.2 A worker's reporting decision

Because the actual harassment status $h_{i,a}$ of workers is typically not observed, employers must proxy the true statistics of interest with reported harassment. One difficulty is that victims are often unwilling to come forward. This may be due to 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 a set-up in which worker i in team a completes a binary survey, meaning that they can submit an intended response $r_{i,a} \in \{0, 1\}$. In our setting, rates of reported harassment 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 false positives, as well as equilibrium responses to garbling by managers, in Section 7. We argue using the framework of Chassang and Padró i Miquel (2018) that getting people to complain is a necessary first step, even if false positives become an issue.

Following Chassang and Padró i Miquel (2018) and Chassang and Zehnder (2019), we consider 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 $\widetilde{r}$ distributed according to $\phi(r) \in \Delta(\{0, 1\})$.

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.

[7] We note that witnesses are exposed to the same retaliation risk as victims, and may derive lower personal benefits than victims from informing about a problem manager who has not harassed them directly.

- *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 always recorded, but reports of 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 of individual reports exist only in the mind of respondents. 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.[8] Leaks are inevitable even under ideal governance.

Worker i's utility U_i associated with an intended report r depends on their true harassment status and consists of direct benefits from reporting as well as potential reputational and/or retaliation costs:

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

where:

- PB is a psychological benefit from taking action such that $\mathsf{PB}(1|1) > 0$ and for simplicity $\mathsf{PB}(1|0) = \mathsf{PB}(0|1) = \mathsf{PB}(0|0) = 0$. Respondents only derive psychological benefits from taking action against a misbehaving manager.

- SB is a social benefit from realized report $\widetilde{r}$ as perceived by the worker, either because 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 $\mathsf{SB}(0|1) = \mathsf{SB}(0|0) = 0$.[9] Recorded complaints only yield social benefits if they are associated with a misbehaving manager.

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

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

- $\mathsf{RC}(\widetilde{r})$ is a reputational and/or retaliation 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 strictly increasing function; $\text{prob}(r_{i,a} = 1|\widetilde{r}_{i,a})$ is the posterior belief that worker i intended to submit a complaint $r_{i,a} = 1$ about manager a, conditional on recorded report $\widetilde{r}_{i,a} = 1$.[10]

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

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

In turn, a harassed worker's payoffs are

$$U_i(1|1) = \mathsf{PB}(1|1) + \mathsf{SB}(1|1) - p \times K(\text{prob}(\text{r}_{\text{i,a}} = 1|\widetilde{\text{r}}_{\text{i,a}} = 1))$$
$$U_i(0|1) = \pi \times [\mathsf{SB}(1|1) - p \times K(\text{prob}(\text{r}_{\text{i,a}} = 1|\widetilde{\text{r}}_{\text{i,a}} = 1))]\,.$$

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

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

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

$$\text{prob}(r_{i,a} = 1|\widetilde{r}_{i,a} = 1) = \frac{1}{1 + \pi \frac{\text{prob}(r_{i,a}=0)}{1-\text{prob}(r_{i,a}=0)}}. \tag{2}$$

Taking as given the share of null reports $\text{prob}(r_{i,a} = 0)$, it follows from (1) and (2) that increasing π increases a worker's propensity to send report $r_{i,a} = 1$. Equation (2) captures the fact that increasing π reduces the impact of a positive recorded report $\widetilde{r}_{i,a}$ in terms of the reputational and/or retaliation cost. In addition, coefficient $(1-\pi)$ in (1) captures the fact that increasing π shrinks the reputational and/or retaliation cost savings associated with sending a null report $r_{i,a} = 0$. As a result, the left-hand side of (1) exhibits single crossing in the garbling rate π: whenever its value is negative, it is increasing in π.

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

[10] This functional form captures concerns over ex post retaliation by managers and related career concerns.

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

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

(iii) equilibrium reporting weakly decreases with 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*.*

The value of survey design. Proposition 1 suggests two approaches to encourage reporting through survey design. First, increase the garbling rate π. Second, reduce the worker's subjective probability p of a leak. In our survey design, we aim to do this in two ways. First, we vary the elicitation of team identifiers that are needed to compute team statistics, and second, we vary the extent of rapport built with enumerators prior to the sensitive module. Removing team identifiers 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 worker's trust that survey enumerators are trustworthy, 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.

3.3 Measurement

Corollary 1 argues that garbling reduces bias in the measurement of policy-relevant statistics based on intended reports. However, intended reports are not directly available to the analyst when garbling is used. We now discuss how to infer S^r_V, S^r_{PM}, and $\mathsf{S}^r_{TV\geq k}$ from garbled reporting data under different garbling schemes.

For simplicity, we state identification results under the assumption that team size is constant. In practice, team size varies, and we use a likelihood framework to draw inferences.

Inference from garbled reports. A key insight of Warner (1965) is that the share of workers reporting harassment S_V^r can be consistently estimated from garbled data, even though it depends on intended reports. The following estimator is consistent

$$\mathsf{S}_V^{\widetilde{r}} \equiv \frac{\frac{1}{n}\sum_{a,i\in M\times I}\widetilde{r}_{i,a} - \pi}{1-\pi}. \tag{3}$$

It turns out that the same is true for other statistics of intended reports, such as $\mathsf{S}_{TV\geq k}^{\widetilde{r}}$, but the precision of estimators depends on the specific garbling scheme used, and some trade-offs may have to be made.

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:

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

We are interested in recovering μ from $\widetilde{\mu}$. Let us express garbled reports as

$$\widetilde{r}_{i,a} = r_{i,a} + (1 - r_{i,a})\eta_{i,a} \tag{4}$$

where $\eta_{i,a} \in \{0,1\}$ is a Bernoulli random variable equal to 1 with probability π. As we discuss below, the correlation structure across shocks $\eta_{i,a}$ will turn out to matter for power. We distinguish three cases:

- i.i.d. garbling, in which $(\eta_{i,a})_{i\in I,a\in M}$ are i.i.d. across i, a.
- population-blocked garbling, in which errors $(\eta_{i,a})_{i\in I,a\in M}$ are exchangeable across workers in the population and $\sum_{i\in I,a\in M}\eta_{i,a} = \pi n$. By exchangeable across workers in the population, we mean that the distribution of $(\eta_{i,a})_{i\in I,a\in M}$ is unchanged by permutations of labels (i,a), and a fraction of messages exactly equal to π is automatically switched to a complaint.
- team-blocked garbling, in which, for every team a, errors $(\eta_{i,a})_{i\in I}$ are exchangeable across workers in team a, and satisfy $\sum_{i\in I}\eta_{i,a} = \pi L$.[11]

[11] In settings with varying team size, or if πL is not an integer, the number of garbled reports by team may vary. In that case, inference with blocked garbling is equivalent to inference with known team-level numbers of reports assigned to be garbled. This is the informational setting under which we perform inference in Section 6. Note that blocking is performed ex ante, independently of participants' intended response.

Proposition 2 (identification)**.** *Under both i.i.d. and population-blocked garbling, the sample distribution of intended reports μ is identified from the sample distribution of recorded reports $\widetilde{\mu}$ as the number of teams m grows large.*

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 $\mathsf{S}^r_{TV\geq k}$, all of which are functions of distribution μ, under i.i.d. and population-blocked garbling.

Identification does not always hold under team-blocked garbling. The reason for this is that μ admits L degrees of freedom, while $\widetilde{\mu}$ admits only $L - \pi L$ degrees of freedom under team-blocked garbling: mechanically, $\widetilde{\mu}(0) = \cdots = \widetilde{\mu}(\pi L - 1) = 0$. In words, team-blocked garbling forces each team to have a minimum of πL recorded reports of complaints, so there is zero probability of observing team profiles with $\pi L - 1$ or fewer recorded reports of complaints. Hence, πL additional restrictions are needed. This is a drawback of using team-blocked garbling. At the same time, this issue can be addressed by what we believe are very reasonable assumptions. Further, as we discuss below, team-blocked garbling considerably improves the precision of inferences, especially when reporting rates are very low.

Proposition 2′ (Appendix B) provides a sufficient condition for μ to be identified from $\widetilde{\mu}$, which is that this is true whenever $\widetilde{\mu}(L) = 0$; i.e. no team has L realized reports of harassment, even as the number of teams m gets large. This implies that $\mu(L) = \mu(L-1) = \cdots = \mu(L - \pi L) = 0$. In words, if we assume that there is zero probability of observing teams with profiles of all recorded reports of complaints, it implies that there are at least $\pi L + 1$ intended reports of "no" on all teams. This reduces the dimensionality of μ because there is zero probability that there are team profiles with $L - \pi L$ or more intended reports of complaints. This assumption may be likely to hold in contexts in which team size is large and there is reason to believe that not all members of a team have been affected.

Another approach is to assume that reporting data is generated by a small-dimensional data-generating process (DGP). A simple and plausible approach is to consider a DGP corresponding to conditionally i.i.d. harassment with three types of managers. This likelihood-based approach has the advantage of extending naturally to data in which team size varies, and it lets us extrapolate statistics of interest to teams of different sizes.

Conditionally i.i.d. harassment with three types (CIH). 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. We assume that $\rho_L = 0$ and $\rho_M \leq \rho_H$. In words, the low-type managers (θ_L) do not harass the workers under their span of control, the intermediate type (θ_M) may have a non-zero probability of harassing workers under their span of control, but this probability is weakly less than for the high-harassment type managers (θ_H). This DGP is entirely specified by the 4 dimensional vector $\gamma = (q_M, q_H, \rho_M, \rho_H)$.

Provided that the observable data exhibit enough degrees of freedom, i.e., that $L - \pi L \geq 4$, the true parameter γ is identified from the distribution of observable data $\widetilde{\mathbf{r}}$, even under team-blocked garbling. Appendix B.2 derives the associated likelihood functions.

An advantage of this approach is that it provides a very intuitive classification of managers: low concern, medium concern, and high concern. This makes it easy to quantify the trade-offs of different policy responses: for example, what would be the cost and impact of firing all high concern managers? In addition, this DGP lets us aggregate reporting data from team of different sizes, whereas Propositions 2 and 2′ operate at a given team size.

Blocked garbling improves precision. The reason team-blocked garbling is of interest, although it requires additional assumptions for identification, is that it can considerably increase precision. This is especially useful when underlying reporting rates are low. The intuition for this is made especially clear by comparing the precision of the estimator $\mathsf{S}_V^{\widetilde{r}}$, defined in (3), under i.i.d. and either population- or team-blocked garbling.

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

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

Take as given a vector of intended reports, $\mathbf{r}$, and denote by $\overline{r}$ its sample mean. When garbling terms η_i are i.i.d. across workers, then the variance of the sum of garbled reports is

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

When the average reporting rate $\overline{r}$ is small, most of the variance is due to sampling error in

aggregate garbling term A (its variance is $\pi(1-\pi)n$).

For this reason, whenever the mean reporting rate $\overline{r}$ is small, blocked garbling lowers the variance of $\sum_{j=1}^{n} \widetilde{r}_j$: term A is a constant so that the only remaining uncertainty is assigned to term B. For instance, under population-blocked garbling $\mathsf{Cov}(\eta_j, \eta_{j'}) = -\frac{\pi(1-\pi)}{n-1}$ and

$$\mathsf{Var}\left(\sum_{j=1}^{n} \widetilde{r}_j \middle| \mathbf{r}\right) = \mathsf{Var}\left(\sum_{j=1}^{n} r_j \eta_j \middle| \mathbf{r}\right) = \overline{r}\left(1 - \frac{\overline{r}n - 1}{n-1}\right)\pi(1-\pi)n.^{12} \tag{5}$$

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

While population-blocking is enough to improve the estimation of the mean number of victims per team, it does not improve the estimation of other team statistics of interest. This is why blocking at the team level is valuable. We illustrate this point using simulations in the paper's Supplementary Materials.

In our application, we attempted to implement team-blocked garbling by ensuring that 2 out of every 10 consecutive reports were recorded as complaints (i.e. $\pi = 2/10$). Because team sizes are not multiples of 10, we do not achieve exact team-blocking. This leads us to track the number of reports actually assigned to be garbled at the team-level and to perform inference given this data.

We note that in settings where retaliation and leakages (through legitimate action, or malicious channels) are an especially high concern, then i.i.d. garbling may be preferred to blocked (or known team-level) garbling. In our context, where leakages are a subjective concern, and where we are especially interested in learning about team level statistics, the benefits of blocked (or known team level) garbling outweighed its costs.

3.4 Alternative indirect survey response methods

Starting with the pioneering work of Warner (1965) on randomized response (RR), many indirect response methods have been developed to guarantee survey respondents plausible deniability, including the unrelated question approach (Greenberg et al., 1969), list experiments (LEs, Raghavarao and Federer (1979)), and most recently, crosswise RR methods (Blair et al., 2015). We use HG instead of these alternatives for several reasons.

A common feature of these alternative approaches, including RR and LE, is that the

[12] See Appendix A for a proof of (5).

respondent has full control of the report being sent, and plausible deniability occurs only if respondents comply with the instructions provided by the survey designer. For this reason, we refer to these approaches as soft garbling approaches: the garbling of responses only occurs when people comply with instructions.

As Tourangeau and Yan (2007) and Chuang et al. (2020) highlight, a difficulty with these approaches is that respondents frequently do not comply with instructions, and they selectively deviate from following instructions when they are supposed to submit a more sensitive answer. Chuang et al. (2020) propose tests of non-compliance and use them to show that non-compliance is large and problematic in both RR and LEs.

This has two implications. First, it is not possible to use standard inference formulas to recover intended response rates. Because the number of randomly induced sensitive responses is not known, one cannot reliably normalize the number of recorded sensitive responses to obtain estimators of intended response rates as in (3). Second, as Chassang and Zehnder (2019) highlight in a laboratory setting, over time, lack of compliance with garbling instructions means that the plausible deniability associated with indirect response methods unravels. This is especially important in an organizational setting, in which participants repeatedly interact with the survey method and learn to play in equilibrium. For instance, under list experiments, if providing a higher number incriminates one's manager, then employees may be told to systematically agree with the smallest plausible number of statements.

HG addresses both concerns. Because the garbling is performed by the survey tool, the nature of the noise is known, permitting inference. Second, plausible deniability does not unravel even if all agents submit non-sensitive reports: some non-sensitive reports are mechanically switched to sensitive reports. For this reason, and especially in view of the evidence provided by Chuang et al. (2020), we think that HG designs are better suited for steady state monitoring of harassment issues in organizations. Another advantage of HG is to allow for blocked designs that deliver more precise estimates than i.i.d. garbling, both across the total participant population and across team members. This is especially valuable when baseline reporting rates are low and sampling error can dwarf the statistic of interest. In contrast, i.i.d. garbling is the only option under RR.[13]

[13]In terms of survey implementation, HG has the advantage over RR that it does not rely on the availability of a randomization aid such as a die. This consideration becomes relevant when surveys are being conducted remotely, and respondents may not be relied upon to have a randomization 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.

This is not to say that hard garbling does not have drawbacks. Because the garbling is performed by the survey tool, respondents need to have some trust that the survey organization follows the protocol it announces. This is feasible in organizational settings where longer-term relationships allow third-parties to build a reputation with respondents (as is the case in our application), but it may be much more difficult in one shot, large scale surveys where the survey organization does not have high trust within the respondent population. In such settings, although compliance is an issue, RR may be preferred since it allows respondents to be in control of the noise.

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 the firm 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 introduced the survey, established a baseline level of trust, and recruited the prospective respondent. The second call completed the main survey. The third call, two weeks later, conducted a follow-up survey. 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 informed participants that aggregated results would be shared with senior management and would inform HR policy. All survey enumerators for the study were women.[14]

[14]Budget constraints prohibited the research team from randomly assigning the sex of the survey enumerator after stratifying respondents by their sex. 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, than the reverse.

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. Respondents were instructed to answer "Yes" if they had experienced *any* of the actions, without revealing which of the specific actions they had experienced. Ex ante, we hypothesized that threatening behavior would be the least sensitive to report and that sexual harassment would be the most sensitive to report.

Table 1: Harassment definitions

Type of harassment	Examples 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?"

4.2 Treatment Conditions

We randomly assigned survey participants to different combinations of treatment conditions. We varied whether the survey method garbled respondents' intended reports. We varied the

extent to which the survey enumerator built rapport with the surveyed individual. Finally, we varied the level of identifiability of a workers' team and manager. As discussed in Section 3.4, the latter two conditions aim to reduce the worker's subjective probability of a leak p. More specifically, the status quo and alternative treatment conditions were as follows.

Survey method for harassment-related questions:

- Direct elicitation (DE): directly ask the survey respondent about sensitive information.
- Hard garbling (HG): for a yes or no question, where *yes* is the more sensitive answer, exogenously flip *no* answers to *yes* with probability $\pi = 2/10$.

DE is the status quo survey method and the control condition. HG is the treatment condition: it provides respondents with plausible deniability if they submit a sensitive answer. We set the flipping rate to 2/10 and use ensure that 2 out of every 10 consecutive reports, after ordering respondents by production team and gender, are garbled. We explained HG to workers as follows.[15]

> "We are now going to ask you several questions about the way your manager treats you and other employees. For instance: 'Has your manager shouted at you in the last month? Yes or No?'
>
> Each of the questions has a Yes or No answer.
>
> Our system is set up so that it's safe to report an issue.
>
> If you choose to respond YES (there is an issue), our system will record it as a YES for sure. Importantly, if someone responds NO, the system will sometimes record the response as YES.
>
> This means that if you respond YES, we can guarantee that you won't be the only person saying YES. For every 5 responses from workers, at least 1 will be recorded as YES."

Rapport-building (RB):

[15] We implemented the blocked garbling ex ante such that the garbling is not conditional on the respondent's intended response. See Supplementary Materials for implementation details.

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

- 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 prespecified manner, beyond the minimum small talk typical in the standard social science approach.[16] 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.

  The status quo approach 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.

  - 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.

Removing personally-identifying information (Low-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.

[16] 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.

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. Note that identifying respondents' teams is necessary to compute team-level statistics such as the number of manager involved in harassment, the number of victims associated to repeat offenders, and the degree of isolation of victims. This represents an unavoidable trade-off.

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

		No Rapport	Rapport 1	Rapport 2	TOTAL
Direct elicitation	PII	Arm 1 412(476)	Arm 2a 190(225)	Arm 2b 188(229)	790(930)
	Low PII	Arm 3 197(226)	Arm 4 189(220)		386(446)
Hard garbling	PII	Arm 5 416(487)	Arm 6a 188(225)	Arm 6b 195(227)	799(939)
	Low PII		Arm 7 270(305)		270(305)
	Total	1025(1189)	837(975)	383 (456)	2245(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.[17] Appendix Figure C.1 displays the survey modules and the treatment interventions' locations in the survey.

There are small variations across HG treatment arms in the realized garbling rate, as it was blocked within team but not within treatment arm. Consequently, we use the realized garbling rate for each HG treatment arm in the analysis to ensure that differences in the realized garbling rate across treatment arms do not affect the treatment effect estimates.

[17]Questions on COVID-prevention behavior were included after harassment-related survey modules for the purpose of measuring surveyor demand effects. See Section 7.2 for a discussion.

4.3 Sampling and 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).[18] 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 group size suitable for stratified assignment (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.

Based on power calculations, we targeted a sample size of 2,620 workers. Because we had access to the complete population 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 2,245 workers (63% 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.[19] We did not achieve our target sample 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.

During the data quality checking and cleaning process, it became apparent that one of the survey enumerators had not adhered to the protocol for recording respondents' responses in the HG arm.[20] Upon further questioning by the BIGD, it was confirmed that the enumerator understood the HG data entry protocol but had not adhered to it. This enumerator

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

[19] 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 listed for some workers were outdated.

[20] We checked all enumerators for systematic data entry issues, and this was the only enumerator that we identified as problematic.

conducted 53 DE and 48 HG surveys, all of which we drop from the analysis. As a result, the sample size is 2,144 observations for the remainder of the paper.

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. To improve balance, we proceeded along the lines suggested by Banerjee et al. (2020). 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.[21]

Table 3 presents summary statistics of our sample. Appendix Table C.1 presents team-level summary statistics for the teams represented in the survey. Appendix Table C.2 shows balance tests for workers' characteristics across the main treatment conditions. Appendix Table C.3 presents balance tests for workers' characteristics separately across no rapport, short rapport, and long rapport treatment arms. Among 48 tests, there are no statistically significant differences across treatment conditions.

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 Low PII treatments are substitutes or complements. Finally, we conduct robustness checks for our results.

[21] We used 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	5.15	17	23	26	30	55
Experience (yrs)	5.19	3.57	0	2.83	4.42	7.17	28.8
Tenure (yrs)	2.89	2.43	0.052	0.65	2.82	4.17	17.0
Tenure in Team (yrs)† [n=1516]	2.57	2.52	0	0.50	1.83	3.92	14.5
Years of Education	6.70	3.39	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	0.17	0.38	0	0	0	0	1
Position: Ironing/Folding	0.086	0.28	0	0	0	0	1
Position: Operator	0.60	0.49	0	0	1	1	1
Position: Packer	0.044	0.20	0	0	0	0	1
Position: Quality	0.097	0.30	0	0	0	0	1

Notes: This table reports summary statistics on workers' characteristics. Unless otherwise noted, the sample includes 2,144 workers who participated in our survey. †This variable is available for the 1516 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.

5.1 Specifications

We aim to estimate coefficients in the following regression:

$$r_{is} = \alpha\, HG_i + \beta\, Rapport_i + \gamma\, LowPII_i + \mu_s + \theta X_i + \epsilon_{is} \tag{6}$$

where r_{is} is the intended reporting outcome of interest for individual i in stratum s. HG_i, $Rapport_i$ and $LowPII_i$ are indicators for hard-garbling, rapport, and not asking for team-related identifying information. 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, referred to as PDS going forward).

Identification using garbled responses. For individuals in the HG arms, we observe garbled response $\widetilde{r}_i$ instead of intended response r_i. However, following Blair et al. (2015),

we note that recorded reports can be expressed as

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

where ε_i is a mean-zero error, equal to $1 - \pi$ with probability π and equal to $-\pi$ with probability $1 - \pi$. We defined normalized recorded reports $\hat{r}_i$ by

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

with $\pi = .2$ for the HG group and $\pi = 0$ for the DE group. Normalized report $\hat{r}_i$ is equal to the intended report of interest plus a heteroskedastic error term.

If r_i satisfies (6), then $\hat{r}_i$ satisfies a similar regression (6b) with heteroskedastic, mean-zero errors (conditional on covariates). Consequently, OLS is consistent, and robust standard errors are correct.[22] We estimate the following equation, in which ξ_{is} is now the residual, and report robust SEs:

$$\hat{r}_{is} = \alpha\, HG_i + \beta\, Rapport_i + \gamma\, LowPII_i + \mu_s + \theta X_i + \xi_{is} \tag{6b}$$

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

$$\begin{aligned}\hat{r}_{is} = {} & \alpha_f\, HG_i * Female_i + \alpha_m\, HG_i * Male_i + \beta_f\, Rapport_i * Female_i + \beta_m\, Rapport_i * Male_i \\ & + \gamma_f\, LowPII_i * Female_i + \gamma_m\, LowPII_i * Male_i + \lambda\, Female_i + \mu_s + \theta X_i + \xi_{is}\end{aligned} \tag{7}$$

Complementarity across treatments. We test for complementarity vs. substitutability across treatments by estimating the effects separately for each arm. The omitted category

[22] In the case of blocked-garbling, error terms ε_i are negatively correlated within blocks, and uncorrelated across. Because of negatively correlated errors, standard errors clustered at the block level turn out to be less conservative than heteroskedastic standard errors for the HG treatment indicators. We report clustered standard errors in Appendix Table C.4, where we cluster by HG block for HG respondents and by respondent for DE respondents. Mean estimates are unchanged, and standard errors are reduced for the HG treatment indicators.

is $\mathbb{1}(\text{DE} \times \text{PII} \times \text{No RB})_i = 1$, which is treatment arm 1, the control condition.

$$\begin{aligned}\hat{r}_{is} = {} & \alpha_1 \mathbb{1}(\text{DE} \times \text{PII} \times \text{RB 1})_i + \alpha_2 \mathbb{1}(\text{DE} \times \text{PII} \times \text{RB 2})_i + \alpha_3 \mathbb{1}(\text{DE} \times \text{Low PII} \times \text{RB 1})_i \\ & + \alpha_4 \mathbb{1}(\text{DE} \times \text{Low PII} \times \text{No RB})_i + \beta_1 \mathbb{1}(\text{HG} \times \text{PII} \times \text{No RB})_i \\ & + \beta_2 \mathbb{1}(\text{HG} \times \text{PII} \times \text{RB 1})_i + \beta_3 \mathbb{1}(\text{HG} \times \text{PII} \times \text{RB 2})_i + \beta_4 \mathbb{1}(\text{HG} \times \text{Low PII} \times \text{RB 1})_i \\ & + \mu_s + \theta X_i + \xi_{is} \end{aligned} \tag{8}$$

5.2 Results

Main effects of survey design on reporting. Table 4 reports the main treatment effects.[23] In regression tables throughout the paper, odd-numbered columns display the results from the baseline specification, while even-numbered columns display the results with PDS lasso-selected controls.

In the control arm, (DE × PII × No RB), 9.9% of workers report experiencing threatening behavior, 1.52% report being physically harassed, and 1.78% report being sexually harassed by their supervisor. Among workers who report being harassed under DE, meaning respondents in arms 1-5, 43% who experienced threatening behavior reported it through one of their factory's internal channels, 52% who were physically harassed reported it, and 68% of those who were sexually harassed did. Based on the mean reporting rates for arms 1-5, from the producer's perspective, it would have detected that 3.7%, 0.98%, and 1.87% of workers, respectively, experienced threatening behavior, physical harassment, and sexual harassment by their supervisor in the past year.

We now turn to the effect of survey design. In percentage points (ppts), HG increases the reporting of threatening behavior, physical harassment, and sexual harassment, respectively, by 4.5 ppts (or 45.8%, p <0.05), 4.4 ppts (or 288%, p <0.05), and 4.8 ppts (or 269%, p <0.05).

Removing questions about respondents' supervisor (Low PII) increases the reporting of physical harassment by a marginally statistically significant 2.8 ppts (or 184%, p =0.134) but has no detectable effect on the reporting of threatening behavior or sexual harassment.

Building rapport appears to have a positive effect on the reporting of threatening behavior (1.28 ppts, or a 12.9% increase) and sexual harassment (1.88 ppts, or a 106% increase), but

[23] Appendix Table C.5 reports the main results with separate indicator variables for short- and long-RB conditions.

it is not statistically significant. Rapport has no detectable effect on physical harassment.

Table 4: Effects of Survey Design on Reporting of Harassment

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
HG Treatment	0.0453**	0.0453**	0.0437**	0.0437**	0.0478**	0.0478**
	(0.0213)	(0.0206)	(0.0189)	(0.0182)	(0.0193)	(0.0186)
Rapport Treatment	0.0128	0.0128	-0.0094	-0.0094	0.0188	0.0188
	(0.0200)	(0.0194)	(0.0176)	(0.0170)	(0.0182)	(0.0176)
Low PII Treatment	0.0092	0.0092	0.0280	0.0280	0.0045	0.0045
	(0.0227)	(0.0219)	(0.0187)	(0.0181)	(0.0190)	(0.0184)
Control Group Mean	.099	.099	.0152	.0152	.0178	.0178
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Yes	No	Yes	No	Yes
Observations	2141	2141	2141	2141	2141	2141

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.

Together, these results make a compelling case for the importance of plausible deniability in the design of information transmission mechanisms in organizational settings. The costs and benefits of removing team-level identifying questions are much less clear. Low PII appears to increase the reporting of physical harassment, but there is no effect on threatening behavior or 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, we cannot reject that RB has no effect on reporting. It is possible that this null result masks heterogeneous effects across respondents or that RB's effect depends tightly on the 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 is different for men and women, we estimate the main effects separately by sex in Table 5.

In our control arm, 19.12% of men report experiencing threatening behavior, 4.41% report experiencing physical harassment, and 1.47% report experiencing sexual harassment.

Reporting rates among women are very different: 7.98% report threatening behavior, 0.92% report physical harassment, and 1.84% 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.

As in the analysis presented Table 4, HG continues to increase reporting across the board. Interestingly, the point estimates of the effects are particularly large for men, but because our sample of men is small, with the exception of threatening behavior, we cannot reject that the effects are the same for men and women. The impact of removing PII appears to be weakly more positive for women, although standard errors increase, and we cannot reject that the effects for both groups are zero or are the same.

Table 5 suggests that the impact of rapport may be different on men and women. For both threatening behavior and sexual harassment, rapport appears to have increased reporting among women and may have backfired for men. This is plausible: survey enumerators were women, and being forced into small talk with an unknown woman may have raised men's suspicion regarding the survey. This suggests that further experimentation, and better tailoring of the RB treatment, is needed to assess its value.

Interactions among treatment conditions. We examine the possibility that the treatment conditions may substitute or complement each other using regression equation (8). Figure 1 summarizes the results, which are presented in Appendix Table C.6. The omitted category is the control arm, DE $\times$ PII $\times$ No RB.

The top 4 treatment conditions illustrated by Figure 1 correspond to reports elicited using DE. To a first order, removing PII and building rapport do not seem to have a large impact either way, in that setting. There maybe an impact of extended rapport on the reporting of sexual harassment. While this individual coefficient is significant, the overall picture invites caution. Appendix Figure C.2, shows that 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. This suggests that there is an impact of RB on reporting, but that it is subtle, and heterogeneous across participants. In all likelihood, RB needs to be carefully tailored to the survey respondent.

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	0.0283	0.0269	0.0405*	0.0414**	0.0365*	0.0385*
	(0.0234)	(0.0226)	(0.0209)	(0.0201)	(0.0214)	(0.0206)
HG Treatment × Male	0.1230**	0.1216**	0.0597	0.0585	0.0938**	0.0930**
	(0.0514)	(0.0498)	(0.0452)	(0.0437)	(0.0452)	(0.0440)
Rapport × Female	0.0215	0.0229	-0.0173	-0.0162	0.0336*	0.0339*
	(0.0222)	(0.0214)	(0.0195)	(0.0187)	(0.0202)	(0.0194)
Rapport × Male	-0.0250	-0.0250	0.0243	0.0193	-0.0512	-0.0510
	(0.0472)	(0.0454)	(0.0416)	(0.0403)	(0.0423)	(0.0413)
Low PII Treatment × Female	0.0119	0.0109	0.0325	0.0324	0.0097	0.0107
	(0.0252)	(0.0242)	(0.0207)	(0.0198)	(0.0215)	(0.0207)
Low PII Treatment × Male	-0.0067	-0.0089	0.0120	0.0169	-0.0237	-0.0211
	(0.0535)	(0.0517)	(0.0443)	(0.0432)	(0.0417)	(0.0404)
Female	-0.0924	-0.0885	-0.0210	-0.0055	0.0624	0.0800
	(0.1039)	(0.0996)	(0.0777)	(0.0749)	(0.0769)	(0.0740)
Control Mean - Female	.0798	.0798	.0092	.0092	.0184	.0184
Control Mean - Male	.1912	.1912	.0441	.0441	.0147	.0147
p(HGxFemale - HGxMale)	[0.094]	[0.084]	[0.700]	[0.724]	[0.253]	[0.264]
p(RapportxFemale - RapportxMale)	[0.373]	[0.340]	[0.366]	[0.425]	[0.071]	[0.063]
p(NoPIIxFemale - NoPIIxMale)	[0.753]	[0.729]	[0.675]	[0.745]	[0.478]	[0.487]
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Lasso	No	Lasso	No	Lasso
Observations	2141	2141	2141	2141	2141	2141

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.

The bottom 4 treatment conditions illustrated by Figure 1 correspond to reports elicited using hard garbling. They entail the iterative introduction of RB and Low PII. Estimated treatment effects appear to be rising as additional trust-enhancing steps are taken. This contrasts with patterns under DE, where trust-enhancing steps are not accompanied with increasing treatment effects. Altogether, this suggests that there exist complementarities between HG and other steps fostering trust in the survey protocol. This is intuitive since HG can only be effective if respondents trust that protocol will be followed.

This visual intuition corresponds to the fact that the point estimate for the effect of (HG × Low PII × RB 1) is larger than the sum of the point estimates for (DE × PII × RB 1)+ (DE × Low PII × No RB) + (HG × PII × No RB) for all three harassment outcomes. We test the null hypothesis of no complementarity among HG, removing team-level identifying information, and RB in the complementarity test reported at the bottom of Appendix Table C.7, focusing on even-numbered columns, which include PDS-lasso-selected controls). The test is rejected for threatening behavior (p=0.033) but is too imprecise to be rejected for physical harassment (p=0.223) and sexual harassment (p=0.290). We thus interpret this as suggestive evidence of complementarity among the design features.

In Section 7, we discuss the robustness of our findings. We first show that confusion among respondents does not explain the impact of HG on reporting. We then discuss how to interpret our findings if there are concerns over false reporting by workers.

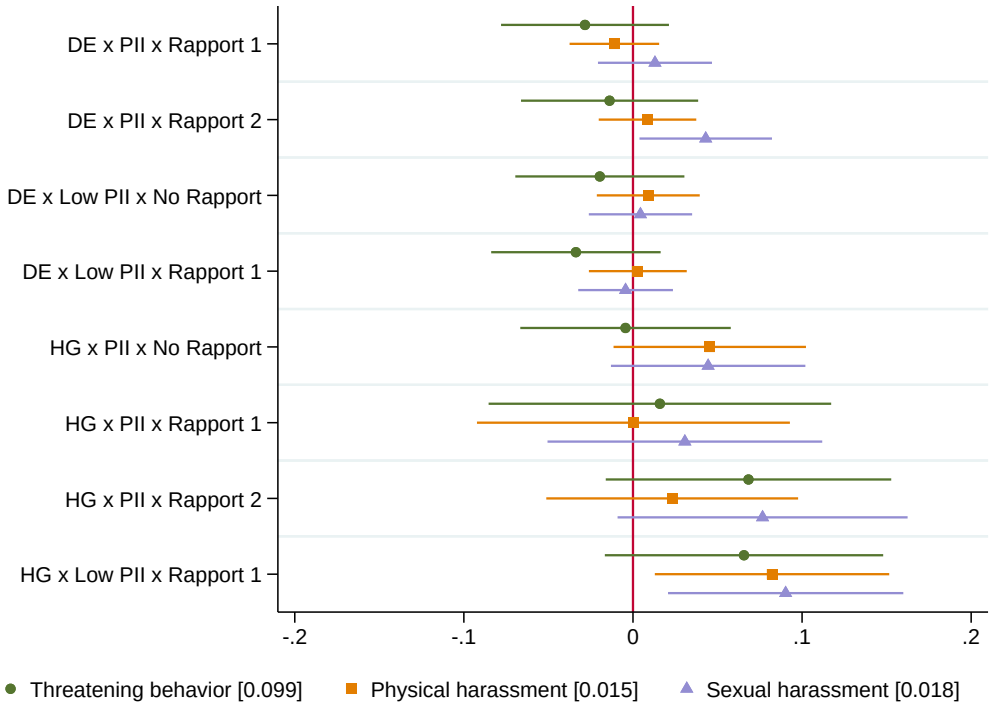
6 Understanding Harassment

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 effect of HG on reporting, we compute team-level statistics using data pooled across treatment arms that use HG and collect PII (when PII are needed).

We begin by describing the patterns of harassment in the organization, and then discuss policy implications for the producer.[24]

[24]One may be concerned that supervisors who engage in more harassment may have pressured workers not to participate the survey. This would not affect the internal validity of the survey experiment results as selection into the sample happens before a respondent knows their treatment assignment. This could affect the external validity of the experiment, though, and 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 C.11, which reports the correlation between the team-level response rate to the survey and the team-level reporting rates for harassment with DE and HG, respectively, as well as the difference between the team-level reporting rates under the two mechanisms. On the whole, the correlations are small or zero, and most are not statistically different from zero, which increases our confidence that our results are externally valid. In the case of threatening behavior, the coefficients are weakly negative, suggesting teams with higher rates of threatening behavior have slightly lower response rates, so our estimates of team-level statistics may be downwardly biased for threatening behavior.

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. 8. 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.

Scale of the issue and potential gains. Figure 2 illustrates the estimated share of victimized workers, computed using (3) under DE and HG. Since this statistic does not require PII, we pool data across all arms.

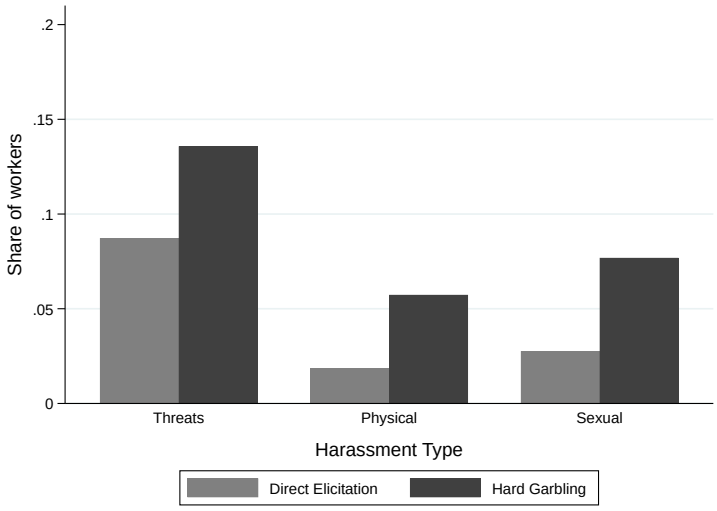
As we have already discussed, HG considerably increases workers' propensity to report harassment: 13.6% reported threatening behavior with HG compared to 8.7% with DE, 5.7% reported physical harassment with HG compared to 1.9% with DE, and 7.7% reported sexual harassment with HG compared to 2.8% with DE.

The primary takeaway is that harassment is meaningfully more widespread than standard surveys, or the firm's internal reporting channels would suggest.[25] This means that

[25]Reporting rates under HG are also higher than the rates that would have been detected by the apparel producer through its reporting channels: 3.7% for threatening behavior, 0.98% for physical harassment, and 1.87% for sexual harassment.

addressing harassment may have a much more positive impact on overall employee welfare than what previously available data would lead one to conclude. We also note that since both men and women report significant levels of harassment under HG, addressing harassment would likely benefit both groups.

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



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.

Problem managers and isolated victims. We now turn to team-level characteristics of interest, $\mathsf{S}_{TV\geq k}$ and $\mathsf{E}_{2V|1V}$, computed by pooling data from treatment arms that use HG and collect PII. Because our data exhibits varying team sizes and varying numbers of garbled reports per team, it is most easily investigated through the conditionally i.i.d. harassment DGP introduced in Section 3.3. As a reminder, we assume that managers fall in one of three types L, M, or H. Conditionally on type, a manager's decisions to harass are independent across workers under their span of control. We denote by $\rho_L \equiv 0 < \rho_M < \rho_H$ the respective probability with which a manager of a given type harasses a worker, and by q_L, q_M, and q_H the respective share of each type in the population of managers. We estimate parameters ρ_M, ρ_H, q_M, q_H and their standard errors by computing a posterior distribution over parameters (starting from a uniform prior over feasible parameters).

We then map this posterior distribution over DGPs to a posterior distribution over the statistics of interest, $\mathsf{S}_{TV\geq k}$ and $\mathsf{E}_{2V|1V}$, for teams of size 7.[26]

Table 6: Posterior estimates of supervisor types, shares, and harassment rates

Parameter	Threatening Behavior (1)	Physical Harassment (2)	Sexual Harassment (3)
ρ_L	0	0	0
	–	–	–
ρ_M	0.111	0.051	0.075
	(0.028)	(0.024)	(0.026)
ρ_H	0.240	0.164	0.180
	(0.174)	(0.181)	(0.154)
q_L	0.051	0.266	0.128
	(0.045)	(0.159)	(0.096)
q_M	0.593	0.468	0.558
	(0.317)	(0.258)	(0.289)
q_H	0.356	0.275	0.314
	(0.316)	(0.242)	(0.283)

Table 6 reports mean parameters for the posterior distribution over DGPs, Table 7 reports team-level statistics S_V, $\mathsf{S}_{TV\geq k}$ for $k \in 1, 2$, and $\mathsf{E}_{2V|1V}$, and Figure 3 illustrates the full distribution of $\mathsf{S}_{TV\geq k}$ for each type of harassment. We note that although the standard errors over estimated DGP parameters are large, the standard errors over implied team-level statistics are small.

The estimated DGP parameters are fairly similar across different types of harassment. Several observations are worth noting. First, supervisors who do not harass any workers (q_L) are a minority, ranging from 5 to 27% depending on the type of harassment. Roughly 50% of supervisors are estimated to harass workers at an intermediate rate. These supervisors harass a worker with probability 5% for physical harassment, 7.5% for sexual harassment, and 11% for threatening behavior. Finally, roughly a third of managers are estimated to

[26]This corresponds to the median number of team-members included in HG/PII treatment arms. Appendix C includes two robustness checks. Table C.12 reports team-level statistics for teams of size 10 and 15, estimated using the same posterior over DGPs as the one used in this section. Because the family of DGPs we posit may be misspecified, we also report in Table C.13 statistics computed using a non-parametric closed-form estimator (Proposition 2′) under the approximation that all teams have size 7, and exactly 2 reports per team are forced complaints. It is reassuring that the estimates from these two approaches are similar.

harass workers at a high rate (36% for threatening behavior, 27.5% for physical harassment, and 31% for sexual harassment). High type supervisors harass a worker with probability 24% for threatening behavior, 16% for physical harassment, and 18% for sexual harassment.

According to these estimates, for threatening behavior, high types (36% of supervisors) are responsible for 56% of the harassment. For physical harassment, high types (27.5% of supervisors) are responsible for 65% of the harassment, and for sexual harassment, high types (31% of supervisors) are responsible for 57.5% of the harassment. This suggests that there is value in targeting high type offenders, especially for physical harassment. However, a significant share of harassment is performed by intermediate offenders. Addressing harassment likely requires changing the behavior of existing managers, rather than simply getting rid of a few bad apples.

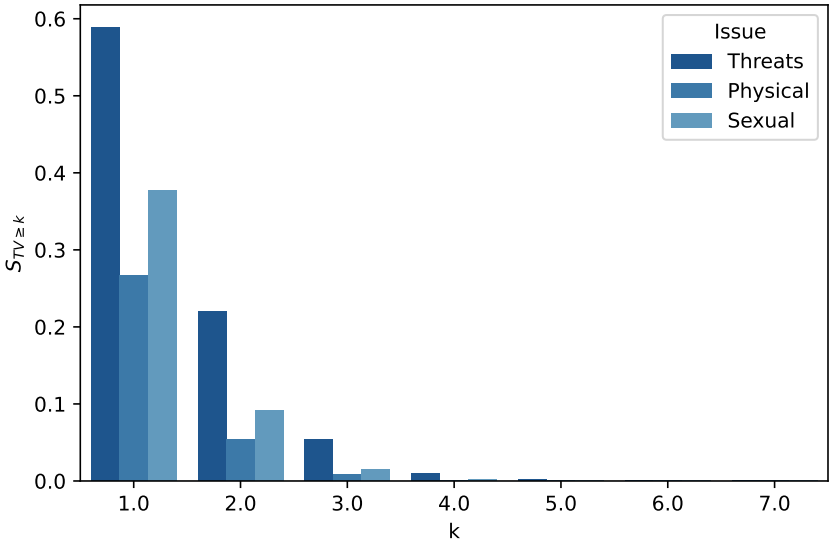
For teams of size 7, the share $\mathsf{S}_{TV\geq 1}$ of problem teams in which at least one worker has been harassed is equal to 59% for threatening behavior, 27% for physical harassment, and 38% for sexual harassment. The share of teams with at least 2 victimized workers is 22% for threats, 5.5% for physical harassment, and 9% for sexual harassment. In turn, as Figure 3 shows, the share of teams with k victimized workers for k greater than 2 is very low, especially for physical and sexual harassment. This confirms that the bulk of the challenge consists in dealing with fairly widespread medium intensity harassment, rather than dealing with a fairly circumscribed group of high intensity offenders: $k\mathsf{S}_{TV\geq k}$ is not high for k large.

Table 7: Team-level statistics – Teams of size 7

Statistic	Threatening Behavior	Physical Harassment	Sexual Harassment
$S_{tv\geq 1}$	0.589	0.266	0.377
	(0.039)	(0.040)	(0.043)
$S_{tv\geq 2}$	0.221	0.055	0.091
	(0.033)	(0.017)	(0.021)
$E_{2V\|1V}$	0.373	0.203	0.240
	(0.035)	(0.047)	(0.037)

Across issues, victims are relatively isolated: conditional on getting a report, the likelihood of getting at least another one is equal to 37% for threats, 20% for physical harassment, and 24% for sexual harassment.

Figure 3: Share of teams with k or more victimized workers – Teams of size 7.



Policy implications. Our findings suggest three policy takeaways. First, harassment is much more prevalent than filed complaints would suggest, and it affects both men and women. Second, harassment occurs at a moderate intensity but is widespread across teams. This means that firing a few bad apples cannot be the sole policy option. Instead, the behavior of existing managers must be changed. Nonetheless, it would be beneficial to prioritize the worst offenders. Third, the extent to which victims are isolated in teams varies substantially by type of harassment. This has implications for setting different burdens of proof for harassment. For types of harassment for which victims are more isolated, requiring multiple victims to come forward, for example, to avoid "he said, she said" situations, may miss the majority of cases; eradicating harassment likely requires having actions available that can be taken in cases when only one victim comes forward.

7 Discussion

7.1 Summary

Our work makes two main contributions. First, 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 real-life organizational context. We then use our improved reporting

data to assess several policy-relevant aspects of harassment.

Our experimental results show that lack of plausible deniability causes severe under-reporting of harassment in this organizational setting. Lack of trust in the integrity of the reporting system may also contribute, though our results suggest that the process of trust-building is highly contextual and may backfire when not well-targeted. In addition, harassment appears to be widespread, a majority of managers exhibit some propensity to harass workers, and victims are frequently isolated.

In the remainder of this section, we address some of the robustness concerns related to our findings and discuss how they might be further explored through replication.

7.2 Robustness

7.2.1 Confusion in the HG condition

One concern with our findings is that HG is a more complicated mechanism than DE. This means that it takes more time to explain HG, which increases survey duration by 4% on average (Appendix Table C.8). 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 6.70 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.[27] Respondents answered these prior to being asked the questions about harassment, and survey enumerators explained the answers to the 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 8.6% of women in HG answer at least 1 question incorrectly ($p = 0.685$), while 6.9% of men and 4.3% of women answer 2 questions incorrectly ($p = 0.133$). 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

[27] 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.

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, and re-estimate our main results, considering all respondents who answer at least 1 comprehension question incorrectly as confused and setting all confused respondents' *recorded* answers to harassment questions equal to "no."

Panel A of Appendix Table C.9 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 3.4 ppt increase in reporting ($p < 0.10$) compared to a 4.5 ppt increase ($p < 0.05$). For physical harassment, the effect is a 3.3 ppt increase ($p < 0.10$) compared to a 4.4 ppt increase ($p < 0.05$). For sexual harassment, the effect is a 3.6 ppt increase ($p < 0.10$) compared to a 4.8 ppt increase ($p < 0.05$). Even under the extremely conservative to set all confused respondents' recorded report to "no" for all questions, the effects of HG are positive, large, and statistically significant.[28] 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.

7.2.2 Strategic misreporting by workers and follow-up actions

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. 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 and 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.

Empirical evidence. We think that this possibility is a priori unlikely in our context given the very low baseline rates of reporting and the stigma that victims face. 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

[28] A less conservative approach would be to exclude these respondents from the analysis, or set their intended response to "no" and simulate out their recorded responses.

used by factories to determine workers' eligibility to become a supervisor.[29] 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 C.10 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 hypothesis that strategic misreporting due to career concerns is driving our results.

Model guidance. Chassang and Padró i Miquel (2018) explicitly study equilibrium whistleblowing in a model where managers make endogenous retaliation choices, and workers may have malicious incentives to submit false accusations. Chassang and Padró i Miquel (2018) show that it is possible to achieve robust bounds on the underlying level of misbehavior by:

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

   In our application, the only action associated with a report of harassment is a change in the aggregate statistic reported to firm's executives.

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

Our paper can be viewed as achieving step (1) in the process described above. Investigating step (2) is a central question for future research.

7.3 Directions for future research

The question of how to scale up enforcement actions taken as a function of reports strikes us as a particularly valuable direction for future research. The choice of an action following a report of harassment is key, and it should be driven by multiple considerations.

First, the action needs to be an acceptable, legitimate response to an inherently noisy signal. For instance, sending the manager to a training seminar, initiating a more thorough yearly review, or moving the worker associated with the report to a new team may be

[29] 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."

appropriate responses to noisy evidence, whereas firing the manager would not be. Note that the need for moderation in responses associated with noisy signals may in fact be a plus from the perspective of the organization's leadership. Stronger evidence may lead to costly repercussions out of the organization's control, leading organizations to avoid information in the first place. Second, some follow up actions are more likely than others to attract the interest of malicious workers. For instance, sending a manager to a training seminar is unlikely to benefit a worker interested in sabotaging a manager's career, whereas linking recorded reports to managers' promotion opportunities would. We think that a natural objective for future research is to experimentally evaluate the impact of different follow up actions to garbled reports.

Another interesting direction is to evaluate the mental health and broader welfare effects of reporting harassment for workers. Sociological research suggests that the act of confiding secrets can improve an individual's well-being through improving one's perceived coping ability and reducing one's mental load associated with the secret (Slepian and Moulton-Tetlock, 2019). To explore this possibility, we resurveyed workers two weeks after the survey experiment to test whether reporting harassment improved workers' mental well-being and job satisfaction. We estimate a 2SLS model with the randomized assignment to the treatments as our instruments for reporting harassment. Appendix D details our empirical strategy and results. It provides suggestive evidence that reporting harassment improves workers' mental well-being and job satisfaction. The effects are large, and consistent across questions, but imprecisely estimated. The effect on job satisfaction also suggests one possible mechanism for beneficial effects to flow to the producer. We think that exploring the welfare and distributional implications of improved governance systems for harassment for workers, managers, and producers – building on recent research on the labor market implications of harassment by Adams-Prassl et al. (2022), Folke and Rickne (2022), and Dahl and Knepper (2021) – is a valuable direction for future research.

References

Adams-Prassl, A., K. Huttunen, E. Nix, and N. Zhang (2022): "Violence Against Women at Work," Tech. rep., mimeo.

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.

Anderson, M. L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481 –1495.

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 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)*.

Ortner, J. and S. Chassang (2018): "Making corruption harder: Asymmetric information, collusion, and crime," *Journal of Political Economy*, 126, 2108–2133.

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.

Raghavarao, D. and W. Federer (1979): "Block Total Response as an Alternative to the Randomized Response Method in Surveys," *Journal of the Royal Statistical Society*, B, 40–45.

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*.

Slepian, M. L. and E. Moulton-Tetlock (2019): "Confiding Secrets and Well-Being," *Social Psychology and Personality Science*, 10, 472–484.

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.

Tourangeau, R. and T. Yan (2007): "Sensitive questions in surveys." *Psychological bulletin*, 133, 859.

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 $\widetilde{\mu}$. The same result holds under population-blocked garbling since the distribution of $\widetilde{\mu}$ conditional on μ under i.i.d. garbling and population-blocked garbling converge as m grows large. ■

Proof of Equation (5).

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

■

B Extensions

B.1 Measurement under alternative frameworks

B.1.1 Blocked garbling

Proposition 2′ (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 2′. 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:

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

In general the system of equations (B.1) is not invertible. However it is invertible whenever $\widetilde{p}_L = 0$. This implies that $p_L = p_{L-1} = p_{L-2} = 0$. As a consequence, 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 define $\widetilde{k}_a \equiv \sum_{i \in I} \widetilde{r}_{i,a}$. Altogether the log likelihood associated with observing data $\widetilde{\mathbf{r}}$ is

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

C Figures & Tables

Figure C.1: Survey Modules & Treatment Interventions

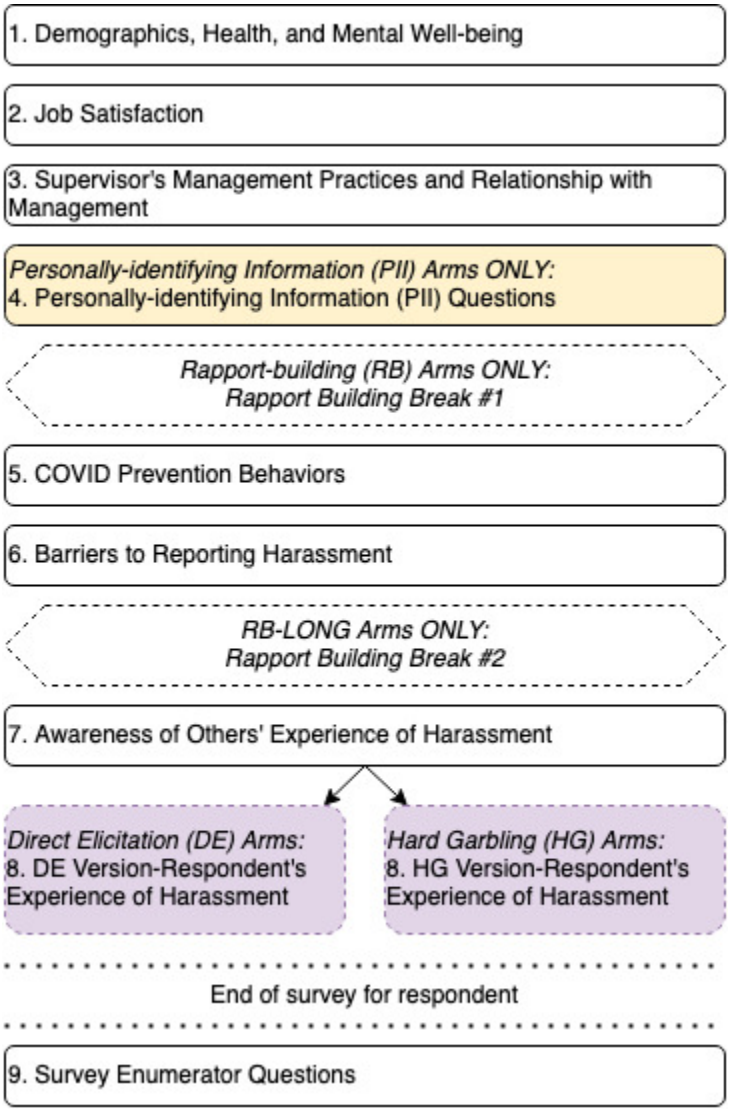
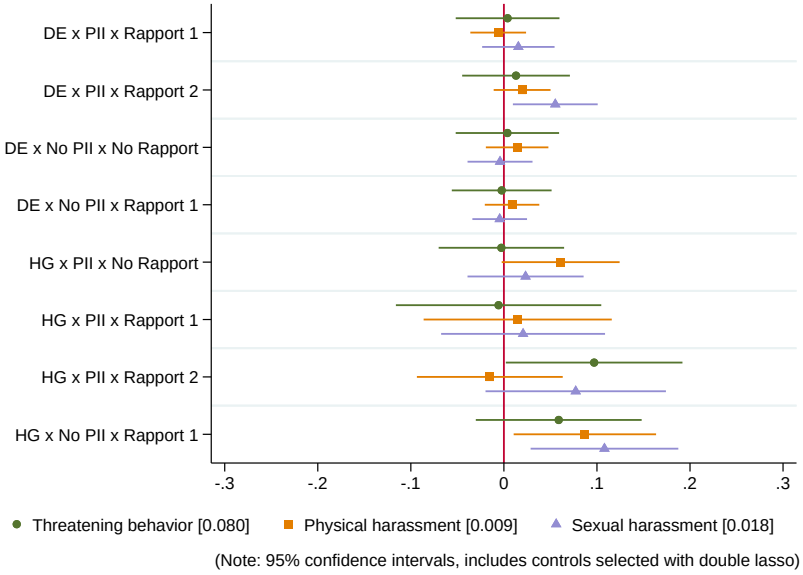


Figure C.2: Treatment effects by survey arm, separately by sex

(a) For Women



(b) For Men

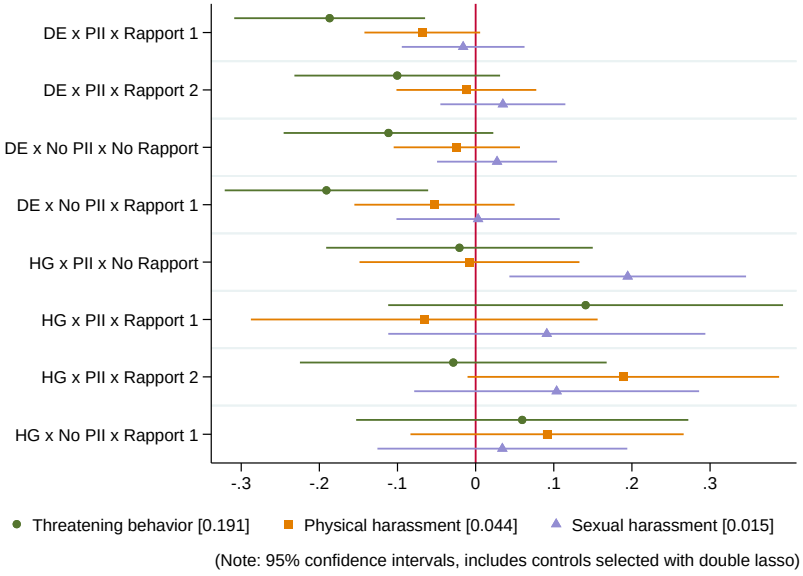


Figure C.3: Team size & gender composition by production section

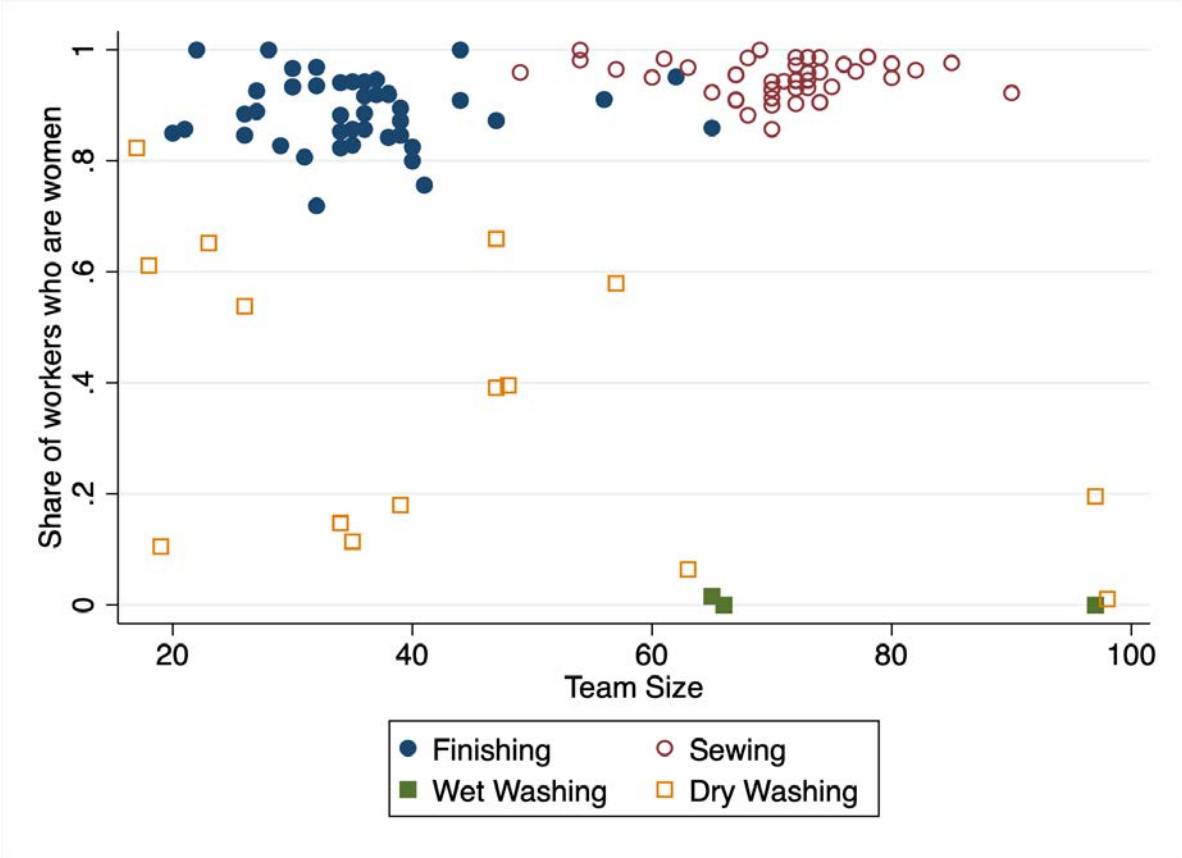


Table C.1: Summary Statistics (Team)

	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 C.2: Balance tests: main treatment conditions

	Mean / (SE)						Difference in means / [p-value]		
Variable	DE	HG	No Rapport	Rapport	PII	Low PII	HG-DE	Diff Rapport	Diff PII
Female	0.811	0.816	0.815	0.812	0.815	0.809	0.007	0.005	-0.001
	(0.392)	(0.388)	(0.388)	(0.391)	(0.388)	(0.393)	[0.152]	[0.280]	[0.855]
Currently Working	0.957	0.961	0.955	0.962	0.960	0.957	0.003	0.005	-0.004
	(0.202)	(0.194)	(0.207)	(0.191)	(0.197)	(0.203)	[0.750]	[0.529]	[0.659]
Age	26.678	26.881	26.662	26.870	26.811	26.686	0.205	0.114	-0.111
	(5.048)	(5.254)	(5.067)	(5.214)	(5.215)	(4.982)	[0.346]	[0.601]	[0.634]
Experience (yrs)	5.170	5.204	5.130	5.234	5.190	5.178	-0.010	0.067	0.010
	(3.632)	(3.510)	(3.535)	(3.607)	(3.591)	(3.536)	[0.946]	[0.646]	[0.951]
Tenure (yrs)	2.879	2.900	2.867	2.907	2.899	2.864	0.035	-0.066	-0.032
	(2.430)	(2.429)	(2.430)	(2.429)	(2.419)	(2.454)	[0.688]	[0.442]	[0.740]
Years of Education	6.760	6.640	6.697	6.708	6.725	6.650	-0.098	0.046	-0.103
	(3.401)	(3.386)	(3.384)	(3.403)	(3.361)	(3.473)	[0.488]	[0.742]	[0.502]
Marital Status (1=Yes)	0.834	0.811	0.824	0.822	0.821	0.830	-0.025	-0.007	0.008
	(0.372)	(0.392)	(0.381)	(0.382)	(0.384)	(0.376)	[0.130]	[0.688]	[0.670]
Children (1=Yes)	0.737	0.744	0.743	0.739	0.739	0.744	0.005	-0.007	0.008
	(0.440)	(0.436)	(0.437)	(0.439)	(0.439)	(0.437)	[0.772]	[0.719]	[0.705]
Observations	1,123	1,021	979	1,165	1,516	628	2,144	2,144	2,144
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table summarizes workers' characteristics in each treatment condition. 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 C.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	0.815	0.820	0.795	0.006	0.003	-0.005
	(0.388)	(0.385)	(0.404)	[0.253]	[0.635]	[0.409]
Currently Working	0.955	0.965	0.956	0.008	-0.000	-0.009
	(0.207)	(0.184)	(0.205)	[0.407]	[0.991]	[0.488]
Age	26.662	26.860	26.891	0.131	0.107	-0.029
	(5.067)	(5.124)	(5.411)	[0.582]	[0.741]	[0.930]
Experience (yrs)	5.130	5.323	5.040	0.168	-0.167	-0.341
	(3.535)	(3.589)	(3.644)	[0.296]	[0.436]	[0.121]
Tenure (yrs)	2.867	2.932	2.854	-0.018	-0.182	-0.115
	(2.430)	(2.419)	(2.452)	[0.849]	[0.132]	[0.354]
Years of Education	6.697	6.683	6.762	0.027	0.112	0.069
	(3.384)	(3.430)	(3.348)	[0.860]	[0.579]	[0.745]
Marital Status (1=Yes)	0.824	0.825	0.817	-0.005	-0.010	-0.007
	(0.381)	(0.380)	(0.387)	[0.802]	[0.676]	[0.787]
Children (1=Yes)	0.743	0.746	0.724	0.002	-0.023	-0.023
	(0.437)	(0.436)	(0.448)	[0.928]	[0.390]	[0.405]
Observations	979	799	366	1,778	1,345	1,165
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table summarizes workers' characteristics in each treatment condition. 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 C.4: Main treatment effects, estimated with standard errors clustered by HG block (HG respondents) or respondent (DE respondents)

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main effects						
HG Treatment	0.0453***	0.0453***	0.0437***	0.0437***	0.0478***	0.0478***
	(0.0149)	(0.0144)	(0.0121)	(0.0117)	(0.0111)	(0.0107)
Rapport Treatment	0.0128	0.0128	-0.0094	-0.0094	0.0188	0.0188
	(0.0205)	(0.0198)	(0.0200)	(0.0193)	(0.0184)	(0.0177)
Low PII Treatment	0.0092	0.0092	0.0280	0.0280	0.0045	0.0045
	(0.0244)	(0.0236)	(0.0184)	(0.0178)	(0.0203)	(0.0196)
Control Group Mean	.099	.099	.0152	.0152	.0178	.0178
Panel B: Heterogeneity by sex						
HG Treatment × Female	0.0283*	0.0269*	0.0405***	0.0414***	0.0365***	0.0385***
	(0.0169)	(0.0163)	(0.0132)	(0.0126)	(0.0131)	(0.0128)
HG Treatment × Male	0.1230***	0.1216***	0.0597*	0.0585*	0.0938***	0.0930***
	(0.0417)	(0.0407)	(0.0347)	(0.0335)	(0.0351)	(0.0342)
Rapport × Female	0.0215	0.0229	-0.0173	-0.0162	0.0336*	0.0339*
	(0.0230)	(0.0220)	(0.0234)	(0.0223)	(0.0204)	(0.0195)
Rapport × Male	-0.0250	-0.0250	0.0243	0.0193	-0.0512	-0.0510
	(0.0467)	(0.0448)	(0.0370)	(0.0360)	(0.0474)	(0.0469)
Low PII Treatment × Female	0.0119	0.0109	0.0325	0.0324	0.0097	0.0107
	(0.0262)	(0.0255)	(0.0207)	(0.0197)	(0.0229)	(0.0223)
Low PII Treatment × Male	-0.0067	-0.0089	0.0120	0.0169	-0.0237	-0.0211
	(0.0550)	(0.0527)	(0.0457)	(0.0451)	(0.0398)	(0.0385)
Female	-0.0924	-0.0885	-0.0210	-0.0055	0.0624	0.0800
	(0.1059)	(0.1018)	(0.0751)	(0.0731)	(0.0779)	(0.0753)
Control Mean - Female	.0798	.0798	.0092	.0092	.0184	.0184
Control Mean - Male	.1912	.1912	.0441	.0441	.0147	.0147
p(HGxFemale - HGxMale)	[0.044]	[0.038]	[0.612]	[0.640]	[0.155]	[0.167]
p(NoPIIxFemale - NoPIIxMale)	[0.750]	[0.725]	[0.690]	[0.759]	[0.455]	[0.461]
p(RapportxFemale - RapportxMale)	[0.376]	[0.338]	[0.350]	[0.413]	[0.107]	[0.101]
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Lasso	No	Lasso	No	Lasso
Observations	2141	2141	2141	2141	2141	2141

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. Standard errors clustered by HG batch (HG respondents) or respondent (DE respondents) are reported in round brackets. *p <0.1; **p <0.05; ***p <0.01.

Table C.5: Effects of Survey Design on Reporting, Differentiating Rapport Treatments

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
HG Treatment	0.0466**	0.0466**	0.0441**	0.0441**	0.0493**	0.0493***
	(0.0212)	(0.0205)	(0.0188)	(0.0181)	(0.0193)	(0.0187)
Low PII Treatment	0.0183	0.0183	0.0308	0.0308	0.0151	0.0151
	(0.0244)	(0.0236)	(0.0204)	(0.0197)	(0.0197)	(0.0191)
Rapport Treatment (Short)	0.0023	0.0023	-0.0126	-0.0126	0.0066	0.0066
	(0.0234)	(0.0226)	(0.0207)	(0.0200)	(0.0199)	(0.0192)
Rapport Treatment (Long)	0.0298	0.0298	-0.0043	-0.0043	0.0385	0.0385
	(0.0289)	(0.0279)	(0.0251)	(0.0242)	(0.0286)	(0.0277)
Control Group Mean	.099	.099	.0152	.0152	.0178	.0178
p(Long − Short Rapport)	[0.403]	[0.387]	[0.773]	[0.766]	[0.302]	[0.286]
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Yes	No	Yes	No	Yes
Observations	2141	2141	2141	2141	2141	2141

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 C.6: 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	-0.0304	-0.0284	-0.0069	-0.0110	0.0156	0.0130
	(0.0257)	(0.0253)	(0.0132)	(0.0135)	(0.0172)	(0.0172)
DE × PII × Rapport 2	-0.0146	-0.0138	0.0091	0.0086	0.0410**	0.0430**
	(0.0275)	(0.0267)	(0.0149)	(0.0147)	(0.0208)	(0.0200)
DE × Low PII × No Rapport	-0.0159	-0.0196	0.0122	0.0091	0.0035	0.0044
	(0.0266)	(0.0255)	(0.0156)	(0.0155)	(0.0154)	(0.0156)
DE × Low PII × Rapport 1	-0.0337	-0.0337	0.0080	0.0029	-0.0022	-0.0044
	(0.0268)	(0.0256)	(0.0153)	(0.0148)	(0.0145)	(0.0143)
HG × PII × No Rapport	-0.0034	-0.0044	0.0487	0.0454	0.0450	0.0445
	(0.0328)	(0.0317)	(0.0300)	(0.0290)	(0.0305)	(0.0293)
HG × PII × Rapport 1	0.0205	0.0159	0.0044	0.0003	0.0306	0.0307
	(0.0532)	(0.0517)	(0.0493)	(0.0472)	(0.0430)	(0.0414)
HG × PII × Rapport 2	0.0661	0.0683	0.0240	0.0232	0.0776*	0.0767*
	(0.0448)	(0.0431)	(0.0392)	(0.0380)	(0.0455)	(0.0438)
HG × Low PII × Rapport 1	0.0653	0.0657	0.0792**	0.0822**	0.0870**	0.0903**
	(0.0437)	(0.0420)	(0.0368)	(0.0354)	(0.0368)	(0.0355)
Control Group Mean	.099	.099	.0152	.0152	.0178	.0178
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Lasso	No	Lasso	No	Lasso
Observations	2141	2141	2141	2141	2141	2141

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 C.7 on next page for p-values.

Table C.7: 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	[0.604]	[0.626]	[0.339]	[0.236]	[0.294]	[0.203]
DExPIIxRB1 − DExNoPIIxNoRB	[0.624]	[0.759]	[0.264]	[0.237]	[0.533]	[0.663]
DExPIIxRB1 − DExNoPIIxRB1	[0.913]	[0.853]	[0.375]	[0.398]	[0.341]	[0.356]
DExPIIxRB1 − HGxPIIxNoRB	[0.446]	[0.485]	[0.069]	[0.056]	[0.369]	[0.320]
DExPIIxRB1 − HGxPIIxRB1	[0.355]	[0.407]	[0.822]	[0.815]	[0.737]	[0.682]
DExPIIxRB1 − HGxPIIxRB2	[0.039]	[0.032]	[0.442]	[0.380]	[0.185]	[0.159]
DExPIIxRB1 − HGxNoPIIxRB1	[0.036]	[0.034]	[0.022]	[0.010]	[0.065]	[0.039]
DExPIIxRB2 − DExNoPIIxNoRB	[0.967]	[0.849]	[0.865]	[0.980]	[0.098]	[0.079]
DExPIIxRB2 − DExNoPIIxRB1	[0.546]	[0.513]	[0.952]	[0.748]	[0.056]	[0.030]
DExPIIxRB2 − HGxPIIxNoRB	[0.762]	[0.792]	[0.208]	[0.224]	[0.910]	[0.966]
DExPIIxRB2 − HGxPIIxRB1	[0.529]	[0.582]	[0.926]	[0.863]	[0.822]	[0.780]
DExPIIxRB2 − HGxPIIxRB2	[0.092]	[0.075]	[0.712]	[0.711]	[0.451]	[0.471]
DExPIIxRB2 − HGxNoPIIxRB1	[0.086]	[0.077]	[0.065]	[0.043]	[0.257]	[0.222]
DExNoPIIxNoRB − DExNoPIIxRB1	[0.554]	[0.622]	[0.817]	[0.729]	[0.740]	[0.609]
DExNoPIIxNoRB − HGxPIIxNoRB	[0.729]	[0.660]	[0.246]	[0.231]	[0.192]	[0.193]
DExNoPIIxNoRB − HGxPIIxRB1	[0.511]	[0.507]	[0.877]	[0.857]	[0.541]	[0.540]
DExNoPIIxNoRB − HGxPIIxRB2	[0.084]	[0.053]	[0.773]	[0.721]	[0.111]	[0.107]
DExNoPIIxNoRB − HGxNoPIIxRB1	[0.078]	[0.054]	[0.082]	[0.049]	[0.027]	[0.019]
DExNoPIIxRB1 − HGxPIIxNoRB	[0.402]	[0.398]	[0.191]	[0.156]	[0.136]	[0.111]
DExNoPIIxRB1 − HGxPIIxRB1	[0.328]	[0.354]	[0.944]	[0.957]	[0.452]	[0.406]
DExNoPIIxRB1 − HGxPIIxRB2	[0.035]	[0.024]	[0.693]	[0.605]	[0.083]	[0.067]
DExNoPIIxRB1 − HGxNoPIIxRB1	[0.032]	[0.025]	[0.063]	[0.031]	[0.017]	[0.009]
HGxPIIxNoRB − HGxPIIxRB1	[0.683]	[0.719]	[0.435]	[0.407]	[0.780]	[0.780]
HGxPIIxNoRB − HGxPIIxRB2	[0.174]	[0.140]	[0.608]	[0.633]	[0.542]	[0.531]
HGxPIIxNoRB − HGxNoPIIxRB1	[0.169]	[0.145]	[0.508]	[0.404]	[0.363]	[0.301]
HGxPIIxRB1 − HGxPIIxRB2	[0.490]	[0.412]	[0.752]	[0.701]	[0.444]	[0.435]
HGxPIIxRB1 − HGxNoPIIxRB1	[0.493]	[0.431]	[0.216]	[0.157]	[0.310]	[0.264]
HGxPIIxRB2 − HGxNoPIIxRB1	[0.989]	[0.963]	[0.295]	[0.245]	[0.871]	[0.804]
Complementarity Test	[0.041]	[0.033]	[0.315]	[0.223]	[0.333]	[0.290]
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: 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.

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

Table C.8: Effects of Survey Design on Survey Duration

	Rapport Treatment (Pooled)		Rapport Treatment	
	(1)	(2)	(3)	(4)
HG Treatment	1.6370***	1.6370***	1.7091***	1.7091***
	(0.5325)	(0.5146)	(0.5340)	(0.5159)
Low PII Treatment	-1.7302***	-1.7302***	-1.1744*	-1.1744*
	(0.5869)	(0.5672)	(0.6420)	(0.6203)
Rapport Treatment (Pooled)	6.1315***	6.1315***		
	(0.5399)	(0.5218)		
Rapport Treatment (Short)			5.4950***	5.4950***
			(0.6196)	(0.5986)
Rapport Treatment (Long)			7.1718***	7.1718***
			(0.7861)	(0.7594)
Control Group Mean	42.1364	42.1364	42.1364	42.1364
p(Long − Short Rapport)			[0.056]	[0.048]
Strata FE	Yes	Yes	Yes	Yes
PDS lasso controls	No	Yes	No	Yes
Observations	2101	2101	2101	2101

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 C.9: Main treatment effects, estimated with *recorded* response = "no" for confused respondents

	Threatening behavior		Physical harassment		Sexual harassment	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main effects						
HG Treatment	0.0339	0.0339*	0.0333*	0.0333*	0.0359*	0.0359*
	(0.0211)	(0.0204)	(0.0187)	(0.0180)	(0.0191)	(0.0184)
No PII Treatment	0.0067	0.0067	0.0291	0.0291	0.0058	0.0058
	(0.0225)	(0.0218)	(0.0185)	(0.0179)	(0.0188)	(0.0182)
Rapport Treatment	0.0124	0.0124	-0.0126	-0.0126	0.0161	0.0161
	(0.0199)	(0.0192)	(0.0174)	(0.0168)	(0.0180)	(0.0174)
Control Group Mean	.099	.099	.0152	.0152	.0178	.0178
Panel B: Heterogeneity by sex						
HG Treatment × Female	0.0184	0.0172	0.0350*	0.0362*	0.0264	0.0285
	(0.0232)	(0.0225)	(0.0208)	(0.0200)	(0.0212)	(0.0204)
HG Treatment × Male	0.1051**	0.1032**	0.0273	0.0249	0.0745*	0.0727*
	(0.0511)	(0.0494)	(0.0436)	(0.0420)	(0.0445)	(0.0434)
No PII Treatment × Female	0.0089	0.0076	0.0322	0.0319	0.0086	0.0092
	(0.0250)	(0.0240)	(0.0205)	(0.0197)	(0.0212)	(0.0205)
No PII Treatment × Male	-0.0086	-0.0104	0.0179	0.0236	-0.0129	-0.0076
	(0.0529)	(0.0510)	(0.0427)	(0.0416)	(0.0414)	(0.0401)
Rapport × Female	0.0245	0.0259	-0.0184	-0.0175	0.0317	0.0317*
	(0.0221)	(0.0213)	(0.0193)	(0.0187)	(0.0201)	(0.0193)
Rapport × Male	-0.0402	-0.0401	0.0127	0.0085	-0.0567	-0.0564
	(0.0469)	(0.0451)	(0.0401)	(0.0388)	(0.0416)	(0.0405)
Female	-0.1005	-0.0981	-0.0262	-0.0113	0.0625	0.0736
	(0.1037)	(0.0994)	(0.0767)	(0.0739)	(0.0764)	(0.0737)
Control Mean - Female	.0798	.0798	.0092	.0092	.0184	.0184
Control Mean - Male	.1912	.1912	.0441	.0441	.0147	.0147
p(HGxFemale - HGxMale)	[0.123]	[0.114]	[0.875]	[0.808]	[0.331]	[0.358]
p(NoPIIxFemale - NoPIIxMale)	[0.765]	[0.751]	[0.763]	[0.858]	[0.644]	[0.709]
p(RapportxFemale - RapportxMale)	[0.212]	[0.186]	[0.486]	[0.547]	[0.056]	[0.049]
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
PDS lasso controls	No	Lasso	No	Lasso	No	Lasso
Observations	2141	2141	2141	2141	2141	2141

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 C.10: HTEs by respondents' schooling qualification for supervisor position

	Threatening behavior	Physical harassment	Sexual harassment
	(1)	(2)	(3)
HG Treatment × Female × Min Grade 8	0.0216	0.0430	0.1032***
	(0.0352)	(0.0325)	(0.0340)
HG Treatment × Female × Below Grade 8	0.0344	0.0360	-0.0085
	(0.0308)	(0.0273)	(0.0271)
HG Treatment × Male × Min Grade 8	0.0971	0.1035	0.0554
	(0.0673)	(0.0636)	(0.0583)
HG Treatment × Male × Below Grade 8	0.1427*	0.0224	0.1232*
	(0.0744)	(0.0642)	(0.0665)
Rapport Treatment	0.0137	-0.0093	0.0176
	(0.0200)	(0.0176)	(0.0182)
No PII Treatment	0.0088	0.0275	0.0060
	(0.0226)	(0.0188)	(0.0191)
Control Mean-Female & Above	.0725	.0072	.0145
Control Mean-Female & Below	.0851	.0106	.0213
Control Mean-Male & Above	.2222	.0278	.0278
Control Mean-Male & Below	.1562	.0625	0
p(HGXFemaleXHigh-HGXFemaleXLow)	[0.783]	[0.871]	[0.010]
p(HGXMaleXHigh-HGXMaleXLow)	[0.642]	[0.369]	[0.435]
Strata FE	Yes	Yes	Yes
Observations	2141	2141	2141

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 C.11: Correlation of team-level reporting rates and response rate to the survey

Correlations	DE	HG	HG-DE
ρ(Threat, Survey Response Rate)	-0.121	-0.150	-0.053
	(0.094)	(0.084)	(0.090)
	[-0.316,0.045]	[-0.304,0.035]	[-0.213,0.140]
ρ(Physical, Survey Response Rate)	-0.097	0.008	0.045
	(0.064)	(0.093)	(0.090)
	[-0.226,0.015]	[-0.182,0.197]	[-0.142,0.217]
ρ(Sexual, Survey Response Rate)	0.069	-0.050	-0.073
	(0.107)	(0.092)	(0.093)
	[-0.126,0.303]	[-0.222,0.135]	[-0.245,0.119]

Notes: This table reports the correlation between the team-level response rate to the survey and the team-level reporting rates of harassment using arms that collect PII. 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).

C.1 Patterns of harassment

As a reminder, we first estimate parameters ρ_M, ρ_H, q_M, q_H and their standard errors by computing a posterior distribution over parameters (starting from a uniform prior over feasible parameters). We discuss this in Section 6 and report the mean parameters for this posterior distribution in Table 6. We then map this posterior distribution over DGPs to a posterior distribution over the statistics of interest, $\mathsf{S}_{TV\geq k}$ and $\mathsf{E}_{2V|1V}$, for teams of size 10 and 15, and report them in Table C.12.

Note that under the conditionally i.i.d. model used to estimate the statistics, $\mathsf{S}_{TV\geq k}$ for $k > 0$ all converge to $q_M + q_H$, the share of managers likely to engage in harassment with positive probability, as team-size grows large, while conditional expectation $\mathsf{E}_{2V|1V}$ converges to 1. Table C.12 directionally reflects these changes, but the qualitative takeaways from Table 7 are not radically changed: victims are frequently isolated, and bulk of harassment cannot be assigned to a few high type offenders.

Evaluating team statistics for large teams also suggests limits of the conditionally i.i.d. model used. For instance, it would poorly capture harassment strategies in which problem managers focus on a small number of specific targets, regardless of team size. The DGP could be enriched to include a random upper bound to the number of victims targeted.[30]

Table C.12: Team-level Statistics for Teams of Size 10 and 15

Statistic	Threatening Behavior	Physical Harassment	Sexual Harassment
Team Size 10			
S_V	0.125	0.047	0.070
	(0.012)	(0.008)	(0.010)
$\mathsf{S}_{TV\geq 1}$	0.709	0.346	0.481
	(0.039)	(0.051)	(0.050)
$\mathsf{S}_{TV\geq 2}$	0.362	0.099	0.163
	(0.046)	(0.028)	(0.034)
$\mathsf{E}_{2V\|1V}$	0.509	0.286	0.337
	(0.042)	(0.063)	(0.048)
Team Size 15			
S_V	0.125	0.047	0.070
	(0.012)	(0.008)	(0.010)
$\mathsf{S}_{TV\geq 1}$	0.827	0.447	0.606
	(0.036)	(0.064)	(0.057)
$\mathsf{S}_{TV\geq 2}$	0.562	0.182	0.288
	(0.033)	(0.044)	(0.052)
$\mathsf{E}_{2V\|1V}$	0.679	0.406	0.473
	(0.045)	(0.082)	(0.06)

Table C.13 reports estimates from applying the closed form non-parametric inference formula of Proposition 2′ to the sample distribution $\widetilde{\mu} \in \Delta(\{0, 1, \cdots, 7\})$ of number of recorded reports across teams. Note that Proposition 2′ only applies when team size is fixed, and the number of preassigned complaints is equal to 2 in each team. In contrast, in our data, team-size varies around a mean and median near 7. Estimates of S_V, $\mathsf{S}_{TV\geq 1}$, $\mathsf{S}_{TV\geq 2}$ and $\mathsf{E}_{2V|1V}$ very similar to those of Table 7 using a likelihood based approach. It is reassuring that, although misspecified in different ways, these two approach yield similar results.

[30] We note that for large teams, even very small rates of false positives would make the interpretation of statistics $\mathsf{S}_{TV\geq 1}$ and $\mathsf{S}_{TV\geq 2}$ problematic, so that we would have to focus of statistics $\mathsf{S}_{TV\geq k}$ for larger values of k to draw meaningful policy inference.

Table C.13: Team-level statistics, with 2 forced complaints & teams of size 7

Statistic	Threatening Behavior	Physical Harassment	Sexual Harassment
S_V	0.119	0.059	0.077
$\mathsf{S}_{TV\geq 1}$	0.635	0.338	0.458
$\mathsf{S}_{TV\geq 2}$	0.195	0.074	0.078
$\mathsf{E}_{2V\|1V}$	0.307	0.220	0.170

D Reporting harassment & mental health

Sociological research suggests that the act of confiding secrets can improve an individual's well-being through improving one's perceived coping ability and reducing one's mental load associated with the secret (Slepian and Moulton-Tetlock, 2019). To explore this possibility, we resurveyed workers two weeks after the survey experiment to test whether reporting harassment improved workers' mental well-being and job satisfaction. We measure mental health and job satisfaction, respectively, using summary index variables following Anderson (2008). We report the variables comprising each index at the end of this appendix.

As per our PAP, we run a 2SLS model with (6) as our first stage and (D.1) as our second-stage regression:

$$W_{is} = \delta Y_{is} + \rho W_{is}^0 + \theta X_i + \mu_s + \epsilon_{is} \tag{D.1}$$

where W_{is} is worker well-being in the follow-up survey for individual i in stratum s, Y_{is} are reports of threats, physical and/or sexual harassment from the main worker survey, and W_{is}^0 is the baseline worker well-being, measured in the main worker survey. We control for stratum fixed-effects μ_s and individual demographic characteristics X_i. Since there is a possibility that some elements of the survey design directly impact worker well-being (notably, RB), we also report the reduced form effect in a regression equivalent to (6), with W_{is} as outcome (and controlling for W_{is}^0).

Because we find no main effect of RB and Low PII on reporting, there is not a first stage between these two instruments and reporting. In other words, these are weak instruments, which will bias our 2SLS results towards the OLS results we would get if regressing mental health on reporting.[31] While we pre-specified that we would use (6) as our first stage, because

[31]The coefficients from the OLS regressions of mental health on reporting are zero or weakly negative (results not reported).

of the weak instruments concern, we also report results only using randomized assignment to HG as the instrument and controlling for assignment to the RB and Low PII arms.

Table D.1 reports the reduced form and 2SLS effects on mental health and job satisfaction, using randomized assignment to HG, RB, and Low PII as instruments, measured in the follow-up survey. Columns (1)-(2) show that the treatments do not directly effect mental health or job satisfaction. Columns (3)-(5) show that reporting harassment improves mental health among those induced to report by the treatment interventions by 8-16% of a standard deviation, although none of the increases is statistically significant. Columns (7)-(10) show that reporting harassment improves job satisfaction among those induced to report by the treatment interventions by 37-68% of a standard deviation on average, although none of the increases is statistically significant.

Table D.2 reports the reduced form and 2SLS effects on mental health and job satisfaction, using randomized assignment to HG as the instrument and including controls for assignment to the RB and Low PII treatment arms. As expected, the estimated coefficients are uniformly more positive than in Table D.1, but they remain imprecise. Column (4) suggests that increasing the reported share of yeses from 0 to 1 improves mental health by 23% of a standard deviation among those induced to report under HG (p=0.291). Column (8) suggests that it improves job satisfaction by 86% of a standard deviation (p=0.170). While imprecise, the large, consistently positive coefficients suggests that the psychological and/or expected social benefits of reporting may be large. We think that more precisely quantifying these benefits, and more broadly exploring the benefits and costs of improved reporting systems for harassment for workers, presents an interesting direction for future research.

Table D.1: Reduced form & 2SLS effects on mental health & job satisfaction, measured in follow-up survey

	Reduced form		2SLS							
	Mental health index	Job satisfaction index	Mental health index				Job satisfaction index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
HG Treatment	0.0107	0.0368								
	(0.0093)	(0.0247)								
Rapport Treatment	0.0019	0.0117								
	(0.0097)	(0.0255)								
No PII Treatment	-0.0132	-0.0407								
	(0.0103)	(0.0278)								
Reported threatening behavior			0.1441				0.6924			
			(0.1806)				(0.5734)			
Reported physical harassment				0.0779				0.3650		
				(0.2114)				(0.6113)		
Reported sexual harassment					0.1608				0.6756	
					(0.1799)				(0.5292)	
Share of reports that are yes						0.1528				0.6766
						(0.1953)				(0.5676)
Control Mean	.06	.296	.06	.06	.06	.06	.296	.296	.296	.296
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1988	1988	1985	1985	1985	1985	1985	1985	1985	1985
Kleibergen-Paap Wald F			1.8	1.8	2.4	4	1.5	1.7	2.2	3.5

Notes: This table reports reduced form and 2SLS results for respondents' mental health and job satisfaction, measured in the follow-up survey. Columns (1)-(2) report reduced form results, and columns (3)-(10) report 2SLS results using the randomized assignment to the HG, RB, and Low PII treatments as the instrumental variables. All regressions include controls for the baseline value of the dependent variable, gender, age, work experience, tenure, schooling, marital status, and whether the respondent has children. Robust standard errors in round brackets. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table D.2: Reduced form & 2SLS effects on mental health & job satisfaction, measured in follow-up survey, only using HG as an instrument

	Mental health index				Job satisfaction index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reported threatening behavior	0.2269				0.8768			
	(0.2321)				(0.7339)			
Reported physical harassment		0.2620				0.9667		
		(0.2665)				(0.8045)		
Reported sexual harassment			0.2003				0.7414	
			(0.1972)				(0.5701)	
Share of reports that are yes				0.2282				0.8561
				(0.2163)				(0.6242)
Control Mean	.06	.06	.06	.06	.296	.296	.296	.296
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1985	1985	1985	1985	1985	1985	1985	1985
Kleibergen-Paap Wald F	3.9	3.8	6.2	10.4	3.4	3.6	5.9	9.6

Notes: This table reports reduced form and 2SLS results for respondents' mental health and job satisfaction, measured in the follow-up survey. All columns report 2SLS results using the randomized assignment to the HG treatment as the instrumental variable. All regressions include controls for the baseline value of the dependent variable, gender, age, work experience, tenure, schooling, marital status, whether the respondent has children, and assignment to the RB and Low PII arms. Robust standard errors in round brackets. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Survey questions used to construct index variables:

1. Mental health:

    - Generalized Anxiety Disorder 7-item (GAD-7) scale): In the past 7 days, how often... (Select one: Not at all or less than 1 day; 1-2 days; 3-4 days; 5-7 days.)
        a have you felt nervous, anxious, or on edge?
        b have you felt depressed?
        c have you felt lonely?
        d have you felt hopeful about the future?
        e have you been so restless that it is hard to sit still?
        f have you become easily annoyed or irritable?
        g have you felt afraid, as if something awful might happen?
    - Please imagine a ladder with steps numbered from zero at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. (Select one: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.)
        a On which step of the ladder would you say you personally feel you stand at this time?
        b On which step do you think you will stand about five years from now?

2. Job satisfaction:

    - How satisfied are you with your job overall? (Select one: Very dissatisfied; Dissatisfied; Neutral; Satisfied; Very satisfied)
    - How satisfied are you with the following aspect of your job: (Select one: Very dissatisfied; Dissatisfied; Neutral; Satisfied; Very satisfied)
        a You are listed to?
        b You are treated with respect?
        c Career opportunities?
        d Job training and support?
        e Pay is fair for your job?

- Which of the following statements best describes your feelings about your job? In my job... (Select one: I only work as hard as I have to; I work hard, but not so that it interferes with the rest of my life; I make a point of doing the best work I can, even if it sometimes does interfere with the rest of my life; I don't know)
- *Main survey experiment only:* For the following statements, please state whether you strongly agree, agree, neither agree nor disagree, disagree, or strongly disagree..
  a For me this is the best of all possible organizations for which to work.
  b I find that my values and the organization's values are very similar.
  c I feel very little loyalty to this organization.
  d Often, I find it difficult to agree with this organization's policies on important matters relating to its employees.
  e I am proud to tell others that I am part of this organization.