Pre-Analysis Plan:

Survey Design for Sensitive Information in Organizations

Laura Boudreau
Columbia Business School
l.boudreau@columbia.edu

Sylvain Chassang
Princeton University
chassang@princeton.edu

Ada González-Torres
Ben Gurion University of the Negev
adagt@bgu.ac.il

Rachel Heath
University of Washington
rmheath@uw.edu

September 17, 2021

Contents

1 Introduction

1.1 Abstract

This research studies how survey design affects transmission of sensitive information within organizations. We conduct a phone-based survey experiment with workers at two garment factories in Bangladesh to study how survey design affects their willingness to report misbehavior by managers, including threats, physical harassment, and sexual harassment. We experimentally vary whether the survey elicitation method provides plausible deniability when asking sensitive questions. In particular, building on Chassang and Padró i Miquel (2018) and Chassang and Zehnder (2019), we use hard garbling to provide plausible deniability by exogenously distorting survey responses. We also experimentally vary the extent to which the survey enumerator builds rapport with the surveyed individual and the level of identifiability of a surveyed worker.

1.2 Motivation

In many organizations, individual members are unwilling to report experience or knowledge of misbehavior because doing so will result in retaliation and/or stigma. As a result, misbehavior often remains under the radar and goes unpunished. Social scientists have developed indirect survey methods that may be able to increase information transmission in organizational settings. In particular, researchers have developed methods that provide survey respondents with plausible deniability regarding their answers, such as hard garbling (HG) and randomized response (RR) techniques (Warner, 1965; Chassang and Zehnder, 2019).

While empirical validation studies are available for some of these methods (Rosenfeld, Imai and Shapiro, 2016; Chuang et al., 2020), the existing evidence remains patchy. Further, the empirical literature comprises of surveys of socially-sanctioned beliefs and behaviors elicited for the purpose of conducting social science research; there is little to evidence from organizational settings in which the transmitted information may trigger real-world actions and payoffs. In this context, we ask how survey designs that provide plausible deniability affect transmission of sensitive information. Further, we ask how respondents' trust in the surveyor influences their responses and whether trust may complement or substitute for plausible deniability. Our context in Bangladesh's garments sector, where we believe that many forms of misbehavior by managers, in particular related to harassment, goes unreported.

Further, sociological research suggests that the act of confiding secrets can

improve individuals' well-being; this benefit is due, at least in part, to improving one's perceived coping ability and reducing one's mental load associated with the secret (Slepian and Moulton-Tetlock, 2019). If our survey design increases reporting of experienced misbehavior, we hypothesize that it may convey similar benefits to workers. Moreover, we will use our survey to understand to what extent do "inexpensive" covariates predict (1) sensitive survey answers; and (2) the types of respondents for whom building trust and/or understanding may be important?

This research will also allow us to calculate *policy-relevant statistics in our setting:* What share of managers are responsible for what share of the misbehavior? What share of workers have experienced misbehavior by their managers? How much do workers know about mistreatment of other workers? We also aim to learn about the types of deterrents to reporting sensitive issues, and whether workers actually reported on different issues through internal factory channels.

We collaborate with a Bangladeshi apparel manufacturer that employs approximately 25,000 workers. In collaboration with the manufacturer, we survey workers in two of their factories, which together employ over 8,700 people. While we will only share aggregated statistics from the survey with the manufacturer's senior management, this information will be used to craft HR policies; in this way, our set-up has the feature that survey reports may trigger real-world actions and payoffs.

1.3 Research Questions

Our main research question is:

1. How does survey design affect transmission of sensitive information in a real-world organizational setting? More precisely:

   (a) Does providing respondents with plausible deniability through hard garbling increase reporting of sensitive information in an organizational setting?

   (b) Can building trust or rapport during a survey increase reporting of sensitive information in an organizational setting?

   (c) Can reducing the amount of personally identifying information elicited from respondents increase reporting of sensitive information in an organizational setting?

Our secondary research questions are:

1. Do different methods for increasing information transmission, mainly, providing plausible deniability, building rapport, and minimizing requested PII, complement or substitute for each other?

2. Do particular aspects of the survey design differentially affect reporting by different types of workers, such as women compared to men?

3. To the extent that our survey design increases reporting of mistreatment by managers, does this improve workers' well-being in the short-run?

1.4 Research Design

We conduct two surveys with workers. In the main survey, we randomly assign workers to different combinations of experimental conditions. In the second, follow-up survey, there is no additional experimental variation. We also conduct a complementary survey with supervisors after both worker surveys were completed, with no additional experimental variation.

The experimental conditions introduced in the first survey are as follows:

Survey method:

1.a) Direct elicitation (DE): entails directly asking the survey respondent about sensitive information. DE is the status quo survey method and the control condition.

1.b) Hard garbling (HG): for a yes or no question, where "yes" is the more sensitive answer, we exogenously flip "no" answers to "yes" with 20% probability. We fix the flipping rate at 20% in groups of 10 HG surveys, which means that the flipping rate is typically fixed across stratas of workers that are located nearby to each other on the same production floor (see Section 2 for more information on sampling and assignment to treatment). In this way, if a respondent's answer is saved as "yes", it is impossible to know whether they actually responded yes or no, however we can still recover the overall ratio of yes to no answers.

Rapport-building (RB):

2.a) Status quo approach: By status quo, we mean that we will use a typical social science research introduction script before beginning the survey and then ask the survey questions. This is the control condition.

2.b) RB approach: we allocate survey time to build "rapport" or trust with the participant, chatting about family, hobbies in a natural but pre-specified manner, beyond the minimum small talk typical in the standard social science approach. We have two RB sub-treatments to test for the possibility that the marginal returns of building rapport decrease quickly:

- RB-short: 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. We develop guidelines summarized in an attached document *"Guidelines for Building Rapport."*
- RB-long: In this extended rapport-building section, the enumerator becomes personable with the worker, who has the chance to ask questions to them. The enumerator also shares a related experience. For details see the attached document *"Guidelines for Building Rapport."*

Personally-identifying information (PII):

3.a) Status quo approach: We ask survey respondents to answer questions that reveal relatively more PII (questions include production line number or section and direct supervisor). We consider this to be the "status quo" because surveys in organizational settings often explicitly or de facto reveal respondents' identities. This is the control condition.

3.b) Low PII approach: We limit the amount of PII requested from the survey respondent (no questions about production line number/section or direct supervisor).

The experimental conditions are introduced in different sections of the main worker survey:

1. Survey elicitation method (DE, HG): experimentally varied in asking questions on *respondents' experience with harassment*

2. Rapport-building: experimentally varied before workers answer questions on *barriers to reporting and Covid behaviors*, which is followed by *observed harassment of co-workers* and then *respondents' experience with harassment*

   2.a Rapport-building (RB1): first part right before *barriers to reporting and Covid behaviors*

2.b Rapport-building (RB2): second part right before *observed harassment of co-workers*

3. Personally-identifying information (PI): experimentally varied whether respondent completes a short module of employment-related questions (section is input immediately before *rapport-building*)

Design definitions:

- Survey method $\in \{DE, HG\}$
- Rapport-building $\in \{0, RB1, RB2\}$
- Personally-identifying information (PI) $\in \{0, 1\}$

In the experiment, we have 9 different treatment arms:

1. (DE, PI) Survey=DE, PI=1, Rapport=0 [Benchmark].

2a. (DE, PI, RB1) Survey=DE, PI=1, Rapport=1.

2b. (DE, PI, RB2) Survey=DE, PI=1, Rapport=2.

3. (DE) Survey=DE, PI=0, Rapport=0.

4. (DE, RB1) Survey=DE, PI=0, Rapport=1.

5. (HG, PI) Survey=HG, PI=1, Rapport=0.

6a. (HG, PI, RB1) Survey=HG, PI=1, Rapport=1.

6b. (HG, PI, RB2) Survey=HG, PI=1, Rapport=2.

7. (HG, RB1) Survey=HG, PI=0, Rapport=1. [Most protective (or with RB2)]

Table 1 summarizes these treatment arms.

Updates with respect to pilot experiment: The differences with the pilot experiment are reported in the Appendix A. Based on the results of the pilot, we removed one indirect survey elicitation mechanism.

We also separated the RB treatment arm in two levels of intensity. This is to test for differences in the effects of shorter vs longer rapport building. We made this decision based on feedback from the pilot that respondents appreciated rapport building, but they started to grow impatient toward the end of the survey, so the marginal returns to RB may decrease fairly quickly.

Table 1: Main Experiment - Treatment Arms

		No Rapport	Rapport 1	Rapport 2
Direct elicitation	PI	Arm 1	Arm 2a	Arm 2b
	No PI	Arm 3	Arm 4	
Hard garbling	PI	Arm 5	Arm 6a	Arm 6b
	No PI		Arm 7	

Notes: Arm 1 = Benchmark; Arm 7 = Most protective ex ante (unless "rapport" removes the benefit of "No PI," as the respondent is asked for more information; we will test this possibility by comparing Arms 3 and 4).

1.5 Treatment effects of interest

Primary treatment effects of interest:

1. Does providing respondents with plausible deniability through hard garbling affect reporting?
2. Does building rapport affect reporting?
3. Does asking for more personally identifying information affect reporting?

Secondary treatment effects of interest:

1. How do different aspects of survey design interact or compare? In particular:
   (a) What is the marginal impact of a lengthier rapport-building section?
   (b) Do different methods for increasing information transmission, mainly, providing plausible deniability, building rapport, and minimizing requested PII, complement or substitute for each other?
   (c) How much does information transmission increase with what we perceive to be the most protective elicitation mechanism ex-ante, which entails three modifications (HG, RB, lower PII), compared to the status quo survey system (DE, no RB, higher PII)?
2. Do particular aspects of the survey design differentially affect reporting by different types of workers, such as women compared to men?

3. Does an experimentally-induced increase in reporting improve workers' well-being in the short-run?

2 Research Strategy

2.1 Sampling

2.1.1 Sampling Frame

Context: Our empirical setting is Bangladesh's apparel sector, which is critical to the country's economy. Most workers in the sector are employed on sewing lines of 40-90 workers. While factories vary, 70-80% of workers in the sewing section are women, while more than 90% of managers are men. In the two factories that we work with, 95% of sewing line workers are women, while more than 86% of sewing section managers are men. The sector has long struggled with an international reputation for poor working conditions and limited labor rights. Workers who experience harassment or other forms of mistreatment by management often have few to no options to obtain resolution: Internal reporting systems are often corrupt or ineffective and legal institutions are weak.

We collaborate with an anonymous apparel manufacturer that aims to improve its internal reporting systems and relations with its workers.[1] The manufacturer employs approximately 25,000 workers.

Eligible population: We conduct a survey experiment with workers at two of the apparel manufacturer's factories, which together employ over 8,700 people. First, we conduct a stratified random selection of workers to participate in the survey experiment. Using the entire list of employees in the two factories, we sample workers from four types of production teams. Among these teams, we chose teams with a sufficient large number of workers (approximately above 15). We are left with 112 eligible teams and a total of 5,948 eligible workers, out of a workforce of 8,727 people, including 1,000 managers and administrative staff members.

We next stratify workers on eligible teams by their sex, which we identify based on name (male, female, uncertain).[2] In some cases, there are teams with very small

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

[2] Names in Bangladesh are highly gendered. As such, we are able to categorize names as male or female for 5,929 out of 5,948 eligible workers. We will verify our categorizations using the survey data.

numbers of one group; in this case, we aggregate these workers to the smallest level that yields a sufficiently large groups size (e.g., production section-floor). We first select 9 workers per strata and then we sample larger strata in proportion to their share of the overall eligible worker population. We take this approach because we aim to have at least one worker per stratum in each treatment arm. We detail workers' assignment to treatment below.

Types and counts of eligible teams:

- Sewing lines: 48
- Finishing teams: 46
- Dry wash teams: 15
- Wet wash teams: 3

Sample size: In total we aim at a sample size of 2620 workers. Because we have access to the complete population of workers at the two factories, we are able to replace workers who decline to participate or who are unreachable. Subsequently, even with an imperfect response rate, we should be able to reach close to our target sample size. The reason why this may not be the case is because we conduct a stratified randomized selection and assignment of workers by factory-production team-gender, which means that in smaller strata, we may run out of workers to replace workers who decline to participate. This sample size represents 30% the workforce (and 34% of workers) in the two factories.

2.1.2 Statistical Power

We used data from a 2017 survey on working conditions in Bangladesh's apparel sector and from our pilot survey to inform our power calculations. We conduct power calculations for a binary outcome variable.

Assumptions:

- Power: $\beta = 0.8$.
- Significance level: We set $\alpha = 0.10$ and apply the Bonferroni adjustment with 3 primary outcomes $0.10/3 = 0.033$ (see Section 3.1 for primary outcome variables).

- Minimum detectable effect (MDE) size: We set the MDE to 0.2 standard deviations (SDs).

Sample size per treatment condition: Based on power calculations detailed above, we require 443 workers per treatment condition. With 198 workers per treatment condition, we are powered to detect an effect size of 0.3 SDs. Given experimental constraints, we design the experiment to have at least 220 workers per treatment condition, and at least 440 workers in key treatment conditions. In a few cases, we pool two treatment arms in order to achieve our target sample size. Table 2 summarizes our final sample size for each treatment condition.

Table 2: Main Experiment - Planned sample sizes

		No Rapport	Rapport 1	Rapport 2	TOTAL
Direct elicitation	PI	Arm 1 476	Arm 2a 225	Arm 2b 229	930
	No PI	Arm 3 226	Arm 4 220		446
Hard garbling	PI	Arm 5 487	Arm 6a 225	Arm 6b 227	939
	No PI		Arm 7 305		305
	TOTAL	1189	975	456	2620

Notes: Arm 1 = Benchmark; Arm 7 = Most protective ex ante.

2.1.3 Assignment to Treatment

The **unit of randomization** is a worker, **stratified** by factory-production team and sex. As detailed under sampling above, in cases where there are too few men or women on a production team, we aggregate to the next highest level that yields a sufficiently large stratum size. The randomization is done in Stata. We first randomly assign one worker per stratum to each treatment arm; this is because we want to ensure that all strata are represented in all treatment arms. For larger strata, we then randomly assign workers to each treatment arm with probabilities of assignment that correspond to the treatment arm's target share of the overall sample

size. We use the *randtreat* package by Carril (2017) to address misfits across strata. We conducted 10 randomizations and selected the one that performed best in terms of balance on two covariates available to the research team (tenure and skill group).

2.1.4 Attrition from the Sample

We are able to replace workers who decline to participate or who are unreachable; for each stratum, we have a randomly-ordered list of back-up workers that we will use to replace workers who decline to participate or who are unreachable. In a pilot survey of workers employed by this apparel manufacturer conducted by Boudreau and Heath in 2020, the overall response rate was 79% (sampling with replacement). As such, we anticipate that we may need to approach around 3,315 workers in order to achieve our target sample size of 2,620. Due to some strata having a limited number of back-up workers, it is possible that we may not achieve our target of 2,620 workers.

There may be attrition between our main survey experiment and our follow-up survey. While this attrition does not threaten our ability to measure the primary treatment effects of interest, it would impact our ability to detect one of the secondary treatment effects of interest, which is the impacts of increased reporting on workers' well-being. As we were unable to pilot the follow-up survey, it is difficult to estimate how much attrition may occur. We expect attrition to be low, however, because we are conducting the follow-up survey two weeks after our main survey. We will report the overall attrition rate in the paper as well any differences in the attrition rates across treatment conditions.

2.2 Fieldwork

2.2.1 Instruments

We use the following data collection instruments:

- Main worker survey
- Follow-up worker survey
- Call and rapport-building scripts
- Supervisor survey

In the pilot, we only ran the main worker survey. The additional surveys were planned for the main experiment.

The main worker survey has a total of 8 sections:

1. Core demographics, health/COVID, and well-being
2. Job satisfaction
3. Supervisor's practices & relationship with management
4. Identification questions
5. Barriers to reporting and Covid behaviors (to measure social desirability)
6. Observed harassment of co-workers
7. Respondents' experience of harassment
8. Interviewer's comments

The experimental conditions apply to different sections of the main worker survey:

1. Survey elicitation method (DE, HG): experimentally varies method for questions in *Section 7: Respondents' experience of harassment*
2. Rapport-building: experimentally inserted before *Section 5: Barriers to reporting and Covid behaviors*
   - 2.a Rapport-building (RB1): first part before *Section 5: Barriers to reporting and Covid behaviors*
   - 2.b Rapport-building (RB2): second part before *Section 6: Observed harassment of co-workers*
3. Personally-identifying information (PII): experimentally included or not as: *Section 4: Identification questions*

The follow-up worker survey has a total of 4 sections:

1. Health/COVID and well-being
2. Job satisfaction
3. Supervisor's practices & relationship with management
4. Reporting of misbehavior by manager

The call and rapport-building scripts have 3 parts:

1. First call to recruit worker (including the informed consent process) and to schedule interview (same for all treatment groups)
2. Second call to initiate interview (same for all treatment groups)
3. Rapport-building script: Block I and II (for groups treated with short- and long-rapport, respectively)

The supervisor survey has a total of 2 sections:

1. Core demographics, health/COVID and well-being
2. Cognitive ability, personality, and other traits

Some of these modules in these surveys will be used for separate project by Boudreau and Heath.

2.2.2 Data Collection

We will conduct the **main worker survey** from early September 2021 to late September or October 2021. We roll out the survey by factory-floor in order to avoid spillovers as workers could discuss the survey details with other workers before they have been treated.

The **follow-up worker survey** is planned two weeks after the first survey, using the same roll-out strategy. The **supervisor survey** is planned a week after workers have completed the follow-up survey.

2.2.3 Data Processing

Data privacy and protection of human subjects: Participation in the survey may involve some risks. First, workers may feel they owe to the factory that they need to participate in the survey. To address this issue, we inform all subjects that participation in the surveys is entirely voluntary in each of the survey rounds. A surveyor will read out loud a consent form that is provided to each worker surveyed and if they decide to participate, they will be asked them to give verbal consent before doing so. The consent form includes a broad description of the purpose of the research, without providing too much detail to avoid biasing the results. Before each data collection round, workers will be reminded that participation is entirely voluntary.

Second, collecting survey data comes with privacy concerns. To address this, we will ensure confidentiality of any personal information acquired. The survey data will be stored on a password protected server and will not be shared with any factories or brands, or anybody outside the research team. Third, most journals require the publication of such data to ensure transparency and replicability of research. Upon this research's publication in a peer-reviewed journal, we will publish de-identified data. To ensure effective anonymity, we will aggregate variables as needed to ensure individuals are not identifiable.

3 Empirical Analysis

3.1 Variables

3.1.1 Primary outcomes

The key outcome variables are the reporting of different types of labor issues, and more precisely of:

- Threats from direct supervisor
- Physical harassment from direct supervisor
- Sexual harassment from direct supervisor

The variables are defined in Table 3. In the survey, we randomize the order of the question asking about each of these experienced behaviors.

Table 3: Primary outcome variables

Variable Name	**Variable Definition**
Threats	Your supervisor has threatened or told you that they will harm you if you do not agree to or fulfill their demands.
Physical harassment	Your supervisor has taken one or more of the following actions toward you against your will: Hit, slapped, or punched; Cut or stabbed; Tripped; Otherwise intentionally caused physical harm.
Sexual harassment	Your supervisor has taken one or more of the following actions toward you against your will: 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.

3.1.2 Secondary outcomes

Additionally, as secondary outcome, if we find that one or more of our experimental methods increases reporting, we will study how increased reporting affects workers' well-being in the short run. We measure well-being using an index of mental health survey questions (questions 4 and 5 in our follow-up survey, which we have preregistered).

3.2 Regression Model

Main regression model:

$$Y_{is} = \alpha\, HG_i + \beta\, Rapport_i + \gamma\, NoPII_i + \mu_s + \theta X_i + \epsilon_{is} \tag{1}$$

where Y_{is} is the outcome of interest for individual i in stratum s. $HG_i, Rapport_i$ and $NoPII_i$ are hard-garbling, rapport, and not asking for personally identifying information, respectively. We control for stratum fixed-effects μ_s. We will also present results including individuals' characteristics X_i (selected using a lasso-based approach). We will also run a regression in which we control for the actual flipping rate for hard garbling.

Secondary regression models: In addition, we are interested in the following variations of model (1):

1. Allowing for differential effects by levels of rapport (RB1 versus RB2).
    - If we find no differential effects, we will keep rapport-levels pooled.
2. Interactions between different aspects of the survey design (survey method, RB, and PII).
    - We will run the fully saturated model, and if certain interactions are not significantly different from zero, we will use a disciplined approach to collapse different interactions.
3. Heterogeneous effects by gender of worker.
4. Further, we are interested in the impact of increased reporting through our survey on worker well-being in the short-run. We run a 2SLS model with (1)

as our first stage and (2) as our second-stage regression:

$$W_{is} = \rho W_{is}^0 + \delta Y_{is} + \theta X_i + \mu_s + \epsilon_{is} \tag{2}$$

where W_{is} is worker well-being in the follow-up survey for individual i in stratum s, W_{is}^0 is the baseline worker well-being, measured in the main worker survey, and Y_{is} are reports of threats, physical and/or sexual harassment from the main worker survey. We control for stratum fixed-effects μ_s and individual demographic characteristics X_i. As above, we will run a regression in which we control for the actual flipping rate for hard garbling.

Since there is a possibility that some elements of the survey design directly impact worker well-being (notably rapport), we also report the reduced form effect in a regression equivalent to (1), with W_{is} as outcome (and controlling for W_{is}^0).

Controls: The stratum fixed effects absorb factory, team, and gender fixed effects. In addition, we plan to include other characteristics as controls using a lasso-based approach.

Standard error adjustments: We use robust standard errors.

4 Research Team

The Principal investigators are:

Laura Boudreau, Columbia Business School

Sylvain Chassang, Princeton University

Ada González-Torres, Ben Gurion University of the Negev

Rachel Heath, University of Washington

References

Carril, Alvaro. 2017. "Dealing with misfits in random treatment assignment." *Stata Journal*, 17(3): 652–667.

Chassang, Sylvain, and Christian Zehnder. 2019. "Secure Survey Design in Organizations: Theory and Experiments."

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

Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban. 2020. "Sex, Lies, and Measurement: Do Indirect Response survey methods work?"

Rosenfeld, Bryan, Kosuke Imai, and Jacob N. Shapiro. 2016. "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science*, 60(3): 783–802.

Slepian, Michael L., and Edythe Moulton-Tetlock. 2019. "Confiding Secrets and Well-Being." *Social Psychology and Personality Science*, 10(4): 472–484.

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

A Appendix: Pilot Experiment

Additional survey elicitation mechanism: In the pilot experiment, we had an additional indirect survey elicitation mechanism compared to the main experiment (Section 1.4):

1.c) Lottery responses (LR): for a yes or no question, we allow "maybe" as a possible answer, which the system saves as "yes" with probability p, and as "no" with probability 1-p. This option is designed to induce individuals who do not feel comfortable directly responding "yes" to still convey some information about their experience. The idea of using a lottery response design comes from the possibility that individuals have a certain degree of ambiguity aversion.

Differences with experiment: We found that none of the workers chose the lottery response "maybe". Therefore, the lottery response design had no impact on reporting, and we decided to remove LR from the main experiment.

In addition, during the pilot we only had one treatment arm with the full-length rapport-building script (RB2).

Design definitions:

- Survey method $\in \{DE, HG, LR\}$
- Rapport-building (RB) $\in \{0, 1\}$
- Personally-identifying information (PI) $\in \{0, 1\}$

where RB refers to the full length rapport building design (RB2).

In the pilot we had five different treatment arms:

1. (DE, PI) Survey=DE, PI=1, Rapport=0 [Benchmark].
2. (DE, PI, RB) Survey=DE, PI=1, Rapport=1.
3. (DE, RB) Survey=DE, PI=0, Rapport=1.
4. (HG, PI, RB) Survey=HG, PI=1, Rapport=1.
5. (LR, PI, RB) Survey=LR, PI=1, Rapport=1.

Table 4 summarizes the treatment arms.

Table 4: Pilot Experiment - Treatment Arms

		No Rapport	Rapport
Direct elicitation	PI	Arm 1	Arm 2
	No PI	Arm 3	
Hard garbling	PI		Arm 4
	No PI		
Lottery response	PI		Arm 5
	No PI		

Notes: Arm 1 = Benchmark.

Main treatment effects of interest:

1. Do respondents understand and trust indirect survey methods? Do they affect reporting?
    - Hard garbling
    - Lottery response
2. Does building rapport affect reporting?
3. Does asking for identifying information affect reporting?

Pilot Sampling Strategy: We had a sample of 316 workers. Table 5 summarizes the sample size for each treatment arm.

Research instruments: We only run the **main worker survey**, and we also used the **call and rapport scripts** described above.

Data collection: The main worker survey ran from April 9 - 16, 2021.

Pilot Regression model:

$$Y_{ij} = \alpha_1 LT_i + \alpha_2 HG_i + \beta PI_i + \gamma Rapport_i + \theta_i X_i + \mu_j + \epsilon_{ij} \tag{3}$$

where Y_{ij} is the outcome of interest for individual i in factory j. We control for individual demographic characteristics X_i and factory fixed-effects μ_j.

Controls: Gender.

Heterogeneous effects: We also studied heterogeneous effects by **gender**.

Standard error adjustments: We cluster standard errors at the level of the factory-team.

Table 5: Pilot Experiment - Sample sizes

<table>
<tr><th></th><th></th><th>No Rapport</th><th>Rapport</th></tr>
<tr><td rowspan="2">Direct elicitation</td><td>PI</td><td rowspan="2">Arm 1
63
Arm 3
61</td><td>Arm 2
64</td></tr>
<tr><td>No PI</td><td></td></tr>
<tr><td rowspan="2">Hard garbling</td><td>PI</td><td></td><td>Arm 4
64</td></tr>
<tr><td>No PI</td><td></td><td></td></tr>
<tr><td rowspan="2">Lottery response</td><td>PI</td><td></td><td>Arm 5
64</td></tr>
<tr><td>No PI</td><td></td><td></td></tr>
</table>

Notes: Arm 1 = Benchmark.