

Wageningen University – Department of Social Sciences

MSc Thesis Development Economics Group

The position of identifying questions and responses to potentially sensitive questions in a face-to-face interview:

Survey experiment in Ghana

May 2017

MSc Programme

International Development Studies

Name of student

David Valenta

Reg. No. 910325-851110

Specialization

Economics of Development

Supervisor

Maarten Voors

Thesis code: DEC-80433



The position of identifying questions and responses to potentially sensitive questions in a face-to-face interview: Survey experiment in Ghana

David Valenta

Master Thesis, Wageningen University

23/5/2017

Abstract

This study investigates whether varying placing of identifying questions prior to potentially sensitive questions, or after these questions, leads to an effect on answers to the potentially sensitive questions, such as attitudinal questions and questions about various institutions. This question is investigated on a case of a face-to-face administrated survey questionnaire in Ghana. The sample includes a total of 3528 respondents. The respondents had been sampled from 120 districts of Ghana and include local leaders (local politicians and traditional leaders) using convenience and local citizens. Randomized survey experiment with two treatment conditions is used to answer this question. In one treatment condition the identifying questions are asked prior to the potentially sensitive questions. In the other treatment condition, no identifying questions are asked prior to the potentially sensitive questions. We hypothesize the effects of different placing of the identifying questions to be driven by changes in perceived anonymity. We find no evidence of effect of this variation on the response-rate to the potentially sensitive questions. On the other hand, we find mean answers to some potentially sensitive questions to be significantly different among the two treatment conditions. We attribute this difference to an increase in social desirability bias in answers caused by asking identifying questions prior to the potentially sensitive questions. This is the case for respondents that have been selected using non-purposeful sampling. No evidence of such effect is observed for the local leaders selected using purposeful sampling. Further, we hypothesize this variation only has effect in the case of a sensitive outcome with clear direction of socially desirable reporting. Overall, our results indicate, that when a researcher is concerned with social desirability in answers, placing identifying questions at the end of the questionnaire should be preferred in case of face-to-face interview in a country like Ghana.

Acknowledgements

The theory presented is based on ex-post interpretation of the findings of this study. The theories as used in the pre-analysis plan are not presented. Only the final mechanism believed to play a role is described in this document. Note on the pre-analysis plan theories can be found in Appendix 9.1. The full pre-analysis plan can be requested at the following email address: maarten.voors@wur.nl

The survey was financed using resources from the International Initiative for Impact Evaluation (3ie) grant TW8.1002: “Examining Transparency and Accountability within the Oil and Gas Sector: Impact evaluation of key provisions in Ghana’s Petroleum Revenue Management Act (2011)” and from TrAc Revenues project, funded by the Research Council of Norway.

I am grateful to Maarten Voors for supervision of this thesis. As well as to the Kumasi Institute of Technology Energy and Environment (KITE), Ghana for allowing me to cooperate on their project, especially to Ishmael Edjekumhene. Further, I express my gratitude to the local research assistants as well as all the team members and who collaborated on the survey, namely Christa Brunnschweiler, Päivi Lujala and Charles Kofi Owosu. I am also thankful to participants of a workshop in Trondheim, as well as to all the colleagues, classmates and friends who provided comments on my thesis, especially to Sabrina McDonald. Next, I am grateful to my parents Helena and Jan for their support. Finally, I am pleased to acknowledge funding from the 3ie grant for my field-work in Ghana and from the TrAc Revenues project for covering the workshop participation in Trondheim, Norway.

Table of contents

Abstract	2
Acknowledgements	3
Section 1 Introduction.....	7
Section 2 Background, mechanism and hypothesis.....	10
2.1 Background.....	10
2.2 Mechanism and hypothesis.....	13
2.2.1 Hypothesis 1:.....	14
2.3 Heterogeneous treatment effects.....	14
2.3.1 Hypothesis 2:.....	15
Section 3 Data, research design and sample	15
3.1 Data	15
3.2 Sample description and sampling strategy	15
3.3 Assurances of confidentiality and informed consent policy	16
3.4 Research design.....	17
3.5 Treatment variation	18
3.6 Outcomes	20
3.6.1 Non-response rate to the potentially sensitive questions	20
3.6.2 Answers to the potentially sensitive questions.....	20
Section 4 Empirical model.....	21
4.1 Permutation test for comparison of means.....	21
4.2 Heterogeneous treatment effects and model with interaction effect	22
4.3 Multiple hypothesis testing adjustment	23
4.3.1 Families of hypotheses	23
4.4 Robustness	23
4.4.1 Research-assistant-specific effects and interpreter-usage effects	24
4.4.2 Control for covariates.....	24
4.4.3 Combination of testing for covariates and RA-specific effects	25
4.4.4 Controlling for observable demographic characteristics and their interaction with treatment	25
Section 5 Results	25
5.1 Treatment effects.....	25
5.1.1 Treatment effects on non-response rate to the potentially sensitive questions.....	25
5.1.2 Treatment effects on answers to the potentially sensitive questions.....	26
5.2 Heterogeneous treatment effects.....	27
5.2.1 Heterogeneous treatment effects on non-response rate	27

5.2.2	Heterogeneous treatment effects on answers to the potentially sensitive questions.	28
5.3	Robustness	30
Section 6	Discussion	30
6.1	Changes in the non-response rate to the potentially sensitive questions.....	31
6.2	Changes in the answers to the potentially sensitive questions	31
6.2.1	Effects on the ‘donating and buying’ outcome	32
6.2.2	Limitations	34
6.2.3	Power and internal validity	35
6.2.4	External validity	35
6.3	Implications for survey design.....	36
6.4	Possibilities for future research.....	38
Section 7	Conclusion	38
Section 8	Reference List	41
Section 9	Appendix.....	43
9.1	Note on the pre-analysis plan	43
9.2	Informed consent	44
9.3	Identifying and demographic questions and answer options	45
9.4	Outcome variables composition.....	48
9.5	Summary statistics.....	52
9.6	Models with interaction effect results	54
9.7	Comparison of permutation test and t-test results for the main outcomes	54
9.8	Treatment effects on individual potentially sensitive questions	55
9.9	Heterogeneous treatment effects on individual potentially sensitive questions	56
9.10	Robustness results.....	57
9.11	Pilot interviews images.....	60
9.12	Map of the interview locations	63
9.13	Power.....	64
9.13.1	Power accounting for multiple hypothesis testing adjustment.....	65

List of tables

- Table 1: Results of experiments introducing comparable survey design variation 12
- Table 2: Groups in the sample..... 16
- Table 3: Assignment to treatment per subgroup 17
- Table 4: The identifying questions 19
- Table 5: Multiple hypothesis testing families and adjusted significance levels..... 23
- Table 6: Treatment effect on non-response rate to the potentially sensitive questions 26
- Table 7: Treatment effect on changes in answers to the potentially sensitive questions..... 27
- Table 8: Heterogenous treatment effect on non-response rate to the potentially sensitive q..... 28
- Table 9: Heterogenous treatment effects on changes in answers to the potentially sensitive q. 29
- Table 10: Treatment questions and answer options..... 45
- Table 11: Outcome variables composition..... 48
- Table 12: Summary statistics – treatment variable, leader variable and their interaction 52
- Table 13: Summary statistics – knowledge variables..... 52
- Table 14: Summary statistics – satisfaction with chief..... 52
- Table 15: Summary statistics – activity and interest variables 52
- Table 16: Summary statistics – donating and buying variables 52
- Table 17: Summary statistics – Non-response to potentially sensitive questions..... 52
- Table 18: Summary statistics – Not-answering to identifying variables 53
- Table 19: Summary statistics – identifying and demographic variables and other covariates..... 53
- Table 20: Summary statistics – Giving name and contact information..... 54
- Table 21: Results – model with interaction effects – effects on answers to the potentially sen. q. 54
- Table 22: Comparison of permutation test and t-test results..... 54
- Table 23: Treatment effects on individual potentially sensitive questions 55
- Table 24: Treatment effects on individual potentially sensitive questions – common citizens 56
- Table 25: Treatment effects on individual potentially sensitive questions – leaders..... 57
- Table 26: Donating and buying outcome robustness 58
- Table 27: Robustness analysis - models with significant treatment or subgroup interaction effects.. 59

Section 1 Introduction

Surveys are undoubtedly one of the most commonly used data collection methods in development research. Surveys can range in size from small for a single propose to large scale multi-country repeated surveys, like the barometer surveys. Equally wide-ranging is the usage of the data collected using surveys, varying from having minor implications to important policy implications. Therefore, survey data quality is critical and has relevant implications for many researchers and policy makers.

The main concerns in response quality of survey answers are an item response rate (or lack thereof which is referred to as the item non-response rate), and a response biases. This study focuses on the item non-response rate, which is defined as the proportion of respondents refused to answer the question, and one specific response bias, the social desirability bias. Social desirability bias in survey answers can be defined as a tendency to over-report socially desirable behavior and under-report socially undesirable behavior (Krumpal, 2011; Nederhof, 1985) . Social desirability is observed especially in answers to sensitive questions (Krumpal, 2011).

For a respondent facing a survey question, there are two decisions to be made when answering. First is the decision whether to answer or not to answer the question¹. The second decision to make if answering the question is how to respond, whether it be with a truthful or inaccurate response. The second decision is conditional on choosing to answer in the first decision. The first decision leads directly to non-response rate. The second decision, in the case of an untruthful answer, leads to a bias in the responses. If the second decision is motivated by self-presentational motives, it leads specifically to a social desirability bias in the answers.

Both non-response rate and social desirability bias can have serious implications on survey data quality. Increased non-response rate is directly related with decreased sample size, and consequently the power of a study. Furthermore, systematic non-response can lead to a form of selection bias in answers by changing the study population.

Social desirability bias, if present, can lead to incorrectly measured outcomes such as attitudes, knowledge or occurrence of behavior. Moreover, social desirability bias can also mask the true relationship between different variables by decreasing variance of the outcomes, or by systematically changing the outcomes. Consequently, it can lead to serious issues when analyzing the data, such as finding spurious relationships between variables in case of systematic social desirability bias² (Podsakoff et al., 2003). Therefore, it is imperative for researchers to mitigate social desirability bias in order to provide accurate data and analysis on the topic being surveyed.

As lack of anonymity is commonly associated with higher social desirability bias (Krumpal, 2011) the methods used to cope with a social desirability bias generally include some form of increasing anonymity or confidentiality. Assurances of confidentiality or anonymity and informing respondents

¹ In our case non-responses are recorded as 'Do not know' or 'Do not want to answer' options, depending on the respondent's reason not to provide the answer.

² Simple example can be used to illustrate mechanism of social desirability bias leading to spurious relationship. Certain religion might prohibit unmarried sex while generally it might be accepted by others in the society. Therefore, when asked about unmarried sex, the respondents holding this religion would have stronger reasons to under-report unmarried sex than the rest of the population. Therefore, the data would show correlation between holding this religion and having less unmarried sex. However, in situation when all respondents regardless of religion would have on average the same amount of unmarried sex, the relation found between the religion and unmarried sex could actually be explained purely by social desirability bias. Non-response induced selection bias could have a similar effect, if for example, only respondents holding the religion and have unmarried sex would refuse to answer the question.

there are no correct or incorrect answers are a standard part of informed consent policies that researchers use. Not having to provide name and contact is found to lead to less socially desirable answers (Ong and Weiss, 2000). Furthermore, the mode of questionnaire delivery can be used to decrease social desirability bias. Self-administrated questionnaires generally lead to lower social desirability bias than when the data is collected using face-to-face interviews (Langhaug et al., 2010; Nederhof, 1985). Moreover, different techniques can be used to decrease social desirability bias when collecting data using a face-to-face interview. This is done by varying anonymity of the response itself. These techniques include randomized response³, unmatched count⁴ or bogus pipeline⁵ techniques. However, these techniques require presence of additional elements during the surveying; that is an additional list of questions, a randomness generating device or a 'true measuring' device. The extra effort needed for each technique makes these techniques more difficult to implement. Further, unmatched count and randomize response lead to loss of information about answer of a particular respondent, as only mean answers for a sample can be inferred when using these techniques.

This study aims to investigate an easy-to-implement variation in a survey design leading to possible change in survey data quality of the potentially sensitive questions. This variation is asking identifying and demographic questions at the very beginning of the survey or asking these questions at the very end of the survey. That is the identifying questions are either asked or they are not asked prior to the potentially sensitive questions. The identifying questions include questions asking for name, phone number, age, education, literacy and tribe as well as questions asking for other demographic information. Different attitudinal questions and questions about various institutions are expected to be potentially sensitive questions in the cultural setting of this study.

Research question:

Does asking identifying questions prior to potentially sensitive questions, as opposed to not asking prior, lead to changes in answers to the potentially sensitive questions?

We hypothesize the potential effect of this variation on the answers to the potentially sensitive questions to be driven by changes in perceived anonymity by the respondents. Asking identifying questions directly violates anonymity of the answers as the identity of the respondent is uncovered. The tendency of humans to change behavior when being observed and facing different levels of anonymity is well suggested by experimental evidence (Bateson et al., 2006; Haley and Fessler, 2005; Hoffman et al., 1996; Ong and Weiss, 2000; Rigdon et al., 2009).

³ When using randomized response, the respondent's answer is connected to a random event (Krumpal, 2011), for example a flip of coin or dice roll. The respondent can in the simplest form be asked to throw a coin in such way only s/he can see it. Further, s/he is told to answer 'yes' (or generally the undesirable way) if tails are thrown and truthfully if heads are thrown. The proportion of answers than can be estimated from the difference of the stated proportion among all respondents and the expected outcome of the random event. For our coin example expected proportion of yes would be 50%, therefore if 80% people would answer 'yes', our estimated proportion of yes in the study population would be $(80\% - 50\%) / 50\% = 60\%$.

⁴ When using unmatched count, also called list experiment, respondents are randomly divided into two groups and presented several statements then asked to answer how many are true (Krumpal, 2011). List presented to one group includes an additional statement of interest, such as mentioned illicit drug use. The difference in mean number of statements said to be true, between the two groups can be used as an estimate of proportion of true answers to the statement of interest among the study population.

⁵ For the bogus pipeline method the respondent is lead to believe an 'objective' measure (for example heart rate monitor or similar) is being used to validate the answer and to detect liars (Nederhof, 1985). An example of this would be polygraph test, however no ability of the device to actually detect lie is required.

Several other studies introduce the variation in position of identifying questions in survey questionnaires (Drummond et al., 2008; Giles and Feild, 1978; Green et al., 2000; Roberson and Sundstrom, 1990; Teclaw et al., 2012). The findings of these studies did not find evidence of a strong common effect. Different effects of this variation on a study response rate were observed by different studies. Additionally, significant effects were almost never found on demographic items response rate, outcome item response rate or answers to the outcomes.

However, unlike any of the previously mentioned studies, this study aims to investigate the effects of the placing of the identifying questions in a face-to-face interview in the setting of a developing country, Ghana. Whereas the previous studies focused on self-administrated, that is handed out, postal or web, surveys in western countries setting. Furthermore, our study additionally includes in the survey variation, aside similar demographic questions as used by the previously mentioned studies, also questions asking for a name of the respondent, a phone number, and a willingness to participate in the next round of a questionnaire. Findings of this study could therefore differ from the previous studies due to several reasons. First, face-to-face interviews generally suffer from higher social desirability bias than self-administrated questionnaires (Langhaug et al., 2010). Second, specific culture might influence desire for social approval and socially desirable reporting (Johnson and van de Vijver, 2003; Krumpal, 2011). Lastly, the provision of name as part of the survey variation leads to potentially higher effects on perceived anonymity.

Finding a significant effect of such variation would provide additional evidence of the tendency of humans to change their behavior with different levels of anonymity. Specifically, in the case of observability of answers in the setting of face-to-face interview in Ghana. Deriving further implications depends on assumption that socially desirable reporting happens in a predictable and common direction for all or majority of the respondents. Further, it is needed to assume that we correctly know this direction. Under these assumptions, a lower proportion of socially desirable behavior indicates higher response accuracy (Ong and Weiss, 2000). Given these assumptions, findings of this study lead to implications for researchers designing surveys. Specifically, surveys with potentially sensitive questions that are to be asked during face-to-face interviews in developing countries.

This research is designed as a survey experiment with two randomly-allocated treatment conditions. The demographic and identifying questions are either asked prior to the potentially sensitive questions, or only after the potentially sensitive questions. To answer the research question, we use a sample population of 3528 respondents of a survey investigating attitudes around natural resources revenues and transparency of their management. Our sample consists of citizens and local leaders from 120 districts across Ghana. All 32 districts classified as mining or oil districts had been selected, further 88 districts had been selected using probability-proportionate to population sampling. The local leaders included in the sample are local politicians and traditional authorities. The responses had been collected using tablet computers and a research assistant-administrated questionnaire.

Our outcomes are constructed using answers to potentially sensitive questions with a hypothesized uniform direction of the potential social desirability bias. The potentially sensitive questions include questions about knowledge of various institutions, claimed and planned activity, willingness to donate to an NGO or satisfaction with traditional leaders. These questions have been collected in a survey used for larger oil revenues transparency research and have not been selected with a purpose of conducting this experiment. Therefore, the type of questions asked should be comparable to other knowledge and attitudinal surveys in developing countries. We construct two outcomes measuring

non-response rate to these questions and four outcomes to measure change in answers and potential increase in social desirability.

We find no effect of asking the identifying questions prior to the potentially sensitive questions as opposed to not asking them prior to the potentially sensitive questions on the non-response rate to the potentially sensitive questions. However, we observe increase in social desirability bias in answers to sensitive questions about donating to an NGO and buying a transparency report, if the identifying questions are asked prior to these questions. There is no effect observed on the remaining three outcomes. This however, is only the case of the non-purposefully selected respondents (the common citizen subsample).

Our findings suggest it might be advisable not asking the identifying questions prior to the potentially sensitive questions in case of face-to-face interview in cultural setting comparable to Ghana.

In Section 2 we provide background for this study and present a mechanism in which we hypothesize the treatment to have an effect. Further, in Section 3 we describe our design in more detail and in Section 4 we explain the empirical model used to test our hypotheses. Subsequently, in Section 5 and Section 6 we present the results and discuss the findings. Finally, Section 7 concludes the study.

Section 2 Background, mechanism and hypothesis

2.1 Background

Anonymity in a survey setting is a state in which the respondent cannot be observed and the answer of that respondent cannot be identified back to them in any way. Provision of name is one of the most direct violations of anonymity. However, any other information increasing the probability of identification can lead to decrease in anonymity of the respondent. For example, in the case of small communities, a simple combination of gender, age and community name can be used to identify a person. Therefore, asking identifying and demographic questions prior to the potentially sensitive questions directly violates anonymity of the responses.

Anonymity is essentially the level towards which the respondents' answer can be observed. Therefore, the effects of anonymity on respondents' behavior derive from the tendency to change behavior when being observed. Change in behavior when one is being observed is observed due to intuitive or deliberate tendency to maintain social reputation, expectations of reciprocity of such action or due to consideration of material benefits deriving from actions of the observer (Hoffman et al., 1996; Krumpal, 2011; Milinski and Rockenbach, 2007). The tendency of humans to change behavior when being observed and less anonymous is well suggested by experimental evidence.

Even very small cues of being watched result in positive effects on cooperation in an economic lab game (Haley and Fessler, 2005; Rigdon et al., 2009) as well as in a field experiment (Bateson et al., 2006).

The tendency of research participants changing behavior also holds when observability of actions by the researchers is manipulated. Contributions in a dictator game are found to significantly decrease with the level in which participants' actions can be observed, or inferred, by the researchers (Hoffman et al., 1996).

Such tendency is also found in non-western contexts. Cilliers et al. (2015) observed significant effects of white foreigner presence on behavior in a dictator game in Sierra Leone. Cilliers et al. (2015) argue, that the participants adhered to 'researcher demand effects'. The respondents were found to share more in the game when in the presence of the foreigner. While they are found to share less in

his presence, if they believed they were being tested for aid suitability. This reflects two directions of the demand effects. First is tendency to present oneself in a favorable way, where the opposite effect is driven by consideration of the potential aid benefits and wanting to look in need of the aid. (Cilliers et al., 2015).

Similar pattern of behavior is prevalent in answering surveys as well. Higher degree of anonymity is found to have a positive effect on decreasing response biases in answers to sensitive questions. Evidence suggests that anonymity leads to considerably less socially desirable responses to sensitive questions, as opposed to confidentiality (Ong and Weiss, 2000). Furthermore, similar results are observed when the level of assurances of confidentiality of answers is varied. Respondents given higher assurances tend to have a small decrease in social desirability in answers to sensitive questions (Singer et al., 1995). However, no such effect is found for non-sensitive questions (Singer et al., 1995). The effects of anonymity on survey answers on sensitive questions are likewise observed in non-western context (Langhaug et al., 2010).

However, the changes in behavior become much less clear once introducing only a small variation in the survey, like changing order of identifying and potentially sensitive questions. Giles and Feild (1978), Roberson and Sundstrom (1990), Green et al. (2000), Drummond et al. (2008) and Teclaw et al. (2012) also investigated the effects of different placings of identifying questions in a survey on responses. These studies tried to determine whether it is preferable to put them prior to the non-identifying questions, or whether to keep them after. There were inconclusive findings from these studies as the effect on a study response rate is concerned. Furthermore, none of these studies observed any effects on the potentially sensitive question. The findings of these studies are summarized in the next few paragraphs and visualized in Table 1.

Giles and Feild (1978), Green et al. (2000) and Drummond et al. (2008) used mail surveys among faculty members, social workers and physicians respectively to conduct their experiments. Roberson and Sundstrom (1990) did so in a survey handed out to office workers. While Teclaw et al. (2012) researched the effects of this variation on a web survey sent out to Veteran Health Administration employees.

Most of the studies focused on effects of the position of demographic questions on the study response rate (participation rate). For surveys that were mailed or handed out this was represented by the proportion of questionnaires that returned to the researchers. Drummond et al. (2008) found positive effect when asking demographic questions first on study response rate. On the other hand Roberson and Sundstrom (1990) observed negative effects. Whereas, Giles and Feild (1978) and Green et al. (2000) concluded there was no statistically significant effect. Drummond et al. (2008) explain their findings with a difficulty to fill in the first questions. For the physicians participating in the Drummond's et al. (2008) survey, it might be easier to fill in questions about themselves and their practice before more complicated topic-specific questions. Therefore, if these questions are asked first, it is easier to start filling in the survey. On the other hand, Roberson and Sundstrom (1990) hypothesized that the effect they observed was due to increased perceived identifiability of the responding employees.

Furthermore, several of the studies focused on the item response rate of the demographic questions. Green et al. (2000) observed no statistically significant effect of the different placing of identifying questions. However, Teclaw et al. (2012) found a positive effect of asking demographic questions in the beginning of a survey on the response rate to these questions in a web-administrated survey. The difference between the two cases is that Teclaw's et al. (2012) respondents could not see the

subsequent questions in the web survey. Whereas the respondents of Green's et al. (2000) mailing survey, could easily take a look at the whole questionnaire before starting to fill it in.

Table 1: Results of experiments introducing comparable survey design variation

Study	Direction of effects of placing the demographic questions prior to the outcome questions in a survey on:			
	Study response rate	Demographic items response rate	Outcome items response rate	Outcome items answer means
Giles and Feild (1978)	≈	X	X	≈
Roberson and Sundstrom (1990)	–	X	X	≈
Green et al. (2000)	≈	≈	X	X
Drummond et al. (2008)	+	X	X	X
Teclaw et al. (2012)	X	+	≈	≈
Our study				
Hypothesis	X	≈	+	+
Results	X	≈	≈	+
<p>This table shows the findings of studies using the same survey design variation. Direction of effects of placing the demographic questions prior to the rest of the questions in a survey found by the study is indicated by the following signs.</p> <p>X : the study does not research this outcome + : significant positive effect of placing demographic questions first found for the outcome – : significant negative effect of placing demographic questions first found for the outcome ≈ : no significant difference observed for the outcome</p>				

Finally, Giles and Feild (1978), Roberson and Sundstrom (1990) as well as Teclaw et al. (2012) concluded no effects of this questionnaire variation on changes in mean answers to the non-demographic questions. Further, Teclaw et al. (2012) observed no effect on item response rate of these items.

All of the studies had been conducted in the United States, with the exception, of Drummond et al. (2008), who used sample from Ireland. Unlike these studies, our study focused on a non-western society sample of respondents in Ghana. Furthermore, these studies all used self-administrated questionnaires, while we use interviewer-administrated questionnaire⁶. This is a major difference compared to the previously mentioned studies because, nobody else directly observes and records the respondents' answers, like our research assistants do. Next, our survey variation includes not only the demographic questions, but also questions asking for a name of the respondent, a phone number, and a willingness to participate in the next round of a questionnaire. These differences are connected to variations that are found to influence social desirability. Therefore, the findings of our experiment could differ from the previously conducted experiments. First, face-to-face interviews

⁶ In Ghana, similarly to many other developing countries, usage of self-administrated, i.e. handed-out, postal, or web, surveys is not feasible, due to lower literacy rate. About 25% of our respondents are illiterate. Audio computer-assisted survey instruments had been used to overcome this issue and deliver self-administrated questionnaires in developing countries. However, these instruments are connected to higher costs of such data collection. These methods have been commonly implemented in investigating sexual behavior in developing countries, see Langhaug et al. (2010) for details.

generally suffer from higher social desirability bias than self-administrated questionnaires (Langhaug et al., 2010). Next, culture might influence desire for social approval and socially desirable reporting (Johnson and van de Vijver, 2003; Krumpal, 2011). Lastly, the provision of name as part of the survey variation leads to potentially higher effects on perceived anonymity.

Furthermore, this study puts emphasis on changes in answers to the potentially sensitive questions then the previous studies did. Focusing on participation rate is not applicable in our case, as participation is agreed on by the respondent and the research assistant prior to any variation in the questionnaire was allocated. Therefore, we focus also on the non-response to items and the changes in answers to the potentially sensitive questions. We expect it to be more likely that respondents would not answer a question or provide an untruthful answer rather than quit the interview completely. Information from research assistants suggests a low dropout rate. However, this information was not possible to record and therefore rigorous investigation of the dropout rate is not possible.

2.2 Mechanism and hypothesis

In the case of face-to-face interview, we identify two distinct dimensions of anonymity. The first comes from the interaction with the research assistant, and any other people who could potentially overhear the answers at the interview site. This dimension is not expected to change with position of the identifying questions. The respondent is directly observed providing the answers and therefore no further identification is needed. Thus, this dimension of anonymity will not be further considered. The second dimension of anonymity comes from the fact that someone reads the respondents' answers in the data. Where completely anonymous data would include no identifying information. Asking the identifying and demographic questions prior to the potentially sensitive questions causes there to be identification with the answers. Thus, the observability of these answers is salient to the respondent and therefore perceived anonymity decreases. In accordance with the presented experimental evidence on effects of observability and anonymity on human behavior, we hypothesize our respondents change behavior when facing different levels of perceived anonymity.

The respondents concerned about perception by, or actions of, the one who they believe might see the data, are more likely to answer in such a way to be perceived desirably by others. The one believed to be seeing the data could be the researcher, government, development organization or similar actors. We believe the motivation of our respondents for changing the answer is the desire to be perceived as a 'good citizen' or a 'good community'. Alternatively, respondents could believe projects funded from oil money, as well as other government or development funding, might be distributed based on answers to questions in the survey. These two motives are essentially the same motives hypothesized by Cilliers et al. (2015) to play a role in their experiment.

The desired image of self can be manipulated by the respondent through not answering the potentially sensitive questions, or by answering such questions untruthfully. This potentially leads to increased non-response rate and social desirability bias in the responses to the potentially sensitive questions. We assume that the socially desirable changes in given responses happen in a predictable and common direction. Given this assumption, less anonymity leads to higher proportion of socially desirable answers, thus giving the researcher less accurate results.

Even though the confidentiality assurances given at the beginning of the interview, it likely is still believed by the respondents that the answers are not anonymous. As the research assistant directly observes the answer. Furthermore, the respondents are likely to realize that the provision of the identifying information leads to possibility of different people observing their answer from the data.

Given the mechanism described, we specify the following hypotheses about the treatment effects:

2.2.1 Hypothesis 1:

The answers to potentially sensitive questions change if the identifying questions are asked prior to the potentially sensitive questions as opposed to not being asked prior to these questions.

Hypothesis 1A: There is higher non-response rate in answers to potentially sensitive questions if identifying questions are asked prior to the potentially sensitive questions as opposed to not being asked prior to these questions.

Hypothesis 1B: There is higher social desirability bias in answers to potentially sensitive questions if identifying questions are asked prior to the potentially sensitive questions as opposed to not being asked prior to these questions.

No previously conducted study observed effects the similar treatment variation corresponding either to the Hypothesis 1A or the Hypotheses 1B. Teclaw et al. (2012) observed no significant effects of this treatment variation on the non-demographic items' non-response rate, this is in contrast with Hypothesis 1A. Furthermore, Giles and Feild (1978), Roberson and Sundstrom (1990) and Teclaw et al. (2012) found no effect of the treatment variation on answer means to the non-demographic items. However, this study is conducted in a different cultural background and with a different mode of administration, both of which are found to influence social desirability (Krumpal, 2011; Langhaug et al., 2010). Therefore, we believe differences in design of our experiment and the cultural context of this study could lead to different findings.

2.3 Heterogeneous treatment effects

The hypothesized effect of asking or not asking the identifying questions prior to the potentially sensitive question arises out of differences in perceived anonymity. However, not all the respondents in our sample are identified to the same extent and therefore should have initially different level of perceived anonymity.

Our sample includes two distinct subgroups, 'leaders' and 'common citizens'. The respondents in the leader category included chiefs, District Assembly members, Unit Committee members and opinion leaders. During the respondent selection, the 'leader' respondents are specifically sought for by the research assistants based on their position in the community. The district assembly member is contacted by phone and asked to participate in an interview. The rest of the leader respondents are introduced to the research team by the District Assembly member. The respondents are told that they have been specifically selected due to their position in the community.

This sampling procedure makes salient the fact that the 'leaders' are already identified. These respondents know they answer in 'their name' or 'in the name of their position' before they even begin the survey. On the other hand, respondents in the 'common citizen' category are selected based on a presence within public areas of the community reflecting a convenience sampling method which is discussed in Section 3.2. Therefore, these respondents have less reason initially to believe they are identified with the responses they provide. This holds true until the moment when they provide the identifying information within the survey.

Consequently, with overall higher identification in the leader subgroups, the treatment had little to no room to influence perceived anonymity within the leader subgroup. This is consistent with the following hypotheses about heterogeneous treatment effects:

2.3.1 Hypothesis 2:

If the identifying questions are asked before the potentially sensitive questions as opposed to not being asked prior to these questions, there will be a less pronounced change in answers to the potentially sensitive questions for the leader subgroup compared to the common citizen subgroup.

Hypothesis 2A: If the identifying questions are asked before the potentially sensitive questions as opposed to not being asked prior to these questions, there will be a less pronounced increase in non-response rate to the potentially sensitive questions for the leader subgroup compared to the common citizen subgroup.

Hypothesis 2B: If the identifying questions are asked before the potentially sensitive questions as opposed to not being asked prior to these questions, there will be a less pronounced increase in social desirability bias in answers to the potentially sensitive questions for the leader subgroup compared to the common citizen subgroup.

Section 3 Data, research design and sample

3.1 Data

Data for this research have been collected as part of a baseline questionnaire of a larger impact evaluation project. The aim of this project is to evaluate impact of meetings organized by the Public Interest Accountability Committee (PIAC) and to test a new ICT information campaign. PIAC is a Ghanaian public body charged with responsibility to inform citizens about usage of oil revenues. In order to do so, the PIAC organizes meetings for local leaders (politicians, traditional leaders, journalist and other). Following this data collection, such meetings have been organized in 60 of the sampled districts. Further ICT platform sending voice and SMS messages to various recipients have been introduced. These messages inform about oil revenues and the PIAC itself. The messages also include information about possibilities to ask questions back.

The questionnaire aims to find base level of knowledge and attitudes towards natural resources revenues and transparency and accountability surrounding their management⁷.

The data were collected using questionnaire administrated using tablet computers in local languages by trained research assistants between May and August 2016. The research assistant team was led by team leaders employed by Kumasi Institute of Technology, Energy and Environment (KITE). The team leaders all had experience with conducting surveys. Further, majority of the research assistants have been students from Department of Planning of Kwame Nkrumah University of Science and Technology and had prior experience with collecting survey data.

Figure 2, Figure 3 and Figure 4 in Appendix 9.11 capture images of interviews conducted during the pilot survey stage.

3.2 Sample description and sampling strategy

The sample includes in total 3528 respondents from 120 out of 216 districts in Ghana. The 32 districts with larger mining activity and districts with offshore oil production have been all included; the remaining 88 districts have been selected randomly with probability proportionate to population sampling.

⁷ The full questionnaire is available on request at the following email address: maarten.voors@wur.nl

In each of the districts, five electoral areas have been selected randomly from a list obtained at the District Headquarters. In each electoral area the chief, the District Assembly member, a Unit Committee member⁸ and an opinion leader⁹ were interviewed. The District Assembly member in each community pointed the team to one of the Unit Committee members and selected an opinion leader. In addition, two common citizens were selected in the communities using convenience sampling based on their presence within public areas of the community with a dispersion created by walking a pre-set distance in two opposite directions. In case some of the ‘leader’ respondents could not be found, another opinion leader or citizen was interviewed.

This sampling strategy lead to sample of 388 to 764 respondents for each of the leader groups, 2315 for leaders in total, and 1213 ‘common citizens’. The precise numbers of respondents in each category can be seen in Table 2.

Locations of the interviews on a map of Ghana can be seen in Figure 5 in Appendix 9.12.

Table 2: Groups in the sample

Group	Frequency	Percentage
<u>Common citizens</u>	<u>1,213</u>	<u>34.38</u>
Unit Committee members	605	17.15
District Assembly members	558	15.82
Traditional authorities (Chiefs)	388	11.00
Opinion leaders	764	21.66
<u>Leaders total</u>	<u>2,315</u>	<u>65.62</u>
Total	3,528	100.00

3.3 Assurances of confidentiality and informed consent policy

We implement informed consent policy designed to meet all the ethical standards and to lead to as low biases in answers as possible, while recording all the information necessarily. This informed consent policy is like other informed policies commonly used when conducting surveys.

Respondents of our survey are briefed about the conditions of participation prior to the start of the surveying. Purpose of the research project (not this experiment) and the use of the data are introduced first. Further, the respondents are ensured the participation is being voluntary and that their data will be safely stored and anonymous. Moreover, participants are told that they can refuse to answer any question or stop completely at any time. It is also emphasized that there are no wrong or correct answers and they should express themselves freely.

Further the participants have been informed that we will ask for their name, contact information and that those will be recorded separately from the data¹⁰ and that they do not have to give us their name and contact, if they prefer not to. This was repeated once more when the name and contact information were asked (position depended on the treatment condition).

⁸ Unit Committee is the lowest level of local government in Ghana. Five elected members are governing electoral areas of about 500 to 2500 inhabitants.

⁹ Opinion leader is a knowledgeable and respected member of the community.

¹⁰ The name and contact were recorded on a paper sheet, with clear distinction being made from recording the survey responses to the tablet computer. The sheets have been stored securely at KITE’s offices in Accra, for the purpose of contacting the respondents for potential participation in the endline survey. The data and the name-sheets are only connected using ID number, preventing (simple) identification of the respondents using only the data file.

All participants had to agree with the informed consent, in case they did not, the interview was ended. Full informed consent can be found in Appendix 9.1.

During the enumerator training, the informed consent policy and the anonymity assurances have been repeatedly explained to the research assistants conducting the interviews. Further field training and pilot data collection were organized prior to the data collection. During this stage, the interviews conducted by the research assistants have been observed and potential mistakes had been discussed.

Further, during the training, the research assistants have been also explained the purpose of the impact evaluation and meaning of all the questions. The questions have been debated both in English and Akan, the most common language in Ghana. Debriefing clarifying any issues was organized after the field-training before the teams continued with the data collection.

However, at last two research assistants that have not undergone training with the rest had to be included in the data collection at later stage due to organizational reasons.

3.4 Research design

This research is designed as a survey experiment with two treatment conditions. Each respondent is assigned to one of the two treatment conditions randomly, with equal probability. This is conducted automatically by the tablet computer using random number generating function. Importantly, research assistant has no control over the selection to treatment. This process results in probabilities of assignment to the two treatment conditions independent on each other among all the respondents and groups of respondents. Therefore, the resulting treatment groups are very close to half of the sample but not exactly half. Randomization also ensures the probability of treatment assignment to be independent on any other observable or unobservable characteristics. Therefore, any significant differences in answers to the questions among the treatment groups can be attributed to the effect of the treatment.

Based on the sampling procedure we define two subgroups in our sample. One is the group of ‘common citizens’, which were sampled using convenience sampling. The second group is the group of ‘leaders’, which includes the purposefully selected local leaders. That is the chiefs, the District Assembly members, the Unit Committee members and the opinion leaders. The resulting assignment to treatment per subgroup can be seen in Table 3.

Table 3: Assignment to treatment per subgroup

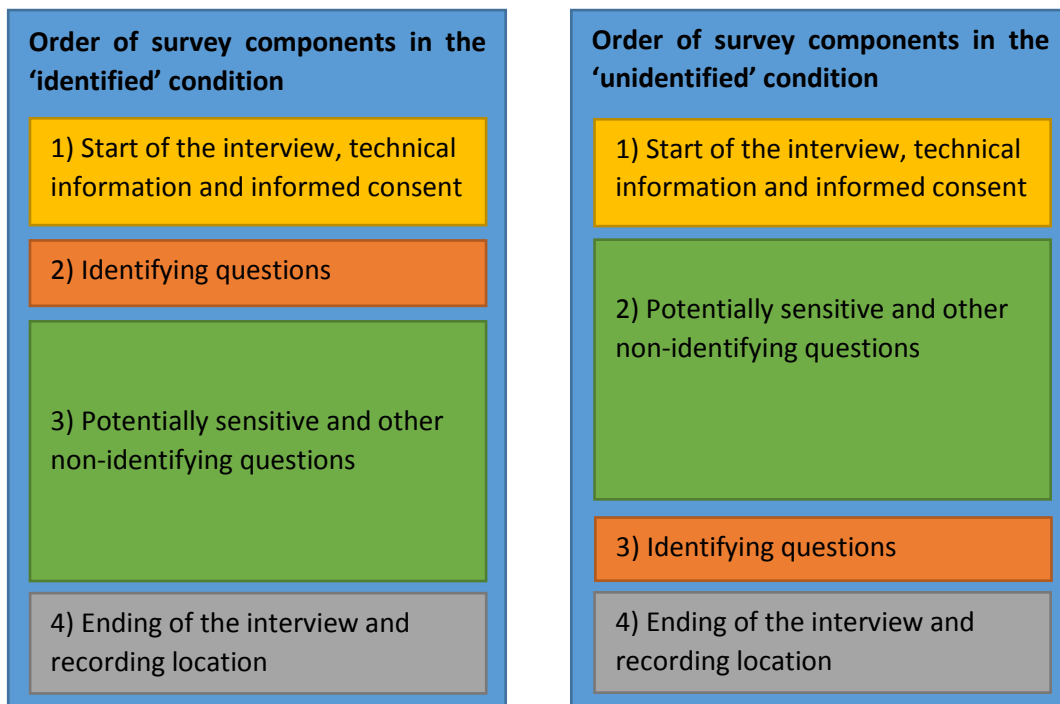
		Treatment condition		
		Identified	Unidentified	Total
Sample subgroup	Common citizen	627 (51.69%)	586 (48.31%)	1,213 (100%)
	Leader	1,152 (49.76%)	1,163 (50.24%)	2,315 (100%)
	Total	1,779 (50.43%)	1,749 (49.57%)	3,528 (100%)
Number of respondents in each treatment group within the sample subgroups. Frequencies of treatment within the subgroup in parentheses.				

3.5 Treatment variation

There are two treatment conditions in this experiment. In the first condition, there are identifying questions asked prior to the potentially sensitive questions. While in the second condition there are no identifying questions asked prior to the potentially sensitive questions. In the first condition the identifying questions are asked at the beginning of the questionnaire. In the second condition, these questions are asked at the end of the questionnaire, only after the potentially sensitive questions. Figure 1 illustrates the survey variation.

When reporting the results, the condition where the identifying questions are asked prior to the potentially sensitive questions is referred to as the 'identified' condition and the other condition is referred to as 'unidentified' condition.

Figure 1: Illustration of the treatment variation



The identifying questions include questions asking for name, phone number, participation in a next round of the survey, age, education and literacy, social status, tribe and language, occupation, frequency of travelling to different places and household characteristics. There are in total 20 identifying and demographic (treatment) questions included in the survey variation. List of the treatment questions as they followed can be seen in Table 4. Together with the answer options the questions can be seen in the Table 10 in Appendix 9.3.

Table 4: The identifying questions

Name of variable	Code	Question
-	S1	<i>PERSONAL BACKGROUND AND HOUSEHOLD INFORMATION SECTION</i>
-	S1 intro	First (Finally), I'll ask you some questions about your personal background and your household
-	S1Q0 note	We would like to ask you for your name and phone number. Your name and phone number will be recorded on separate sheet of paper and kept aside of the rest of the data, so your anonymity would be ensured. You can also refuse to give these to us. If you give us your phone number, we would potentially contact you. You might also receive information about oil revenues governance. Moreover, would you be interested to participate in a subsequent round of this research in about a year's time? If yes, please indicate this and fill your contact details on the same sheet of paper.
Name	S1Q0a ¹¹	Did the respondent agree to give her/his name and did you record the name and the respondents assigned ID on the separate paper?
Phone number	S1Q0b ¹¹	Did the respondent agree to give her/his phone number and did you record the phone number on the separate paper?
Endline participation	S1Q0c	Agreed to participate in the next round of this questionnaire?
Age	S1Q1	How old are you?
Gender	S1Q2	Are you a male or female?
Relation to household head	S1Q3	What is your relation to household head?
Language	S1Q4	What language do you most commonly speak at home?
Tribe	S1Q5	What is your tribe?
Literacy in English	S1Q6	Can you read and write in your local language?
Literacy in local language	S1Q7	Can you read and write in English?
Education	S1Q8	What is your highest level of education?
Occupation	S1Q9	What is your main occupation?
Social positions	S1Q10	Do you hold any of the following positions in your community?
-	S1Q11	During the last 6 months, did you go to ____?
Traveling to regional capital	S1Q11a	Accra or regional capital
Traveling to district capital	S1Q11b	District capital
Travelling to other settlement	S1Q11c	Other settlement in your district
-	S2intro	Now I will ask you some questions about your household. A household is a group of people who presently eat together from the same pot, excluding temporary visitors.
Number of adults in the household	S2Q1	How many adults (people over 18 years old) currently live in your household?
Number of kids in the household	S2Q2	How many children (people under 18 years old) currently live in your household?
Living conditions	S2Q3	In general, how would you describe your households' present living conditions?
Electricity	S2Q4	On average, how many hours a day is electricity available at your home?

¹¹ This, question was not read out loud, it was simply evaluated by the research assistant.

3.6 Outcomes

This study uses two outcomes to measure the change in non-response rate to the potentially sensitive questions and four outcomes to measure the change in answers to these questions.

We identify 16 questions as being potentially sensitive with assumed direction of social desirability bias.

Overall, the questionnaire consists of over 100 questions that are asked as outcomes. However, many of these questions are only asked conditionally on answer to previous questions. Therefore, only the 16 selected questions are used. These are the questions that have been asked unconditionally, every respondent was asked these questions. Further, only the questions where we are able to identify which answers might be perceived as more socially desirable are selected.

3.6.1 Non-response rate to the potentially sensitive questions

We define two outcomes to measure the non-response to the potentially sensitive questions. First one (*V1: Non-response to potentially sensitive questions (pot. sen. q.) – dummy*) is a dummy variable, capturing the occurrence of “Do not know” and “Do not want to answer” answers among the questions comprising outcomes A to D. The second outcome (*V2: Non-response to potentially sensitive questions - count*) is defined as the number of “Do not know” and “Do not want to answer” responses among these questions.

3.6.2 Answers to the potentially sensitive questions

Four outcomes measuring changes in answers to the potentially sensitive questions had been created using answers to the 16 identified potentially sensitive questions.

We group the selected questions to four outcomes, based on their similarity. These were constructed in such way they so they would resemble the types of outcomes commonly used in survey research, such as knowledge, satisfaction or actions taken.

First outcome is ‘knowledge of institutions’ (A), which is created from questions about knowledge of five different institutions connected to oil revenues management and transparency. We assume higher knowledge is more socially desirable answer, as to seem ‘knowledgeable’ is positively perceived.

Second outcome is ‘satisfaction with chief’ (B). This outcome is likely the most sensitive one in our questionnaire. Chiefs hold prominent social position locally and relation to chief is sensitive question in Ghanaian society. Disagreeing or being dissatisfied with chief is likely to be perceived negatively by other community members. We assume agreeing with a chief is more socially desirable answer than disagreeing with a chief.

Third outcome is ‘activity and interest’ (C). It comprises of questions directed at election behavior as well as questions investigating interest in and actions taken around natural resources revenues. We assume that the respondents believe the researchers would like them to be active.

Last of the four outcomes is ‘donating and buying’¹² (D), which is created out of answers to questions asking about interest to buy an oil revenues management report and willingness to potentially donate to a transparency-promoting NGO, the local chapter of Transparency International. We

¹² Originally, as stated in the PAP, the outcome ‘Donating and buying’ was created using a level of the donation. However the level of donation is very influenced by small number of outliers. Therefore, we decided to use the dummy variable for the donation instead.

assume respondents wishing to present themselves in a good light to the researchers are going to answer they would be interested in the report and in donating more often, as they believe this is how the researchers would like them to behave.

The outcomes ‘knowledge of institutions’ (A) and ‘activity and interest’ (C) and ‘donating and buying’ (D) are constructed as a sum of normalized answers to the individual questions. Where $Normalized_{Xi} = \frac{Xi - mean(X)}{SD(X)}$ and X_i is an answer to the question for respondent i . This is done in order to sum up answers to questions of different types and means, while giving them the same weight. Finally, these three outcomes are scaled to have a minimum at 0 and a maximum at 1. Therefore, the value of these outcomes can be seen as a ‘proportion’ of maximal knowledge, a ‘proportion’ of maximal activity and interest, and as the ‘proportion’ of willingness to donate and buy report the respondent shows.

‘Satisfaction with chief’ (B) is answer to question asking whether the respondent agrees the chief does ‘a good job’. The answers are on scale from 0 to 4, where 0 is a complete disagreement and 4 is a complete agreement.

Furthermore, all the questions comprising the outcomes A-D are also analyzed individually during the subsequent stage of the analysis.

Detailed description of the composition of the outcome variables as well as a precise formulation of the questions and answers can be seen in Table 11 in Appendix 9.4.

Section 4 Empirical model

In this section, we discuss the empirical model chosen to evaluate the presence of the treatment effect. Various specifications of the model are developed in order to test the robustness of the results. Permutation tests are used to test the hypothesis in the simple mean comparison models. Parametric models are used to test the hypotheses for the more complicated models of robustness analysis.

4.1 Permutation test for comparison of means

The presence of treatment effect is tested using permutation mean-comparison test, also called randomization test or randomization inference. The permutation test tests the two-sided hypothesis, where H_0 is equivalence of means and under H_A the means are not equal.

Permutation tests can generally evaluate treatment effects on any test statistic. For mean-comparison test, this test statistic is simply an absolute value of a mean difference between the two treatment groups.

The permutation test tests a sharp null hypothesis, i.e. the null hypothesis is ‘no treatment effect on any respondent’. While t-test tests a null hypothesis of ‘no average treatment effect among all the respondents’ and thus t-test allows for heterogeneity of treatment effects among respondents (Young, 2015).

In the permutation test, treatment values are randomly reassigned large amount of times, in our case at last 10000 times. The newly created treatment groups are then compared same as the originally assigned groups were. Resulting p-value is the proportion of times when the new randomly assigned treatment lead to higher mean difference between treatment and control than our true treatment assignment (Ernst, 2004). The reassignment of treatment is done in the exact same way as

assignment of treatment in the first place. Each unit is assigned with 50% probability into one of the treatment conditions independently on the other units¹³.

Advantage of permutation test is that, unlike t-test, it relies on no distributional assumptions. As we use survey data with discrete scales, in which people often answer the extreme options, the assumption of normality is possibly violated. However, as implied by the central limit theorem, the distribution of sample mean converges to normal distribution as the sample size approaches infinity. Therefore, on large enough sample, t-test performs equally well as permutation test. Our sample is likely large enough in order for t-test to also perform reasonably well.

Disadvantage of permutation tests is that, as only a subset of the possible permutations of treatment assignment is used, this test results only in an estimation of a p-value. A confidence interval of the p-value can be calculated. We calculate it using exact binomial confidence interval¹⁴. The substantial number of permutations used ensures high precision of the estimated p-value.

Finally results of the permutation tests are compared to results of t-test in order to increase robustness of the results.

4.2 Heterogeneous treatment effects and model with interaction effect

To estimate heterogeneous treatment effects on different subgroups we use the mean-comparison permutation test, as described in the previous section, on the leader and common citizen subsample separately.

However, individual mean-comparisons fail to test for significance of the difference in the treatment effects on the leader and the common citizen groups. Therefore, in order to test the significance of the difference between the two groups, we also estimate the following linear regression model with treatment*subgroup interaction:

$$y_{id} = \alpha + \beta T_i + \gamma L_i + \delta(T \times L)_i + \varepsilon_{id}$$

Where y_{id} is an outcome variable, $i = (1, \dots, n)$ denotes respondent, T_i is the treatment dummy, L_i is leader dummy and ε_{id} denotes error term clustered over district.

In this model, the coefficient δ is the difference between treatment effect on leaders and non-leaders.

Our classification of a respondent as a leader or as a common citizen comes out of the research design. The different respondent categories have been needed for the impact evaluation questionnaire. Further, the selection of the respondent within the category was done prior to the random treatment assignment and thus they are independent of each other. Therefore, both the treatment variable as well as the leader*treatment interaction are exogenous.

¹³ This is in fact a Monte Carlo test, as only a subset of possible treatment assignments is used. The possible amount of treatment assignments in our case is higher than 10^{1000} , far beyond computational possibilities. Own Stata program is used to conduct this test.

¹⁴ P-value resulting from the permutation test is a binomial outcome. In each permutation, we compare the original test statistic (the absolute value of mean difference in our case) against the newly observed one. Result of each permutation is labelled 1 if it is higher or equal and 0 if it is lower than the originally observed test statistic. During each permutation, the treatment labels are randomly assigned, therefore the outcomes of individual rounds are statistically independent to each other. Therefore, exact binomial confidence interval can be used. Stata function `-cii-` is used to perform this calculation.

4.3 Multiple hypothesis testing adjustment

This study uses multiple outcomes and subgroups, therefore using standard significance levels for testing our hypothesis would lead to much higher overall family-wise error rate (FWER), than a desired significance level (e.g. Shaffer, 1995). Hence, the statistical tests used, must be adjusted to correct for the increase in FWER. To control for family-wise error rate we adjust the p-values for number of hypothesis tested using a Holm-Bonferroni correction. The Holm-Bonferroni correction implements a Bonferroni correction in a step-down testing procedure (e.g. Ludbrook, 1998). Where the number of hypotheses in a family is reduced (by one) every time a hypothesis is rejected. Bonferroni adjusted p-value is $\tilde{p} = mp$. Where m is the number of hypothesis tested and p is the unadjusted p-value of the hypothesis. This adjustment controls for FWER regardless the data structure (e.g. Ludbrook, 1998).

4.3.1 Families of hypotheses

Each stage of our analysis is considered as one hypothesis family (i.e. hypotheses considered as testing the same hypothesis). This is done in order not to decrease the power of the main parts of analysis by subsequent, more detailed, analysis. Further, separate families are used for tests conducted on the full sample and the two sub-groups.

The outcomes considered in a first family are the six outcomes, as described in Section 3.6 and in the Table 11 in Appendix 9.4. While the outcomes in the subsequent analysis family are the 17 potentially sensitive questions. Table 5 shows families of hypotheses used and resulting Bonferroni-adjusted significance levels for these families.

Table 5: Multiple hypothesis testing families and adjusted significance levels

Outcome family	Sample(s)	Outcomes in the family		Number of hypotheses in the family	Bonferroni adjusted α corresponding to $\alpha = 0.05$
Main outcomes	All	A, B, C, D, V1, V2		6	0.0083
Main outcomes – heterogeneous treatment effects (HTE)	Citizens and leaders			12	0.0042
Effects on answers to individual potentially sensitive questions	All	S9Q1 S9Q4 S9Q8 S9Q12	S4Q3b S6Q7 S6Q8 S8Q4	17	0.00294
Effects on answers to individual potentially sensitive questions - HTE	Citizens and leaders	S9Q16 S4Q1 S4Q2a S4Q2b S4Q3a	S14Q2 S14Q2_d S14Q3 S14Q4	34	0.00147

4.4 Robustness

We test robustness of our findings for research-assistant specific effects, controlling for covariates and controlling for treatment*observable characteristics interactions. Our design included selection of respondents within selected districts. Therefore, we use robust standard errors with clustering over district in all the models.

Robustness analysis is only conducted for the main outcomes (A to V2), not for all the individual variables.

4.4.1 Research-assistant-specific effects and interpreter-usage effects

Our interview is research-assistant administrated as well as the questions are translated from English to different local languages. Therefore, answers to the potentially sensitive questions potentially depend on research assistant (RA, enumerator)-specific effects. Research assistant influences the way the question is asked, how respondent feels about answering the questions or provides slightly different assurances of anonymity. All of which could have an effect on the answer. Similarly, it is the case also for usage of an interpreter during the interview. Therefore, we correct for RA-specific effects and effect of usage of interpreter in the following econometric model:

$$y_{id} = \alpha + \beta T_i + \gamma L_i + \delta(T \times L)_i + \mathbf{RA}_i \boldsymbol{\tau} + \sigma IN_i + \varepsilon_{id}$$

Where \mathbf{RA}_i are the RA dummy variables, IN_i is a dummy variable stating usage of interpreter during the interview, $i = (1, \dots, n)$ denotes respondent and $d = (1, \dots, 120)$ district, y_{id} is an outcome variable, T_i is the treatment dummy, L_i is leader dummy and ε_{id} denotes the error term clustered over district.

4.4.2 Control for covariates

Randomized allocation to treatment conditions does not guarantee that there are no differences between the treatment groups. However, it guarantees any such differences are occurring simply due to chance.¹⁵ But these differences can still result in biased results, so call ‘chance bias’ (Altman, 1985; Roberts and Torgerson, 1999; Schulz et al., 2010).

We do expect various socio-economic variables to have effect on the way the questions are answered. Among the most suspected to have an effect would be education or literacy.

Therefore, we also estimate a regression model correcting for all such observable characteristics. We believe the following variables could influence the answers to the questions and therefore are controlled for: Age; Gender; Literacy in local language and literacy in English; Education; Occupation; Number of social positions hold; Frequency of traveling to Accra/regional capital and to a district capital; Classification of the area as rural or urban; Language of the interview; Classification of the district as oil or mining; Ownership of radio, television, mobile phone, internet and access to newspaper¹⁶. The resulting econometric model is specified in the following way:

$$y_{id} = \alpha + \beta T_i + \gamma L_i + \delta(T \times L)_i + \mathbf{X}_i \boldsymbol{\tau} + \varepsilon_{id}$$

¹⁵ Apart by chance, this difference could be caused by a selective dropout. As this study fails to trace respondents that quit the questionnaire before it was finished, there could potentially be unequal assignment of treatment due to selective drop out which we are unable to trace. However, information from enumerators suggests that drop out from questionnaire was rather rare. Two-sided t-test of mean of the treatment condition variable to be equal to 0.5 results in $p = 0.61$. Therefore, there is no evidence of unequal dropout based on treatment condition.

¹⁶ Variables containing access to different media are technically one of the potentially sensitive questions of the survey experiment, and therefore potentially endogenous to treatment. However, we assume access to different media should not be sensitive question and therefore we do not expect it to have high variation in social desirability in answers. Therefore, the variables are assumed exogenous to the treatment.

Where \mathbf{X}_i is vector of the controlling variables, y_{id} is an outcome variable, $i = (1, \dots, n)$ denotes respondent and $d = (1, \dots, 120)$ district, T_i is the treatment dummy, L_i is leader dummy and ε_{id} denotes the error term clustered over district.

4.4.3 Combination of testing for covariates and RA-specific effects

Further, the following model combining controlling for covariates and RA-specific effects is estimated:

$$y_{id} = \alpha + \beta T_i + \gamma L_i + \delta(T \times L)_i + \mathbf{X}_i \boldsymbol{\rho} + \mathbf{RA}_i \boldsymbol{\tau} + \sigma TRA_i + \varepsilon_{id}$$

Where \mathbf{RA}_i are the RA dummy variables, TRA_i is a dummy variable stating usage of interpreter during the interview, \mathbf{X}_i is vector of controlling variables, y_{id} is an outcome variable, $i = (1, \dots, n)$ denotes respondent and $d = (1, \dots, 120)$ district, T_i is the treatment dummy, L_i is leader dummy and ε_{id} denotes the error term clustered over district.

4.4.4 Controlling for observable demographic characteristics and their interaction with treatment

Last, model controlling for observable respondent characteristics and their interaction with treatment is estimated. As stated in the Section 2.3 we hypothesize heterogeneous treatment effects among the leader and common citizen groups. These effects are hypothesized to be a result of different levels of identifiability among the two groups.

However, treatment effects could potentially be also caused by different observable or unobservable (demographic) characteristics of the respondents in leaders and non-leaders category and their interaction with treatment.

To eliminate the possibility of treatment effects being explained by the observable characteristics, we estimate a linear regression model including the following demographic variables and their interactions with treatment variable: Age; Gender; Literacy in local language and literacy in English; Education; Occupation (farmer, own business or wage job); Number of social positions hold; Frequency of traveling to Accra/regional capital and to district capital. This model allows us to distinguish whether the heterogeneous treatment effects are coming from 'being part of the leader category' (and to it connected difference in perceived anonymity) or whether it can be explained by observable demographic characteristic differing among the leaders and the common citizens.

The model controlling for effect of demographics and their interaction with the treatment can be specified as follows:

$$y_{id} = \alpha + \beta T_i + \gamma L_i + \delta(T \times L)_i + \mathbf{Z}_i \boldsymbol{\tau} + (T \times \mathbf{Z})_i \boldsymbol{\varphi} + \varepsilon_{id}$$

Where \mathbf{Z}_i are the demographic characteristics, $i = (1, \dots, n)$ denotes respondent and $d = (1, \dots, 120)$ district, y_{id} is an outcome variable, T_i is the treatment dummy, L_i is leader dummy and ε_{id} denotes the error term clustered over district.

Section 5 Results

The following section presents the results of the empirical models as presented in the previous section.

5.1 Treatment effects

5.1.1 Treatment effects on non-response rate to the potentially sensitive questions

Table 6 shows the treatment effect on non-response rate to the potentially sensitive questions. Neither the extensive margin (the number of people not responding to at least one question, outcome

V1) not the intensive margin (the number of non-responses recorded, outcome V2) statistically differ with the treatment conditions.

Table 6: Treatment effect on non-response rate to the potentially sensitive questions

Outcome	N ¹⁷	Mean identified (SD)	Mean unidentified (SD)	Diff	p-value	95% confidence interval		p-value MHT adjusted ¹⁸
						LB ¹⁹	UB ¹⁹	
Non-response to pot. sen. q. – dummy (V1)	3227	0.20 (0.40)	0.19 (0.39)	0.01	0.349	0.340	0.359	1
Non-response to p. s. q. – count (V2)	3227	0.23 (0.50)	0.24 (0.54)	-0.00	0.868	0.861	0.875	1

Permutation test results with 10000 repetitions.

5.1.2 Treatment effects on answers to the potentially sensitive questions

The effects of treatment on the outcomes measuring changes in answers to the potentially sensitive questions can be seen in Table 7. Treatment effects on all the four main outcomes are found insignificant at significance level $\alpha = 0.05$. ‘Satisfaction with chief’ is significantly lower in the identified condition at significance level $\alpha = 0.1$. However, it does not stay significant after multiple hypothesis adjustment and it is likely such effect was observed simply by chance.

¹⁷ The data for non-response to potentially sensitive questions do not include responses gathered by one research assistant (N=300). He recorded considerably higher number of non-responses than the rest of the research assistants. More than 50% of his respondents refused to answer at last one question, compared to 20% for the rest of the RA. Further, his respondents have been allocated to the identified condition quite often (57%). The non-response rate is on average higher for his respondents in the identified condition than in the end condition, however not significantly ($p=0.3$). When the data from this research assistant are used the effect of treatment is found significantly higher when identifying questions are asked prior to the outcomes, both for full sample ($p=0.08$) and for leader subsample ($p=0.03$). Significance of this effect does not hold when controlling for RA-specific effects. The respondents interviewed by this RA have been almost only from the leader subgroup.

¹⁸ Bonferroni-Holm adjusted p-values are shown. Instead of controlling for family wise-error rate (probability of Type I error), we can also control for false-discovery rate (the proportion of falsely rejected hypotheses). This can be done (among other) using Benjamini–Hochberg procedure. In this procedure we adjust each α in such way it reflects the proportion of hypotheses falsely rejected, see Benjamini and Hochberg (1995) for details. Importantly, for all of our analysis, both procedures lead to the same inference at $\alpha=0.1$ and at $\alpha=0.05$.

¹⁹ LB: lower bound, UB: upper bound

Table 7: Treatment effect on changes in answers to the potentially sensitive questions

Outcome	N	Mean identified (SD)	Mean unidentified (SD)	Diff	p-value	95% confidence interval		p-value MHT adjusted
						LB	UB	
Knowledge (A)	3526	0.16 (0.21)	0.17 (0.23)	-0.01	0.216	0.208	0.224	1
Satisfaction with chief (B)	2843	2.98 (1.41)	3.08 (1.36)	-0.09	0.072	0.071	0.074	0.433
Activity and interest (C)	3330	0.66 (0.11)	0.66 (0.12)	0.00	0.983	0.981	0.986	1
Donating and buying (D)	2887	0.66 (0.42)	0.64 (0.43)	0.02	0.137	0.131	0.144	0.824

Permutation test results with 10000 repetitions.
For 'Satisfaction with chief' outcome with 100000 repetitions.

5.1.2.1 Treatment effects on answers to individual potentially sensitive questions

No significant treatment effects on answers to the individual potentially sensitive questions, except for one. Only variable S14Q1 ('Would the respondent buy the report?') differs significantly ($p=0.029$) between the treatment conditions. Respondents in who are asked the identifying questions prior to the potentially sensitive questions are more likely to answer they would buy the report. Detailed results can be seen in Table 23 in Appendix 9.8.

5.2 Heterogeneous treatment effects

The following paragraphs elaborate on treatment effects on leader and common citizen subsamples separately.

5.2.1 Heterogeneous treatment effects on non-response rate

Table 8 shows the treatment effects on non-response rate to the potentially sensitive questions for the citizen and leader subsample separately. Neither the extensive margin (V1) nor the intensive margin (V2) statistically differ with the treatment conditions in either of the subgroups.

Table 8: Heterogeneous treatment effect on non-response rate to the potentially sensitive questions

Sample	Outcome	N	Mean identified (SD)	Mean unidentified (SD)	Diff	p-value	p-value 95% confidence interval		p-value MHT adjusted
							LB	UB	
Common citizens	Non-response to pot. sen. q.– dummy (V1)	1208	0.20 (0.40)	0.18 (0.39)	0.02	0.473	0.463	0.482	1
	Non-response to pot. sen. q.– count (V2)	1208	0.23 (0.49)	0.24 (0.55)	-0.01	0.788	0.780	0.796	1
Leaders	Non-response to pot. sen. q.– dummy (V1)	2019	0.21 (0.40)	0.19 (0.40)	0.01	0.499	0.489	0.509	1
	Non-response to pot. sen. q.– count (V2)	2019	0.24 (0.50)	0.24 (0.53)	0.00	0.999	0.998	0.999	1

Permutation test results with 10000 repetitions.

5.2.2 Heterogeneous treatment effects on answers to the potentially sensitive questions

Table 9 presents treatment effects on the outcomes constructed out of potentially sensitive questions for the citizen and the leader subsamples. The outcome donating and buying is found significantly higher ($p = 0.0023$) in the identified condition for the common citizen subsample. Significance of this effect holds even when controlled for multiple hypothesis testing using Bonferroni correction ($p = 0.032$)^{20 21}. No significance effect for this outcome is observed among the leader subgroup. Further, the difference between the two subgroups (leader*treatment interaction) is statistically significant at $\alpha = 0.01$, as can be seen in Table 21 in Appendix 9.6. No other outcomes have been observed to be significantly different at significance level $\alpha = 0.05$.

²⁰ Outcome donating and buying constructed from a level of intended donation instead of the dummy variable for donation, as was originally stated in the pre-analysis plan, has p-value of 0.0065 and multiple hypothesis testing adjusted $p = 0.084$. Therefore, even this outcome would be found statistically significant, however with less confidence. Further, the treatment effect on this outcome is not significant for the leader subsample.

²¹ Similarly, outcome created by combining not just from donating and buying variables (S14Q1 and S14Q2_d) but also willingness to provide contact for controlling project (S14Q3) and to participate in a meeting about natural resources revenues (S14Q4) is also found significantly higher in the identified condition. This outcome differs between the treatment conditions for common citizen sub-sample with high confidence. With $p = 0.0005$ and multiple hypothesis testing adjusted $p = 0.008$. Further, the treatment effect on this outcome is not significant for the leader subsample.

Table 9: Heterogeneous treatment effects on changes in answers to the potentially sensitive questions

Sample	Outcome	N	Mean identified (SD)	Mean unidentified (SD)	Diff	p-value	p-value 95% confidence interval		p-value MHT adjusted
							LB	UB	
Common citizens	Knowledge (A)	1213	0.11 (0.18)	0.12 (0.19)	-0.00	0.809	0.801	0.816	1
	Satisfaction with chief (B)	1045	2.91 (1.40)	3.04 (1.33)	-0.13	0.117	0.115	0.119	1
	Activity and interest (C)	1142	0.61 (0.13)	0.60 (0.15)	0.01	0.378	0.368	0.387	1
	Donating and buying (D)	1035	0.62 (0.44)	0.53 (0.45)	0.09	0.002	0.002	0.003	0.032
Leaders	Knowledge (A)	2313	0.18 (0.23)	0.19 (0.24)	-0.01	0.275	0.266	0.284	1
	Satisfaction with chief (B)	1798	3.03 (1.42)	3.09 (1.37)	-0.07	0.310	0.301	0.319	1
	Activity and interest (C)	2188	0.69 (0.09)	0.69 (0.09)	-0.00	0.810	0.802	0.818	1
	Donating and buying (D)	1852	0.68 (0.41)	0.69 (0.41)	-0.01	0.703	0.694	0.712	1

Permutation test results with 10000 repetitions.

For 'Donating and buying' outcome in common citizen subgroup with 100000 repetitions

5.2.2.1 Heterogeneous treatment effects on answers to individual potentially sensitive questions

Effects of treatment on answers to individual outcomes for common citizens and leader samples separately are shown in Table 24 and Table 25 in the Appendix 9.9.

For the common citizen subsample treatment effect is found significant for both the outcomes comprising 'donating and buying'. Both are significantly higher when the identifying questions are asked prior to them. For buying a report – S14Q1 ('Would the respondent buy the report?') – it is strongly significant ($p=0.0012$) and is found significant even after adjusting for multiple hypothesis testing ($p=0.0428$). For donating dummy – S14Q2_d ('Would the respondent consider donating?' dummy) – the treatment effect is found significant only before adjusting for multiple hypothesis testing ($p=0.0120$)²². Therefore, the main factor driving the significance of treatment effect for 'donating and buying' outcome is the buying a report component. Further, variables S14Q3 ('Gave contact information to be contacted about attending a meeting?') and S14Q4 ('Gave contact information to be contacted about monitoring a project?'), are also found significantly higher in the identified condition, with $p=0.069$ and $p=0.032$ respectively.

For the subsample of leaders, the treatment effect is significant only for variable S9Q8 ('Have you heard about the Public Interest Accountability Committee (PIAC). The effect of treatment is negative, with $p = 0.0124$. However, this effect is not found significant when adjusted for multiple hypothesis testing.

²² The level of intended donation (S14Q2), as used in pre-analysis plan, is not found significant ($p=0.69$).

5.3 Robustness

The robustness analysis included five regression models, all with treatment-subgroup interaction. These five models included a model controlling for research assistants-specific effects, a model controlling for covariates, a model controlling for both covariates and RA-specific effects and a model controlling for treatment*demographic variables interactions. The models are described in Section 4.4 in more detail.

The ‘Donating and buying’ outcome generally stays significant throughout the robustness analysis, as can be seen in Table 26 in Appendix 9.10. In model with only interaction (1) and in model controlling for baseline covariates (2) both the effect of treatment and the treatment*leader interaction are found significant with $\alpha=0.01$. In the models controlling for RA-specific effects (3) and RA-specific effects and covariates (4) the effect of treatment is significant with $\alpha=0.05$ and the effect of treatment*leader interaction is significant with $\alpha=0.1$. Only in the model controlling for interactions of demographic variables and treatment (5) the effect of treatment is not significant ($p=0.17$) while the treatment*leader interaction stays significant at $\alpha=0.1$. In this model, the demographic characteristics (S1Q1-S1Q11) are jointly significant ($p < 0.001$), while the interactions of these variables with treatment are not ($p = 0.4$).

In all the models testing robustness of ‘donating and buying’ outcome, the treatment and interaction effects have the same direction as in the simple mean comparison models. That is, the effect of asking the identifying questions prior to the potentially sensitive questions is positive and the respondents claim they would donate to a NGO or buy a transparency report more often. While this effect is lower for leader subgroup, i.e. the treatment*leader interaction has a negative effect.

For the remaining outcomes²³ the results generally remain insignificant throughout the robustness analysis. Only three models show significant effects of the treatment or the treatment-subgroup interaction. Each for a different outcome. One for ‘knowledge’, one for ‘satisfaction with chief’ and one for ‘Activity and interest’. However, all the effects have been significant only at significance level $\alpha=0.1$. Due to high number of outcomes and models and considering the fact only one model for each variable shows significant treatment effects, it is likely such effect was observed simply due to chance. The detailed results of these three models can be seen in Table 27 in Appendix 9.10.

Further, for all the nine main outcomes, results of permutation test and t-test are very close to each other and lead to the same inference at $\alpha = 0.1$, $\alpha = 0.05$ and at $\alpha = 0.01$. For comparison of resulting p-values see Table 22 in Appendix 9.7.

Section 6 Discussion

This study investigates effect of a simple question placement variation within a questionnaire on response quality. In one treatment condition identifying questions are placed prior to potentially sensitive questions. In the other treatment condition, no identifying questions are placed prior the potentially sensitive questions, they are placed after. Outcomes include the non-response rate and answers to the identifying and demographic questions. The effects of the treatment variation are investigated on a case of face-to-face interviews in Ghana. A survey experiment on a sample of total 3528 local leaders and common citizens is used to answer the research question. Heterogeneity of

²³ Robustness analysis was conducted for the following outcomes: Knowledge (A), Satisfaction with chief (B), Activity and interest (C), Donating and buying (D), Non-response to potentially sensitive questions – dummy (V1), Non-response to potentially sensitive questions – count (V2)

treatment effects is investigated for two sub samples, the local leaders and citizens. The subgroups are differentiated by the sampling procedure. The common citizens are selected using a convenience sampling and the leaders using a purposive sampling.

The hypothesis of treatment effects is derived based on differences in perceived anonymity of the responses. And heterogeneous treatment effects are hypothesized to result from overall lower level of perceived anonymity for leaders resulting from the different sampling procedure.

Non-response rate and social desirability in answers to the potentially sensitive questions are hypothesized to be higher if the identifying questions are asked before the potentially sensitive questions as opposed to being asked after these questions.

This section provides interpretation of our results and links these to our initial hypothesis. Further strengths and limitations of this study are discussed. Next, possible directions for future research are outlined. Finally, implications of our findings for other researchers are derived.

6.1 Changes in the non-response rate to the potentially sensitive questions

Findings of this study suggest no effect of the different placement of identifying and demographic questions on non-response rate to the potentially sensitive questions. This is the case for both the overall sample and when investigating the heterogeneous treatment effects among leaders and common citizens. We cannot support the Hypothesis 1A that asking identifying questions prior to the potentially sensitive questions leads to change in behavior in the form of not-answering to the potentially sensitive questions. This finding is in accordance with findings of Teclaw et al. (2012), who also found no effect of the same treatment variation on the non-identifying (outcome) questions in a web survey in the United States.

6.2 Changes in the answers to the potentially sensitive questions

Significant effects of the survey variation are found on one of the four outcomes created from the answers to the potentially sensitive questions for the citizen subgroup. No significant effect is observed for the leader subgroup. For the common citizen subgroup the outcome 'donating and buying' is significantly higher when the identifying questions are asked prior to the potentially sensitive questions, as opposed to when they are asked after.

When asked identifying questions first, the respondents in the common citizen group, claimed they would like to buy the report and that they would like to donate to the transparency-seeking NGO more often. Significant treatment effects with the same direction are observed on both the variables comprising this outcome, the donating to an NGO (S14Q2) and the buying of a report (S14Q1). Furthermore, significant treatment effects of the same direction are observed for answers to questions about willingness to participate in a PIAC meeting and to potentially control an oil-funded project in the area (questions S14Q3 and S14Q4).

On the other hand, there is no such difference found for any of these outcomes for the leader subgroup. Further, the difference in treatment effect between the two sub-groups is statistically significant. The effect of asking the identifying questions before outcomes is significantly lower for the leaders than for the common citizens. These findings are robust to different model specifications.

This finding supports the Hypothesis 1A about the effect of identifiability on answers to the potentially sensitive questions. Where (perceivably) more identified individuals answer on average in a more socially desirable way. Further, this effect is not significant for respondents in leader subgroup, which are hypothesized to already be identified due to different sampling procedure. Therefore, our Hypothesis 2B about heterogeneous treatment effects is supported as well.

Combining these findings with the findings of effects on non-response rate to the potentially sensitive questions, as discussed in the previous section, it seems likely that if our respondents change their behavior because of our survey variation, the preferable option is to change the answer to the potentially sensitive question, rather than choosing not to respond.

Finding a significant effect of our treatment variation on answer means is consistent with findings of Hoffman et al. (1996) who found effects of observability by researchers to have significant impacts on behavior in a dictator game, as well as with findings of Ong and Weiss (2000) and Singer et al. (1995) who find significant effects of anonymity and confidentiality on answers to sensitive questions in surveys. However, this finding is contrasting with findings of Giles and Feild (1978), Roberson and Sundstrom (1990) and Teclaw et al. (2012), who concluded no effect of changing position of identifying questions on answers to questions in mailed and web-based surveys in the US. Our design differed from the experiments of Giles and Feild (1978), Roberson and Sundstrom (1990) and Teclaw et al. (2012) in several ways. The two major differences are that we use face-to-face interviews as a mode of the survey delivery and our study is conducted in very different cultural setting. Both culture and mode of delivery are found to influence social desirability (Johnson and van de Vijver, 2003; Krumpal, 2011; Langhaug et al., 2010) and could be causing the differences in effects found. Further, our variation included asking for name, which could highlight the treatment effect as well.

6.2.1 Effects on the ‘donating and buying’ outcome

The donating and buying outcome is comprised of two questions. First question asks whether the respondents would buy a PIAC report, costing 3 GHC²⁴, if it would be available. The second question asks, how much would the respondent donate to a transparency NGO if they would have the chance.

The effect of our treatment variation on the ‘donating and buying’ significantly significant with a considerable magnitude of difference. When asked identifying questions prior to the questions about donating and buying report, 67% of the respondents in the citizen category stated they would buy the report and 57% stated they would donate. While the respondents that have been asked the identifying questions only after answering these two questions, only 58% of the common citizens stated they would buy the report and 50% stated they would donate to the NGO. Therefore, asking the identifying questions first, leads to an overall increase of about 16% in both the number of common citizens claiming they would buy the report and the number claiming they would donate to the NGO.

Why donating and buying?

The question that arises is, why did we found the significant effect for ‘donating and buying’ outcome and not any other outcomes?

Prior research suggests social desirability is consistently present among answers to sensitive questions, however no such presence is concluded for non-sensitive outcomes (Singer et al., 1995). Therefore, sensitivity of different questions is likely a confounding factor for presence of the treatment effect on a social desirability. However, we possess no data about sensitivity of our questions to be able to investigate this directly.

Nevertheless, according to Krumpal (2011), higher level of sensitivity of a survey question is typically associated with higher non-response rate. Therefore, we can use the number of non-responses as a proxy for judging sensitivity level of a question.

²⁴ At the time of the survey 3 Ghanaian cedi were roughly 0.7 USD.

Questions about donating to an NGO and buying a report are among the questions with the highest number of non-responses among the evaluated questions. This is the case for both leader and common citizen sub-samples. For the common citizen sub sample, asking about buying report has the third highest non-response rate and asking about donation has the second highest one. With only slightly less non-responses than the question asking about one's satisfaction with chief. For details refer to Table 24 in Appendix 9.9. For comparison, the question about satisfaction with chief was without any doubt highly sensitive and almost all the Ghanaian members of our research team raised concerns about sensitivity of this question and the way it should be asked²⁵.

The high non-response rate suggests the 'donating and buying' outcome could simply be more sensitive than the other outcomes, and be even as sensitive as the satisfaction with chief outcome. Therefore, the difference between treatment effects to these outcomes compared to the 'knowledge' and 'activity and interest' outcomes could be driven by different levels of sensitivity. However, if the sensitivity would be an only factor to play a role, significant effect should be present for satisfaction with chief as well.

Additionally, our ability to detect social desirability bias depends on an assumption that the social desirability bias in answers from different respondents happens in a uniform direction. While our ability to judge the response, quality depends on an assumption that our knowledge about this direction is correct. Violation of the first of the assumptions for part of the outcomes can lead to difference in treatment effects observed on different outcomes.

Further, our treatment varies only the salience of identifiability with the answers in the data. While the answer itself is always observable by the research assistant, or anybody else who might overhear it, regardless the treatment condition. Therefore, the concern about perception by the researcher or whoever might read the data is triggering the potential social desirable reporting. Thus, the effect of this dimension on the answers therefore must be (sufficiently) uniform in order to be detectable.

We believe this assumption is met for the 'donating and buying outcome' while it is likely violated for the 'Satisfaction with chief' outcome.

It is likely that our respondents believe that for us to see them as a 'good citizens' they should answer that they would like to donate and that they would like to buy the report. In a society such as the Ghanaian, there are conflicting motivations to answer these questions in a certain way. On one hand, in the hierarchical Ghanaian society, donating to an NGO could be seen as undermining the ruling authorities. On the other hand, in front of many other actors, such as NGO workers or (foreign) researchers it could be the opposite. The respondent could feel an obligation to answer 'yes', to seem as a 'good citizen' concerned with good governance, while having no intention to spare money on the cause. The second dimension is the one triggered by our treatment. Furthermore, the respondents could believe that answers to the donating and buying questions could have implications for chances of allocation of natural-resources revenues funded projects to their area. Therefore, as well inflate their claimed willingness to donate to NGO and buy the report. These beliefs and resulting change in behavior are similar to the researcher demand effects observed by Cilliers et al. (2015) in their experiment in Sierra Leone. In the presence of white foreigner, people overall behave more altruistically, while when they believed they are being tested for aid suitability they behaved less altruistically.

²⁵ This question was included especially in order to have a certainly sensitive question present in our questionnaire.

The knowledge of which answer might be desired by the researchers for the 'donating and buying' is in sharp contrast with the situation for the 'satisfaction with chief' question. Chiefs hold a prominent position in Ghanaian society, so it is very likely that other community members would expect one to show high satisfaction with chief. However, it could be possibly much less clear to the respondents what the researcher might see as a desirable answer. Only the dimension of concern of researchers seeing the data is highlighted by our treatment and the concern about other community members' views should stay constant with our treatment. Therefore, with no clear idea what the researchers would like to see, the respondents have no reason to change their answer to appeal to them. Alternatively, different respondents could have different ideas about which answer is desired and therefore, the effect would not be uniform and the changes in answer would cancel out. Therefore, the assumption of clear uniform effect of social desirability is likely violated for the 'satisfaction with chief'. Consequently, we would find no treatment effects on this outcome.

However, we lack data about sensitivity of different questions and lack of data about respondents' beliefs about researchers' (and others who might see the data) ideas about desirability of different answers and patterns of behavior. Therefore, we are unable to claim with certainty whether this is the mechanism driving the difference between 'donating and buying' and the rest of the outcomes.

Is the change in perceived anonymity responsible for the treatment effect?

Our Hypothesis 2 as stated in Section 2.3 depends on an assumption that citizens in the leader category are less anonymous because of different sampling procedures. This in turn leads to the lower effect of asking identifying questions prior to the potentially sensitive questions on changes in answers for leaders than for citizens.

However, possible alternative explanation is that the heterogeneous treatment effect is caused by differences among the respondents in the two subgroups and interaction of these characteristics with the treatment. Leaders are statistically significantly older, more often male, higher educated and more literate as well as they differ on many other characteristics.

In the case this is to be the explanation of heterogeneous treatment effects; the effect of leader*treatment interaction should become much less significant when controlling for the observable demographic characteristics of respondents and their interactions with the treatment. Furthermore, the interactions of treatment and demographic variables should be found statistically significant, if this alternative explanation holds.

The results, as discussed in Section 5.3, show the effect of leader*treatment interaction stays statistically significant when controlling for demographic characteristics and their interactions with treatment ($p = 0.084$). Furthermore, the effects of interactions of demographic variables with treatment are jointly as well as each individually statistically insignificant.

Therefore, these results imply that the difference between the treatment effect on the leader and the common citizen subgroups cannot be explained by the observable demographic characteristics. This suggests that there is a significant effect of the differences in sampling procedure for the two groups and therefore the assumed different levels of perceived anonymity. However, it is still possible, that the effect was caused by unobservable characteristics of the respondents differing among the two subgroups.

6.2.2 Limitations

Main shortcoming of this study is a lack of information to support a mechanism in which the treatment variation leads to the effect found. We possess no data about (perceived) levels of

sensitivity of different questions, therefore we can only assume the ‘donating and buying’ outcome is sensitive. Further, we have no information on respondents’ beliefs which answers to the questions are socially desirable and believed to be viewed positively by the other community members or the researchers. We are left to assume that the respondents believe the researchers and other people looking at the data would want them to respond positively to the ‘donating and buying’ questions. Such information could be gathered by asking direct questions about sensitivity of different questions and desirability of various answers on a distinct sample of respondents in the same area. Furthermore, question about who the respondent believes sent the interviewers, would provide valuable information about beliefs the respondents hold about the interview. Last, we have no data about perceived anonymity of the interview. However, this information would be difficult to measure for practical reasons. Such question would have to be asked prior to the identifying questions. Furthermore, asking how anonymous one feels and later asking for name and other information could face understandable negative reactions from the respondents.

6.2.3 Power and internal validity

Major advantage of this study is high internal validity resulting from the randomized assignment to treatment. Consequently, the treatment assignment is independent on any observable or unobservable characteristics. Furthermore, the permutation test used relies on no distributional assumptions and therefore the validity of our inference is high.

Another main advantage of this study is large sample size and resulting high power of the experiment²⁶. Prior to adjusting for multiple hypothesis testing our minimum detectable difference (with $\alpha = 0.05$ and $\beta = 0.8$) range from 0.1 to 0.17 standard deviations (SD), depending on the (sub-)sample used. While after adjusting for multiple hypotheses testing, the minimum detectable difference ranges from about 0.13 to 0.23 SD depending on the hypothesis family and sub-sample used. Cohen (1988) characterizes ‘small effect’ as effect of a size of about 0.2 SD. The power analysis suggests our study can capture similar effect in all model specifications with high probability. In some model specifications, it can capture even much smaller effect, especially prior to multiple-hypothesis testing adjustment.

Furthermore, for the six main outcomes in the two subgroups, no other significant results apart for ‘donating and buying’ would be observed when using significance level of $\alpha = 0.05$, even prior to the multiple hypothesis adjustment. The power of our study without MHT adjustment was high. Therefore, if there is another treatment effect present among the variables, it is likely only very small.

6.2.4 External validity

Our study is the first study to research this treatment variation in a face-to-face interview in a developing country. While all studies previously investigating the same treatment variation had been conducted on a self-administrated questionnaires in developed countries (Drummond et al., 2008; Giles and Feild, 1978; Green et al., 2000; Roberts and Torgerson, 1999; Teclaw et al., 2012). Unlike any of these studies, we do find effect of the treatment variation on social desirability in answers to potentially sensitive questions.

The sample used by this study includes respondents from 120 out of 216 districts in Ghana. These districts have been selected with probability-proportionate to population sampling, with inclusion of

²⁶ The power was calculated prior to the data analysis. Only sizes of hypothesis families for multiple hypothesis testing adjustment have been changed accordingly to changes in the research design.

all districts classified as oil or mining districts. Among our respondents, we have both respondents selected using convenience sampling, the common citizens, as well as purposefully selected respondents, the local leaders. We believe the results of the study should be generalizable for similar samples of respondents in Ghana. However, culture is found influence desire for social approval and socially desirable reporting (Johnson and van de Vijver, 2003; Krumpal, 2011). Therefore, the more different cultural setting, the less likely are our findings to be applicable.

Further, our questionnaires are administrated by research assistants using tablet computers. Results of similar study could differ with different method of administration of the questionnaires. Mode of questionnaire delivery is found to influence level of social desirability in answers (Langhaug et al., 2010). Therefore, effects of this treatment variation would likely differ with self-administrated questionnaire²⁷. When no research assistant is collecting the answers, and therefore no one directly observes the answers of the respondent, the effects of asking identifying questions prior to the potentially sensitive questions could potentially be even greater, given the same setting as a higher proportion of perceived anonymity is influenced by the variation in survey.

Last, we only find significant treatment effects for the respondents selected using purposeful sampling. Identification of a respondent resulting from convenience sampling and random sampling should be similar as they are both non-purposeful sampling procedures. Therefore, as perceived anonymity is assumed to play a role, a similar effect is believed to be present for respondents selected using a random sampling as well.

6.3 Implications for survey design

We find no effect of asking identifying questions prior to potentially sensitive questions on non-response rate to the potentially sensitive question. However, we find one of four outcomes constructed from answers to the potentially sensitive questions to significantly differ among the two treatment conditions. Asking the identifying questions prior to the potentially sensitive questions, leads to significantly higher answers to questions measuring claimed willingness to donate to an NGO and to buy a transparency report. We attribute this effect to an increase in social desirability bias caused by a change in perceived anonymity of the interview and a clear idea of the respondents consider to socially desirable by the researchers. Assuming higher donation and willingness to buy the report is believed to be socially more desirable by the respondents, than the lower proportion stated equates to higher data quality. Therefore, asking identifying questions prior to the outcome leads to increase in social desirability and connected decrease in data quality. This works only for the subsample of the respondents selected using non-purposeful sampling (common citizen subsample). No effect is observed on the respondents selected using purposeful sampling, arguably due to higher initial identifiability. However, the other three outcomes measuring change in answers to the potentially sensitive outcomes are not found significantly different among the treatment conditions. We hypothesize this is due these outcomes either not being sensitive or not having a clear socially desirable level.

For a researcher concerned with survey quality, there however is one more factor to consider, the non-response rate to the identifying questions. Teclaw et al. (2012) found that the non-response rate to the identifying questions increased when those questions had been asked after the rest of the questions. However, there is no indication of this in our data. Of the variables constructed out of answers to the identifying questions, only the variable stating whether the respondent provided

²⁷ Giles and Feild (1978), Roberson and Sundstrom (1990), Green et al. (2000), Drummond et al. (2008) and Teclaw et al. (2012) answered this question in western country settings.

phone number (S1Q0b) differs significantly ($p=0.028$) among the two treatment conditions and only for the leader subgroup. The non-response rate is significantly lower when the identifying questions are asked prior to the potentially sensitive questions. However, this effect is only marginal and is driven by a very small number (20) of opinion leaders. The non-response rate in answers to the identifying questions is very small. When not considering questions about name and phone number, only 31 respondents refused to answer one or more of the identifying questions. For detail, see Table 19 in Appendix 9.5 containing summary statistics of the identifying questions.

Overall, placing the identifying questions specifically after the potentially sensitive questions comes at no cost in form of negative effect on the sensitive questions or on the response-rate to the identifying questions. However, for non-purposefully selected respondents in case of sensitive outcome with an assumed clear social desirability direction, placing the identifying questions prior to the sensitive question leads to increased social desirability bias in the sensitive questions.

Therefore, it is preferable, during a face-to-face interview in cultural setting comparable to Ghana, to place the identifying questions at the end of the survey, only after the potentially sensitive questions had been answered.

This recommendation has to be considered together with other recommendations researcher might receive when designing a survey. Such as recommendation to start with interesting, easy to answer questions applicable to most respondents (e.g. Czaja and Blair, 2005 p. 93-96). The identifying questions might satisfy many of these characteristics. However, there are likely many other questions satisfying these criteria, while not allowing identification of the respondent. Therefore, if other such questions can be found, our recommendation does not need to be in conflict with this recommendation. Therefore, the potentially negative effect of asking identifying questions prior to sensitive questions should be at last considered, and weighted against other recommendations, by any researchers designing similar surveys as our survey.

Finally, the role of an informed consent in our survey is relevant for thinking about survey methodology as well. Our informed policy (see it in Appendix 9.2), briefly informed that we will ask for a name of the respondent. If this information would have been adequately processed by the respondents, there should be no room for our variation to have an effect. Therefore, finding the significant effect indicates there might be incomplete understanding of the informed consent. This could be caused by overwhelming amount of information about the research presented in the beginning of the interaction with the respondent. Further, such interaction might not be familiar to respondents in Ghana. The issue of potential incomplete understanding of informed consent needs to be researched in more detail.

Furthermore, if the respondents are indeed motivated by desire to be seen as a 'good citizen' or by hope of obtaining some oil revenue funded projects, our findings would indicate misunderstanding about the purpose of the study and the way the data are analyzed and used. In case the respondents correctly believe the purpose of the study (to evaluate the level of knowledge, the attitudes and the situation) and know the way data are analyzed, there should be no motivation to answer untruthfully. However, respondents of our survey have likely little prior experience with surveying or the way such data might be analyzed. Our observations indicate the respondents might incline to believe their data are investigated individually, with the answers being connected to the particular respondent. However, in reality, such large samples are rarely analyzed by investigating answers individual respondents. Therefore, there might be a potential for improvement in informed consent policies and the respondents understanding of data usage. However, further investigation of effects of such variation in informed consent on a social desirability in answers is needed.

Unfamiliarity of participants with research might cause lower understanding of informed consent and more misunderstanding about the purpose of the data usage. Therefore, familiarity of respondents with surveying could potentially be another reason why, unlike Giles and Feild (1978), Roberson and Sundstrom (1990) and Teclaw et al. (2012), we do find an effect of the survey variation on social desirability bias. Their respondents in US are much more likely familiar with surveys than our respondents in Ghana.

6.4 Possibilities for future research

This study hypothesizes that identification and resulting changes in perceived anonymity are the drivers of the treatment effect. However, to provide conclusive evidence, the mechanism needs to be researched in more detail.

Future research could be focused in several ways. One, it could focus on finding the effect of the same treatment variation on perceived anonymity. This would provide useful insight about the hypothesized main driver of the treatment effect.

Second, characteristics of questions and answer options should be investigated. This can be done by gaining more knowledge about how sensitive the respondents consider the different questions. As well as by finding out which answer options to these questions are perceived as the most socially desirable ones by different groups of people. Knowledge about perceptions of answers desired by other community members, local politicians, the government, NGOs or researchers would be beneficial. Together with respondents' beliefs about who will read the data, this would provide valuable information to judge the mechanism of the treatment effects and why the unequal treatment effects on the different outcomes might have arisen.

Further, investigating effect of the different position of demographic questions, without changing the position of asking for name and phone number, preferably in a 2x2 design, would provide indication about relative effects of different parts of the variation.

Next, replications in distinct cultural context, on different topics and with different modes of administration, would provide evidence about external validity of our findings.

Finally, effects of informed consent policy on the treatment effect and on social desirability, as described in implications of the findings, should be investigated.

Section 7 Conclusion

This study investigated the possibility of increasing survey data quality by varying the position of identifying and demographic questions in a survey questionnaire. Name, phone number, age, gender, ethnicity and other demographic variables are asked either in the beginning of the questionnaire, prior to the potentially sensitive questions, or at the end of the questionnaire, only after the potentially sensitive questions on a during of face-to-face interviews. These interviews were conducted in a developing country, in Ghana.

The research is designed as a randomized survey experiment with two treatment conditions, using a large sample. The data includes responses from a total of 3528 face-to-face interviews. The interviews had been conducted with local leaders (respondents holding a political or social position in the community) and common citizens from 120 districts in Ghana. All 32 districts classified as mining and oil had been included, remaining 88 districts had been selected using probability-proportionate to population sampling. The common citizens are selected using convenience sampling, while the leaders are selected using purposive sampling.

Giles and Feild (1978), Roberson and Sundstrom (1990), Green et al. (2000), Drummond et al. (2008) and Teclaw et al. (2012) also investigated the same treatment variation experimentally. However, this study is the first to do during a face-to-face interview in developing country setting.

The treatment effects are hypothesized to be driven by changes in perceived anonymity of the responses. The fact that somebody investigating the responses can identify who provided the answers is made salient by asking the identifying questions. Therefore, respondents who are asked these questions prior to the potentially sensitive questions are aware of being identified when answering the questions, while the rest of the respondents do not have to be aware of being identified until the identifying questions are asked at the end of the survey.

Therefore, the differences between the treatment conditions are expected to be a result of change in the behavior of the respondents due to observability of the answers by the researchers, or any other person looking at the data.

Further, unequal treatment effect among the leader and common citizen subgroups are hypothesized. The heterogeneous treatment effects are assumed to arise out of overall higher initial identification of the leaders as resulting from the sampling procedure. The 'leader' respondents are selected using purposeful sampling and are selected because of their particular position in the community. Therefore, they have no reason to believe they are not identified with their answers from the very beginning of the survey. On the other hand, the common citizens are selected using convenience sampling and have no initial reason to believe they are identified.

This study researches treatment effects on non-response rate and answers to potentially sensitive questions. To investigate changes in the answers to the potentially sensitive questions this study uses four outcomes. The outcomes are constructed out of responses to potentially sensitive questions in a questionnaire investigating attitudes and knowledge surrounding natural resources revenues in Ghana. The four outcomes group the questions in categories labelled as 'knowledge of institutions', 'satisfaction with chief', 'activity and interest' and 'donating and buying'.

Significant treatment effect is only found on 'donating and buying' outcome for the common citizen subgroup, no effect is found for leader subgroup or any other of the four outcomes.

The 'donating and buying' and the 'satisfaction with chief' are the only outcomes showing high number of non-responses, which is commonly associated with sensitive outcomes (Krumpal, 2011). Furthermore, we expect these two outcomes to differ in the clarity of which response would be perceived well by the researcher. We believe the respondents' think that the researchers would be in favor of the respondent answering that they will donate an NGO and to buy the report. However, this is not as clear for the 'satisfaction with chief' outcome. Common direction of social desirability is needed to detect changes in social desirability bias. Therefore, we hypothesize the effect to be present for the 'donating and buying' outcome and no other outcomes due to a combination of the 'donating and buying' outcome being both sensitive and with a uniform direction of the social desirability bias.

The treatment effect on this outcome is statistically significant throughout a robustness analysis. Furthermore, the effect on answers to this outcome is of a considerable magnitude. With an approximately 16% increase in the number of respondents claiming they would donate to a transparency supporting NGO. Likewise, a 16% increase can as well be seen in the number of respondents claiming they would buy an oil revenues management report if they would have a chance.

The results are consistent with the hypothesis of asking identifying questions prior to the potentially sensitive question leading to change in answers. However, this is only applicable for the common citizen subgroups, and has no observed effect on the leader subgroup. Therefore, the hypothesis about lower effects of treatment on the leader subgroups is supported by this finding.

We assume that socially desirable answer to the donating and buying questions, as considered by the respondents, is to be willing to buy the report and to donate. Under this assumption, we conclude that asking the identifying question first leads to less socially desirable answers and thus a higher quality data for the common citizen subsample.

Moreover, this finding supports the prior evidence that humans change behavior when being observed and less anonymous.

On the other hand, there is no evidence of treatment effect on non-response rate to the potentially sensitive questions. This suggests placing the identifying questions only after the potentially sensitive questions comes at no cost as there are no negative effects of this placing.

Therefore, we derive the following recommendation for researchers preparing a survey questionnaire which is to be administrated using face-to-face interviews in a similar cultural setting to Ghana:

Implications for survey design:

It is preferable to place the identifying questions at the end of a survey, ensuring they are placed after the potentially sensitive questions.

Section 8 Reference List

- Altman, D.G., 1985. Comparability of Randomised Groups. *J. R. Stat. Soc. Ser. Stat.* 34, 125–136. doi:10.2307/2987510
- Bateson, M., Nettle, D., Roberts, G., 2006. Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* 2, 412–414. doi:10.1098/rsbl.2006.0509
- Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.
- Cilliers, J., Dube, O., Siddiqi, B., 2015. The white-man effect: How foreigner presence affects behavior in experiments. *J. Econ. Behav. Organ., Economic Experiments in Developing Countries* 118, 397–414. doi:10.1016/j.jebo.2015.03.015
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.
- Czaja, R., Blair, J., 2005. *Designing survey: A guide to decisions and procedures*, 2nd ed. ed. Pine Forge Press.
- Drummond, F.J., Sharp, L., Carsin, A.-E., Kelleher, T., Comber, H., 2008. Questionnaire order significantly increased response to a postal survey sent to primary care physicians. *J. Clin. Epidemiol.* 61, 177–185. doi:10.1016/j.jclinepi.2007.04.012
- Duval, S., Wicklund, R.A., 1972. *A theory of objective self awareness*. Academic Press.
- Ernst, 2004. Permutation methods: a basis for exact inference. *Stat. Sci.* 19, 676.
- Giles, W.F., Feild, H.S., 1978. Effects of amount, format, and location of demographic information on questionnaire return rate and response bias of sensitive and nonsensitive items. *Pers. Psychol.* 31, 549–559.
- Green, R.G., Murphy, K.D., Snyder, S.M., 2000. Should demographics be placed at the end or at the beginning of mailed questionnaires? An empirical answer to a persistent methodological question. *Soc. Work Res.* 24, 237.
- Haley, K.J., Fessler, D.M.T., 2005. Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* 26, 245–256. doi:10.1016/j.evolhumbehav.2005.01.002
- Hoeffding, W., 1952. The Large-Sample Power of Tests Based on Permutations of Observations. *Ann. Math. Stat.* 23, 169–192.
- Hoffman, E., McCabe, K., Smith, V.L., 1996. Social Distance and Other-Regarding Behavior in Dictator Games. *Am. Econ. Rev.* 86, 653–660.
- Johnson, T.P., van de Vijver, F.J.R., 2003. Social desirability in cross-cultural research, in: *Cross-Cultural Survey Methods*. Wiley, pp. 193–209.
- Kettle, K.L., Häubl, G., 2011. The Signature Effect: Signing Influences Consumption-Related Behavior by Priming Self-Identity. *J. Consum. Res.* 38, 474–489. doi:10.1086/659753
- Krumpal, I., 2011. Determinants of social desirability bias in sensitive surveys: a literature review. *Qual. Quant.* 47, 2025–2047. doi:10.1007/s11135-011-9640-9
- Langhaug, L.F., Sherr, L., Cowan, F.M., 2010. How to improve the validity of sexual behaviour reporting: systematic review of questionnaire delivery modes in developing countries. *Trop. Med. Int. Health* 15, 362–381. doi:10.1111/j.1365-3156.2009.02464.x
- Ludbrook, J., 1998. Multiple Comparison Procedures Updated. *Clin. Exp. Pharmacol. Physiol.* 25, 1032–1037. doi:10.1111/j.1440-1681.1998.tb02179.x
- Milinski, M., Rockenbach, B., 2007. Spying on Others Evolves. *Science* 317, 464–465. doi:10.1126/science.1143918
- Nederhof, A.J., 1985. Methods of coping with social desirability bias: A review. *Eur. J. Soc. Psychol.* 15, 263–280. doi:10.1002/ejsp.2420150303
- Ong, A.D., Weiss, D.J., 2000. The Impact of Anonymity on Responses to Sensitive Questions. *J. Appl. Soc. Psychol.* 30, 1691–1708. doi:10.1111/j.1559-1816.2000.tb02462.x
- Rigdon, M., Ishii, K., Watabe, M., Kitayama, S., 2009. Minimal social cues in the dictator game. *J. Econ. Psychol.* 30, 358–367. doi:10.1016/j.joep.2009.02.002
- Roberson, M.T., Sundstrom, E., 1990. Questionnaire design, return rates, and response favorableness in an employee attitude questionnaire. *J. Appl. Psychol.* 75, 354–357.

- Roberts, C., Torgerson, D.J., 1999. Baseline imbalance in randomised controlled trials. *BMJ* 319, 185.
- Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group, 2010. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 340, c332–c332. doi:10.1136/bmj.c332
- Shaffer, J.P., 1995. Multiple hypothesis testing. *Annu. Rev. Psychol.* 46, 561–584.
- Singer, E., Von Thurn, D.R., Miller, E.R., 1995. Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature. *Public Opin. Q.* 59, 66–77.
- Teclaw, R., Price, M.C., Osatuke, K., 2012. Demographic Question Placement: Effect on Item Response Rates and Means of a Veterans Health Administration Survey. *J. Bus. Psychol.* 27, 281–290. doi:10.1007/s10869-011-9249-y
- Weber, M., Sawilowsky, S., 2009. Comparative Power Of The Independent t, Permutation t, and WilcoxonTests. *J. Mod. Appl. Stat. Methods* 8, 3.
- Young, A., 2015. Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. Working paper, October.

Section 9 Appendix

9.1 Note on the pre-analysis plan

Prior to the analysis of our data, we developed a pre-analysis plan (PAP). The theories and hypothesis stated in the PAP were based on different psychological mechanisms. However, after the data had been analyzed, the learning process still continued and our belief about the mechanism changed. The following two paragraphs briefly indicate the theories used in the PAP and the reasons why we anymore do not believe they play a role in our treatment effect. The theory of change in perceived identifiability with answers as driving the treatment effect was used in pre-analysis plan only as an alternative mechanism and in much less extended form.

The full pre-analysis plan can be requested at the following email address: maarten.voors@wur.nl

Hypothesis of treatment effects

Originally, in the pre-analysis plan, Duval and Wicklund's (1972) Theory of objective self-awareness was used to derive the hypothesis about the treatment effect. This theory implies that individuals focusing their attention on themselves (i.e. being in the 'objectively self-aware state') compares the self to own standards of correctness more than when focusing on aspects external to the self. Where, the hypothesized effect of our treatment was that asking the identifying questions prior to the potentially sensitive questions will lead to less social desirability in answers. This is exactly the opposite effect than hypothesized now. The identifying questions were hypothesized to direct the attention of the respondent to himself and his moral standards of answering honestly, where the state would remain throughout the interview. This was to lead to (higher) negative benefits from lying (discrepancy with honesty standards) and therefore more honest answers. However, when developing this theory, we neglected the fact that the discrepancy of self and the societal standards will as well be highlighted in this state. This in turn leads to higher motivation for socially desirable answers. Furthermore, the respondents during the interview are expected to already be in highly objectively self-aware state, which is rather unlikely to be significantly influenced by asking the identifying questions. This objectively self-aware state is rather caused by the research-assistant asking the questions, being highly attentive and interested in the respondent.

Hypothesis of heterogeneous treatment effects

Similarly, also the thinking about heterogeneous treatment effects changed considerably since the pre-analysis plan. Initially, identity priming was used to derive the hypothesis about heterogeneous treatment effects. Connecting identity priming to Gibson's theory of affordances (Greeno, 1994), we hypothesized that for the leader subgroup; a 'leader' identity will be primed by our treatment. In this view, the 'leader' identity is primed because being a leader allowed for the interaction (which is also salient to the individuals). Similarly to Kettle and Häubl (2011), this thinking was used to derive hypothesis of heterogeneous effect of the treatment on leader subgroup. The treatment was expected to lead to more social desirability bias in answers as leaders are expected to be more knowledgeable or active. However, even if the 'leader' identity would be primed in this way, it should be primed in the very beginning of research assistant's interaction with the leaders, prior to the treatment assignment. Therefore, all leaders have the 'leader' identity primed the same way regardless of the treatment condition they are in and the identity priming cannot result in heterogeneous treatment effects.

9.2 Informed consent

Principle Investigator: Ishmael Edjekumhene

My name is _____. I am a research associate hired by the Kumasi Institute of Technology Energy and Environment (KITE) and Norwegian University of Science and Technology (NTNU). I would like you to participate in a survey in order to understand natural resource management in Ghana and your about knowledge and attitudes towards natural resource revenues. Before we begin, I would like to take a minute to explain why I am inviting you to participate and what I will be doing with the information you provide to me. Please stop me at any time if you have any questions. After I've told you a bit more about my project, you can decide whether or not you would like to participate.

This research is being conducted by researchers from KITE, Ghana and NTNU, Norway. We will be interviewing in 120 districts across Ghana. The researchers will use the information we collect in articles that might be published, as well as in academic presentations. No publications will include names or other identifiable information on participants. Your name will be registered separately from your responses to the survey to make sure that you remain anonymous. You can also refuse to give us your name.

Participation should take about an hour. Participation is on a purely voluntary basis. You will be asked a series of questions about yourself and your household. There are no risks to you from answering these questions. The information we collect today is private and confidential. We will not share any details from the survey about your friends or family with anyone besides the research team. These data will be stored on KITE computers.

If at any time and for any reason, you would prefer not to answer any questions, please feel free not to. If at any time you would like to stop participating, please tell me. We can take a break, stop and continue at a later date, or stop altogether. You will not be penalized in any way for deciding to stop participation at any time.

There are no correct or incorrect responses, so please express your opinions freely.

If you have any questions regarding this research or your rights as a research study participant, you may contact us at the phone numbers or email addresses on this card: Akosua Amaka-Otchere on telephone number 1, telephone number 2 or KITE email (The participant will be given a business card showing the information)

9.3 Identifying and demographic questions and answer options

Table 10: Treatment questions and answer options

Code	Question	Answers / units and coding
S1	<i>PERSONAL BACKGROUND AND HOUSEHOLD INFORMATION SECTION</i>	-
S1intro	First (Finally), I'll ask you some questions about your personal background and your household	-
S1Q0 note	We would like to ask you for your name and phone number. Your name and phone number will be recorded on separate sheet of paper and kept aside of the rest of the data, so your anonymity would be ensured. You can also refuse to give these to us. If you give us your phone number, we would potentially contact you. You might also receive information about oil revenues governance. Moreover, would you be interested to participate in a subsequent round of this research in about a year's time? If yes, please indicate this and fill your contact details on the same sheet of paper.	-
S1Q0a ²⁸	Did the respondent agree to give her/his name and did you record the name and the respondents assigned ID on the separate paper?	1 Yes 0 No
S1Q0b ²⁸	Did the respondent agree to give her/his phone number and did you record the phone number on the separate paper?	1 Yes 0 No
S1Q0c	Agreed to participate in the next round of this questionnaire?	1 Yes 0 No
S1Q1	How old are you?	<i>Years</i>
S1Q2	Are you a male or female?	1 Male 2 Female
S1Q3	What is your relation to household head?	1 the head 2 Spouse 3 Child 4 Parent 5 Brother or sister 6 Friend 7 Foster child .o Other (specify) .a Do not know (DNR ²⁹) .b Do not want to answer (DNR)
S1Q4	What language do you most commonly speak at home?	1 English 2 Akan 3 Ewe 4 Ga 5 Dagbani

²⁸ This, question was not read out loud, it was simply evaluated by the research assistant.

²⁹ DNR stands for 'Do not read'.

		6 Dangbe 7 Kasem 8 Gonja 9 Nzema .o Other
S1Q5	What is your tribe?	1 Akan 2 Mole-Dagbon 3 Ewe 4 Ga-Dangme 5 Gurma 6 Guan 7 Grusi 8 Mande .o Other (specify) .a Do not know (DNR) .b Do not want to answer (DNR)
S1Q6	Can you read and write in your local language?	0 No
S1Q7	Can you read and write in English?	1 Only read 2 Yes .a Do not know (DNR) .b Do not want to answer (DNR)
S1Q8	What is your highest level of education?	0 None 1 Incomplete Primary School 2 Completed Primary School 3 Incomplete Junior High School 5 Completed Junior High School 4 Incomplete Middle School 6 Completed Middle School 7 Incomplete Secondary/Technical School 8 Completed Secondary/Technical School 9 Incomplete tertiary 10 Completed tertiary .a Do not know (DNR) .b Do not want to answer (DNR)
S1Q9	What is your main occupation?	1 Farming 2 Fishing 3 Mining or Galamsey 4 Forestry, timber and non-timber products 5 Wage job (salary workers) in other sectors 6 Working for own or family owned business in other sectors 7 Student 9 Religious authority 0 None
S1Q10	Do you hold any of the following positions in your community?	1 Chief 2 Chief's wife 3 Queen mother

		4 Unit Committee member 5 District assembly member 6 Youth leader 7 Women’s leader 8 Religious leader 9 Tribal authority (or ethnic leader) 10 Teacher 11 Opinion leader 0 None .a Do not know (DNR) .b Do not want to answer (DNR)
S1Q11	During the last 6 months, did you go to ____?	-
S1Q11a	Accra or regional capital	1 Yes
S1Q11b	District capital	0 No
S1Q11c	Other settlement in your district	.a Do not know (DNR) .b Do not want to answer (DNR)
S2intro	Now I will ask you some questions about your household. A household is a group of people who presently eat together from the same pot, excluding temporary visitors.	-
S2Q1	How many adults (people over 18 years old) currently live in your household?	<i>Number</i>
S2Q2	How many children (people under 18 years old) currently live in your household?	<i>Number</i>
S2Q3	In general, how would you describe your households’ present living conditions?	0 Very bad 1 Fairly bad 2 Neither good nor bad 3 Fairly good 4 Very good .a Do not know (DNR) .b Do not want to answer (DNR)
S2Q4	On average, how many hours a day is electricity available at your home?	<i>Hours (0-24)</i> .a Do not know (DNR) .b Do not want to answer (DNR)

9.4 Outcome variables composition

Table 11: Outcome variables composition

Out-come code	Name of outcome (scale)	Questions composing the outcome		
		Variable code	Question	Answers and their coding
A	Knowledge of institutions (0-1)	S9Q1	Have you heard about the Petroleum Revenue Management Act (PRMA)?	1 Yes, heard about it 0 No, have not heard about it .a Do not know (DNR) .b Do not want to answer (DNR)
		S9Q4	Have you heard about the Annual Budget Funding Amount (ABFA) projects?	
		S9Q8	Have you heard about the Public Interest Accountability Committee (PIAC)?	
		S9Q12	Have you heard about the Ghana Extractive Industries Transparency Initiative (GHEITI)?	
		S9Q16	Have you heard about the Ghana Agency for Mining, Oil and Gas (GAMGO)?	
B	Satisfaction with chief ³⁰ (0-4)	S4Q7	To what extent do you agree with the following statement: My chief is doing a very good job in fulfilling his responsibilities.	0 Completely disagree 1 Somewhat disagree 2 Neither agree nor disagree 3 Somewhat agree 4 Completely agree .a Do not know (DNR) .b Do not want to answer (DNR)
C	Activity and interest (0-1)	S4Q1	How often do you discuss political matters and public affairs with friends, family or colleagues?	0 Never 1 Rarely 2 Sometimes 3 Often 4 Very often 5 All the time .a Do not know (DNR)

³⁰ Chiefs were not asked this question.

				.b Do not want to answer (DNR)
		S4Q2	<i>Did you vote in last ___ elections?</i>	
		S4Q2a	Presidential / Parliamentary	1 Yes
		S4Q2b	District Assembly / Unit Committee	0 No .x Not allowed to vote .a Do not know (DNR) .b Do not want to answer (DNR)
		S4Q3	<i>Do you plan to vote in the next ___ elections?</i>	
		S4Q3a	Presidential / Parliamentary	3 Yes, surely
		S4Q3b	District Assembly / Unit Committee	2 Most likely 1 Not likely 0 No, definitely not .a Do not know (DNR) .b Do not want to answer (DNR)
		S6Q7	In the past 12 months have you discussed usage of revenues from oil, gas and mining with your family, friends or colleagues?	1 Yes 0 No
		S6Q8	In the past 12 months have you contacted anybody to ask about how revenues from oil, gas or mining are handled in Ghana or in your area?	.a Do not know (DNR) .b Do not want to answer (DNR)
		S8Q4	During the last 12 months, have you or a member of your household engaged in an action mentioned in the previous question, or any other action, to improve the handling of revenues from oil, gas and mining?	
D	Donating and buying ³¹ (0-1)	S14Q1note	<i>Public Interest Accountability Committee (PIAC) is independent public body with a goal to inform citizens of Ghana about the usage of revenues from oil and gas extraction.</i> <i>The Public Interest Accountability Committee (PIAC) publishes an annual report about the utilization of Ghana's oil and gas resources. It includes general information as well as the list of the projects supported with this</i>	<i>Only text show – not a question</i>

³¹ In the pre-analysis plan this outcome was defined sum of normalized S14Q1 and S14Q2. However, the variable stating intended donations (S14Q2) is highly influenced by small amount of outliers and therefore a dummy variable was chosen to be used instead. Differences among results of the two specifications is pointed out in the results section.

			<i>money. This report is accessible freely on the PIAC website. However, due to budget issues it is not distributed to citizens in any other way. If you had the option, would you be interested in such a report for the printing cost of 3 Cedi³²?</i>	
		S14Q1 ³³	Would the respondent buy the report?	1 Yes 0 No .a Do not know (DNR) .b Do not want to answer (DNR)
		S14Q2note	<i>One of the ways one can contribute to better governance of oil and gas revenues is supporting a non-governmental organization with such a goal. To mention one: Ghana Integrity Initiative, the local chapter of Transparency International, is a non-partisan, non-profit civil organisation focused on addressing corruption. Their mission is: "to fight corruption and promote good governance in the daily lives of people and institutions by forging strong, trusting and effective partnership with government, business and civil society and engagement with the people". It functions thanks to contributions from donors. If you were offered the possibility to donate a few Cedi to the Ghana Integrity Initiative (GII), would you consider doing so? In case yes, how much would you be willing to donate?</i>	Only text show – not a question
		S14Q2 ^{33 34}	Would the respondent consider donating?	Decimal, Ghana Cedi
		S14Q2_d	Dummy variable stating whether the respondent decided to donate any money.	= 1 if S14Q2 >0 = 0 if S14Q2 =0
-	Questions not used as part of any outcome but looked at individually	S14Q3note	The Public Interest Accountability Committee (PIAC) is thinking of organizing meetings where interested citizens would be informed about the usage of oil revenues. If such a meeting would be organized in your district, would you like to participate? In case you would, please give us your name and phone number or email, we will give it to PIAC so they	

³² At the time of the survey 3 Ghanaian cedi were roughly 0.7 USD.

³³ This, question is not read out loud, it is simply evaluated by the research assistant.

³⁴ Outcome S14Q2 only used as below stated dummy - S14Q2_d.

			could potentially contact you.	
		S14Q3	Gave contact information to be contacted about attending a meeting?	1 Yes 0 No
		S14Q4note	The Public Interest Accountability Committee (PIAC) is also considering creating a citizen engagement platform, where volunteer citizens would regularly monitor progress of projects in their communities. In case the platform will be implemented and there will be a project to be monitored in your area, would you be interested to participate and help PIAC? In case you would, please give us your name and phone number or email, we will give it to PIAC so they could potentially contact you.	
		S14Q4	Gave contact information to be contacted about monitoring a project?	1 Yes 0 No
V1	Non-response to potentially sensitive questions – dummy (0-1)	Extensive margin of “Do not know” and “Do not want to answer” to the questions comprising outcomes A, B, C and D i.e. the following questions: S9Q1, S9Q4, S9Q8, S9Q12, S9Q16, S4Q7, S4Q1, S4Q2, S4Q2a, S4Q2b, S4Q3, S4Q3a, S4Q3b, S6Q7, S6Q8, S8Q4, S14Q, S14Q2. V1 = 1 if there is at last one “Do not know” or “Do not want to answer” among answers to these questions for the respondent. V1 = 0 otherwise.		
V2	Non-response to potentially sensitive questions – count (0+)	Intensive margin of “Do not know” and “Do not want to answer” to the questions comprising outcomes A, B, C and D. V2 = number of the “Do not know” and “Do not want to answer” to these questions recorded by the respondent.		

9.5 Summary statistics

Table 12: Summary statistics – treatment variable, leader variable and their interaction

variable	count	mean	sd	min	max
identified	3528	0.504	0.500	0	1
leader	3528	0.656	0.475	0	1
identified*leader	3528	0.327	0.469	0	1

Table 13: Summary statistics – knowledge variables

variable	count	mean	sd	min	max
Knowledge (A)	3526	0.16	0.22	0	1
S9Q1	3527	0.37	0.48	0	1
S9Q4	3528	0.13	0.33	0	1
S9Q8	3527	0.22	0.42	0	1
S9Q12	3526	0.05	0.23	0	1
S9Q16	3527	0.16	0.36	0	1

Table 14: Summary statistics – satisfaction with chief

variable	count	mean	sd	min	max
Satisfaction with chief (B)	2843	3.03	1.39	0	4

Table 15: Summary statistics – activity and interest variables

variable	count	mean	sd	min	max
Activity and interest (C)	3330	0.66	0.12	0	1
S4Q1	3504	2.38	1.68	0	5
S4Q2a	3441	0.96	0.20	0	1
S4Q2b	3440	0.93	0.26	0	1
S4Q3a	3420	2.91	0.41	0	3
S4Q3b	3421	2.89	0.47	0	3
S6Q7	3510	0.20	0.40	0	1
S6Q8	3514	0.06	0.24	0	1
S8Q4	3476	0.05	0.22	0	1

Table 16: Summary statistics – donating and buying variables

variable	count	mean	sd	min	max
Donating and buying (D)	2887	0.65	0.43	0	1
S14Q1	3424	0.70	0.46	0	1
S14Q2_d	2929	0.61	0.49	0	1
S14Q2	2929	15.67	81.29	0	2500
S14Q3	3528	0.97	0.18	0	1
S14Q4	3528	0.94	0.23	0	1

Table 17: Summary statistics – Non-response to potentially sensitive questions

variable	count	mean	sd	min	max
Non-response to pot. sen. q. – dummy (V1)	3227	0.20	0.40	0	1
Non-response to pot. sen. q. – count (V2)	3227	0.23	0.52	0	4

Table 18: Summary statistics – Not-answering to identifying variables

variable	count	mean	sd	min	max
Non-response to identifying questions – dummy	3527	0.01	0.09	0	1
Non-response to identifying questions – count	3527	0.01	0.13	0	5

Table 19: Summary statistics – identifying and demographic variables and other covariates

variable ³⁵	count	mean	sd	min	max
S1Q1	3464	45.98	15.07	17	89
S1Q2	3528	0.22	0.41	0	1
S1Q3_hhh	3524	0.71	0.45	0	1
S1Q4_akan	3528	0.01	0.09	0	1
S1Q5_akan	3528	0.58	0.49	0	1
S1Q6	3527	1.21	0.93	0	2
S1Q7	3523	1.48	0.85	0	2
S1Q8	3523	5.97	3.31	0	10
S1Q9_farm	3528	0.33	0.47	0	1
S1Q9_own_b	3528	0.30	0.46	0	1
S1Q9_wage_j	3522	0.17	0.38	0	1
S1Q10_nu	3528	1.18	0.42	1	4
S1Q11a	3525	0.71	0.45	0	1
S1Q11b	3525	0.96	0.20	0	1
S1Q11c	3524	0.97	0.17	0	1
S2Q1	3447	4.74	3.15	0	20
S2Q2	3460	3.60	3.19	0	20
S2Q3	3515	1.99	1.14	0	4
S2Q4	3316	16.26	7.51	0	24
S0Q7	3528	0.46	0.50	0	1
S0Q8a_english	3528	0.31	0.46	0	1
S0Q8a_akan	3528	0.67	0.47	0	1
S0Q8b	3528	0.11	0.31	0	1
S0Q9_mining	3528	0.21	0.41	0	1
S0Q9_oil	3528	0.05	0.22	0	1
S3Q1a	3525	0.93	0.26	0	1
S3Q1b	3526	0.85	0.36	0	1
S3Q1c	3526	0.97	0.17	0	1
S3Q1d	3523	0.32	0.47	0	1

³⁵ The following variables are dummy variables capturing certain response to the question:

S1Q3_hhh – the respondent is the household head;

S1Q4_akan – the respondent speaks Akan at home;

S1Q5_akan – the respondent is of Akan ethnicity;

S1Q9_farm – the respondent's main occupation is farming;

S1Q9_own_b – the respondent's main occupation is working for own or family owned business in non-farming, fishing, forestry or mining related fields;

S1Q9_wage_j – the respondent has a wage job in in non- farming, fishing, forestry or mining related fields;

S1Q10_nu – number of social positions hold in the community;

S0Q8a_english – the interview is conducted in English;

S0Q8a_akan – the interview is conducted in Akan language;

S0Q9_mining – the district is classified as mining;

S0Q9_oil– the district is classified as oil district

Table 20: Summary statistics – Giving name and contact information

variable	count	mean	sd	min	max
Giving name and contact (W)	3528	0.98	0.13	0	1
S1Q0a	3528	0.98	0.13	0	1
S1Q0b	3528	0.97	0.17	0	1
S1Q0c	3528	0.98	0.13	0	1

9.6 Models with interaction effect results

Table 21: Results – model with interaction effects – effects on answers to the potentially sensitive questions

VARIABLES	(1) Knowledge (A)	(2) Satisfaction with chief (B)	(3) Activity and interest (C)	(4) Donating and buying (D)	(5) Non-response to pot. sen. q.– dummy (V1)	(6) Non-response to pot. sen. q.– count (V2)
identified	-0.00271 (0.0115)	-0.134 (0.0849)	0.00722 (0.00840)	0.0851*** (0.0296)	0.0162 (0.0237)	-0.00797 (0.0320)
identified*leader	-0.00780 (0.0154)	0.0673 (0.101)	-0.00816 (0.00943)	-0.0926*** (0.0350)	-0.00419 (0.0304)	0.00799 (0.0420)
leader	0.0734*** (0.0113)	0.0488 (0.0785)	0.0879*** (0.00942)	0.161*** (0.0280)	0.0120 (0.0226)	-0.000149 (0.0318)
Constant	0.116*** (0.0104)	3.044*** (0.0712)	0.599*** (0.00908)	0.532*** (0.0250)	0.181*** (0.0177)	0.236*** (0.0256)
Observations	3,526	2,843	3,330	2,887	3,227	3,227
R-squared	0.023	0.002	0.116	0.020	0.000	0.000

Standard errors clustered over district in parentheses
*** p<0.01, ** p<0.05, * p<0.1

9.7 Comparison of permutation test and t-test results for the main outcomes

Table 22: Comparison of permutation test and t-test results

Outcome	Sample					
	All		Common citizens		Leaders	
	p-value permuta- tion test	p-value t-test	p-value permuta- tion test	p-value t-test	p-value permuta- tion test	p-value t-test
Non-response to pot. sen. q. – dummy (V1)	0.349	0.339	0.473	0.472	0.499	0.498
Non-response to pot. sen. q. – count (V2)	0.868	0.866	0.788	0.790	0.999	0.999
Knowledge (A)	0.216	0.220	0.809	0.801	0.275	0.274
Satisfaction with chief (B)	0.072	0.073	0.117	0.115	0.310	0.314
Activity and interest (C)	0.983	0.984	0.378	0.380	0.810	0.810
Donating and buying (D)	0.137	0.137	0.002	0.002	0.703	0.698

P-values resulting from mean comparison permutation test and t-test, comparing the two treatment conditions for different samples and outcomes.
Means and standard deviations of the outcomes can be seen in 0.
Permutation test with 10000 or 100000 repetitions, depending on a variable.

9.8 Treatment effects on individual potentially sensitive questions

Table 23: Treatment effects on individual potentially sensitive questions

Outcome	Variable	N	Mean identified (SD)	Mean unidentified (SD)	Diff	p-value	p-value 95% confidence interval		p-value MHT adjusted
							LB	UB	
Knowledge (A)	S9Q1	3527	0.36 (0.48)	0.38 (0.49)	-0.02	0.153	0.146	0.160	1
	S9Q4	3528	0.13 (0.33)	0.13 (0.34)	-0.00	0.810	0.802	0.818	1
	S9Q8	3527	0.21 (0.41)	0.23 (0.42)	-0.02	0.145	0.138	0.152	1
	S9Q12	3526	0.05 (0.22)	0.06 (0.23)	-0.01	0.216	0.208	0.224	1
	S9Q16	3527	0.16 (0.37)	0.15 (0.36)	0.01	0.659	0.649	0.668	1
Activity and interest (C)	S4Q1	3504	2.34 (1.68)	2.42 (1.68)	-0.08	0.172	0.165	0.179	1
	S4Q2a	3441	0.96 (0.20)	0.96 (0.20)	0.00	0.901	0.895	0.907	1
	S4Q2b	3440	0.93 (0.25)	0.92 (0.27)	0.01	0.137	0.131	0.144	1
	S4Q3a	3420	2.92 (0.39)	2.91 (0.42)	0.01	0.468	0.458	0.477	1
	S4Q3b	3421	2.89 (0.45)	2.89 (0.48)	0.01	0.634	0.625	0.644	1
	S6Q7	3510	0.21 (0.40)	0.20 (0.40)	0.00	0.891	0.885	0.897	1
	S6Q8	3514	0.06 (0.24)	0.07 (0.25)	-0.01	0.243	0.235	0.252	1
	S8Q4	3476	0.05 (0.22)	0.05 (0.22)	0.00	0.878	0.871	0.884	1
D.and b.(D)	S14Q1	3424	0.72 (0.45)	0.68 (0.46)	0.03	0.029	0.028	0.030	0.493
	S14Q2_d	2929	0.62 (0.49)	0.60 (0.49)	0.02	0.336	0.327	0.345	1
Not used	S14Q3	3528	0.97 (0.17)	0.96 (0.31)	0.01	0.310	0.307	0.313	1
	S14Q4	3528	0.95 (0.22)	0.94 (0.32)	0.01	0.321	0.318	0.323	1
Permutation test results with 10000 repetitions. For outcome S14Q1 with 100000 repetitions.									

9.9 Heterogeneous treatment effects on individual potentially sensitive questions

Table 24: Treatment effects on individual potentially sensitive questions – common citizens

Outcome	Variable	N	Mean identified (SD)	Mean unidentified (SD)	Diff	p-value	p-value 95% confidence interval		p-value MHT adjusted
							LB	UB	
Knowledge (A)	S9Q1	1213	0.23 (0.42)	0.23 (0.42)	-0.00	0.927	0.921	0.932	1
	S9Q4	1213	0.08 (0.27)	0.09 (0.29)	-0.02	0.280	0.272	0.289	1
	S9Q8	1213	0.19 (0.39)	0.16 (0.37)	0.03	0.153	0.146	0.160	1
	S9Q12	1213	0.02 (0.14)	0.03 (0.18)	-0.01	0.209	0.201	0.217	1
	S9Q16	1213	0.15 (0.36)	0.15 (0.36)	-0.00	0.922	0.917	0.927	1
Activity and interest (C)	S4Q1	1206	1.92 (1.50)	1.93 (1.49)	-0.01	0.866	0.859	0.873	1
	S4Q2a	1207	0.89 (0.31)	0.89 (0.31)	0.00	0.897	0.891	0.903	1
	S4Q2b	1206	0.83 (0.37)	0.80 (0.40)	0.03	0.127	0.120	0.133	1
	S4Q3a	1197	2.84 (0.55)	2.80 (0.60)	0.04	0.261	0.253	0.270	1
	S4Q3b	1197	2.79 (0.62)	2.74 (0.69)	0.05	0.198	0.190	0.206	1
	S6Q7	1208	0.10 (0.30)	0.12 (0.32)	-0.02	0.372	0.363	0.382	1
	S6Q8	1211	0.01 (0.11)	0.02 (0.12)	-0.00	0.664	0.655	0.673	1
	S8Q4	1176	0.02 (0.15)	0.03 (0.17)	-0.01	0.447	0.438	0.457	1
D.and b.(D)	S14Q1	1173	0.67 (0.47)	0.58 (0.49)	0.09	0.0012	0.0010	0.0014	0.040
	S14Q2_d	1057	0.57 (0.50)	0.50 (0.50)	0.08	0.012	0.011	0.013	0.397
Not used	S14Q3	1213	0.93 (0.25)	0.90 (0.29)	0.03	0.069	0.067	0.071	1
	S14Q4	1213	0.89 (0.32)	0.85 (0.36)	0.04	0.032	0.031	0.033	1

Permutation test results with 10000 repetitions.
For S14Q1, S14Q2_d, S14Q3 and S14Q4 with 100000 repetitions.

Table 25: Treatment effects on individual potentially sensitive questions – leaders

Outcome	Variable	N	Mean identified (SD)	Mean unidentified (SD)	Diff	p-value	p-value confidence interval		p-value MHT adjusted
							LB	UB	
Knowledge (A)	S9Q1	2314	0.43 (0.49)	0.45 (0.50)	-0.03	0.174	0.167	0.182	1
	S9Q4	2315	0.15 (0.36)	0.15 (0.35)	0.01	0.659	0.650	0.668	1
	S9Q8*	2314	0.23 (0.42)	0.27 (0.45)	-0.04	0.012	0.012	0.013	0.410
	S9Q12	2313	0.06 (0.25)	0.07 (0.26)	-0.01	0.500	0.490	0.509	1
	S9Q16	2314	0.16 (0.37)	0.15 (0.36)	0.01	0.536	0.526	0.546	1
Activity and interest (C)	S4Q1	2298	2.58 (1.73)	2.67 (1.72)	-0.09	0.212	0.204	0.220	1
	S4Q2a	2234	0.99 (0.09)	0.99 (0.09)	0.00	0.905	0.899	0.911	1
	S4Q2b	2234	0.99 (0.11)	0.98 (0.14)	0.01	0.153	0.146	0.161	1
	S4Q3a	2223	2.96 (0.26)	2.96 (0.27)	0.00	0.994	0.993	0.996	1
	S4Q3b	2224	2.95 (0.31)	2.96 (0.30)	-0.01	0.534	0.525	0.544	1
	S6Q7	2302	0.26 (0.44)	0.25 (0.43)	0.02	0.398	0.389	0.408	1
	S6Q8	2303	0.08 (0.28)	0.10 (0.29)	-0.01	0.345	0.336	0.355	1
	S8Q4	2300	0.07 (0.25)	0.06 (0.24)	0.01	0.519	0.509	0.529	1
D.and b.(D)	S14Q1	2251	0.75 (0.44)	0.74 (0.44)	0.01	0.676	0.673	0.679	1
	S14Q2_d	1872	0.64 (0.48)	0.65 (0.48)	-0.01	0.524	0.520	0.527	1
Not used	S14Q3	2315	0.99 (0.11)	0.99 (0.09)	-0.00	0.397	0.394	0.400	1
	S14Q4	2315	0.98 (0.14)	0.99 (0.11)	-0.01	0.183	0.180	0.185	1

Permutation test results with 10000 repetitions.
For S9Q8, S14Q1 and S14Q2_d with 100000 repetitions.

9.10 Robustness results

Table 26: Donating and buying outcome robustness

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Donating and buying (D)	Donating and buying (D)	Donating and buying (D)	Donating and buying (D)	Donating and buying (D) ³⁶
identified	0.0851*** (0.0296)	0.0888*** (0.0269)	0.0600** (0.0272)	0.0590** (0.0258)	0.162 (0.117)
identified*leader	-0.0926*** (0.0350)	-0.0913*** (0.0321)	-0.0619* (0.0316)	-0.0592* (0.0302)	-0.0676* (0.0388)
leader	0.161*** (0.0280)	0.0971*** (0.0292)	0.0363 (0.0268)	0.0174 (0.0275)	0.0948*** (0.0304)
S1Q1		-0.00231*** (0.000665)		-0.00150** (0.000631)	-0.00244*** (0.000892)
S1Q2		-0.0349 (0.0283)		-0.00327 (0.0252)	-0.0471 (0.0345)
S1Q6		0.0157 (0.0112)		0.00197 (0.0101)	0.0374** (0.0145)
S1Q7		0.0241 (0.0151)		0.0350*** (0.0133)	0.0297 (0.0219)
S1Q8		0.0193*** (0.00442)		0.0162*** (0.00403)	0.0199*** (0.00643)
S1Q9_farm		0.0250 (0.0234)		0.00210 (0.0225)	0.0356 (0.0365)
S1Q9_own_b		0.0251 (0.0228)		-0.00204 (0.0205)	0.00636 (0.0315)
S1Q9_wage_j		0.0620** (0.0238)		-0.00696 (0.0253)	0.0624* (0.0345)
S1Q10_nu		0.0400** (0.0198)		0.0213 (0.0175)	0.0271 (0.0231)
S1Q11a		-0.0436** (0.0185)		-0.0253 (0.0188)	-0.0172 (0.0280)
S1Q11b		0.0535 (0.0411)		0.0632 (0.0408)	0.0821 (0.0560)
S0Q*		YES ***		YES **	
S3Q1*		YES *		YES	
i.RA & S0Q8b			YES ***	YES ***	
beg_S1Q1					-0.000127 (0.00120)
beg_S1Q2					0.0165 (0.0465)
beg_S1Q6					-0.0462** (0.0202)
beg_S1Q7					-0.00886 (0.0296)
beg_S1Q8					0.00334 (0.00843)
beg_S1Q9_farm					-0.0138 (0.0592)
beg_S1Q9_own_b					0.0270 (0.0495)
beg_S1Q9_wage_j					0.00507 (0.0515)
beg_S1Q10_nu					0.0308 (0.0350)
beg_S1Q11a					-0.0406 (0.0365)
beg_S1Q11b					-0.0529 (0.0769)
Constant	0.532*** (0.0250)	0.290*** (0.0732)	0.266*** (0.0393)	0.0958 (0.0823)	0.361*** (0.0814)
Observations	2,887	2,828	2,887	2,828	2,832
R-squared	0.020	0.110	0.242	0.282	0.105

Standard errors clustered over district in parentheses; *** p<0.01, ** p<0.05, * p<0.1
beg_ denotes interaction of the variable with treatment variable;
S0Q* is S0Q7, S0Q8a_english, S0Q9_mining and S0Q9_oil; S3Q1* is S3Q1a, S3Q1b, S3Q1c, S3Q1d;
beg_S3Q1* is beg_S3Q1a, beg_S3Q1b, beg_S3Q1c, beg_S3Q1d

³⁶ For outcome constructed from the two variables of 'Donating and buying' outcome plus questions about willingness to control a project (S14Q4) and participate in meeting informing about oil revenues (S14Q3), both effect of the treatment (p=0.077) and the treatment*leader interaction (p=0.017) are found statistically significant.

Table 27: Robustness analysis - models with significant treatment or subgroup interaction effects

VARIABLES	(1) Knowledge	(2) Satisfaction with chief	(3) Activity and interest
identified	0.00488 (0.00968)	-0.133* (0.0790)	-0.0544* (0.0299)
identified*leader	-0.0223* (0.0130)	0.0523 (0.0955)	-0.0176* (0.0100)
leader	0.0364*** (0.0123)	0.0946 (0.0951)	0.0734*** (0.00952)
S1Q1	1.87e-06 (0.000309)		-0.000404* (0.000229)
S1Q2	-0.0294*** (0.0101)		-0.0321*** (0.00952)
S1Q6	0.0110** (0.00443)		0.00367 (0.00337)
S1Q7	0.0115* (0.00606)		-0.0129** (0.00503)
S1Q8	0.0133*** (0.00201)		0.00202 (0.00164)
S1Q9_farm	0.0122 (0.0107)		0.00245 (0.00874)
S1Q9_own_b	-0.00413 (0.00962)		-0.00159 (0.0103)
S1Q9_wage_j	0.0355*** (0.0120)		0.00459 (0.0107)
S1Q10_nu	0.00379 (0.00817)		0.00983* (0.00509)
S1Q11a	0.0103 (0.00832)		0.0156** (0.00750)
S1Q11b	0.0114 (0.0178)		0.0244 (0.0198)
S0Q7	-6.15e-05 (0.00744)		
S0Q8a_english	-0.00128 (0.00957)		
S0Q9_mining	0.0453*** (0.0124)		
S0Q9_oil	0.0601*** (0.0203)		
S3Q1*	YES***		
i.RA & S0Q8b	YES***	YES***	
beg_S1Q1			0.000734** (0.000281)
beg_S1Q2			0.00715 (0.0116)
beg_S1Q6			0.00224 (0.00454)
beg_S1Q7			0.00958 (0.00645)
beg_S1Q8			0.000467 (0.00187)
beg_S1Q9_farm			0.00242 (0.0119)
beg_S1Q9_own_b			0.00233 (0.0131)
beg_S1Q9_wage_j			0.00518 (0.0152)
beg_S1Q10_nu			-0.00667 (0.00850)
beg_S1Q11a			0.000436 (0.00861)
beg_S1Q11b			0.0191 (0.0261)
Constant	-0.0376 (0.0314)	2.752*** (0.114)	0.590*** (0.0267)
Observations	3,444	2,843	
R-squared	0.228	0.056	3,268

Standard errors clustered over district in parentheses

*** p<0.01, ** p<0.05, * p<0.1

beg_ denotes interaction of the variable with treatment variable
and S3Q1* is S3Q1a, S3Q1b, S3Q1c, S3Q1d

9.11 Pilot interviews images

Figure 2: Pilot interview



Figure 3: Pilot interview with a chief

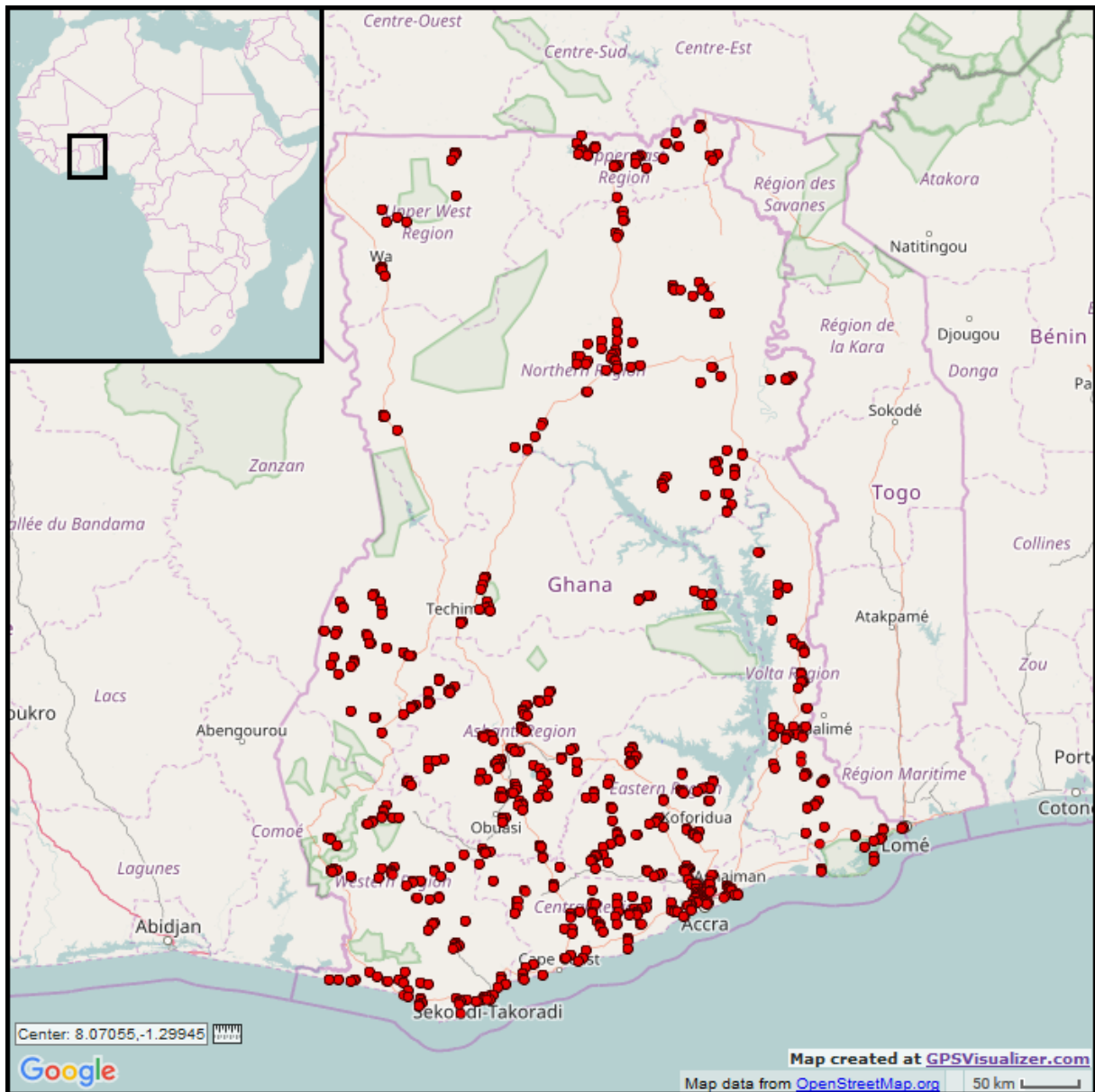


Figure 4: Pilot interview



9.12 Map of the interview locations

Figure 5: Map of the interview locations



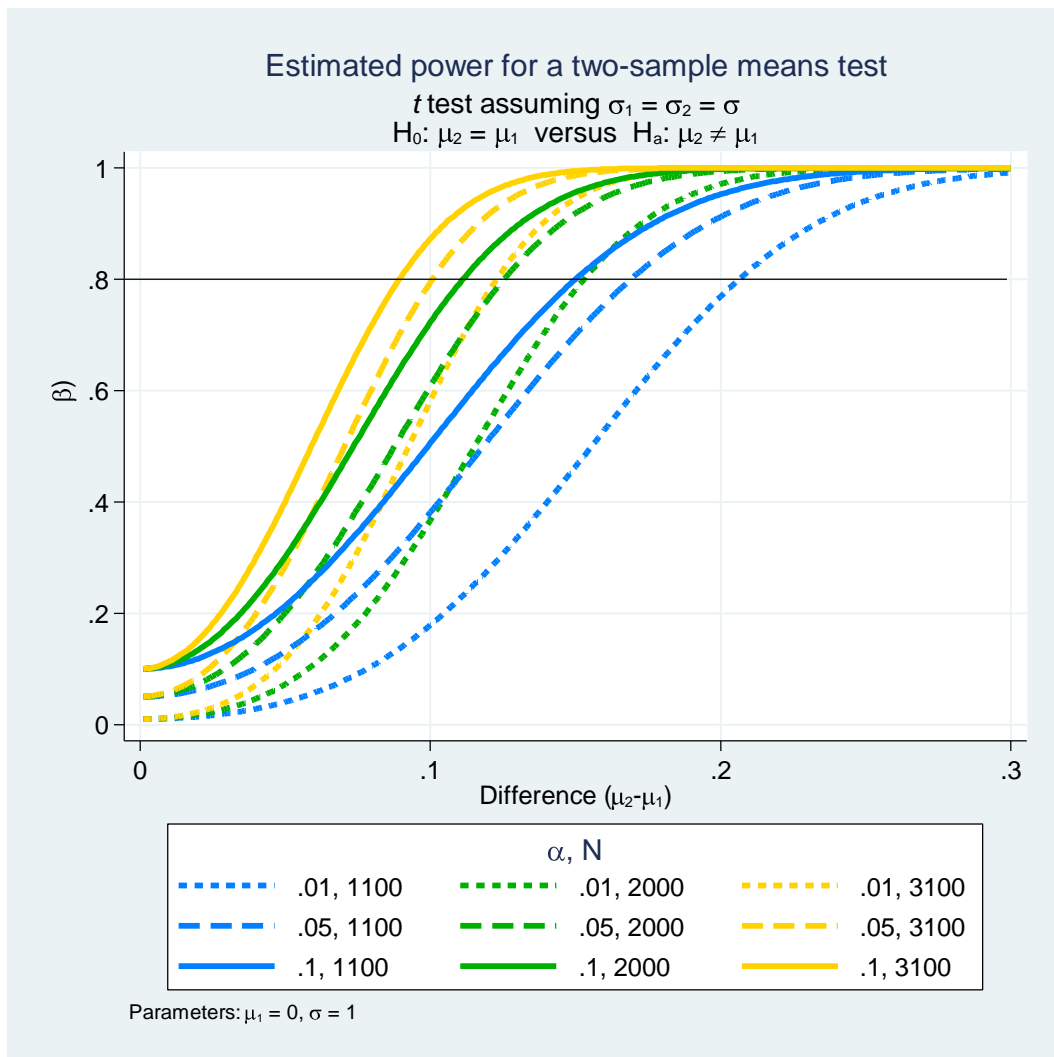
9.13 Power

To estimate power of our experiment we use formula for calculation power of t-test. Permutation test is used to test the hypothesis, however t-test and permutation tests are shown to have similar power for both large samples and, under certain assumptions, small samples (Hoeffding, 1952; Weber and Sawilowsky, 2009), therefore we chose to estimate power of our experiment as if t-test would be used.

Sample sizes used in the power calculations are rounded values close to mean number of observations for different outcomes in our sample (1100 for citizen subsample, 2000 for leaders subsample and 3100 for the whole sample). Further, the treatment arms are for simplicity assumed to be of an equal size.

Figure 6 shows the relation of power and minimum detectable effect size for significance levels of $\alpha = 0.1, 0.05$ and 0.01 . Resulting minimum detectable difference with 80% power, depending on sample size and significance levels, ranges between about 0.09 and about 0.2 SD.

Figure 6: Relation of minimum detectable difference and power for different sample sizes and significance levels



9.13.1 Power accounting for multiple hypothesis testing adjustment

However, calculating power for simple mean difference with standard significance levels would result in over-estimation of power of our experiment. As multiple hypotheses testing (MHT) adjustment is used on the p-values resulting from our tests, we have to also adjust the power calculation for such change as well. As in our analysis, we use Bonferroni correction to adjust the significance level.

Figure 7 and Figure 8 show the relation between different levels of minimum detectable difference and power, for different families of MHT adjustment and different sample sizes (subgroups)

For analysis of the whole sample, the resulting minimum detectable difference, with $\beta = 0.8$ and $\alpha = 0.05$, ranges from 0.13 to 0.14 SD, depending on the MHT adjustment. For subgroup analysis, the minimum detectable difference ranges around 0.16 to 0.17 SD for leader subgroup and around 0.21 to 0.23 for common citizen subsample depending on the MHT adjustment. Generally, the resulting power is about the power needed to distinguish 'small effect' (0.2 SD) as characterized by Cohen (1988).

Figure 7: Relation of minimum detectable difference and power for overall analysis with different MHT adjustment levels

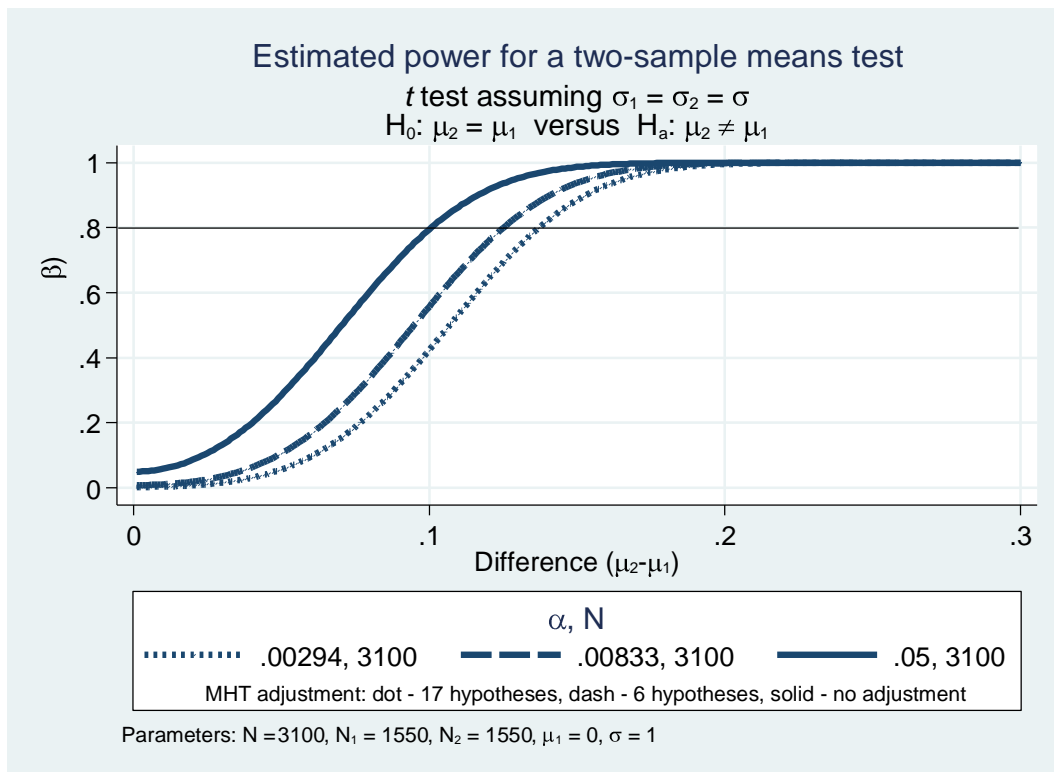


Figure 8: Relation of minimum detectable difference and power for a subgroup-analysis with different MHT adjustment levels

