Asking Survey Respondents about Reasons for Their Behavior: A Split Ballot Experiment in Ethiopia

Survey Methods: Insights from the Field

Charles Q. Lau | Gretchen McHenry

How to cite this article : Lau C.Q., & McHenry G. (2014). Asking Survey Respondents about Reasons for Their Behavior: A Split Ballot Experiment in Ethiopia, Survey Methods: Insights from the Field. Retrieved from https://surveyinsights.org/?p=2914

DOI : 10.13094/SMIF-2014-00001

Abstract : When policymakers design programs and policies, they often want to understand why individuals engage in particular behaviors. Collecting survey data about respondents' reasons for their behavior presents important challenges, and there is little methodological research on this topic. We conducted an experiment to investigate the best practices for asking questions about respondents' reasons for their behavior. We embedded a split ballot experiment in a face-to-face survey of 608 entrepreneurs in Ethiopia. Respondents were asked questions about why they did not engage in three business practices (advertising, sharing product storage, and switching suppliers). When asked these questions, respondents were randomly assigned to one of three conditions: close-ended questions, open-ended questions with interviewer probing, and open-ended questions without probing. Respondents endorsed more responses when asked close-ended (versus open-ended) questions. Close-ended responses produced higher rates of socially undesirable responses and fewer "other" responses. Notably, probing had no effect on the number or types of responses given. Our results suggest some best practices for asking respondents questions about reasons for their behavior.

Introduction

When policymakers design programs and policies, they often want to understand why individuals act in particular ways. Although some researchers caution against asking respondents to cite reasons why they do (or do not) engage in behaviors (Pasek and Krosnick 2010: 41; Wilson 2010), data about respondent motivations for their behavior are analytically useful. By understanding the causes of people's behavior, policymakers can take steps to reduce undesirable behaviors or encourage desirable behaviors. For example, questions in the National Health Interview Survey ask respondents why they delayed seeking medical care, allowing researchers to understand barriers to healthcare access (Centers for Disease Control and Prevention, 2012). The Current Population Survey also asks individuals why they did not vote or register to vote, shedding light on mechanisms underlying political participation (United States Census Bureau, 2010).

Given the value of data about individuals' motivations for behavior, it is notable that there is little research on best practices for designing these questions. To address this gap, we embedded a split ballot experiment in a face-to-face survey of 608 entrepreneurs in Ethiopia. Conducting the survey in a

developing country allowed us to study this topic in a context that poses additional challenges to asking such questions. In our survey, we randomly assigned one of three methods for asking respondents about reasons for their behavior. The methods differ in whether questions are close-ended versus open-ended, and whether interviewers probed respondents. Our analysis evaluates the three methods by comparing numbers of endorsed responses and the number of socially desirable responses in particular.

Background

There are many ways to collect survey data about reasons for respondent's behavior. The method we adopt in this paper involves pre-specifying a list of possible reasons for behavior on the instrument, and then having interviewers record whether each response applies or not (using yes/no responses for each item).[1] When designing this type of question, researchers must make two key decisions (Wilson 2010). First, should interviewers ask close-ended questions—reading each possible response and then recording a yes/no response for each? Or should interviewers ask open-ended questions and then record yes/no responses based on the respondent's open-ended answer? Second, if open-ended questions are used, should interviewers probe respondents for clarification? In the following sections, we draw from previous literature to develop expectations about the advantages and disadvantages of different types of close-ended and open-ended questions.

Close-Ended Versus Open-Ended Questions

Asking close-ended questions (rather than open-ended ones) is a form of standardized interviewing, in which each respondent hears the exact same question and response options, regardless of the interview flow or tone (Converse and Schuman 1974; Schober and Conrad 2002). This approach has the advantage of encouraging respondents to consider reasons they had not previously thought about. It also encourages respondents to think about the issue from a variety of perspectives, which may result in a greater number of endorsed responses and may also limit "don't know" responses. Further, close-ended questions may reduce respondents' concerns about reporting socially undesirable answers. Reading response options may give tacit approval for socially undesirable answers and may help develop a sense of trust between a respondent and the interviewer. It also means that respondents do not have to verbally state a socially undesirable admission about themselves, which is the case in an open-ended question format.

Close-ended questions also have disadvantages, many of which are rectified by open-ended formats. Reading close-ended response options during the interview can be time consuming and feel repetitive to the respondent. Open-ended questions, in contrast, may be more engaging for respondents because they comport more with conversational norms and allow respondents to better communicate the reasoning behind their behavior (Fowler 1995). Open-ended questions have also been shown to solicit meaningful, salient information from respondents (Geer 1988; Geer 1991). In addition, close-ended questions may suffer from primacy or recency effects, where the first (or last) response options are more likely to be endorsed, whereas primacy and recency effects are eliminated with open-ended questions. Finally, reading response options may implicitly convey the researcher's values or preferences, potentially biasing respondents in a particular direction. Open-ended questions, in contrast, do not have this limitation, and also provide an opportunity to collect data about issues researchers had not previously considered.

Probing

If open-ended questions are used, interviewers could simply select the pre-specified reasons that apply to the respondents' open-ended answer (without probing), or probe for a more complete or detailed response. Probing can facilitate respondent comprehension of the question and may reduce errors in the interviewer's coding of responses. An exchange with the interviewer also may encourage respondents to think more deeply about their answers. This increased engagement with the question-answer process, as well as with the interviewer, may yield more endorsed answers, reduce respondent satisficing, and increase reports of socially undesirable behaviors. Schaeffer and Maynard (2008) show that directive probes or requests for confirmation from interviewers increase a respondent's likelihood of reporting embarrassing or incriminating responses.

There are three potential drawbacks of probing. First, probing gives interviewers more discretion and may lead to increased interviewer errors and variance. For example, Fowler and Mangione (1990) find that the number of probes, directive probes used, and occasions where an interviewer failed to probe are associated with increased error. They also suggest that probing may introduce interviewer-level variance, which decreases the efficiency of survey estimates. However, Schober and Conrad's (1997) small-scale experimental study finds no evidence that probing increases interviewer error or variance. Second, probing may increase the number of "other" responses if the interviewer cannot code the response into one of the pre-existing categories due to the nuanced response from the respondents. Third, the conversational nature of the interview may increase administration time, increasing survey costs and field data collection time.

In sum, the literature suggests that there are advantages and disadvantages of using close-ended versus open-ended questions, as well as probing versus not probing. Given the lack of research in this area, we designed a split ballot experiment to investigate the quality of data produced by three methods.

Experimental Design

Data

We analyze data from the *Kal Addis* Business Survey (KABS), a paper-and-pencil interview of 608 entrepreneurs in the Ethiopian capital of Addis Ababa. Eligible respondents were owners or senior managers of small and medium businesses (between 3 and 99 employees) based in Addis Ababa. Examples of businesses in the sample include a restaurant, car repair shop, and a textile manufacturer. The purpose of KABS was to improve sampling and questionnaire design methodologies in developing countries, particularly for surveys of entrepreneurs. The survey measured entrepreneurs' attitudes and business practices, and included questions about purchasing raw materials from suppliers, advertising, product storage, among other topics. Professional Ethiopian interviewers with at least three years of interviewing experience administered the survey in the Amharic language in the summer of 2012. All interviewers also participated in a three day training and pre-test of the instrument. Throughout data collection, survey managers held quality review meetings with interviewers to enhance standardization and to answer questions about field implementation. The mean administration time was 29 minutes (standard deviation = 9 minutes).

Because a sampling frame of entrepreneurs was not available in Addis Ababa, KABS used respondent-driven sampling (RDS). RDS is a method of chain referral sampling that combines a snowball sample with a mathematical model that adjusts for the non-random selection of the initial set of respondents (Heckathorn 1997). To implement RDS in KABS, we initially recruited a convenience sample of 24 individuals through personal networks. These individuals were interviewed and then provided with three

invitations to recruit up to three individuals to participate in the study. Each additional wave of recruits was asked to recruit up to three additional individuals. Recruited individuals contacted the field data collection teams, who then scheduled and conducted the interview in a location of the respondent's choosing. We provided a leather wallet to respondents for completing the survey and mobile phone airtime for referring others to the study. Because our focus is on the internal validity of the split ballot experiment, we do not apply weights from the RDS in the paper. Characteristics of the sample are presented in Table 1.

Three-quarters of respondents are male with an average age of 31 years old, reflecting the young age of the Ethiopian population. The majority of respondents are owners of the business (82%) versus managers (18%). The sample is comprised of businesses in the manufacturing (14%), service (48%), and trade (39%) sectors. The vast majority of businesses were profitable in the past year, and on average, businesses had eight employees and were six years old.

Table 1. Sample characteristics

Respondent Characteristics			**Business Characteristics**	
Gender (n = 608)			**Sector (n = 608)**	
Male	76%		Manufacturing	14%
Female	24%		Service	48%
Total %	100%		Trade	39%
			Total %	100%
Educational attainment (n = 608)				
Did not complete secondary	15%		**Annual revenue in dollars (n = 539)**	
Secondary school	34%		Less than $2778	32%
Vocational or some university	31%		$2778 – $5,555	21%
Graduate degree or higher	20%		$5,556 – $13,889	13%
Total %	100%		$13,890 – $41,667	17%
			Over $41,667	16%
Position in business (n = 608)			Total %	100%
Owner	82%			
Senior day-to-day manager	18%		**Profit last year (n = 583)**	
Total %	100%		Made money	70%
			Lost money	7%
Age in years (n = 608)	31.2		Broke even	23%
Standard deviation	6.9		Total %	100%
Hours worked/week (n = 596)	55.0		**Number of employees (n = 608)**	7.9
Standard deviation	20.0		Standard deviation	12.3

			Mean business age (years) (n = 606)	6
			Standard deviation	6
Note: The total sample size for KABS sample is 608. The valid sample size for each variable is indicated in table. Percentages may not sum to 100 due to rounding.				

KABS included questions about three business practices: advertising, switching to a new supplier to buy raw materials, and sharing product storage with another business. We present the exact question wording for these questions in the Appendix. These three practices facilitate economic growth and are practices that policymakers would like to encourage in developing countries. Therefore, understanding why individuals do not engage in these practices is important for policymakers who design interventions to stimulate economic growth. In different parts of the interview, respondents were asked if they engaged in these business practices. Those who said they did not take part in each business practice were asked why not. We generated pre-specified reasons for each behavior during formative research, which involved in-depth interviews with entrepreneurs, as well as a review of literature on entrepreneurship in Ethiopia. We modified these reasons throughout the pre-testing process.

Split Ballot Design

We developed three separate instruments, each with a different method of asking questions about reasons for respondent's behavior. Respondents were randomly assigned to one of three methods (Table 2). Each respondent was assigned to the same method across all three business practices based on their respondent ID number (itself randomly assigned). The randomization was successful in that there were no significant correlations between questionnaire version and respondent or business characteristics. Full tables are available from the authors upon request.

Table 2 shows that in the *close-ended* method, interviewers read every pre-specified response option while asking respondents a series of yes/no questions about whether the option applied or not. This method is the norm in social surveys and reflects standardized interviewing practices (Groves et al. 2009). We read the potential reasons orally (rather than using a showcard) because of the survey population's lower levels of literacy and unfamiliarity with showcards.

Table 2. Three methods of asking about reasons for behavior

	Close-ended	Open-ended with probing	Open-ended without probing
Interviewer reads response options	Yes	No	No
Interviewer probes	If needed	Yes	No
Number of respondents	203	203	202

In the *open-ended with probing* method, interviewers asked an open-ended question instead of reading the response options. The interviewer then coded the respondent's open-ended answer into the pre-specified options, and probed the respondent as needed. The interviewer did not record the verbatim open-ended response. Interviewers were trained to adopt conversational interviewing practices when

probing (Schober 1998), and used non-directive, neutral probes to clarify unclear or inadequate responses. Examples of probes included repeating the question, asking a general question, or asking a respondent to clarify a response. In the *open-ended without probing* method, interviewers asked an open-ended question, coded the open-ended data into the pre-specified response options, and did not probe. Again, the interviewer did not collect the verbatim response. This method combines elements of standardization (i.e., no interviewer-respondent discussion) and conversational interviewing (i.e., interviewer has discretion to select the appropriate response.) All three methods contained an "other (specify)" response. During preliminary analysis, we recoded some "other" responses into existing pre-specified categories or created new categories when the other (specify) meaning was unambiguous.

Hypotheses

Our analysis seeks to identify the method that produces the most useful data about why individuals do not engage in the three business practices. Because obtaining validation data for this type of information is difficult, we focus on the number of endorsed responses and socially undesirable responses in particular. Another possible indicator is timing data, but because KABS used paper-and-pencil interviewing (like most surveys in developing countries), timing data on individual questions were not available. Below, we describe each indicator and present hypotheses.

Number of endorsed responses: The number of responses that respondents select is indicative of greater engagement with the subject matter. A greater number of responses is also analytically useful because it helps analysts understand multiple influences on behavior.

Hypothesis 1: The *close-ended* method will result in greater number of endorsed reasons than either *open-ended* method because respondents must consider each option separately. Support for this reasoning comes from the web survey literature, which shows that respondents endorse more responses when presented with a yes/no matrix (that requires an answer for each response) rather than a "check all" list (Smyth et al. 2006).

Hypothesis 2: *Open-ended with probing* will lead to a greater number of endorsements than *open-ended without probing*. During probing, interviewers may encourage respondents to think about the issue from multiple angles and therefore provide more responses.

Socially undesirable reporting: We assume that respondents are reluctant to endorse responses that are socially undesirable, and that increases in socially undesirable reporting reflect a more preferable method. This logic has been widely used in other areas, such as mode effects on reports of sexual activity (Tourangeau and Smith 1996) and smoking (Currivan et al. 2004). We include a range of socially undesirable measures in our study, ranging from more sensitive (e.g., reporting distrust of others) to less sensitive (e.g., reporting lack of knowledge about an issue).

Hypothesis 3: The *close-ended* method will yield more socially undesirable reporting than either *open-ended* method because the interviewer-supplied responses give tacit approval to the possibility of the response. In addition, the respondent only has to say "yes" to endorse a socially undesirable behavior in the *close-ended* method, whereas the respondent must verbalize the socially undesirable behavior in the *open-ended* methods.

Hypothesis 4: *Open-ended with probing* will lead to more socially undesirable reports than *open-ended without probing*. Probing may help an interviewer build rapport with a respondent and uncover issues that

respondents do not immediately discuss.

Results

Number of Endorsed Reasons

In Table 3, we present the number of reasons endorsed by each experimental group, separately for the three business practices. We report the percentage of respondents that endorsed more than one reason, the percentage distribution of the number of reasons endorsed, and the mean number of reasons. For "more than one reason" and "mean number of reasons," we use superscripts to highlight statistically significant differences ($p < .05$) that were obtained through *post-hoc* t-tests.

Table 3: Number of Reasons Endorsed (Percentages unless noted)

A. Reasons for Not Advertising	**Close-ended** (n = 167)	**Open-ended with probing** (n = 165)	**Open-ended without probing** (n = 163)
More than one reason	34[b, c]	18[a]	23[a]
Number of reasons			
0	1	2	4
1	66	80	73
2	25	15	21
3	7	2	2
4	1	1	0
Mean number of reasons (std. dev)	1.4 [b, c](.69)	1.2 [a](.56)	1.2 [a](.53)
B. Reasons for Not Switching Supplier	**Close-ended** (n = 58)	**Open-ended with probing** (n = 67)	**Open-ended without probing** (n = 62)
More than one reason	17	7	6
Number of reasons			
0	3	3	3
1	79	90	90
2	7	7	3
3	9	0	3
4	2	0	0
Mean number of reasons (std. dev)	1.3 [b](.74)	1.0 [a](.32)	1.1(.44)

C. Reasons for Not Sharing Storage	**Close-ended**(n = 37)	**Open-ended with probing**(n = 41)	**Open-ended without probing**(n = 35)
More than one reason	16	5	9
Number of reasons			
0	3	0	0
1	81	95	91
2	11	5	3
3	5	< 1	3
4	0	< 1	3
Mean number of reasons (std. dev)	1.2(.57)	1.0(.22)	1.2(.62)
[a] Statistically significant difference from close-ended ($p < .05$)[b] Statistically significant difference from open-ended with probing ($p < .05$)[c] Statistically significant difference from open-ended without probing ($p < .05$)			

In the reasons for not advertising panel, the results show that the close-ended design yielded more reasons than both open-ended methods. In the close-ended group, 34% of respondents provided more than one response, compared to 18% and 23% for the open-ended groups with and without probing, respectively. The differences between the close-ended group and both open-ended groups were statistically significant ($p < .05$). The full distribution shows that the close-ended group reported two reasons 25% of the time, compared to 15% for the open-ended with probing and 21% for the open-ended without probing group. The close-ended group also provided a higher mean number of reasons than both open-ended groups ($p < .05$). There was no statistically significant difference, however, between the two open-ended groups in the number of endorsed reasons.

We observed a similar pattern in the “switching supplier” panel. The close-ended group reported more than one reason in 17% of cases, higher than the open-ended groups with probing (7%) and without probing (6%), though these differences were only marginally statistically significant ($p < .10$). However, the close-ended group had a significantly higher ($p < .05$) mean number of reasons endorsed (1.3) compared to the open-ended with probing group (1.0). The results in the “sharing storage” panel follow the same pattern. The differences, however, are not statistically significant, likely due to the small sample sizes.

In sum, the close-ended method produced endorsements of more options compared to open-ended methods, supporting Hypothesis 1. The results, however, do not provide support for Hypothesis 2: probing had no effect on the number of reasons respondents endorse.

Type of Responses Provided

Next, we investigated how question design affected the number of socially undesirable responses provided, separately by the three business practices.

Reasons for Not Advertising

In Table 4, we show the reasons respondents provided for not advertising, separately by experimental group. Several of these reasons are socially undesirable, such as the reason that advertising might lead to an "increase government inspections or auditing." This reason is socially taboo because it indirectly refers to bribes: In developing countries such as Ethiopia, advertising increases a business' prominence, making it an easier target for government officials to demand bribes through unnecessary inspections or audits. Respondents may not endorse this reason because they prefer to avoid discussing about the sensitive topic of bribes, and also to minimize being perceived as having paid bribes. Of respondents in the close-ended group, 14% cited this reason, twice as high as the open-ended with probing group (7%); this difference was statistically significant. Nine percent of the open-ended without probing group mentioned this reason.

Table 4. Reasons for Not Advertising, by Experimental Group (Percentages)

	Close-ended (n = 167)	**Open-ended with probing** (n = 165)	**Open-ended without probing** (n = 163)	**Overall χ^2**
Too expensive	51	50	50	$\chi^2(2) = 0.0$; p = .97
Wouldn't help business	44	35	44	$\chi^2(2) = 4.0$; p = .13
Would increase government inspections or auditing	14 [b]	7 [a]	9	$\chi^2(2) = 5.1$; p = .08
Business is too new or small	13	13	8	$\chi^2(2) = 3.0$; p = .22
Never thought about it	8 [b]	3 [a]	5	$\chi^2(2) = 4.8$; p = .09
Other	7	12 [c]	5 [b]	$\chi^2(2) = 6.5$; p = .04
Too complicated or takes too much time	5 [b]	1 [a]	2	$\chi^2(2) = 6.1$; p = .05

[a]Statistically significant difference from close-ended ($p < .05$)[b]Statistically significant difference from open-ended with probing ($p < .05$)[c] Statistically significant difference from open-ended without probing ($p < .05$)

Table 4 also contains two other reasons that, while not socially undesirable, may be sensitive to the method of questioning. These reasons include not advertising because it is too complicated or because the respondent had never thought of advertising. Although these reasons are not socially taboo, respondents may hesitate to report these reasons because the reasons suggest that respondents have low levels of sophistication in running a business. Never thinking of advertising was mentioned by 8% percent of respondents in the close-ended group, more than the open-ended groups with probing (3%) and without probing (5%). Similarly, 5% of the close-ended group said advertising was too complicated, higher than both open-ended groups.

These three results support Hypothesis 3, that close-ended questions will yield more socially undesirable responses. However, there is no support for Hypothesis 4, that probing allows interviewers to build a

rapport with respondents and is more likely to encourage socially undesirable reporting.

Reasons for Not Switching Supplier

Table 5 shows the reasons respondents provided for not switching the business from whom the respondent buys supplies or raw materials. The vast majority of respondents in all groups reported not switching suppliers because they were satisfied with their current supplier. There were no statistically significant differences in the reasons provided by the three experimental groups. It is possible that the highly skewed distribution of these reasons may account for the absence of an effect.

Table 5. Reasons for Not Switching Supplier, by Experimental Group (Percentages)

	Close-ended (n = 58)	**Open-ended with probing** (n = 67)	**Open-ended without probing** (n = 62)	**Overall χ^2**
Satisfied with current supplier	86	85	84	$\chi^2(2) = 0.0$; p = .94
Quality is too poor	10	4	3	$\chi^2(2) = 3.1$; p = .21
Finding a new supplier takes too long	9	4	3	$\chi^2(2) = 1.9$; p = .39
Too expensive	7	1	5	$\chi^2(2) = 2.3$; p = .32
Too complicated to switch	7	1	2	$\chi^2(2) = 3.7$; p = .16
Not available	5	6	8	$\chi^2(2) = 0.0$; p = .80
Other	2	1	2	$\chi^2(2) = 0.0$; p = .99

Reasons for Not Sharing Storage

In Table 6, we show the reasons that respondents provided for not sharing product storage with another business. The sample sizes in this table are small because only respondents who reported using storage (22% of the entire sample) were asked subsequent questions about sharing storage.

Not trusting other businesses is a socially taboo topic because community cohesion is valued in Ethiopia and openly discussing distrust of others is discouraged. This reason was endorsed by 30% of the close-ended group, significantly higher than the open-ended with probing group (10%). Only 14% of the open-ended without probing group cited this reason. This result supports Hypothesis 3 (close-ended responses will increase socially undesirable reporting), but there is no support for Hypothesis 4 (that probing increases socially undesirable reports). The experimental manipulation did not affect reports about “never thought about it,” which contrasts with the results above for the reasons for not advertising.

Table 6. Reasons for Not Sharing Storage, by Experimental Group (Percentages)				
	Close-ended (n = 37)	**Open-ended with probing** (n = 41)	**Open-ended without probing** (n = 35)	**Overall χ^2**
Never thought about it	38	34	51	$\chi^2(2) = 2.5$; p = .28
Can't trust other businesses	30[b]	10 [a]	14	$\chi^2(2) = 5.7$; p = .06
Don't need to share	22	27	29	$\chi^2(2) = 0.5$; p = .78
Can't find other businesses to share with	16	5	9	$\chi^2(2) = 2.9$; p = .23
Cost savings are not worth the effort	8	15	6	$\chi^2(2) = 1.9$; p = .39
Laws prohibit sharing	3	0	9	$\chi^2(2) = 4.2$; p = .12
Other	3	15[c]	0 [b]	$\chi^2(2) = 8.1$; p = .02
[a] Statistically significant difference from close-ended ($p < .05$)[b] Statistically significant difference from open-ended with probing ($p < .05$)[c] Statistically significant difference from open-ended without probing ($p < .05$)				

Discussion

Our goal was to investigate best practices for asking respondents about reasons for their behavior. Respondents endorsed more responses when asked close-ended (versus open-ended) questions. This finding suggests that close-ended questions may spark greater engagement with the subject matter because respondents are forced to consider each option on its own, rather than reporting "top of mind" responses. Close-ended responses also produced higher rates of socially undesirable responses, suggesting that close-ended responses may help to elicit attitudes on sensitive topics. Providing socially undesirable reasons through close-ended questions may reduce the stigma of the response. An alternative hypothesis is that respondents had simply never thought of that reason before. We leave it to future research to distinguish between these explanations.

Second, probing did not affect the number of overall responses or the number of socially undesirable responses provided. This lack of an effect is notable, particularly because the professional interviewers had experience and training in probing. In fact, for two out of three questions, probing leads to more "other" responses that could not be classified into existing or new categories. It is possible that spending time on probing may not be an efficient use of interviewers' efforts, particularly because interviewer probing may also introduce additional variance into estimates. However, additional studies based on larger sample sizes are needed to replicate this null finding, particularly for more difficult questions where probing might be more effective. Future research could also investigate what types of probes are most productive at eliciting sensitive data from respondents.

In sum, our results provide tentative support for the idea that close-ended questions without probing are

the preferred method of asking respondents to provide reasons for their behavior, at least for this population and topic. We are limited, however, in that we do not have a gold standard that could specify which of the three designs produces the most valid data. Future research should investigate the validity of different methods, particularly the assumption that the additional responses provided by the close-ended questions are meaningful. Researchers should also consider the possibility is that there is no method that provides “true” reports, but simply that the three methods collect different types of data. For example, open-ended questions may produce reasons that are immediately accessible in respondent’s minds, whereas close-ended questions can obtain reactions to issues respondents rarely consider in their day-to-day lives. Cognitive interviewing could be useful in understanding how respondents approach the response task of questions that ask about reasons for behavior.

Our results demonstrate the feasibility of asking respondents questions about reasons for their behavior, but also raise questions about the validity of these data. Rates of item non-response were less than 1% for these items, suggesting that individuals are willing to provide reasons for their behavior. However, our results cannot show whether the reasons provided are accurate. Respondents may intentionally misreport reasons for their behavior or otherwise rationalize their behavior. Citing reasons for behavior may be cognitively burdensome, particularly when respondents do not often consciously think about why they do (or do not) engage in behaviors. For example, asking about topics a respondent has not considered before may encourage the respondent to create an answer on the spot, leading to inconsistent responses across items designed to measure similar concepts (Wilson 2013). Alternatively, interviewers may make errors when classifying responses. Because of these limitations, we view respondent self-reports about why they engage in behaviors as one of several possible research methods (in addition to qualitative research and experiments) that policymakers could use when designing policies or programs to modify behavior.

We encourage future research on the best practices and validity of asking respondents questions about reasons for their behavior, particularly in larger samples in different populations and substantive content areas. Research could fruitfully investigate various designs for designing these questions, such as using showcards for close-ended lists or coding verbatim open-ended responses. Another promising avenue for future research is to study how to ask questions about why individuals *do* engage in behaviors rather than why they *do not.* We believe that future research about if and when it is appropriate to ask these questions will ultimately benefit policymakers who rely on social scientists to explain why individuals engage in particular behaviors.

Appendix_Question Wording

---

[1] An alternative method is collecting verbatim responses to open-ended questions and then coding the verbatim responses *post-hoc*, ideally with multiple coders. Although this method may reduce coding errors, it is also time and labor intensive. The survey we analyze in this paper was designed to be a rapid and low-cost survey for use in settings such as Ethiopia, so it used immediate classification methods by interviewers.

References

1. Centers for Disease Control and Prevention (2012). National Health Interview Survey 2012

Questionnaires, English. Retrieved electronically on 20 December, 2013 from http://www.cdc.gov/nchs/nhis/quest_data_related_1997_forward.htm.

2. Converse, J. M., and H. Schuman. 1974. *Conversations at Random: Survey Research as Interviewers See It.* Hoboken, NJ: John Wiley & Sons.

3. Currivan, D., A. Nyman, C. F. Turner, and L. Biener. 2004. Does Telephone Audio Computer-Assisted Self-Interviewing Improve the Accuracy of Prevalence Estimates of Youth Smoking? Evidence from the UMass Tobacco Study. *Public Opinion Quarterly* 68:542-564.

4. Fowler, F. J., and T. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park: Sage.

5. Fowler, F. J. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.

6. Geer, J. G. 1988. What Do Open-Ended Questions Measure? *Public Opinion Quarterly* 52: 365-367.

7. Geer, J. G. 1991. Do Open-Ended Questions Measure "Salient" Issues? *Public Opinion Quarterly* 55: 360-370.

8. Groves, R.M., F.J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology*. Hoboken, NJ: John Wiley & Sons.

9. Heckathorn, D. D. 1997. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, 44:174-199.

10. Pasek, J., and J. A. Krosnick. 2010. Optimizing Survey Questionnaire Design in Political Science: Insights From Psychology. In *Oxford Handbook of American Elections and Political Behavior,* ed. J. Leighley, 27-50. Oxford: Oxford University Press.

11. Schober, M. F. 1998. Making Sense of Questions: An Interactional Approach. In *Cognition and Survey Research*, eds. M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, and R. Tourangeau. New York: Wiley.

12. Schober, M. F., and F. G. Conrad. 1997. Does Conversational Interviewing Reduce Survey Measurement Error? *Public Opinion Quarterly* 61:576-602.

13. Schober, M. F., and F. G. Conrad. 2002. A Collaborative View of Standardized Survey Interviews. In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, eds. D. Maynard, H. Houtkoop-Steenstra, N. C. Schaeffer, and J. van der Zouwen. New York: Wiley.

14. Schaeffer, N. C., and D. W. Maynard. 2008. The Contemporary Standardized Survey Interview for Social Research. In *Envisioning the Survey Interview of the Future*, ed. F. G. Conrad and M. Schober. New York: Wiley.

15. Smyth, J.D., D.A. Dillman, L. M. Christian, and M.J. Stern. 2006. Comparing Check-all and Forced-Choice Question Formats in Web Surveys. *Public Opinion Quarterly*, 70:66-77.

16. Tourangeau, R. and Smith, T. W. 1996. Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, 60: 275-305.

17. United States Census Bureau (2010). Current Population Survey Voting and Registration Supplement Questionnaire, November, 2010. Retrieved electronically on 20 December, 2013 from http://www.census.gov/cps/methodology/techdocs.html.

18. Wilson, S. 2010. Cognitive Interview Evaluation of the 2010 National Health Interview Survey Supplement on Cancer Screenings & Survivorship: Results of interviews conducted October – December, 2008. National Center for Health Statistics. Hyattsville, MD.

19. Willson, S. 2013. Cognitive Interview Evaluation of the Federal Statistical System Trust Monitoring Survey, Round 1: Results of interviews conducted in October, 2011. National Center for Health Statistics. Hyattsville, MD.