What I Say Depends on How You Ask: Experimental Evidence of the Effect of Framing on the Measurement of Attitudes*

Jeffrey R. Bloem† and Khandker Wahedur Rahman‡

October 12, 2023

Abstract

We conduct a survey experiment to study the effects of statement framing in the measurement of attitudes. Half of our respondents received a questionnaire with three positively framed statements and three negatively framed statements aiming to measure attitudes toward mobile banking. The other half received statements framed in the converse. Respondents recorded their level of agreement or disagreement with the statements using a standard Likert scale. We find framing effects between 7 and 21 percentage points in the probability a respondent agrees with positively framed statements or disagrees with negatively framed statements. Additionally, framing of the first statement influences responses to subsequent statements. Moreover, using standard techniques for generating aggregate indices, we find that the framing of the underlying statements can meaningfully influence the relationship of the index with relevant covariates—in some cases changing the magnitude, statistical significance, and even the sign of the estimated relationship. We conclude by discussing how randomizing statement framing across respondents can help address bias in the measurement of attitudes.

Keywords: Survey design, attitudes, response bias, framing effects, non-classical measurement error, data collection

JEL Classification Codes: C83, D91, G41

*Authors are listed in alphabetical order. The authors thank Marc Bellemare, David Evans, Mattia Bertazzini, Abelardo De Anda Casas, Jason Kerwin, Mahreen Khan, Pramila Krishnan, Julien Labonne, Sanghamitra Mukherjee, Andrew Oswald, Munshi Sulaiman, Victor Pouliquen, Chris Woodruff, and Bruce Wydick for constructive comments on a previous draft of this paper. The authors also thank Mohammad Raied Arman and Farhana Kabir for excellent research assistance in coordinating the fieldwork and data collection of the data used in this paper. This project is funded by the WEE-DiFine Initiative at the BRAC Institute of Governance and Development. All remaining errors are our own.

†International Food Policy Research Institute. E-mail: `j.r.bloem@cgiar.org`.

‡University of Oxford, and BRAC Institute for Governance and Development, BRAC University. Email: `khandkerwahedur.rahman@oxmfordmartin.ox.ac.uk`.

1 Introduction

In the social sciences, we are often interested in quantitatively measuring attitudes within a population. In most cases, we try to capture these attitudes by asking survey respondents to indicate whether they agree or disagree with a given statement, often using a Likert scale (Likert, 1932). This statement always implicitly carries a positive or negative framing. The framing of these statements could lead to differential responses due to, for example, social norms or desirability, a reluctance to report indifference, or a tendency to acquiesce. This measurement error can carry meaningful consequences on empirical results from quantitative analysis, particularly when using measures of attitudes as a dependent variable (Bertrand and Mullainathan, 2001). Error in the measurement of attitudes is important because attitudes play both instrumental and intrinsic roles in explaining human behavior, and, therefore, quantitative measures of attitudes are frequently used as both explanatory and dependent variables in applied social science research.[1]

We use a survey experiment to test whether a statement's positive or negative framing influences how respondents report their attitudes. The influence of framing is important because choices about statement framing are unavoidable by researchers designing surveys that aim to measure attitudes. Therefore, bias stemming from how a statement is framed can be unavoidable, unless it is accounted for within the design of the questionnaire.

We embed the survey experiment within the baseline questionnaire of a larger study on transaction costs and demand for digital financial services in Bangladesh (Rahman and Bloem, 2020), randomly assigning respondents into one of two groups.[2] The control group received a questionnaire with three positively framed and three negatively framed statements. The treatment group also received three positively framed and three negatively framed statements, but the framing of each statement is the converse of the framing in the control group. These statements reflect the primary objective of the larger study and aim to measure attitudes toward the use of mobile banking among a sample of respondents from Bangladesh. As shown in Table 1, these statements indicate the trustworthiness and safety of mobile banking, along with attitudes indicating common constraints on the use of mobile banking.

Our survey experiment yields three key findings. First, we find that statement framing influences responses. Across each of the statements in our questionnaire, we find strong evidence that framing changes how respondents report their attitudes. Our results document framing effects

[1] For example, instrumentally, attitudes toward stigma relating to HIV/AIDS can influence testing rates and overall community health (Thornton, 2012; Yang, Allen IV, Mahumane, Riddell IV and Yu, 2023; Yu, 2023), the gendered attitudes held by teachers about the latent ability of students can influence student assessment outcomes (Rakshit and Sahoo, 2023), racial attitudes can influence academic outcomes (Corno, La Ferrara and Burns, 2022), and general attitudes can lead to inefficient gaps in the diffusion of technology (BenYishay, Jones, Kondylis and Mobarak, 2020). Moreover, intrinsically, attitudes reflecting subjective evaluations of well-being relate to changes in income in important ways (Easterlin, 2003; Kahneman and Deaton, 2010; Killingsworth, 2021; Lindqvist, Östling and Cesarini, 2020), attitudes reflecting subjective life evaluation vary over the life-cycle (Stone, Schwartz, Broderick and Deaton, 2010), and exposure to social media can influence attitudes reflecting life satisfaction for adolescents (Orben, Dienlin and Przybylski, 2019)).

[2] Treatment assignment of survey experiment is independent of the treatment status of the larger study.

Table 1: Statement Framing

Treatment (N = 1,930)	Framing	**Control** (N = 2,001)	Framing
Mobile banking is not trustworthy	Negative	Mobile banking is trustworthy	Positive
Mobile banking is unsafe for saving money	Negative	Mobile banking is safe for saving money	Positive
Mobile banking is unsafe for transactions	Negative	Mobile banking is safe for transactions	Positive
Mobile banking is not too expensive	Positive	Mobile banking is too expensive	Negative
Mobile banking is easy to use	Positive	Mobile banking is hard to use	Negative
Mobile banking is for someone like me	Positive	Mobile banking is not for someone like me	Negative

Notes: This table reports each of the statements included in our survey to measure attitudes toward mobile banking in either the treatment or the control group.

ranging between 7 and 21 percentage points ($p < 0.01$) in responses indicating "strongly agree" or "agree" for positively framed statements and "strongly disagree" or "disagree" for negatively framed statements.

Second, we find that statement framing not only influences responses to a given statement but also influences the overall rate of "favorable" responses. Specifically, the framing of the first statement a respondent receives can affect responses to subsequent statements. Respondents who received a negatively framed statement first, due to the random assignment to our treatment, change the probability of responding favorably (i.e., responding "strongly agree" or "agree" to positively framed statements and "strongly disagree" or "disagree" to negatively framed statements) to subsequent statements. This finding further complicates the quantitative measurement of attitudes given that standard practices cannot avoid providing either a positively or negatively framed statement to a respondent first and this choice can meaningfully influence the overall rate of favorable responses.

Third, we demonstrate the possible consequences of framing bias by showing that the bias driven by statement framing can change the magnitude, statistical significance, and even the sign of estimated correlations between an aggregate index of attitudes and relevant baseline covariates. Using standard techniques for generating aggregate indices (i.e., the approach used by Kling, Liebman and Katz (2007) and principal component analysis), we generate aggregate indices of attitudes toward mobile banking. We then examine if and how the association between scores on these aggregated indices and relevant covariates (i.e., whether the respondent is the household head, has completed class 9, has a mobile money account, or has a bank account) differ between the treatment and control groups. We find instances where the framing of the underlying statements within the aggregated index meaningfully influences the magnitude and statistical significance of the correlation. Strikingly, we also observe cases where the sign of the correlation changes based on the framing of the underlying statements.

Our survey experiment demonstrates a useful proof-of-concept for the design of future surveys. Given the observation of meaningful differential responses driven by statement framing, our experimental design can serve as a useful way to address concerns of bias in the measurement of attitudes in at least two ways. First, randomly varying statement framing generates data where the researcher's choices about statement framing are not a systematic feature and the researcher can directly control for statement framing in formal regression analysis. Second, this approach also allows researchers to estimate and report bounds on attitudes within a sample of survey respondents, facilitating an approach to measure attitudes akin to partial identification (Manski, 2003; Molinari, 2020; Tamer, 2010).

This paper is closely related to and builds upon analysis by Dunsch, Evans, Macis and Wang (2018), who find evidence of acquiescence bias in satisfaction surveys administered to patients at healthcare centers in Nigeria. They implemented a survey experiment that randomly assigned respondents into one of three treatment groups: (i) receiving a questionnaire with all positively framed statements, (ii) receiving a questionnaire with all negatively framed statements, and (ii) receiving a questionnaire with a mix of positively and negatively framed statements. The authors find evidence of acquiescence bias. Respondents in the positively framed group were 11 to 19 percentage points more likely to agree with the statement than respondents in the negatively framed group to disagree with the statement, and respondents in the mixed group reported rates of agreement between the positively and negatively framed groups. We build on these results by highlighting how the framing of one statement can influence responses to subsequent statements and by documenting the possible consequences of bias driven by framing effects not only when estimating key population parameters, but also when using measures of attitudes to construct an aggregate index.[3] The latter is important because aggregating responses from multi-item survey instruments represents an analytical approach that dates back to Likert (1932), and remains a popular approach within the social sciences (Krosnick, Judd and Wittenbrink, 2018).

More generally, we contribute to a growing literature reporting results from experiments embedded within larger data collection or survey efforts to test whether the reported measures in surveys vary by different salient features of a survey. This includes randomizing the order of survey modules within a household survey (Abay, Berhane, Hoddinott and Tafere, 2022; Jeong, Aggarwal, Robinson, Kumar, Spearot and Park, 2023; Laajaj and Macours, 2021; Rahman, Bloem and Bellemare, 2023), randomizing the order in which household members appear within a module of a survey (Ambler, Herskowitz and Maredia, 2021), randomly assigning households an in-person or phone survey (Abate, De Brauw, Hirvonen and Wolle, 2023; Anderson, Lybbert, Shenoy, Singh and Stein, 2023), and surveying the same respondent in-person and over the phone but randomizing the order (Rahman and Hossain, 2023).

We also contribute to recent efforts to understand and address some of the statistical and practical challenges of measuring important concepts that cannot be directly observed and must be quantitatively measured on an ordinal scale. This includes an ongoing methodological debate about the scientific value of data measuring feelings, attitudes, or subjective perceptions (Bond and Lang, 2019; Ferrer-i Carbonell and Frijters, 2004; Kaiser and Oswald, 2022b; Kaiser and Vendrik, 2020;

[3] While Dunsch et al. (2018) use the term "acquiescence bias," we prefer to use the more general term "framing bias" because differential responses due to statement framing might be driven by several mechanisms, including a tendency to acquiesce, social norms, or social desirability.

Ravallion, Himelein and Beegle, 2016; Voutilainen, Pitkäaho, Kvist and Vehviläinen-Julkunen, 2016), empirical investigation into the shape of the "reporting function" between subjective perception and objective reality (Bloem, 2022; Kaiser and Oswald, 2022a; Oswald, 2008; Schröder and Yitzhaki, 2017), robustness tests that aim to address inherent challenges in the use of data measured on an ordinal scale (Bloem and Oswald, 2022; Chen, Oparina, Powdthavee and Srisuma, 2022), and calls to question some of the standard practices and assumptions of data analysis using variables measured on an ordinal scale (Killingsworth, Kahneman and Mellers, 2023).

The remainder of this paper is organized as follows. In the next section, we introduce and discuss the design of our survey experiment and describe the data. In Section 3, we discuss our estimation strategy. In Section 4, we present our results and discuss how implementing our experimental design in future surveys can help mitigate framing bias. In Section 5, we conclude.

2 Experimental Design

We focus on the use of the Likert scale in this experiment. Likert (1932) proposed this technique for the measurement of attitudes which requires a researcher to prepare a list of statements expressing positions that a respondent can either agree or disagree with, with various levels of intensity. In the standard application of this technique, the respondent chooses from several response categories (e.g., "strongly disagree," "disagree," "undecided," "agree," or "strongly agree"). For statements framed with a positive view of some object, the researcher codes the response categories in order 1, 2, 3, 4, and 5, respectively. On the contrary, for statements framed with a negative view of some object, the researcher codes the response categories in reverse order 5, 4, 3, 2, and 1, respectively. Mixing positively framed statements with negatively framed statements within a questionnaire is a common practice used to mitigate acquiescence bias (i.e., the tendency for respondents to agree with any statement) and allows the researcher to generate an aggregate index incorporating multi-dimensional constructs in the quantitative measurement of attitudes. This approach, however, assumes that the positive or negative framing of the statement does not affect how respondents report their attitudes—an assumption we dub the "symmetry assumption." This technique remains a common method to measure attitudes by psychologists, political scientists, sociologists, economists, market researchers, and national governments around the world (Krosnick et al., 2018).

Our experiment is embedded within the baseline questionnaire of a larger study that aims to estimate the effect of reducing transaction costs on the demand for digital financial services in Bangladesh (Rahman and Bloem, 2020), randomly assigning roughly half of the respondents to one of two groups.[4] Study participants are borrowers of a local microfinance organization, the Shakti Foundation for Disadvantaged Women. The study covers 300 Shakti Foundation centers in

[4] The baseline survey was conducted in November-December, 2022.

two districts of Bangladesh (i.e., Barisal and Cumilla).

All respondents in our study received, among other questions, six questions that aim to measure their attitudes towards mobile money. The questions measure the following six dimensions: trust, safe savings, safe transaction, whether it is expensive to use, whether it is hard to use, and whether they feel it is for persons like themselves.

We provided our control group with three positively framed statements to measure how much they trust mobile money, how safe they perceive mobile money to save money, and how safe they feel about transacting money using mobile money. We also ask them three negatively framed statements regarding their attitudes towards the cost of using mobile money, their perceived difficulty of using mobile money, and whether they feel mobile money is for someone like themselves.

Our treatment group also received three positively framed statements and three negatively framed statements, but the framing is in the converse of the control group, i.e., the treatment group received negative statements on trust and safety while receiving positive statements on the cost and difficulty of use and whether they feel mobile money is for someone like themselves. Table 1 reports the specific language and framing used for the statements across the treatment and control groups.

As the objective of this experimental design is to test for framing effects directly, each group received a mix of both positively and negatively framed statements, as recommended practice (Dunsch et al., 2018). Differential responses between the treatment and control groups on the provided Likert scale represents bias due to framing effects, invalidating the symmetry assumption underlying the standard technique for generating aggregate indices, and presenting a complication for researchers who aim to quantitatively measure attitudes.

Our experimental design also provides a proof-of-concept for a solution to framing effects in measuring attitudes. Future studies can address bias due to framing effects by effectively integrating our experimental design into their surveys. Electronic survey tools make randomly assigning respondents to two groups with differently framed statements a relatively easy task. The benefit of this approach is that the choice of how to frame specific statements by the researcher becomes an independent feature of the data. Additionally, this approach also allows researchers to generate bounds on measured attitudes within a sample of survey respondents, an approach we demonstrate in Section 4.3.

Table A.1 in the Supplemental Appendix reports summary statistics about our sample and shows the balance of these variables between the treatment and the control groups. Over 99 percent of the respondents are female and almost 93 percent are married. About 66 percent of the respondents have completed class 5, while only 26 percent have completed class 9. Roughly 34 percent of the respondents are the head of their households. About 51 percent of the respondents have no job, while about 42 percent work for pay, and 10 percent owns a business. About 47 percent of the respondents have a mobile money account.

3 Estimation Strategy

We estimate framing effects using the following linear regression.

$$Y_i = \alpha + \beta Treatment_i + \epsilon_i \tag{1}$$

In equation (1), Y_i represents a binary variable indicating if the respondent chooses "completely agree" or "agree" to positively framed statements or "completely disagree" or "disagree" to negatively framed statements. The variable $Treatment_i$ represents the randomized treatment assignment for each respondent (as described in Table 1). The coefficient β represents the estimated effect of statement framing. Finally, ϵ_i is an error term orthogonal to treatment assignment, given our individual-level randomization within the household survey. We use heteroskedasticity-robust standard errors (Abadie, Athey, Imbens and Wooldridge, 2023).

Our null hypothesis is that $\hat{\beta} = 0$, indicating the independence of response to statement framing. Rejecting this null hypothesis indicates the presence of bias due to framing effects.

4 Results

We present three sets of results. First, we document evidence of framing effects within the measurement of attitudes in a household survey and show how the framing of the first statement a respondent receives influences responses to subsequent statements. Second, we show the possible consequences of framing bias by using standard techniques to generate aggregate indices and using these indices to estimate correlations with relevant covariates. Finally, we demonstrate how randomizing statement framing when designing surveys can help address framing bias in the measurement of attitudes.

4.1 Framing Effects

Our core finding is that responses are not independent of statement framing. Figure 1 plots histograms illustrating the percent of respondents indicating each response category. If responses are independent to statement framing, then we would observe similar shares of respondents indicating, for example, "completely agree" in the control group as "completely disagree" in the treatment group. Across each of the panels within Figure 1, we can see that responses vary as statement framing varies. For example, panel A shows responses to the statements assessing if the respondent considers mobile banking to be trustworthy. We find that over 60 percent of the control group completely agree with the statement "mobile banking is trustworthy," while less than 50 percent of the treatment group completely disagrees with the statement "mobile banking is not trustworthy." Similar findings hold, to varying degrees, in the remaining panels of Figure 1.

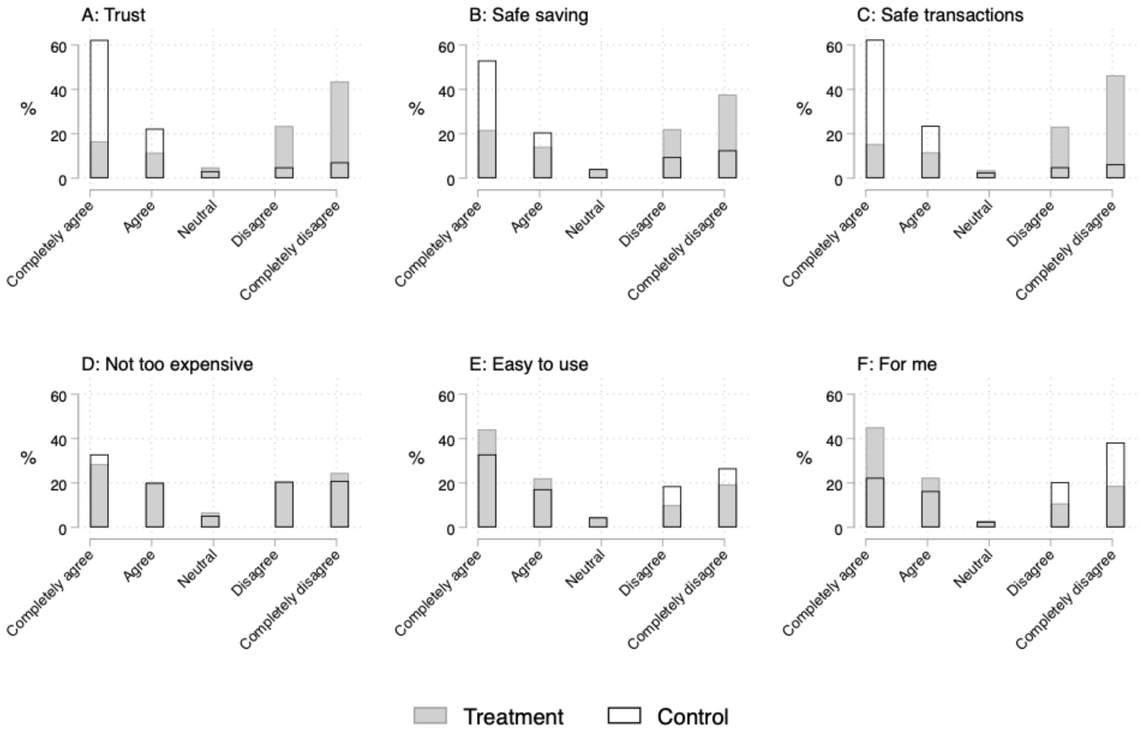

Figure 1: Histograms of Responses by Treatment Group
Notes: This figure shows histograms reporting the percent of respondents reporting each of the response categories by treatment group.

Testing for Framing Effects—We directly test for framing effects, and, across each of the six statements, find strong evidence of the existence of framing effects. We implement the estimation approach specified in equation (1) and present results in Table 2. These effects range in magnitude from 21 percentage points ($p < 0.01$) in column (5) to 7 percentage points ($p < 0.01$) in column (4). Additionally, framing effects persist among both positively and negatively framed questions. In columns (1) through (3), we find that the treatment led respondents to be 14–18 percentage points ($p < 0.01$) less likely to indicate that mobile banking is trustworthy, safe for savings, or safe for transactions. Conversely, in columns (4) through (6), we find that treatment led respondents to be 7–21 percentage points ($p < 0.01$) more likely to indicate that mobile banking is not too expensive, easy to use, or for a person like themselves. In all columns, we reject the null hypothesis of the independence of responses to statement framing.

It is important to discuss that the objective of our survey questions, which ask about attitudes toward mobile money, can be an emotive topic. Therefore, the framing effects identified in Table 2 might reflect this underlying emotion of respondents, and this might influence the external validity of our results. It is useful to emphasize, however, that we find evidence of framing effects in all columns, including column (6) of Table 2, which relates less to a respondent's notion of safety concerns relating to their money, savings, or transactions and more directly relates to a respondent's general attitude toward their own use of mobile banking.

Table 2: Framing Effects on Reported Attitudes

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment group receives:	Negatively framed statements			Positively framed statements		
	Trust	Safe saving	Safe transactions	Not too expensive	Easy to use	For me
Treatment	-0.177***	-0.139***	-0.165***	0.067***	0.210***	0.088***
	(0.013)	(0.015)	(0.013)	(0.016)	(0.016)	(0.015)
Observations	3,931	3,931	3,931	3,931	3,931	3,931
R-squared	0.043	0.022	0.039	0.005	0.045	0.008

Notes: Robust standard errors in parenthesis *** p<0.01, ** p<0.05, * p<0.1

Effects on Overall "Favorable" Response Rate—We now turn to estimate the effect of our treatment on overall on the number of "favorable" responses, rather than narrowly on the response to a single statement. Given that the number of both positively and negatively framed statements is balanced between our treatment and control groups, this effectively tests whether the framing of the first statement a respondent receives can influence responses to subsequent statements. To do so, we estimate equation (1) but use the following binary outcome measures: (i) whether all responses are favorable, i.e., the respondent chooses "completely agree" or "agree" to *all* positively framed statements *and* "completely disagree" or "disagree" to *all* negatively framed statements, (ii) whether no responses are favorable, i.e., the respondent chooses "completely agree" or "agree" to *none* of the positively framed statements *and* "completely disagree" or "disagree" to *none* of the negatively framed statements, (iii) whether more than three responses are favorable, and (iv) whether less than three responses are favorable. Our treatment *effectively* randomly assigns whether the respondent received a positively or negatively frame statement first in the list of statements designed to measure attitudes toward the use of mobile banking. In particular, respondents in the treatment group receive a negatively framed statement first while respondents in the control group received a positively framed statement first.

Table 3 reports results of the effect of our treatment on the number of favorable responses. We find that treatment, i.e., receiving a negatively framed statement first, can influence how the respondent respond to subsequent statements. Column (1) of Table 3 shows that the treatment group is 2.7 percentage points more likely to respond favorably to all six statements. With a control group mean of nearly 14 percent, this effect translates to nearly a 20 percent increase in the probability a respondent answers favorably to all six statements. While there is no systematic difference between the treatment and control groups in their probability of responding favorably to none of the statements, reported in column (2), the results in column (3) show that the treatment group is 4.2 percentage points less likely to respond favorably to more than three statements. With a control group mean of 64 percent, this effect translates to a 6.5 percent decrease in the probability of responding favorably to more than three statements. Finally, column (4) shows that treatment increases the probability of responding favorably to less than three statements by 7.1 percentage points. With a control group mean of 15 percent, this effect translates to a 46 percent increase in

Table 3: Framing Effects on the Number of Favorable Responses

	(1) All (6) favorable responses	(2) No (0) favorable responses	(3) More than 3 favorable responses	(4) Less than 3 favorable responses
Treatment	0.027**	0.001	-0.042***	0.071***
	(0.011)	(0.006)	(0.016)	(0.013)
Control mean	0.137	0.031	0.644	0.153
Observations	3,931	3,931	3,931	3,931
R-squared	0.001	0.000	0.002	0.008

Notes: Favorable response means agreeing in positively framed statements or disagreeing in negatively framed statements. Robust standard errors in parenthesis *** p<0.01, ** p<0.05, * p<0.1

the probability of responding favorably to less than three statements.

Results from Table 3 demonstrate that statement framing not only influences responses to a given statement but also influences responses to the subsequent statements. In the context of our survey experiment, receiving a negatively framed statement first increases the probability of responding very favorably (i.e., all favorable responses) and increases the probability of responding to fewer statements favorably (i.e., less than three favorable responses). Figure A.1 in the Supplementary Appendix presents a visual illustration of these results.

The finding that statement framing can influences the overall response demonstrates that framing bias can spillover across statements. This further complicates the quantitative measurement of attitudes given that the standard practices cannot avoid providing either a positively or negatively framed statement to a respondent first. Our survey experiment highlights that this choice can meaningfully influence the overall response rate of respondents.

Robustness Tests—We conduct a series of robustness tests and report results from these tests in Table A.2 in the Supplemental Appendix. In panel A, we control for treatment status fixed effects from the larger study. Given that assignment to treatment in this survey experiment was independent of assignment to treatment in the larger study, it is no surprise to find that the results are nearly identical.

In panel B, we control for whether the respondent has a mobile banking account and interact this variable with the treatment assignment variable. Our findings are robust to account for whether the respondent has a mobile banking account. However, whether the respondent has a mobile banking account nearly entirely explains the treatment effect on whether mobile banking is too expensive. This finding demonstrates that framing effects can be correlated with individual characteristics and imply that framing bias represents a possible source of non-classical measurement error (Bertrand and Mullainathan, 2001; Bound, Brown and Mathiowetz, 2001).

Finally, in panels C through G in Table A.2, we control for and interact the following variables with the treatment assignment variable: whether the respondent is the household head, whether the respondent has completed class 9, whether the respondent works for pay, whether the respondent has no job, log of food expenditure, and whether the respondent owns a business. Across each of our outcomes within each of these panels, our results are robust to the inclusion of these covariates.

What is the cognitive mechanism that drives framing effects? One possibility is that survey respondents get confused when the statement is already framed negatively (i.e., mobile banking is trustworthy) and the converse statement uses the word "not" to negate a negative statement (i.e., mobile banking is not trustworthy), creating an implied double negative. Although our survey experiment is not designed to test possible cognitive mechanisms directly, our results show patterns that are inconsistent with this mechanism. Respondents in our treatment group received a questionnaire with three negatively framed questions, only one of which includes an implied double negative. If respondents were confused by the use of an implied double negative and mistakenly agreed when they intended to disagree, then we would observe our largest treatment effects in column (1) with the implied double negative. We do not observe this pattern in Table 2. Instead, we find evidence of substantial framing effects in all columns which do not include statements with an implied double negative. Nevertheless, future research could aim to directly test possible cognitive mechanisms that drive framing effects.

4.2 Possible Consequences of Framing Effects

We now demonstrate the possible consequences of bias due to framing effects. To do so we generate aggregate indices indicating attitudes toward mobile banking using two standard techniques commonly used by applied quantitative researchers: (i) the approach implemented by Kling et al. (2007) and (ii) principal component analysis (PCA). To generate the Kling index, we take each of the six binary variables used as dependent variables in Table 2 and standardize each variable so that the distribution of responses has a mean of zero and a standard deviation of one. Next, we sum the standardized values of the six components. Finally, we standardize this aggregated index so that the distribution of the index has a mean of zero and a standard deviation of one. To generate the principal component analysis index we again take each of the six binary variables measuring attitudes toward mobile banking and take the first principal component to generate an aggregated index. The correlation between these two aggregated indices is $r = 0.95$ indicating that these indices are highly correlated with each other. We then estimate the following linear regression:

$$Y_i = \gamma + \delta Treatment_i + \lambda Covariate_i + \theta(Treatment_i \times Covariate_i) + \eta_i \quad (2)$$

Our goal is to compare the conditional mean of the aggregated index associated with a given covariate of interest between the treatment and control groups. With a binary covariate the conditional mean of the index when $Treatment_i = 1$ and $Covariate_i = 1$ is the sum of $\gamma+\delta+\lambda+\theta$ and the conditional mean of the index when $Treatment_i = 0$ and $Covariate_i = 1$ is the sum of $\gamma+\lambda$.

Meaningful differences in the magnitude, statistical significance, or sign of these conditional means by treatment status illustrates the possible consequences of framing effects when constructing aggregate indices.

Table 4 presents results from estimating variations on equation (2), and demonstrates the possible consequences of framing effects. Column (1) uses the Kling index and column (2) uses the principal component analysis index. In panel A of Table 4, we include a binary variable indicating if the respondent is the head of their household and interact this variable with our treatment variable. We find meaningful differences in the conditional mean of each aggregated index associated with being a household head between the treatment and control groups. In column (1), when using the Kling index, we find that the conditional means range between -0.11 standard deviations in the treatment group and -0.06 standard deviations in the control group. Although the sign of this observed difference is robust, the magnitude and statistical significance both meaningfully differ by treatment status. Additionally, a formal test of difference in these conditional means by treatment status shows that this observed difference is statistically significant. In column (2), when using the PCA index, we find that the conditional means range between -0.33 standard deviations in the treatment group and 0.05 standard deviations in the control group. Here, we find that not only do the magnitude and statistical significance of these conditional means differ by treatment status, but sign differs as well. Again, a formal test of difference in these conditional means by treatment status shows that this observed difference is statistically significant. These results demonstrate that the framing of the underlying statements within an aggregate index measuring attitudes can not only change the magnitude and statistical significance of the correlation between the aggregate index and a relevant covariate, but also the sign of the estimated correlation. This is a striking finding and demonstrates the possible severity of bias due to framing effects on empirical results from applied research. Namely, the framing of the underlying statements used to quantitatively measure attitudes can meaningfully change the research conclusions used to test and validate theories or inform policy choices.

In panel B of Table 4, we include a binary variable indicating if the respondent has completed class 9. In column (1), when using the Kling index, we find that the conditional means range between 0.10 standard deviations in the treatment group and 0.25 standard deviations in the control group. Here the sign and statistical significance of this observed difference is robust, however, the magnitude differs by more than a factor of two and a formal test of difference in these conditional means by treatment status shows that this observed difference is statistically significant. In column (2), when using the PCA index, we find that the conditional means range between -0.07 standard deviations in the treatment group and 0.36 standard deviations in the control group. Although the magnitude, statistical significance, and sign of these conditional means differ by treatment status, a formal test of difference in these conditional means by treatment status shows that this observed difference is not statistically significant. These results demonstrate that the consequences

Table 4: Aggregated Index Analysis

	(1) Kling index	(2) PCA index
Panel A:		
Treatment	-0.134***	-0.455***
	(0.039)	(0.055)
Household head	-0.166***	-0.240***
	(0.045)	(0.061)
Treatment × Household head	0.079	0.072
	(0.068)	(0.098)
Constant	0.109***	0.292***
	(0.025)	(0.034)
Treatment = 1 & Household head = 1	-0.113***	-0.331***
Treatment = 0 & Household head = 1	-0.057	0.052
Difference (p-value)	0.018	0.031
Observations	3,931	3,931
R-squared	0.007	0.027
Panel B:		
Treatment	-0.087**	-0.425***
	(0.038)	(0.054)
Completed class 9	0.271***	0.203***
	(0.045)	(0.061)
Treatment × Completed class 9	-0.061	-0.004
	(0.070)	(0.100)
Constant	-0.022	0.155***
	(0.025)	(0.034)
Treatment = 1 & Completed class 9 = 1	0.101**	-0.071
Treatment = 0 & Completed class 9 = 1	0.249***	0.358***
Difference (p-value)	0.002	0.155
Observations	3,931	3,931
R-squared	0.014	0.027
Panel C:		
Treatment	-0.172***	-0.504***
	(0.045)	(0.065)
Has mobile money account	0.082**	0.075
	(0.042)	(0.056)
Treatment × Has mobile money account	0.142**	0.163*
	(0.063)	(0.090)
Constant	0.014	0.176***
	(0.030)	(0.042)
Treatment = 1 & Has mobile money account = 1	0.066*	-0.091*
Treatment = 0 & Has mobile money account = 1	0.096***	0.250***
Difference (p-value)	0.536	0.508
Observations	3,931	3,931
R-squared	0.010	0.026
Panel D:		
Treatment	-0.116**	-0.438***
	(0.052)	(0.075)
Has bank account	0.061	0.060
	(0.044)	(0.059)
Treatment × Has bank account	0.018*	0.017
	(0.066)	(0.094)
Constant	0.015	0.174***
	(0.035)	(0.048)
Treatment = 1 & Has bank account = 1	-0.022	-0.188***
Treatment = 0 & Has bank account = 1	0.076***	0.234***
Difference (p-value)	0.666	0.757
Observations	3,931	3,931
R-squared	0.004	0.023

Notes: Column (1) uses an aggregated index constructed using the technique of Kling et al. (2007). Column (2) uses an aggregated index constructed using principal component analysis. The "difference (p-value)" row in each panel tests the difference in the estimated conditional means in the preceding two rows. Robust standard errors in parenthesis *** p<0.01, ** p<0.05, * p<0.1

of framing effects on estimated conditional means of aggregated indices can vary by the technique used to construct the aggregated index.

Finally, in panels C and D of Table 4, we include binary variables indicating if the respondent has a mobile money account (in panel C) and if the respondent has a bank account (in panel D). In both panels, although we observe instances of conditional means differing by magnitude, statistical significance, and sign formal tests of difference in these conditional means by treatment status all show that these differences are not statistically significant.

We also generate two additional pairs of Kling and PCA indices: one for statements one (trust), two (safe saving), and three (safe transactions) grouped together, and the other for statements four (not too expensive), five (easy to use), and six (for me) grouped together. This dis-aggregation of these aggregated indices allows us to examine the possible consequences of statement framing by grouping the statements which had the same framing for each respondent, but different framing by treatment group.

Table A.3 in the Supplementary Appendix presents the results for these sub-indices. These results demonstrate the same broad set of findings observed in Table 4. Even among sub-sets of these aggregated indices, we observe cases where the magnitude and sign of the conditional mean differs by treatment status. Although, the each of the conditional means estimated in Table A.3 are all statistically significant themselves, we continue to observe cases where formal tests of differences in these conditional means by treatment status show that these observed differences are statistically significant as well.

Taken together, these findings demonstrate the possible consequences of framing effects. The results shown in Table 2 already demonstrate that framing effects can influence the estimation of key population parameters (i.e., the share of respondents who believe that mobile banking is trustworthy). The results shown in Table 4 show that bias due to framing effects can lead to meaningful differences in the magnitude, statistical significance, and sign of correlations between aggregated indices, generated using standard techniques, and relevant covariates. Therefore, framing effects can influence results that use variables measuring attitudes generated from surveys as the dependent variable in a linear regression context (such as in the work of Easterlin (2003), Kahneman and Deaton (2010), Stone et al. (2010), Orben et al. (2019)).

4.3 Randomizing Statement Framing as a Solution

Having shown evidence of framing effects and the potential consequences of these framing effects, we now turn to demonstrate how randomizing statement framing, as we did in our survey experiment, can help address the presence framing effects in the measurement of attitudes. We view our survey experiment as a proof-of-concept that can be implemented in future surveys that include questions designed to measure attitudes.

Framing effects persist when researchers make choices about how to frame specific statements used to measure attitudes in a survey. When all respondents to the survey receive statements framed in the same way, framing effects become a systematic feature of the data and cannot be accounted for in any analysis using the survey data. To limit this systematic bias, we propose that future surveys randomize the framing of statements as we do in this survey experiment. Doing so creates two sub-groups of respondents, one group that receives statements framed both positively and negatively and a second group that receives statements framed in the converse.

Implementing the experimental design in future surveys carries two specific benefits. First, by construction, this approach makes statement framing independent from responses within the full sample and allows researchers to identify statement framing and account for it directly when analyzing the data. For example, in regression analysis, researchers can control from the randomly assigned statement framing group. Second, when estimating population parameters researchers can do so separately for both groups and present estimates as bounds on the true value of the parameter. This bounding approach is akin to partial identification (Manski, 2003; Molinari, 2020; Tamer, 2010).

The results in Table 4 demonstrate the benefits of implementing our survey experiment in formal regression analysis. Controlling for treatment assignment and interacting this variable with a covariate of interest allows the researcher to obtain two estimates of the relationship between a variable of interest and an attitude measure. Taken together these two estimates can be used as bounds on the correlation of interest. As shown in panels A and B in Table 4, relatively wide bounds—especially those that include zero—indicate that a particular relationship is relatively sensitive to the framing of the statements used to measure attitudes. To the contrary, as shown in panels C and D, relatively narrow bounds indicate a particular relationship is robust to statement framing. Without the ability to directly account for statement framing estimates from formal regression analysis will likely fall somewhere within the bounds estimated in Table 4 and the researcher will not have the ability to assess sensitivity of the magnitude, statistical significance, or sign to statement framing. Importantly, without the ability to assess sensitively to statement framing research results and policy conclusions might be biased and the researcher will have no way of addressing this bias.

We further demonstrate this bounding approach in Figure 2, which reports the mean value of each of the binary variables used as the dependent variable in equation (1) for both the control and treatment groups. For example, we can credibly report that between 67 and 85 percent of our sample feel that mobile banking is trustworthy. Randomizing statement framing within a survey allows us to reliably report bounds on population parameters without concern of framing bias and can be easily incorporated in future surveys.

In both formal regression analysis and the estimation of important population parameters, the existence of framing effects calls to question the validity of point identification when using quantitative measures of attitudes. Credibly estimating a single point estimate with a quantitative measure of attitudes requires that the researcher make assumptions about the influence of the

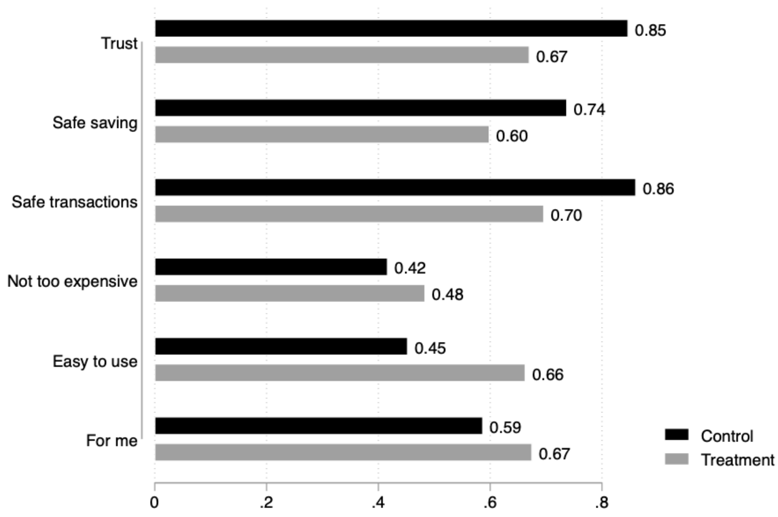

Figure 2: Bounded Estimates of Attitudes toward Mobile Banking
Notes: This figure shows bar graphs representing bounds on reported attitudes toward mobile banking.

framing of the underlying statements. In particular, the researcher will need to follow the seminal work of Likert (1932) and assume that symmetry in responses between positively and negatively framed statements holds. This assumption, however, is strongly rejected by the results from this survey experiment. Given this, applied researchers who use quantitative measures of attitudes will produce more credible results if they implement our experimental design and pursue a bounding approach when reporting empirical results. Such an approach is akin to partial identification (Manski, 2003; Molinari, 2020; Tamer, 2010).

5 Conclusion

The standard approach to measuring attitudes, and generating aggregate indices, requires that statement framing does not influence how respondents report their attitudes. We implement a simple survey experiment designed to test for the presence of framing effects in the measurement of attitudes. We randomly assign half of the respondents to one of two groups. The first group received a questionnaire that included, among other questions, three positively framed statements and three negatively framed statements regarding attitudes toward the use of mobile banking. The second group also received three positively framed statements and three negatively framed statements, but the framing is in the converse of the control group.

This survey experiment allows us to directly estimate the effect of statement framing. We find strong evidence that framing influences how respondents report their attitudes. We further demonstrate the possible consequences of framing effects by generating aggregate indices of attitudes toward mobile banking, using standard techniques. We find instances where the framing of the underlying statements within the aggregated index meaningfully influences the magnitude and statistical significance of estimated correlations between these indices and relevant covariates. Most strikingly, we also observe instances where the framing of the underlying statements within the aggregate indices changes the sign of the estimated correlation.

Finally, we discuss how implementing our experimental design in future surveys can help address the possible presence of framing effects. We demonstrate that framing effects exist in the measurement of attitudes, framing bias can be meaningful, and our experimental design combined with a bounding approach can help address the problem of framing effects biasing research conclusions and policy choices based on empirical analysis using quantitative measures of attitudes.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge (2023) "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics*, 138 (1), 1–35.

Abate, Gashaw T, Alan De Brauw, Kalle Hirvonen, and Abdulazize Wolle (2023) "Measuring consumption over the phone: Evidence from a survey experiment in urban Ethiopia," *Journal of Development Economics*, 161, 103026.

Abay, Kibrom A, Guush Berhane, John Hoddinott, and Kibrom Tafere (2022) "Respondent Fatigue Reduces Dietary Diversity Scores Reported from Mobile Phone Surveys in Ethiopia during the COVID-19 Pandemic," *The Journal of Nutrition*, 152 (10), 2269–2276.

Ambler, Kate, Sylvan Herskowitz, and Mywish K Maredia (2021) "Are we done yet? Response fatigue and rural livelihoods," *Journal of Development Economics*, 153, 102736.

Anderson, Ellen, Travis J. Lybbert, Amish Shenoy, Rupika Singh, and Daniel Stein (2023) "Does Survey Mode Matter? Comparing In-Person and Phone Agricultural Surveys in India," 10.2139/ssrn.4408594.

BenYishay, Ariel, Maria Jones, Florence Kondylis, and Ahmed Mushfiq Mobarak (2020) "Gender gaps in technology diffusion," *Journal of Development Economics*, 143, 102380.

Bertrand, Marianne and Sendhil Mullainathan (2001) "Do people mean what they say? Implications for subjective survey data," *American Economic Review*, 91 (2), 67–72.

Bloem, Jeffrey R (2022) "How much does the cardinal treatment of ordinal variables matter? An empirical investigation," *Political Analysis*, 30 (2), 197–213.

Bloem, Jeffrey R and Andrew J Oswald (2022) "The analysis of human feelings: a practical suggestion for a robustness test," *Review of Income and Wealth*, 68 (3), 689–710.

Bond, Timothy N and Kevin Lang (2019) "The sad truth about happiness scales," *Journal of Political Economy*, 127 (4), 1629–1640.

Bound, John, Charles Brown, and Nancy Mathiowetz (2001) "Chapter 59 - Measurement Error in Survey Data," in Heckman, James J. and Edward Leamer eds. *Handbook of Econometrics*, 5, 3705–3843: Elsevier, 10.1016/S1573-4412(01)05012-7.

Ferrer-i Carbonell, Ada and Paul Frijters (2004) "How important is methodology for the estimates of the determinants of happiness?" *The Economic Journal*, 114 (497), 641–659.

Chen, Le-Yu, Ekaterina Oparina, Nattavudh Powdthavee, and Sorawoot Srisuma (2022) "Robust ranking of happiness outcomes: A median regression perspective," *Journal of Economic Behavior & Organization*, 200, 672–686.

Corno, Lucia, Eliana La Ferrara, and Justine Burns (2022) "Interaction, stereotypes, and performance: Evidence from South Africa," *American Economic Review*, 112 (12), 3848–75.

Dunsch, Felipe, David K Evans, Mario Macis, and Qiao Wang (2018) "Bias in patient satisfaction surveys: a threat to measuring healthcare quality," *BMJ Global Health*, 3 (2), e000694.

Easterlin, Richard A (2003) "Explaining happiness," *Proceedings of the National Academy of Sciences*, 100 (19), 11176–11183.

Jeong, Dahyeon, Shilpa Aggarwal, Jonathan Robinson, Naresh Kumar, Alan Spearot, and David Sungho Park (2023) "Exhaustive or exhausting? Evidence on respondent fatigue in long surveys," *Journal of Development Economics*, 161, 102992.

Kahneman, Daniel and Angus Deaton (2010) "High income improves evaluation of life but not emotional well-being," *Proceedings of the National Academy of Sciences*, 107 (38), 16489–16493.

Kaiser, Caspar and Andrew J Oswald (2022a) "Inequality, well-being, and the problem of the unknown reporting function," *Proceedings of the National Academy of Sciences*, 119 (50).

——— (2022b) "The scientific value of numerical measures of human feelings," *Proceedings of the National Academy of Sciences*, 119 (42).

Kaiser, Caspar and Maarten C M Vendrik (2020) "How Threatening Are Transformations of Happiness Scales to Subjective Wellbeing Research?" IZA Discussion Paper 13905, https://dx.doi.org/10.2139/ssrn.3743129.

Killingsworth, Matthew A (2021) "Experienced well-being rises with income, even above $75,000 per year," *Proceedings of the National Academy of Sciences*, 118 (4), e2016976118.

Killingsworth, Matthew A, Daniel Kahneman, and Barbara Mellers (2023) "Income and emotional well-being: A conflict resolved," *Proceedings of the National Academy of Sciences*, 120 (10), e2208661120.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz (2007) "Experimental analysis of neighborhood effects," *Econometrica*, 75 (1), 83–119.

Krosnick, Jon A, Charles M Judd, and Bernd Wittenbrink (2018) "The Measurement of Attitudes," in Albarracin, Dolores and Blair T. Johnson eds. *The Handbook of Attitudes*, 45–105: Routledge.

Laajaj, Rachid and Karen Macours (2021) "Measuring skills in developing countries," *Journal of Human Resources*, 56 (4), 1254–1295.

Likert, Rensis (1932) "A technique for the measurement of attitudes.," *Archives of Psychology*.

Lindqvist, Erik, Robert Östling, and David Cesarini (2020) "Long-Run Effects of Lottery Wealth on Psychological Well-Being," *The Review of Economic Studies*, 87 (6), 2703–2726, 10.1093/restud/rdaa006.

Manski, Charles F (2003) *Partial Identification of Probability Distributions*, 5: Springer.

Molinari, Francesca (2020) "Microeconometrics with partial identification," *Handbook of econometrics*, 7, 355–486.

Orben, Amy, Tobias Dienlin, and Andrew K Przybylski (2019) "Social media's enduring effect on adolescent life satisfaction," *Proceedings of the National Academy of Sciences*, 116 (21), 10226–10228.

Oswald, Andrew J (2008) "On the curvature of the reporting function from objective reality to subjective feelings," *Economics Letters*, 100 (3), 369–372.

Rahman, Khander Wahedur and Jeffrey R Bloem (2020) "Digital Finance and Economic Empowerment: Experimental Evidence on the Role of Transaction Costs," *BRAC BIGD Project*.

Rahman, Khandker Wahedur, Jeffrey R Bloem, and Marc F Bellemare (2023) "Survey Ordering and the Measurement of Welfare," *Working Paper*, http://marcfbellemare.com/wordpress/wp-content/uploads/2023/04/RahmanBloemBellemareSurveyOrdering06April2023.pdf.

Rahman, Khandker Wahedur and Marjan Hossain (2023) "Does The Measurement of Learning Level Depend on The Mode of Testing?: Evidence from Bangladesh," *BRAC BIGD Project*.

Rakshit, Sonali and Soham Sahoo (2023) "Biased teachers and gender gap in learning outcomes: Evidence from India," *Journal of Development Economics*, 103041.

Ravallion, Martin, Kristen Himelein, and Kathleen Beegle (2016) "Can subjective questions on economic welfare be trusted?" *Economic Development and Cultural Change*, 64 (4), 697–726.

Schröder, Carsten and Shlomo Yitzhaki (2017) "Revisiting the evidence for cardinal treatment of ordinal variables," *European Economic Review*, 92, 337–358.

Stone, Arthur A, Joseph E Schwartz, Joan E Broderick, and Angus Deaton (2010) "A snapshot of the age distribution of psychological well-being in the United States," *Proceedings of the National Academy of sciences*, 107 (22), 9985–9990.

Tamer, Elie (2010) "Partial Identification in Econometrics," *Annual Review of Economics*, 2 (1), 167–195.

Thornton, Rebecca L (2012) "HIV testing, subjective beliefs and economic behavior," *Journal of Development Economics*, 99 (2), 300–313.

Voutilainen, Ari, Taina Pitkäaho, Tarja Kvist, and Katri Vehviläinen-Julkunen (2016) "How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale," *Journal of Advanced Nursing*, 72 (4), 946–957, 10.1111/jan.12875.

Yang, Dean, James Allen IV, Arlete Mahumane, James Riddell IV, and Hang Yu (2023) "Knowledge, stigma, and HIV testing: An analysis of a widespread HIV/AIDS program," *Journal of Development Economics*, 160, 102958.

Yu, Hang (2023) "Social stigma as a barrier to HIV testing: evidence from a randomized experiment in Mozambique," *Journal of Development Economics*, 161, 103035.

Supplemental Appendix

The Supplemental Appendix includes the following additional results.

- Table A.1 reports basic summary statistics about our sample and shows balance between the treatment and control groups.
- Table A.2 reports a set of robustness tests supporting the main results reported in Table 2 in the main manuscript.
- Figure A.1 illustrates the influence of our experimental treatment on the number of favorable responses. This figure complements the results reported in Table 3 in the main manuscript.
- Table A.3 reports a set of results that dis-aggregate the aggregated indices by the first and last three statements included in our questionnaire. These results complement the results reported in Table 4 in the main manuscript.

Table A.1: Balance Table and Summary Statistics

	(1) Control		(2) Treatment		(1)-(2) Pairwise t-test	
	N/Clusters	Mean/(SE)	N/Clusters	Mean/(SE)	N/Clusters	Mean difference
Female (= 1)	2001	0.992 (0.002)	1930	0.993 (0.002)	3931	-0.001
Married (= 1)	2001	0.926 (0.006)	1930	0.925 (0.006)	3931	0.001
Household size	2001	4.763 (0.042)	1930	4.881 (0.046)	3931	-0.118*
Household head (= 1)	2001	0.337 (0.011)	1930	0.336 (0.011)	3931	0.002
Has mobile money account (= 1)	2001	0.475 (0.011)	1930	0.461 (0.011)	3931	0.015
Has bank account	2001	0.626 (0.011)	1930	0.589 (0.011)	3931	0.037**
Completed class 9 (= 1)	2001	0.276 (0.010)	1930	0.256 (0.010)	3931	0.020
Worked for pay (= 1)	2001	0.419 (0.011)	1930	0.413 (0.011)	3931	0.006
No job (= 1)	2001	0.510 (0.011)	1930	0.518 (0.011)	3931	-0.008
Has savings (= 1)	2001	0.951 (0.005)	1930	0.951 (0.005)	3931	-0.000
Receives remittances (= 1)	2001	0.372 (0.011)	1930	0.395 (0.011)	3931	-0.023
Has loans (= 1)	2001	0.863 (0.008)	1930	0.852 (0.008)	3931	0.011
Food expenditures	2001	11885.982 (134.928)	1930	12109.663 (134.908)	3931	-223.681
Education expenditure	2001	3224.529 (91.270)	1930	3169.913 (85.456)	3931	54.616
Health care expenditure	2001	2638.149 (70.183)	1930	2710.948 (63.997)	3931	-72.799
Household utilities expenditure	2001	2688.893 (55.295)	1930	2672.547 (52.659)	3931	16.346
Own a business	2001	0.102 (0.007)	1930	0.107 (0.007)	3931	-0.005

Notes: This table reports basic summary statistics and shows the balance in these statistics between the negative framing and positive framing groups. The expenditure figures report monthly expenditures at the household level. T-test uses robust standard errors. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Robustness Tests

	(1) Trust	(2) Safe saving	(3) Safe transactions	(4) Not too expensive	(5) Easy to use	(6) For me
Panel A:						
Treatment	-0.176***	-0.138***	-0.165***	0.0669***	0.211***	0.0888***
	(0.0134)	(0.0149)	(0.0130)	(0.0159)	(0.0155)	(0.0153)
Larger study treatment FEs	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.044	0.023	0.040	0.005	0.046	0.011
Panel B:						
Treatment	-0.184***	-0.160***	-0.179***	0.0144	0.196***	0.0960***
	(0.0191)	(0.0198)	(0.0183)	(0.0218)	(0.0215)	(0.0214)
Has a mobile banking account	0.0559***	-0.0842***	0.0212	-0.0639***	0.110***	0.0883***
	(0.0160)	(0.0197)	(0.0155)	(0.0220)	(0.0222)	(0.0219)
Treatment × Has a mobile banking account	0.0179	0.0445	0.0317	0.112***	0.0357	-0.0147
	(0.0266)	(0.0299)	(0.0260)	(0.0317)	(0.0306)	(0.0305)
R-squared	0.049	0.027	0.042	0.008	0.062	0.015
Panel C:						
Treatment	-0.166***	-0.163***	-0.160***	0.0559***	0.199***	0.0745***
	(0.0161)	(0.0179)	(0.0156)	(0.0195)	(0.0191)	(0.0188)
Household head	-0.0345**	-0.0952***	-0.0377**	-0.0246	-0.0230	-0.0448*
	(0.0175)	(0.0214)	(0.0169)	(0.0232)	(0.0235)	(0.0234)
Treatment × Household head	-0.0308	0.0724**	-0.0132	0.0333	0.0336	0.0398
	(0.0289)	(0.0320)	(0.0282)	(0.0335)	(0.0327)	(0.0325)
R-squared	0.046	0.027	0.042	0.005	0.045	0.009
Panel D:						
Treatment	-0.183***	-0.147***	-0.171***	0.0441**	0.238***	0.141***
	(0.0158)	(0.0172)	(0.0154)	(0.0185)	(0.0181)	(0.0181)
Completed class 9	0.0217	-0.0598***	0.0174	0.0130	0.231***	0.223***
	(0.0175)	(0.0227)	(0.0169)	(0.0247)	(0.0243)	(0.0227)
Treatment × Completed class 9	0.0246	0.0281	0.0268	0.0909**	-0.0906***	-0.188***
	(0.0298)	(0.0343)	(0.0289)	(0.0358)	(0.0334)	(0.0331)
R-squared	0.043	0.023	0.040	0.005	0.048	0.016
Panel E:						
Treatment	-0.187***	-0.157***	-0.162***	0.0499**	0.203***	0.0666***
	(0.0177)	(0.0194)	(0.0172)	(0.0207)	(0.0204)	(0.0203)
Worked for pay	0.0135	-0.0199	0.0182	-0.00175	0.0267	0.0151
	(0.0162)	(0.0200)	(0.0156)	(0.0223)	(0.0226)	(0.0223)
Treatment × Worked for pay	0.0251	0.0435	-0.00661	0.0417	0.0178	0.0519*
	(0.0270)	(0.0302)	(0.0263)	(0.0321)	(0.0313)	(0.0309)
R-squared	0.044	0.022	0.040	0.005	0.046	0.011
Panel F:						
Treatment	-0.164***	-0.136***	-0.173***	0.0947***	0.209***	0.113***
	(0.0187)	(0.0214)	(0.0182)	(0.0227)	(0.0221)	(0.0216)
No job	-0.0270*	0.00825	-0.0356**	0.00676	-0.0406*	-0.0289
	(0.0161)	(0.0197)	(0.0154)	(0.0221)	(0.0223)	(0.0220)
Treatment × No job	-0.0248	-0.00503	0.0172	-0.0533*	0.00391	-0.0471
	(0.0267)	(0.0298)	(0.0260)	(0.0317)	(0.0310)	(0.0306)
R-squared	0.045	0.022	0.041	0.006	0.046	0.012
Panel G:						
Treatment	-0.176***	-0.137***	-0.162***	0.0677***	0.212***	0.0839***
	(0.0143)	(0.0158)	(0.0139)	(0.0167)	(0.0164)	(0.0162)
Own a business	0.0835***	0.0307	0.0625***	0.0168	0.0259	-0.0256
	(0.0207)	(0.0315)	(0.0211)	(0.0366)	(0.0369)	(0.0366)
Treatment × Own a business	-0.0114	-0.0138	-0.0248	-0.00634	-0.0160	0.0383
	(0.0388)	(0.0477)	(0.0390)	(0.0519)	(0.0506)	(0.0501)
R-squared	0.046	0.022	0.041	0.005	0.045	0.008

Notes: N=3,931 in all regressions. Robust standard errors in parenthesis *** $p<0.01$, ** $p<0.05$, * $p<0.1$

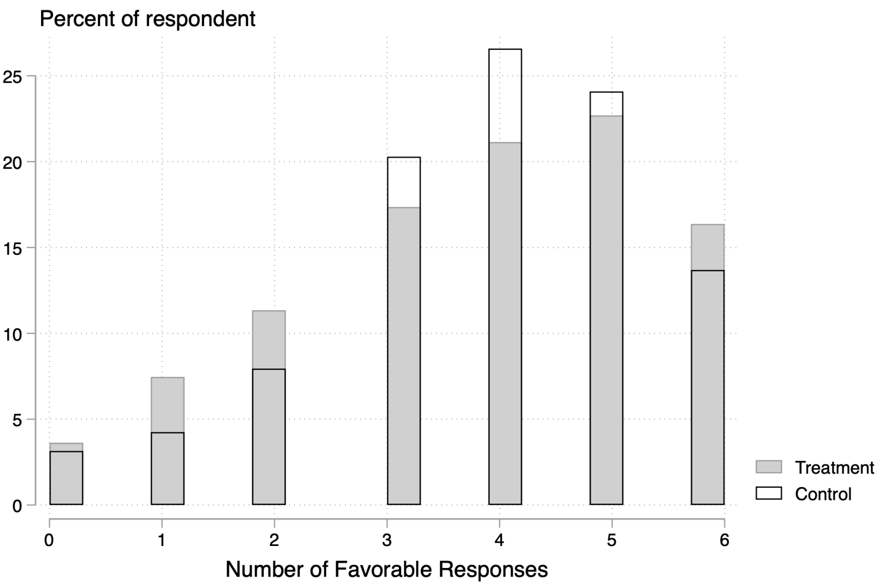

Figure A.1: Distribution of Total Number of Favorable Responses by Treatment Group
Notes: This figure shows the distribution of total number of favorable responses by treatment group. Favorable response means agreeing in positively framed statements or disagreeing or in negatively framed statements.

Table A.3: Indices Based on Sub-sample of Statements

	(1)	(2)	(3)	(4)
	Kling index		PCA index	
	Statements 1-3	Statements 4-6	Statements 1-3	Statements 4-6
Panel A				
Treatment	-0.465***	0.327***	-0.648***	0.408***
	(0.037)	(0.039)	(0.052)	(0.046)
Household head	-0.155***	-0.092*	-0.216***	-0.108*
	(0.042)	(0.048)	(0.058)	(0.057)
Treatment × Household head	0.021	0.106	0.029	0.124
	(0.067)	(0.066)	(0.094)	(0.078)
Constant	0.277***	-0.147***	0.386***	-0.184***
	(0.023)	(0.028)	(0.032)	(0.033)
Treatment = 1 & Household head = 1	-0.323***	0.194***	-0.449***	0.240***
Treatment = 0 & Household head = 1	0.122***	-0.239***	0.170***	-0.293***
Difference (p-value)	0.075	0.060	0.075	0.064
Observations	3,931	3,931	3,931	3,931
R-squared	0.057	0.034	0.057	0.037
Panel B				
Treatment	-0.478***	0.421***	-0.665***	0.530***
	(0.037)	(0.037)	(0.051)	(0.043)
Completed class 9	-0.014	0.466***	-0.020	0.593***
	(0.041)	(0.048)	(0.058)	(0.057)
Treatment × Completed class 9	0.076	-0.191***	0.105	-0.267***
	(0.069)	(0.068)	(0.096)	(0.080)
Constant	0.229***	-0.307***	0.319***	-0.384***
	(0.023)	(0.026)	(0.032)	(0.031)
Treatment = 1 & Completed class 9 = 1	-0.188***	0.390***	-0.261***	0.471***
Treatment = 0 & Completed class 9 = 1	0.215***	0.160***	0.299***	0.208***
Difference (p-value)	0.368	0.000	0.369	0.000
Observations	3,931	3,931	3,931	3,931
R-squared	0.053	0.062	0.053	0.069
Panel C				
Treatment	-0.499***	0.304***	-0.695***	0.391***
	(0.044)	(0.044)	(0.062)	(0.051)
Has mobile money account	0.001	0.135***	0.001	0.195***
	(0.038)	(0.045)	(0.053)	(0.053)
Treatment × Has mobile money account	0.088	0.131**	0.123	0.134*
	(0.062)	(0.062)	(0.087)	(0.074)
Constant	0.224***	-0.242***	0.313***	-0.314***
	(0.028)	(0.031)	(0.039)	(0.037)
Treatment = 1 & Has mobile money account = 1	-0.185***	0.328***	-0.258***	0.406***
Treatment = 0 & Has mobile money account = 1	0.226***	-0.107***	0.314***	-0.118***
Difference (p-value)	0.338	0.967	0.333	0.600
Observations	3,931	3,931	3,931	3,931
R-squared	0.053	0.044	0.053	0.049
Panel D				
Treatment	-0.459***	0.350***	-0.640***	0.432***
	(0.051)	(0.050)	(0.071)	(0.059)
Has a bank account	0.016	0.082*	0.022	0.101*
	(0.040)	(0.046)	(0.056)	(0.055)
Treatment × Has a bank account	0.003	0.026	0.004	0.036
	(0.064)	(0.064)	(0.090)	(0.076)
Constant	0.215***	-0.230***	0.300***	-0.284***
	(0.033)	(0.036)	(0.045)	(0.043)
Treatment = 1 & Has mobile money account = 1	-.226***	0.229***	-0.314***	0.285***
Treatment = 0 & Has mobile money account = 1	0.231***	-0.147***	0.321***	-0.183***
Difference (p-value)	0.886	0.586	0.896	0.593
Observations	3,931	3,931	3,931	3,931
R-squared	0.053	0.035	0.053	0.039

Notes: Columns (1) and (2) use an aggregated index constructed using the technique of Kling et al. (2007). Columns (3) and (4) use an aggregated index constructed using principal component analysis. The "difference (p-value)" row in each panel tests the difference in the estimated conditional means in the preceding two rows. Treatment group receives negative framing in statements 1-3, and positive framing in statements 4-6. Control group positive negative framing in statements 1-3 and negative framing in statements 4-6. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1