Being surveyed can change later behavior and related parameter estimates

Alix Peterson Zwane[a], Jonathan Zinman[b,c,d,1], Eric Van Dusen[e], William Pariente[c,f], Clair Null[g], Edward Miguel[b,c,h,i], Michael Kremer[b,c,h,j,k], Dean S. Karlan[b,c,l], Richard Hornbeck[h,j], Xavier Giné[b,m], Esther Duflo[b,c,n], Florencia Devoto[b,c], Bruno Crepon[c,o], and Abhijit Banerjee[b,c,n]

[a]The Bill & Melinda Gates Foundation, Seattle, WA 98102-3706; [b]Innovations for Poverty Action, New Haven, CT 06510; [c]Jameel Poverty Action Lab, Cambridge, MA 02142; [d]Department of Economics, Dartmouth College, Hanover, NH 03755; [e]Department of Agricultural and Resource Economics, University of California, Berkeley, CA 94720-3310; [f]Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium; [g]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA 30322; [h]National Bureau of Economic Research, Cambridge, MA 02138; [i]Department of Economics, University of California, Berkeley, CA 94720-3880; [j]Harvard University, Cambridge, MA 02138; [k]Brookings Institution, Washington, DC 20036; [l]Yale University, New Haven, CT 06520; [m]World Bank, Washington, DC 20433; [n]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142-1347; and [o]Centre de Recherche en Economie et Statistique, 92245 Malakoff Cedex, France

Does completing a household survey change the later behavior of those surveyed? In three field studies of health and two of microlending, we randomly assigned subjects to be surveyed about health and/or household finances and then measured subsequent use of a related product with data that does not rely on subjects' self-reports. In the three health experiments, we find that being surveyed increases use of water treatment products and take-up of medical insurance. Frequent surveys on reported diarrhea also led to biased estimates of the impact of improved source water quality. In two microlending studies, we do not find an effect of being surveyed on borrowing behavior. The results suggest that limited attention could play an important but context-dependent role in consumer choice, with the implication that researchers should reconsider whether, how, and how much to survey their subjects.

measurement effects | question–behavior effects | Hawthorne effects | survey methodology | models of attention

Many data collection efforts in the social and clinical sciences rely on surveys. Psychologists and marketing and survey experts have long held that surveying a subject can draw attention to risks or choices with ordinarily little salience and thereby induce changes in subsequent behavior (1, 2), but conclusive evidence on this question from which a causal relationship can be clearly established has been limited. We provide evidence from a variety of settings that the act of being surveyed can affect behavior and confound estimates of parameters that initially motivated the data collection.

"Survey" or "interview" effects may occur even when the survey does not ask specifically about intent to engage in the behavior of interest or provide new information, and even when subjects do not know that their later behavior is being observed by researchers. Hence survey effects are conceptually distinct from, but closely related to, "question–behavior" (i.e., mere measurement or self-prophecy) and Hawthorne effects. Question–behavior effects arise when behavior changes as a result of asking subjects for intentions or predictions regarding future behavior effects (3–6). Hawthorne effects occur when behavior changes as a result of a subject responding to being treated and observed, as part of an experiment (7). These effects are also related to "push polling" and other efforts to manipulate subject behavior by posing hypothetical questions (8).

We describe results from five different field experiments, in four developing countries, on whether being surveyed affects subsequent behavior. The first experiment randomizes the frequency of surveys on health behavior and diarrhea incidence, in a context in which other water quality interventions are also randomly provided. More-frequent surveying leads to lower reported child diarrhea and cleaner water (as measured by the presence of detectable chlorine in household drinking water). We posit that frequent surveying serves as a reminder to invest in water purification, although because subjects know they are be-

ing observed it may also (or instead) be the case that subjects in the frequent-surveying condition experience more intense Hawthorne effects.

Frequent surveying also significantly affects the parameter estimate of the treatment effect of a source water quality intervention. An outcome measurement strategy that relied exclusively on frequent surveying would have led to the erroneous conclusion that the source water quality intervention was ineffective. This result suggests that much of the existing evidence on the effects of various environmental health interventions on diarrhea may be biased, because high-frequency measurement of health outcomes is the standard approach used in the epidemiology literature (9, 10).

The remaining four studies use a three-stage design: first a survey firm attempts to complete a survey on a random subset of the full sample, then a microfinance institution offers a medical insurance or loan product to the full sample, and last, take-up data are collected from administrative records. Hawthorne effects are unlikely here because a subject's take-up decision is not observed by the surveyor, nor do subjects know that their take-up is observed subsequently by researchers.

Despite the absence of Hawthorne effects, we still find that being surveyed significantly increases take-up in the two medical insurance experiments. However, in contrast to the water quality experiment, we find that being surveyed does not change estimates of another treatment effect parameter (in this case, the price elasticity of demand for insurance in experiment 2); nor do we find any significant effects of being surveyed on subsequent loan take-up in the two microloan experiments.

We speculate that survey effects arise in some settings but not others because some subjects typically ignore certain elements in their choice or consideration set, instead focusing on more pressing, tempting, and hence salient claims on resources (e.g., those that can be satisfied by borrowing or that are salient). Ignored elements may include health-promoting behaviors (e.g., water purification tasks), certain risks (e.g., accidents), and related products (e.g., insurance) (11). A survey makes neglected needs or opportunities more salient and spurs a more active decision (1).

The potential for asking questions that affect subsequent behavior has substantive, methodological, and ethical implications for the social and medical sciences.

ECONOMIC SCIENCES

MEDICAL SCIENCES

Substantively, survey effects point to the importance of limited attention in individual cognition, and manipulation thereof, in important decisions such as whether to make preventative health investments or buy insurance. A basic assumption of standard economic models may not hold: individuals may *not* consider important options in their choice set unless (subtly) prompted to do so, by a survey or some other external stimulus (12).

Methodologically, our findings suggest that survey effects can bias parameter estimates (e.g., estimates of other treatment effects). Consequently it may be preferable in some contexts to do infrequent measurement of outcomes in a large sample rather than frequent surveying of a small sample (see our experiment 1). It may be preferable in other contexts to forgo or circumscribe a baseline survey that would typically be used to estimate heterogeneous treatment effects (see our experiments 2–5).

Ethically, survey effects raise concerns that researchers or firms may inadvertently or deliberately manipulate the later behavior of subjects, with potentially deleterious effects (13, 14).

Our design is unique in how it combines several key features (the references in this paragraph highlight some prior studies using a subset of these features). First, the survey questions do not involve intentions or prediction measurement; that is, our questions are less directly related to subsequent behaviors of interest than in most studies of how surveying affects behavior (15, 16). Second, the relevant survey questions are only a small fraction of a larger survey (17). Third, we measure subsequent behavior using administrative data (on product take-up) or other objective measures (e.g., chlorine tests) rather than relying solely on respondent self-reports (15, 18–21). Fourth, we investigate whether survey effects dissipate or intensify over time (15, 19). Fifth, we look at the effects of being surveyed more or less frequently (7, 17). Sixth, in two of our studies, the survey effect design is laid over randomized program interventions (source water quality improvement through spring protection in experiment 1, and insurance pricing in experiment 2), enabling us to estimate whether survey effects bias estimates of the program treatment effects [7, 22 (p 24)]. Seventh, in three of our five studies (all except experiments 1 and 5), there is no clear link between the survey and the subsequent product offer or behavior of interest, and our subjects plausibly do not even know they are being observed by a third party on an ongoing basis.

Results

Experiment 1: Point-of-Use Home Water Treatment Take-Up in Kenya.
This experiment examines the effect of the frequency of surveying about health status on the later use of WaterGuard, a chlorine solution used to disinfect home drinking water. WaterGuard is marketed widely by Population Services International in Kenya. The cost of a month's supply is 20 Kenyan Shillings (or approxi-

mately $0.30 [US]), or roughly one quarter of the daily wage for agricultural labor in the local area ($1.26).

This experiment also measures the effect of varying the frequency of survey contact on estimates of a treatment effect of primary interest: the effect of improved source water quality on diarrhea incidence.

The sample for this experiment is composed of 330 households in rural western Kenya who were randomly selected from a frame of 1,500 households involved in a larger randomized evaluation of spring protection and WaterGuard use (23). Of these, 170 households were randomly assigned to be surveyed about health status biweekly [to accord with common practice in surveys of child diarrhea in epidemiology (9, 10)] for 18 2-wk rounds beginning in May 2007, although there was a 4-mo gap between rounds 15 and 16 owing to the postelection violence in early 2008. A final survey (round 19) was conducted 7 mo later in December 2008. The remaining 160 households were randomly selected to get the same survey just three times, or every 6 mo, during the same period: in biweekly survey rounds 9 (September 2007), 16 (April 2008), and 19 (December 2008). More than 97% of both the biweekly and low-frequency groups completed at least one of their survey rounds; 90% of the biweekly group completed at least 17 of the 19 surveys, and 90% of the low-frequency group completed at least 2 of the 3 surveys. Dataset S1 (panel 1) presents some descriptive statistics.

Questions on chlorine treatment constituted a small fraction (three questions) of the 20- to 30-min survey (Dataset S2), but the rest of the survey design in this experiment is such that subjects knew that their behavior was of some interest to the surveyors. Health and diarrhea questions constituted much of the surveys. When a respondent reported that the household's drinking water supply had been treated with chlorine, a DPD (*N*-dimethyl-p-phenylenediamine) test was conducted for total chlorine in the water stored in the household using a Hach Pocket Colorimeter II. Thus subjects knew their WaterGuard use was being observed in a way that could lead to a Hawthorne effect (from being observed more frequently) as well as a "survey" effect (from being surveyed more frequently).

Table 1, columns 1–4, shows that more-frequent survey visits led to lower reports of child diarrhea and to more chlorine (WaterGuard) presence in stored home drinking water. We focus on outcomes in rounds 9, 16, and 19 because in those rounds both the biweekly surveys and the low-frequency surveys were conducted. The survey completion rate was 90% in round 9 and 95% in rounds 16 and 19. Columns 1 and 2 show decreases in child diarrhea prevalence (as reported by mothers or other care-givers) for the biweekly survey group. Column 1 uses round-9 data (with the unit of observation being the child) and finds that reported diarrhea is 7 percentage points lower in the more-frequently measured biweekly group (vs. the low-frequency group). Column

Table 1. Impact of more-frequent surveys on chlorine use, diarrhea, and spring protection treatment effect stability in Kenya

Parameter	Child diarrhea (self-report) in the past week—survey round 9 (1)	Child diarrhea (self-report) in the past week—panel data (2)	Chlorine in home drinking water container—survey round 9 (3)	Chlorine in home drinking water container—survey round 16 (4)	Child diarrhea (self-report) in the past week—panel data (5)
Surveyed more frequently	−0.068* (0.022)	−0.149* (0.026)	0.150* (0.050)	0.053† (0.028)	−0.182* (0.033)
Protected spring					−0.104* (0.034)
Surveyed more frequently × Protected spring					0.135* (0.048)
Mean of dependent variable	0.081	0.140	0.164	0.068	0.140
No. of observations	713	4,296	293	309	4,296

Ordinary Least Squares estimates with Huber-White SEs in parentheses, clustered at the spring level. Specifications reported in columns 2 and 5 include child fixed effects, because the data used there are an unbalanced panel.
*Significant at 1%.
†Significant at 10%.

2 adds data from rounds 16, 19, and four prior annual surveys and finds an even larger effect. Column 2 (and column 5) adds child fixed effects as additional controls, because in these specifications we have multiple observations per child and an unbalanced panel.

Column 3 shows that households in the biweekly survey group were 15 percentage points more likely to use chlorine in round 9. In the raw data (not controlling for month of interview) the difference is 16 percentage points (0.24, vs. 0.08 in the low frequency group). Round-16 data were collected immediately after a 4-mo suspension due to postelection violence after the December 2007 Kenyan election. Chlorine use fell in both groups, but column 4 shows that usage was 5 percentage points higher ($P = 0.06$) in the biweekly survey group. This smaller treatment effect (compared with round 9) is consistent with the hypothesis that a survey serves as a reminder to chlorinate, with an effect that falls over time in the absence of reminders.

Column 5 estimates the impact of being surveyed more frequently on stability of a key parameter estimate. We take our specification from column 2 and add variables for whether the household's nearest spring was protected (randomly, as part of the evaluation described in ref. 23) and for the interaction between spring protection and being surveyed at biweekly frequency. The interaction term delivers an estimate of parameter instability.

The large, positive, and statistically significant ($P = 0.007$) coefficient estimate on the interaction term suggests that more-frequent surveying can indeed change estimates of how spring protection affects child diarrhea. This result differs from ref. 7, in which more frequent contact increases the magnitude of the estimated treatment effect of *Gingko biloba* on cognitive functioning among dementia patients but does not change the statistical significance of the Gingko effect. It may be the case that spring protection and survey reminders are substitutes in producing improved water quality; indeed, the spring protection effect is not statistically significant in the subsample of more-frequently surveyed households. Methodologically, analyzing the biweekly households alone would lead to the erroneous conclusion that source water quality improvement does not have a statistically significant effect on diarrhea.

Experiment 2: Hospitalization Insurance Take-Up in the Philippines. This experiment examines the effect of being surveyed on the later take-up of hospitalization insurance. The insurance provides coverage for inpatient care and loan obligations in the event of vehicle or work-related accidents. Green Bank, a large rural bank in the Philippines, sells the insurance as a voluntary add-on product to its microloans (insurance add-ons are common in credit markets throughout the world). The decision to take up the insurance was probably nontrivial for most subjects (e.g., a typical premium is approximately 3% of monthly income). We measure take-up using administrative data from Green Bank. Forty-six percent took up the insurance in the 9-mo period after the baseline survey.

This experiment also measures the effect of being surveyed on the estimate of a treatment effect of primary interest: the effect of insurance pricing on take-up.

The sample for this experiment includes 1,463 individual liability borrowers, in good standing on their microenterprise loan repayments, drawn from the Northern Mindanao and Caraga regions in southern Philippines. Dataset S1 (panel 2) presents some descriptive statistics. Sixty-two percent of subjects had some sort of health insurance coverage at baseline, but only approximately 5% had the type of broad accident coverage offered by the product studied here. We randomly assigned 80% of the sample to get a baseline survey (94% of the assigned group completed a survey). The survey was administered by an independent firm with no affiliation to Green Bank: the Research Institute for Mindanao Culture at Xavier University. The survey team made no mention of Green Bank or the possibility of follow-up visits for additional surveys. The survey included hundreds of questions on individual and household demographics, health status, family health history, risk-taking behaviors and preferences, and time preferences. Six questions mentioned insurance (Dataset S3). The survey did not ask about intent or likelihood of purchasing insurance.

After the baseline survey, Green Bank marketed the insurance product door to door. Only 87% of the sample was reached for marketing, but we include all 1,463 in our analysis. Marketers reached clients who had been surveyed at a median time of 50 d after the survey (with a range of 9–143 d).

Table 2 presents our estimate of the effects of completing the baseline survey (instrumented by the random assignment to survey status) on subsequent insurance take-up. The point estimate on the "surveyed" variable in column 1 shows a 5 percentage point increase that is statistically insignificant ($P = 0.14$) over the 9-mo window for which we have take-up data. Columns 2 and 4 examine whether being surveyed changes a parameter estimate of interest: the price sensitivity of insurance purchase. The primary purpose of the study was to measure (heterogeneity in) price sensitivity, and Green Bank randomized the insurance premium it offered clients to be between 0 and 10 pesos per 1,000 pesos of outstanding loan amount. We find that subjects are price sensitive on average (from the results on the "premium" variable) and that the survey does not affect price sensitivity (i.e., the point estimate on the "Surveyed × Premium" interaction variable is insignificant). Column 3 shows a smaller and still insignificant survey effect on whether subjects had insurance 6–9 mo after the survey was completed.

Experiment 3: Health Insurance Take-Up in the Philippines. This experiment examines the effect of being surveyed on the later take-up of PhilHealth, a government-sponsored health insurance product in the Philippines. Green Bank wholesales PhilHealth as an add-on to its loan products, similar to the hospitalization insurance offered in experiment 2. The decision to take up PhilHealth was probably nontrivial for most clients. Premiums are

Table 2. Impact of baseline survey on hospitalization insurance take-up, and on price elasticity stability, in the Philippines

Parameter	Ever purchased product offered subsequent to treatment (1)	Ever purchased product offered subsequent to treatment (2)	Owned product 6–9 mo after treatment (3)	Owned product 6–9 mo after treatment (4)
Surveyed	0.051 (0.034)	0.070 (0.067)	0.010 (0.021)	0.021 (0.041)
Initial premium assigned (0–10)	−0.021* (0.004)	−0.018[†] (0.009)	−0.002 (0.003)	−0.000 (0.006)
Surveyed × Premium		−0.004 (0.010)		−0.002 (0.006)
Mean of dependent variable	0.462	0.462	0.115	0.115
No. of observations	1,463	1,463	1,463	1,463

Results estimated using two-stage least squares, with random assignment to survey status instrumenting for survey completion, and controls for randomization strata.
*Significant at 1%.
[†]Significant at 10%.

Table 3. Impact of baseline survey on health insurance take-up in the Philippines

Parameter	Ever purchased product offered subsequent to treatment (1)	Owned product 6–9 mo after treatment (2)	Owned product 15–18 mo after treatment (3)
Surveyed	0.067* (0.033)	0.036† (0.022)	0.023 (0.021)
Mean of dependent variable	0.264	0.094	0.086
No. of observations	1,224	1,224	1,224

Results estimated using two-stage least squares, with random assignment to survey status instrumenting for survey completion, and controls for randomization strata.
*Significant at 5%.
†Significant at 10%.

1,200 pesos per year ($25), which is approximately 12% of the median loan amount and 10.5% of median monthly income. We measure take-up using PhilHealth administrative data. Twenty-six percent of our sample signed up for PhilHealth in the 18-mo period after the baseline survey.

The sample for this experiment includes 1,224 individual liability borrowers, in good standing on their microenterprise loan repayments, drawn from Mindanao and Leyte, Philippines (no overlap with those in experiment 1). Dataset S1 (panel 3) presents some descriptive statistics. Note that only 25% of surveyed households had any health insurance coverage at baseline. We randomly assigned 80% of the sample to get a baseline survey (87% of this group completed a survey).

The survey was administered by independent firms with no affiliations to Green Bank: The Office of Population Studies at San Carlos University for Leyte, and Research Institute for Mindanao Culture at Xavier University for Mindanao. The survey teams made no mention of Green Bank or the possibility of follow-up visits for additional surveys. The survey included hundreds of questions on individual and household demographics, health status, family health history, risk-taking behaviors and preferences, and time preferences. Six questions mentioned insurance (Dataset S4). The survey did not ask about intent or likelihood of purchasing insurance. The only mention of PhilHealth was in the menu of possible responses to the question: "Are you currently covered by any of the following types of health insurance?"

After the baseline survey, Green Bank marketed PhilHealth door to door. Only 73% of the targeted 1,224 borrowers were reached for marketing, but we include all 1,224 in our analysis.

Table 3 shows effects of taking the baseline survey (instrumented by the random assignment to survey status) on subsequent health insurance purchase. Those taking the survey are 6.7 percentage points (25%) more likely to take up the insurance (column 1). This effect dissipates over time, as shown in columns 2 and 3, which analyze insurance coverage 6–9 mo and 15–18 mo after the survey. These findings, and the similar result in experiment 1, are similar to the finding of Chandon et al. (19) but contrast with those of Dholakia and Morwitz (15), whereby the effect first increases and then decreases over time. Whether these contrasting dynamics are due to different methodologies, different subject populations/settings, and/or different treatments is an important consideration for future research.

Pooling the Insurance Results. Given the similarity between the two insurance studies, we also pool their results. Table 4 shows that we find a statistically significant 5.8 percentage point (16%) increase in take-up at any point after the survey. As indicated above, this effect is driven by take-up in months immediately after the survey. Column 2 shows the dissipation of this effect 6–9 mo after the survey.

Experiment 4: Microcredit Take-Up in Morocco. This experiment examines the effect of completing a baseline survey on later take-up of group liability microcredit in rural Morocco. The loan product is offered by Al Amana, the largest microfinance institution in Morocco. The decision to take up a microloan was probably nontrivial for most households, because the median loan size ($600) is large relative to annual household income. We measure take-up using administrative data from Al Amana. Seventeen percent of households took up a loan at some point during our 2-y sample period.

The sample frame includes all 1,099 households in seven small villages that Al Amana was entering for the first time. Dataset S1 (panel 4) presents some descriptive statistics.

We stratified the random assignment of the baseline survey to target 100 households per village. Thus 60% of the 1,099 households were assigned to be surveyed. Nearly 100% of those assigned to be surveyed were actually reached and completed a survey. The survey was administered by an independent firm with no affiliation to Al Amana: Team Maroc.

Surveys asked hundreds of questions on demographics, asset ownership, consumption and income, activities of the households, and the use of and need for financial services. Questions on credit constituted approximately 15% of the survey, and we list them in Dataset S5. The survey's only mention of Al Amana was in a list of 15 potential credit sources on one of the questions. The survey did not directly ask about intent or likelihood of using credit, although one of the questions is effectively a likelihood question, in hypothetical form: "If you could get a credit of 3000 dhs, with installments of 400 dhs per month during 9 months, would you take it?"

After the baseline survey, Al Amana marketed the products in villages using group meetings, flyers, and door-to-door visits. Al Amana marketed every week or two for at least 1 y after entering a village.

Table 5 shows that we do not find any evidence that completing a baseline survey significantly affects loan take-up or use, over any horizon. The estimates are imprecise, however, and the confidence intervals include effects that are large relative to the sample mean loan take-up rates.

Experiment 5: Microloan Renewal in India. This experiment examines the effect of being surveyed on later renewal of group liability microloans from SKS Microfinance, a large lender in rural India. The loan renewal decision was probably nontrivial for most households, because the median loan size was 20% of the median household's annual expenditure. SKS administrative data show that 77% of households renewed a loan within our 2-y sample period.

The sample frame includes 10,780 borrowers from 200 villages. Dataset S1 (panel 5) presents some descriptive statistics.

In each village, we set a target number of households to be surveyed and then randomly assigned the order in which surveyors attempted to survey households. This randomization was done at the household level, with the constraint that someone from each five-person loan group in the village must appear in the survey

Table 4. Impact of baseline survey on insurance take-up: Pooling two samples from Table 3

Parameter	Ever purchased product offered subsequent to treatment (1)	Owned product 6–9 mo after treatment (2)
Surveyed	0.058* (0.024)	0.022 (0.015)
Mean of dependent variable	0.372	0.108
No. of observations	2,687	2,687

Results estimated using two-stage least squares, with random assignment to survey status instrumenting for survey completion, and controls for randomization strata.
*Significant at 5%.

Zwane et al.

Table 5. Impact of baseline survey on microcredit take-up in Morocco

Parameter	Ever purchased product offered subsequent to treatment (1)	Owned product 6–9 mo after treatment (2)	Owned product 15–18 mo after treatment (3)
Surveyed	−0.009 (0.024)	−0.007 (0.024)	−0.024 (0.024)
Mean of dependent variable	0.166	0.141	0.162
No. of observations	1,099	1,099	1,099

Results estimated using two-stage least squares, with random assignment to survey order instrumenting for survey completion, and controls for randomization strata.

order (i.e., be chosen to be surveyed) before an additional member of any group could appear in the order. Surveyors were instructed to make several attempts to reach a household before moving to the next household on the list and to continue surveying in a village until the target number of completed surveys was reached. In all, 41% of the 10,780 households completed a survey.

The surveys asked several hundred questions on demographics, assets, income, expenditure, health, and credit. Questions on credit constituted a small fraction of the survey (Dataset S6). The survey did not ask about intent or likelihood of using credit.

Surveyors were hired and managed by the research team, independently of SKS. However, in this experiment there were explicit connections between the surveys and the subsequent product offer: SKS loan officers and some clients introduced surveyors to other clients, and all survey subjects were informed that they had been selected to participate in the study because they were members of SKS. SKS was also listed as a potential source of credit in six of the survey questions.

After the baseline survey, most clients faced the decision of whether to renew their loan within the next year (because most clients had loans with a 50-wk maturity). We estimate the effect of taking a baseline survey on subsequent loan renewal from SKS by instrumenting for survey completion with the household's randomly assigned survey order. We define renewal as taking a new loan within 10 wk of having paid off one's previous loan (results are not sensitive to alternative definitions).

Table 6 shows that we do not find any significant effects. Column 1 includes the full sample (and hence all renewals subsequent to the time of survey completion). The point estimate on the effect of the baseline survey is small and not significant, and the confidence intervals contain only small effects (a change of 3 percentage points or less, relative to the sample mean of 0.77). Columns 2–4 estimate the survey effect in subsamples that account for the fact that, because of repayment schedules, not all households were actually eligible to renew their loan immediately after the survey. Column 2 includes only those with a loan coming up for renewal within 2 mo of the survey ending in all villages. The point estimate is again small and not significant, and the upper and lower bounds of the confidence interval imply a small effect (approximately a 7 percentage point change in renewal, relative to a sample mean of 0.67). We find similar results when expanding

the sample to include those eligible to renew during other windows subsequent to the survey period (columns 3 and 4).

Discussion

We analyze five field experiments that randomize whether or how often a household was surveyed and then track subsequent use of a related product using data that do not rely on respondent self-reports. In the three health studies, we find that being surveyed increased product use. In the two studies on microloans, we do not find any effect of being surveyed on borrowing behavior. We also find some evidence, in one of the two studies with appropriate data, that being surveyed more frequently generates heterogeneous selection into a behavior of interest and thereby biases the parameter estimate that primarily motivated the data collection in the first place.

Methodologically, our results suggest that researchers should rely on the use of unobtrusive data collection when possible and consider the tradeoffs between potential biases introduced from surveying and the benefits from having baseline data to identify heterogeneous treatment effects not possible to estimate without implementation of a baseline survey. In cases in which obtrusive data collection is essential (e.g., to collect data on certain outcomes, as in our experiment 1), infrequent survey visits on large samples may be preferable to smaller samples with higher-frequency data collection and equivalent statistical power.

Unpacking the cognitive and behavioral mechanisms behind interview effects has implications for model specification in consumer choice and for the design of optimal policy, marketing, and social marketing, as well as for survey methodology. Our designs here are not rich enough to identify definitively different cognitive mechanisms, but we offer some speculation that is informed by findings from related literatures.

The two largest related literatures on how surveys affect behaviors—on mere measurement and self-prophecy effects—point to potentially important roles for baseline attitudes toward, and experience with, the subsequent behavior of interest in "moderating" (producing heterogeneous) treatment effects (17, 24). In contrast, our results do not offer much support for the hypothesis that variation in attitudes and experience produces heterogeneous survey effects. In the one study in which we have the relevant baseline measures (experiment 1), we do not find strong evidence that survey effects vary with several different measures of baseline attitudes toward or experience with the chlorine product or brand (Dataset S7). Less directly, we find significant main survey effects in the insurance studies, for which baseline attitudes may be uniformly indifferent ("insurance is sold, not bought") and for which baseline experience is probably low (hospitalization insurance is a new product). We do not find effects in one setting in which baseline attitudes and experience are plausibly both quite high (experiment 5, conducted on existing microcredit customers in a setting with high renewal rates even in the control group). Some question–behavior studies find that another factor—cognitive dissonance—plays a moderating role, but that seems unlikely in our surveys because our treatments do not ask respondents to state intentions or likelihoods.

So which cognitive mechanisms *do* underlie our results? In experiment 1 it seems likely that each survey serves as a sort of reminder to use WaterGuard; previous studies have shown that both patients and clinicians respond to repeated reminders to en-

Table 6. Impact of baseline survey on microcredit renewal in India

Parameter	Ever renewed subsequent to treatment (1)	Renewed, if eligible to renew within 2 mo of treatment (2)	Renewed, if eligible 6–9 mo subsequent to treatment (3)	Renewed, if eligible 15–18 mo subsequent to treatment (4)
Surveyed	−0.005 (0.013)	−0.004 (0.036)	−0.016 (0.025)	−0.006 (0.032)
Mean of dependent variable	0.769	0.674	0.635	0.712
No of observations	10,780	1,944	4,111	3,116

Ordinary Least Squares estimates, with Huber-White standard errors in parentheses, randomly assigned surveys, and controls for randomization strata.

gage in health-promoting behaviors (25). In experiments 2–5, we speculate that our relatively subtle treatments (being surveyed about topics related to the target behavior) work through non-conscious, low-effort cognition: there is evidence of this even in the more directed questioning used in the mere measurement effect literature (8, 26). If this automatic, "System I" processing (27) is indeed important, then survey designers may have something to learn from the literatures on priming, persuasive advertising, and information provision, which find that small changes in stimuli can produce substantial and lasting (but not always uniform) effects on behavior (28–32). The economics literature on limited attention also seems relevant; as noted at the outset, one could model our survey effects as shocks that draw attention to choice set elements that usually have relatively low salience. Insurance decisions (which bear on future contingencies) may be farther from "top of mind" than borrowing decisions (which bear on immediate needs or opportunities) and hence respond more to the subtle attention prods (33).

Our findings raise several other questions for future research. Do surveys work directly on attention per se or indirectly through intent or goal formation (28, 29, 34, 35)? Do our results generalize to wealthier societies? Are nontargeted behaviors affected as well? Is it more directly related or less directly related questions that drive survey effects on behavior (e.g., in our insurance studies, is it questions on insurance coverage and/or questions on health risks)? How can one use evidence on the magnitude and nature of survey effects to remove bias from survey data (36, 37)? Do longer lags between surveying and subsequent choices mitigate survey effects? Do less-obtrusive methods of measuring outcomes of interest yield sufficiently precise estimates to be feasible (e.g., with a sufficiently large sample, could one estimate the effects of spring protection using a single follow-up survey on diarrhea incidence as an outcome)?

Much work remains. In the meantime, researchers might reconsider whether, how, and how much to survey their subjects. There is a risk of asking.

Methods

In experiments 2–5, we estimate whether and how much subsequent product take-up is influenced by the baseline survey using two-stage least squares with the random assignment of survey frequency status as an instrument for survey completion. Instrumenting avoids any confounding correlations between product take-up and survey completion by using only the randomly engineered component of survey completion to identify the effects of being surveyed on product take-up. It also rescales the results to adjust for differences in survey completion rates, making the magnitudes comparable across studies. We use completed surveys only in experiment 1, because everyone was assigned to a survey condition (more- vs. less-frequent), and we only have outcome data (e.g., chlorine test results) on those who completed a given survey. All regressions also control for any conditions used to stratify the random assignments. For experiment 5, we report results using a binary instrument that equals 1 if the household was assigned a survey order below 40 (where, for example, a survey order of 1 means that we told surveyors to try that household first). Results are robust to other definitions and functional forms for the instrument(s).

1. Bridge G, Reeder L, Kanouse S, Kinder D, Nagy V (1977) Interviewing changes attitudes—sometimes. *Public Opin Q* 41:56–64.
2. Waterton J, Lievesley D (1989) *Panel Surveys* (John Wiley and Sons, New York).
3. Sprott D, et al. (2006) The question-behavior effect: What we know and where we go from here. *Soc Influence* 1:128–137.
4. Dholakia U (2010) *A Critical Review of Question-Behavior Effect Research* (Review of Marketing Research), vol 7, pp 145–197.
5. Sherman S (1980) On the self-erasing nature of errors of prediction. *J Pers Soc Psychol* 39:211–221.
6. Feldman J, Lynch J (1988) Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *J Appl Psychol* 73:421–435.
7. McCarney R, et al. (2007) The Hawthorne effect: A randomised, controlled trial. *BMC Med Res Methodol* 7:30.
8. Fitzsimons G, Shiv B (2001) Nonconscious and contmination effects of hypothetical questions on subsequent decision making. *J Consum Res* 334:782–791.
9. Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S (2007) Interventions to improve water quality for preventing diarrhoea: Systematic review and meta-analysis. *BMJ* 334:782–792.
10. Waddington H, Snilstveit B (2009) Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhoea. *J Development Effectiveness* 1: 295–335.
11. Alba J, Hutchinson JW, Lynch JG (1991) *Handbook of Consumer Behavior* (Prentice Hall, Edgewood Cliffs, NJ), pp 1–49.
12. DellaVigna S (2009) Psychology and economics: Evidence from the field. *J Econ Lit* 47: 315–372.
13. Fitzsimons G, Moore S (2008) Should I ask our children about sex, drugs, and rock & roll? Potentially harmful effects of asking questions about risky behaviors. *J Consum Psychol* 18:82–95.
14. Williams P, Fitzsimons G, Block L (2004) When consumers do not recognize 'benign' intention questions as persuasive attempts. *J Consum Res* 31:540–550.
15. Dholakia U, Morwitz V (2002) The scope and persistence of mere measurement effects: Evidence from a field study of customer satisfaction measurement. *J Consum Res* 29:159–167.
16. Fitzsimons G, Morwitz V (1996) The effect of measuring intent on brand-level purchase behavior. *J Consum Res* 23:1–11.
17. Morwitz V, Johnson E, Schmittlein D (1993) Does measuring intent change behavior? *J Consum Res* 20:46–61.
18. Spangenberg E, Sprott D, Grohmann B, Smith R (2003) Mass communicated prediction requests: Practical application and a cognitive dissonance explanation for self prophecy. *J Mark* 67:47–62.
19. Chandon P, Morwitz V, Reinartz W (2004) The short- and long-term effects of measuring intent to repurchase. *J Consum Res* 31:566–572.

20. Morwitz V, Fitzsimons G (2004) The mere measurement effect: Why does measuring intentions change actual behavior? *J Consum Psychol* 14:64–74.
21. Smith J, Gerber A, Orlich A (2003) Self-prophecy effects and voter turnout: An experimental replication. *Polit Psychol* 24:593–604.
22. Campbell DT, Stanley JC (1963) *Experimental and Quasi-experimental Designs for Research* (Rand McNally College Publishing, Chicago, IL).
23. Kremer M, Leino E, Miguel E, Zwane A (2010) Spring cleaning: Rural water impacts, valuation, and property rights institutions. *Q J Econ*.
24. Levav J, Fitzsimons GJ (2006) When questions change behavior: The role of ease of representation. *Psychol Sci* 17:207–213.
25. van Dulmen S, et al. (2007) Patient adherence to medical treatment: A review of reviews. *BMC Health Serv Res* 7:55.
26. Fitzsimons GJ, Williams P (2000) Asking questions can change choice behavior: Does it do so automatically or effortfully? *J Exp Psychol Appl* 6:195–206.
27. Kahneman D (2003) Maps of bounded rationality: Psychology for behavioral economics. *Am Econ Rev* 93:1449–1475.
28. Chartrand T, Huber J, Shiv B, Tanner R (2008) Nonconscious goals and consumer choice. *J Consum Res* 35:189–201.
29. Sela A, Shiv B (2009) Unraveling priming: When does the same prime active a goal versus a trait? *J Consum Res* 36:418–433.
30. Bertrand M, Karlan D, Mullainathan S, Shafir E, Zinman J (2010) What's advertising content worth? Evidence from a consumer credit marketing field experiment. *Q J Econ* 125:263–305.
31. Ayres I, Raseman S, Shih A (2009) *Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. NBER Working Paper No. 15386* (National Bureau of Economic Research, Cambridge, MA).
32. Allcott H (2010) *Social Norms and Energy Conservation. Working Paper* (New York University, New York, NY).
33. Karlan D, McConnell M, Mullainathan S, Zinman J (2010) *Getting to the Top of Mind: How Reminders Increase Saving.* National Bureau of Economic Research Working Paper no. 16205.
34. Bargh JA, Gollwitzer PM, Lee-Chai A, Barndollar K, Trötschel R (2001) The automated will: Nonconscious activation and pursuit of behavioral goals. *J Pers Soc Psychol* 81: 1014–1027.
35. Webb TL, Sheeran P (2006) Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychol Bull* 132:249–268.
36. McFadden D, et al. (2005) Statistical analysis of choice experiments and surveys. *Mark Lett* 16:183–196.
37. Chandon P, Morwitz V, Reinartz W (2005) Do intentions really predict behavior? Self-generated validity effects in survey research. *J Mark* 69:1–14.