NBER WORKING PAPER SERIES

CAN SIMPLE PSYCHOLOGICAL INTERVENTIONS INCREASE PREVENTIVE HEALTH INVESTMENT?

Johannes Haushofer
Anett John
Kate Orkin

Working Paper 25731
http://www.nber.org/papers/w25731

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2019

Supported by grant NIH UH2 NR016378 to JH. We thank Jane Dougherty, Daniel Mellow, Moritz Poll, and the Busara Center for Behavioral Economics for research assistance; Clair Null, Michael Kremer, and the WASH Benefits Kenya team for collaboration and advice; and Nava Ashraf, David Laibson, Xavier D'Haultfoeuille, and Alessandro Iaria for comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Can Simple Psychological Interventions Increase Preventive Health Investment?
Johannes Haushofer, Anett John, and Kate Orkin
NBER Working Paper No. 25731
April 2019
JEL No. D91,I12,O12

ABSTRACT

Behavioral constraints may explain part of low demand for preventive health products. We test the effects of two light-touch psychological interventions on water chlorination and related health and economic outcomes using a randomized controlled trial among 3750 women in rural Kenya. One intervention encourages participants to visualize alternative realizations of the future; one builds participants' ability to make concrete plans to achieve goals. Both interventions include information on health benefits of chlorination. After twelve weeks, both interventions increase the share of households who chlorinate drinking water and reduce child diarrhea episodes. Analysis of mechanisms suggests both interventions increase self-efficacy – beliefs about one's ability to achieve desired outcomes – as well as the salience of chlorination. They do not differentially affect beliefs and knowledge about chlorination (compared to a group who receive only information), nor affect lab measures of time preferences or planning ability. Results suggest simple psychological interventions can increase use of preventive health technologies.

Johannes Haushofer
Woodrow Wilson School
Princeton University
427 Peretsman-Scully Hall
Princeton, NJ 08540
and Busara Center for Behavioral Economics,
Nairobi, Kenya
and also NBER
haushofer@princeton.edu

Kate Orkin
University of Oxford
Merton College
Merton Street
Oxford, OX14JD
United Kingdom
kate.orkin@merton.ox.ac.uk

Anett John
CREST
5 Avenue Henry Le Chatelier
91120 Palaiseau
FRANCE
anett.john@ensae.fr

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/2850/history/27566

1. Introduction

Individuals often fail to invest in preventive healthcare, even when such investments cost little and individuals are aware of their benefits.[1] An estimated two thirds of deaths of children under 5 could be averted with cheap preventive technologies (Dupas 2014a). A prominent example is chlorination of drinking water, which is highly effective in reducing prevalence of diarrhea, particularly among young children (Arnold and Colford Jr 2007). Diarrhea is the second leading cause of death worldwide among children aged 1–5, contributing to nearly half a million deaths in 2015 (Wang et al. 2016). It is a leading cause of morbidity (Walker et al. 2013), and stunts healthy growth in children through enteric dysfunction (Richard et al. 2013). In many settings, chlorine for water is readily and cheaply available, but infrequently used by individuals without access to clean water. In our study areas in Kenya, only 3 percent of households used chlorine before any intervention (Null et al. 2018), although a month's supply costs only KES 25 (USD 0.25). Interventions which provide chlorine for free, often in combination with information campaigns, increase usage to between 23 and 60 percent, but even after these interventions, take-up remains far below complete (Kremer et al. 2011a; Dupas et al. 2016; Null et al. 2018; Luoto et al. 2014).

A growing body of evidence suggests that behavioral or psychological constraints may explain some of the low demand for preventive health products. For example, demand for commitment products suggests a role of present bias in health decisions (Schilbach 2018; Bai et al. 2017; Giné, Karlan, and Zinman 2010; DellaVigna and Malmendier 2006; Royer, Stehr, and Sydnor 2015); the success of planning interventions argues for a role of planning and follow-through (Stadler, Oettingen, and Gollwitzer 2009; Milkman et al. 2013; Milkman et al. 2011); and the effectiveness of reminders in medication adherence suggests that limited attention may be important (Pop-Eleches et al. 2011; Lester et al. 2010).

In this paper, we present evidence from a field experiment in rural Kenya which studies the role of behavioral constraints in limiting the use of chlorine to treat drinking water, by targeting these constraints directly with simple psychological and informational interventions. We allocate 3750 young women to four treatment groups. The first group received a two-session group intervention where participants visualized alternative realizations of the future, depending on their behavior in the present ("Visualization" or "V"). The intervention aimed to make future outcomes more vivid and tangible in participants' minds, thus increasing their perceived value relative to the more

[1]For instance, providing insecticide-treated bed nets reduces incidence of malaria, which increases both farmworker productivity (Dillon, Friedman, and Serneels 2014) and farm yields (Fink and Masiye 2015). Bleakley (2010) estimates that a child who grew up malaria-free earns 50 percent more throughout their adult life. However, demand for bed nets is low and price sensitive (Cohen and Dupas 2010; Dupas 2014b). Similarly, treating intestinal parasites improves child health and school attendance (Miguel and Kremer 2004), growth (Bobonis, Miguel, and Puri-Sharma 2006), test scores (Ozier 2018), and adult wages (Baird et al. 2016), yet take-up is low at non-zero prices (Kremer and Miguel 2007).

immediate costs required to attain them. The second group received a two-session intervention that helped participants to undertake activities that they were struggling to do regularly, by making concrete, specific plans, anticipating potential obstacles, and establishing routines ("Planning" or "P"). To isolate the effects of the psychologically active elements of these treatments, a third, active control group received all elements of the intervention except the psychologically active components ("Active Control" or "AC" intervention): these participants also gathered as a group, but to discuss "psychologically inactive" topics. In addition, all three groups received a short information module about the benefits of chlorination ("Information" or "INF" intervention), to hold beliefs about chlorination constant across groups.[2] Finally, we compare these treatments to a fourth, pure control group ("PC"), who were simply surveyed at endline. Thus, our groups are "V+INF", "P+INF", "AC+INF", and "PC". The comparison between the three "active" treatment groups and the pure control group gives the total effect on targeted behaviors of providing interventions such as ours in other, similar settings. The comparison between the two "psychological" treatment groups and the AC+INF group tests the effect of interventions specifically targeting the ability to visualize the future or to make and execute plans, respectively, over and above those targeting lack of information.

We report economically large and statistically significant effects of our interventions on our main outcome of interest, chlorination of household drinking water, measured objectively through a chemical test in unannounced household visits. Specifically, in the Visualization and Planning groups relative to the pure control group, we find significant increases of 27 and 18 percent (6 and 4 percentage points), respectively, in the share of households whose drinking water contains chlorine twelve weeks after the interventions. This measure is objective and collected in unannounced household visits. In line with these findings, both the Visualization and Planning treatments significantly reduce the number of diarrhea episodes in children: Visualization reduces diarrhea by 35 percent (8 percentage points) relative to the pure control group and 38 (9 percentage points) percent relative to the active control group, and Planning by 26 and 25 percent (6 percentage points), respectively, with all effects statistically significant. In contrast, the effect in the Active Control group relative to the pure control group is small and not statistically significant for both chlorination and diarrhea.

These interventions are extremely cost-effective by the standards of the World Health Organization (WHO). Delivery of the interventions cost approximately USD 4 per household, or USD 1.33 per child. If we assume, conservatively, that the treatment effects of the visualization and planning interventions are 35 and 26 percent reductions (respectively) in diarrhea for 3 months, and that the effects then immediately go to zero, the cost per Disability-Adjusted Life Year (DALY) saved is USD 121 for the visualization and USD 161 for the planning intervention.[3] The WHO classifies an intervention as "cost-effective" for a cost per DALY saved below USD 4525,

[2]Our information module is similar in format to Jensen (2010), who gives information on returns to education.
[3]See Appendix A for detailed cost effectiveness calculations.

and "highly cost-effective" below USD 1508. By comparison, deworming projects have a cost per DALY saved of USD 28 – USD 70/DALY, and malaria bed nets of USD 100/DALY.[4]

The effect of our interventions is not limited to the health domain, but also extended to economic behavior: the Visualization intervention caused a statistically significant 26 percent increase in the amount saved per week (pre-specified as a secondary outcome). Effects were larger than those of the Planning intervention (significant at the ten percent level). In addition, the Planning intervention decreased the total self-reported hours of work in the last three months by 22 percent (although earnings did not decline, consistent with increased efficiency of work). We see no effects on use of other widely available preventive health products, such as vaccines or antenatal checkups, nor in non-health investments (such as in children's education), all pre-specified as exploratory outcomes. We conclude that light-touch psychological interventions may be most effective when participants also receive information emphasizing the benefits of a particular behavior change.

We also test if our psychological interventions have larger effects in villages where infrastructure lowers the effort and time costs of chlorination. We conducted our intervention in the study sites of a previous trial, the "WASH Benefits" study (Null et al. 2018). In this study, villages were randomly assigned to receive chlorine dispensers placed at the water source, and dispensers had been maintained after the trial. Our treatment effects on chlorination are somewhat larger in villages with dispensers, although differences are mostly statistically insignificant. Thus, we observe weak evidence that, when psychological, information and time cost/effort constraints are alleviated simultaneously, effects on behavior may be larger than when cost and access barriers remain.

Finally, we use lab-in-the-field methods and psychological questionnaires to study the mechanisms through which these treatment effects operate. We show, first, that differences in behavior are not explained merely by differences in beliefs about the efficacy of chlorine, or increased knowledge about chlorination. All three "active" treatment groups received the information treatment and show similar improvements in their belief that chlorination can prevent diarrhea, and in knowledge about using chlorine. But the effects of psychological treatments on chlorination and diarrhea are larger than the effects of pure information (i.e. the Active Control arm).

We then provide suggestive evidence in favor of two potential psychological channels accounting for the increased efficacy of the psychological interventions compared to the Active Control. First, we show large and significant effects of both the Visualization and Planning treatments, of 0.12–0.17 SDs, on a self-efficacy scale, which was pre-specified as a secondary outcome.[5] Second, in a "salience" task, we find that the psychological treatments focus participants' attention on

[4]WHO figures: https://www.who.int/bulletin/volumes/93/2/14-138206/en/.
Deworming: https://www.givewell.org/international/technical/programs/deworming/cost-effectiveness.
Bed nets: https://forum.effectivealtruism.org/posts/HbunzTyFPRwcYihg6/long-lasting-insecticide-treated-nets-usd3-340-per-life

[5]Self-efficacy captures beliefs about one's ability to achieve desirable outcomes (Bandura 1977).

chlorination more than the information treatment alone, even 10 weeks after the intervention. This effect is not present for other future-oriented behaviors (savings and farm investment).[6] While the Visualization and Planning interventions were designed to be domain-general, they use water treatment as an example behavior, often in the context of engaging stories or exercises. Together, these findings suggest that psychological interventions which build self-efficacy and successfully focus attention on a target behavior may be more effective than standard information campaigns.

Our psychological interventions do not significantly affect laboratory measures of time preferences, neither using an incentivized real-effort task, nor using multiple price lists over money. They also have no measurable impact on cognitive functions linked to planning, nor on a self-reported psychological scale measuring whether people took actions to achieve their goals. Furthermore, they do not affect lab measures of risk preference. These results suggest that deeper underlying preferences and cognitive functions may not respond strongly to light-touch interventions. In contrast, attention and self-efficacy appear more malleable.

Although we cannot completely rule out that treatment effects may reflect social desirability bias in answering questions related to the content of the intervention, it is unlikely that such effects account for our findings. Most importantly, we observe increases in objectively measured chlorine content of stored household drinking water during unannounced household visits. The visits were closely concentrated in time within each village (thus making them hard to anticipate), and the treatment effects are robust to controlling for within-village chlorine test delay. We also observe no treatment effects on self-reported scales capturing other socially desirable behaviors specifically mentioned in the workshops, such as scales measuring whether one has achieved one's goals in the past week.

Our paper makes several contributions to the existing literature. First, we show that addressing psychological barriers can lead to economically significant increases in the use of preventive health products, and these effects go beyond information alone. Our study builds on a small literature which shows that such light-touch interventions can increase long-term investments, including in health. The Visualisation intervention drew on theoretical and empirical studies in economics and psychology suggesting that future visualization techniques may be effective at increasing patient behavior (Gabaix and Laibson 2017; Alan and Ertac 2018; Duckworth et al. 2011), including smoking cessation (Stein et al. 2018) and savings (Hershfield et al. 2011). The Planning intervention was adapted from a low-intensity psychotherapy approach called Behavioral Activation (Lejuez et al. 2011; Ekers et al. 2014) and a behavioral intervention known as "implementation intentions" (Gollwitzer and Sheeran 2006; Duckworth et al. 2013; Morisano et al. 2010). Interventions which help people to execute plans have been shown to improve immunization (Milkman et al. 2011), take-up of preventive screening (Milkman et al. 2013), and exercise (Stadler, Oettingen, and Gollwitzer 2009).

[6]This finding is in line with Zwane et al. 2011, who find people chlorinate water more if they are surveyed 18 times compared to three times, making chlorination more salient.

Relatedly, reminders (which counter limited attention) improve compliance with medical treatment regimens (Pop-Eleches et al. 2011; Lester et al. 2010) and increase productivity (Hanna, Mullainathan, and Schwartzstein 2014; Beaman, Magruder, and Robinson 2014). Showing videos of role models increases aspirations, savings, educational as well as business investment (Bernard et al. 2018), and reduces prevalence of sexually transmitted disease (Banerjee, la Ferrara, and Orozco 2018). [7]

Another literature documents that limited information may affect economic decisions. People are known to increase investment in preventive health or in high-return opportunities, such as education, after receiving information about benefits and returns (Dupas 2011; Jensen 2010; Jensen 2012; Dinkelman and Martínez 2014). This is also true in the context of drinking water, where information that water is unsafe can lead to substitution to different water sources (Madajewicz et al. 2007; Jalan and Somanathan 2008). However, in other cases, information campaigns have not induced behavior change (Dupas 2009; Duflo, Dupas, and Kremer 2015). In our setting, providing information and changing beliefs alone does not result in behavior change. In contrast, psychologically targeted elements of these interventions, which improve self-efficacy and the salience of the targeted behavior, are necessary to achieve measurable improvements in outcomes.[8]

More narrowly, our paper contributes to research that aims to increase chlorination of drinking water, which has examined price mechanisms (Ashraf, Berry, and Shapiro 2010; Dupas et al. 2016; Kremer et al. 2011b), communal vs. individual persuasion (Kremer et al. 2011a), and combinations of information and subsidies (Ashraf, Jack, and Kamenica 2010; Null et al. 2018; Kremer et al. 2011a). Other researchers working on chlorination have noted that psychological constraints may affect choices about water treatment.[9] We show that psychological interventions have causal effects on behavior and provide evidence that they work through psychological channels. These findings complement other work on increasing demand for chlorination. Our interventions could therefore be deployed in combination with these other programs, to deliver clean water to the 2 billion

[7] In addition, more involved, multi-session interventions which target self-esteem and self-efficacy improve sex workers' savings and attendance at health checkups (Ghosal et al. 2016) and produce increases in employment McKelway (2018). Multisession interventions that resemble psychotherapy improve patience (Blattman, Jamison, and Sheridan 2017) and reduce automatic decision-making (Heller et al. 2017), reducing violence and antisocial behavior. With depressed mothers, they improve financial empowerment and investment in children (Baranov et al. 2017).

[8] Our results suggest a possible reinterpretation of past findings: some successful "information" interventions may work by combining pure information with psychological elements. For example, information about chlorination provided by promoters may also operate through psychological mechanisms (Kremer et al. 2011b; Kremer et al. 2011a). Indeed, messaging around free safe water products is particularly effective in increasing adoption when it is behaviorally inspired (Luoto et al. 2014).

[9] For example, in Zambia Ashraf, Berry, and Shapiro (2010) note that women often mentioned being "busy" or "distracted" after gathering water as a reason for buying chlorine but not using it. In Bangladesh, Luoto et al. (2011) note the importance of understanding behavioral barriers to use, after finding that none of four water treatment products had usage rates of above 30 percent, despite the provision of subsidies and extensive information.

people who currently use drinking water contaminated with feces.[10] They could also be adapted to encourage use of other preventive health products, as well as productive investments more generally.

The remainder of this paper is structured as follows. Section 2 describes the study design. Section 3 describes the interventions. Section 4 describes the outcome variables. Section 5 describes the estimation approach. Section 6 reports results. Section 7 discusses potential mechanisms. Section 8 concludes.

2. Experimental design

2.1 Sampling and randomization

Our study areas are Bungoma and Kakamega counties in rural Western Kenya. These counties were included in the WASH Benefits study (henceforth WASH) of Null et al. (2018). The WASH study was a cluster-randomized controlled trial, conducted from 2012 to 2014, which included 1,226 villages (Null et al. 2018). Villages were grouped into clusters of one to three neighboring villages. Clusters were then randomized to eight treatment arms, six of which tested household-level water, sanitation, handwashing, and nutrition interventions in isolation and different combinations of interventions.[11]

We sampled 205 villages from the full sample of 1,226 villages. In one treatment group of the WASH study, villages were provided with chlorine dispensers next to water sources, as well as community health promoters. We randomly selected 67 villages from this "Water Quality" (henceforth "dispenser") treatment arm and 67 villages from the "Passive Comparison" arm, which received no interventions.[12] In the dispenser villages, chlorine dispensers were installed at an average of five community water points per village cluster, and refilled as needed. The "Dispensers for Safe Water" program operated by the NGO *Evidence Action* has since maintained these dispensers, ensured they are filled with chlorine, and retained a local promoter in each community. The randomization has remained intact. An additional 71 villages were selected uniformly from the remaining six WASH Benefits arms in Mumias district, Kakamega county. The inclusion of participants from villages treated by the remaining six arms of the WASH intervention was not planned in our study, but occurred as a result of a coding error at the sampling stage. During the WASH Benefits study, 48 of these 71 villages received one of sanitation, handwashing or nutrition interventions, but no dispensers. However, these additional interventions all took place at the household level, and we exclude households which participated directly in the original WASH study from participation in

[10] https://www.who.int/news-room/fact-sheets/detail/drinking-water

[11] For more information on the WASH Benefits study, see Appendix B.

[12] There were two comparison groups in the original WASH study: an active comparison group, where children's arm circumference was measured, and a passive comparison group, where children were not measured but other outcomes were measured.

the present study. In addition, all interventions other than dispensers finished three to four years before our study. For these reasons, we group these 48 villages with the passive comparison villages in our analysis.[13] The other 23 villages received dispensers, as well as combinations of sanitation, handwashing or nutrition interventions. We group these villages with dispenser villages, as *Evidence Action* continued to maintain dispensers in these villages. As a robustness check, we exclude these 71 villages from the heterogeneity analysis by "Water Quality" assignment in Appendix F.

In the 205 study villages, we recruited 3750 women aged 18–35 between October 2017 and January 2018. We focus on women because they are primarily responsible for household chores, including collecting water, and thus for water chlorination. We recruit women aged 18–35 as they are most likely to have small children, who in turn are the most vulnerable to water-borne illnesses. With the help of local guides, enumerators visited all households in each included village (see Section B) and conducted a census to determine household eligibility. Enumerators collected demographic information on women that met the screening criteria: i) aged 18–35 inclusive, ii) within this age range, the most senior woman in their household, and iii) their household did not participate in the WASH Benefits study. The target sample of 3750 women represents all eligible women in the 205 villages. As shown in Table 1, the women in our sample are on average 26 years old, 89 percent are married or cohabiting, and they have on average 6 years of education.

We split our sample into three "active" treatment groups and one pure control group. We randomly assigned 992 participants to the "Visualization" (V+INF) group, 991 to the "Planning" (P+INF) group, 992 to the "Active Control" (AC+INF) group, and 775 to the "Pure Control" group. We stratified the randomization on village of residence, and a wealth index collected during the census.[14] Redundant observations during stratification were equally distributed across treatment groups.

Participants were assigned randomly to attend baseline and intervention sessions either in the morning or in the afternoon. While participants were encouraged to attend the session type assigned to them, they were allowed to switch to the other session time if necessary to minimize attrition. Within a geographical region and within each treatment group, participants were invited to sessions in alphabetical order, based on the first letter of their last name.

[13]Specifically, we exclude women who self-report having participated in the WASH study, and households with children aged either 3–4 or 4–5, depending upon the village's WASH Benefits timing. The reason for this second exclusion is that the WASH Benefits study recruited women in their second or third trimester of pregnancy in 2012. As a result, our sample is composed of women who were exposed to village-level, but not household-level interventions through the WASH Benefits study.

[14]The wealth index consisted of the total value of a limited set of assets (bicycles, cellphones, gas stoves, all livestock, radios, sofas and televisions). Participants were split at the median into a "high" and "low" wealth group.

2.2 Design and timeline

The project was approved by the Princeton (#7376) and KEMRI (#536) IRBs. Our pre-analysis plan was registered at https://www.socialscienceregistry.org/trials/2850/history/27566 (April 2, 2018). Participants in the three "active" treatment groups participated in three sessions: a combined baseline and first intervention session; a second intervention session one week later; and an endline 11 weeks after the first session. Participants in the pure control group did not participate in the baseline and the first and second intervention sessions, but participated in the endline 11 weeks after the beginning of the study in their village. In addition, households in all four groups received an unannounced visit 13 weeks after the beginning of the study, where enumerators collected a drinking water sample to be assayed for the presence of chlorine.

The baseline and endline sessions lasted about two hours each and consisted of behavioral tasks, psychological questionnaires, and a socioeconomic survey. These measures were collected in "mobile labs" operated by the Busara Center for Behavioral Economics in Bungoma and Kakamega counties, which each accommodated up to 25 participants at a time. The behavioral tasks and psychological questionnaires were presented using touch screen computers, audio instructions, color-coded response buttons using the zTree experimental interface (Fischbacher 2007), which allowed both computer-illiterate and entirely illiterate individuals to participate. Enumerators read instructions to the respondents in Kiswahili to maximize comprehension.[15] The socioeconomic survey was administered one-on-one by enumerators using tablets.

Each intervention session in the three "active" treatment groups lasted about two hours. Participants were split into cohorts of five for each session, which were run by locally-trained female facilitators. Participants were reconvened in the same cohorts for the second session. No participant was invited for the second session without having already participated in the first session.

Participants received KES 200 (USD 2) for participating in the baseline survey and first intervention session, KES 200 for the second session, and KES 300 for participating in the endline session.[16] They additionally received a KES 50 bonus for arriving on time for each appointment. Participants were reimbursed for their transport costs, using public transport rates from their village of residence to the mobile laboratory. Finally, participants received payments from the experimental tasks, as described in Section 4. Participants were informed of the fixed amounts that they would be paid during the phone call inviting them to a session, and told that they would have the opportunity to earn some extra money during the session. All participants recruited to the sample were invited to attend endline sessions, regardless of whether they attended the baseline

[15]Most Kenyans speak a tribal "mother tongue" at home, Kiswahili as a lingua franca, and English as the language of education and business. The Busara Center uses Kiswahili as the medium of oral communication in most studies with this population.

[16]USD 1 was equivalent to approximately KES 100 at the time of the study.

and/or intervention sessions. Table A.1 shows the number of participants at each stage of the study.

2.3 Background on chlorination use

Child diarrhea is relatively high in our study area, as it is in many parts of the developing world. In the original WASH study control group, diarrhea prevalence in the past 7 days was 27 percent among children aged 1 and 2 (Null et al. 2018); in our pure control group at endline, there have been 0.2 incidences of diarrhea per child under 15 in the last 3 months. Fecal contamination of drinking water is a likely cause of these episodes. Most of the population relies on communal water sources, usually wells with pumps or springs (Null et al. 2018). Women and children collect drinking water in plastic jerry cans, which is then kept in clay storage pots in the home. Water can be contaminated easily as it is removed from storage pots.

Chlorinating water kills many of the pathogens that cause diarrhea.[17] Absent point-of-collection chlorine dispensers (as installed by WASH), households can purchase dilute chlorine. The main brand of dilute chlorine is *WaterGuard*, which has been distributed and heavily marketed by the NGO *Population Services International* (PSI) in Kenya since 2003. *WaterGuard* is available in most local shops in the study area, and costs KES 25 (USD 0.25) per 150 ml bottle, about 0.2 percent of average monthly earnings. Each bottle treats 1000 l of water (approximately one month to 50 days of household drinking water) (Dupas et al. 2016; Lantagne et al. 2010) and comes with instructions in Swahili and in pictures. There is widespread awareness of this product: even at baseline of the WASH study in 2012, 89 percent had heard of *WaterGuard,* and 29 percent had used it at least once (Null et al. 2018). Households also believe chlorine is effective: at endline, the pure control group believe, on average, that 71 percent of incidences of diarrhea can be avoided by using chlorine.

However, take-up of chlorination is low. Even using chlorine from a free dispenser requires repeated, proactive behavior. Dispensers are not present at every water point, so people might need to plan their journeys to go past the correct water point. They might also need to make alternative arrangements if the dispenser is not filled, or remind children fetching water to chlorinate it. In non-dispenser villages, women would need to purchase *WaterGuard,* paying both time and monetary costs, and remember to add it to each batch of water. In our sample at endline, 27 percent of pure control households report having chlorinated their current drinking water supply, 18 percent had detectable free chlorine in their water, and 22 percent had detectable total chlorine.[18]

Taste is mentioned as a potential barrier to chlorination use (Dupas et al. 2016; Ashraf, Berry, and Shapiro 2010). Indeed, it is possible to taste if water has been chlorinated: in a blind taste test

[17] A meta-analysis finds point-of-use methods of chlorinating drinking water reduce child diarrhea by an average of 29 percent across 21 randomized controlled trials, suggesting unpurified water causes at least some episodes of diarrhea (Arnold and Colford Jr 2007).

[18] See Section 4.1 on the distinction between free and total chlorine. We report free chlorine here for comparison with other studies, but focus on total chlorine as our primary outcome measure.

among staff of the Busara Center for Behavioral Economics, on average participants correctly identified when drinking water had been treated with chlorine 75 percent of the time. However, we found no statistically significant differences in the taste rating participants gave to chlorinated vs. unchlorinated water (unchlorinated water mean score $=6.21/10$, chlorinated water mean score $=6.06/10$, $p=0.66$, $n=90$). While taste may prevent complete take-up of the product, taste differences are present to the same extent in all arms, so do not confound estimates of treatment effects.

3. Interventions

The structure of each session was held constant across treatment groups: each included a short lecture, followed by a story of a woman like them, reflections of how the themes relate to participants' own lives, and finally drawing and list-writing exercises and activities.

3.1 Treatment 1: Visualization + Information module ("V+INF")

The objective of the future visualization intervention was to increase internal motivation for patient behavior. The intervention is based on the idea that present utility is easier to imagine than future utility. A substantial body of evidence in psychology shows that people imagine future events in much less detail than immediately upcoming events, focusing on abstract qualities rather than details of execution (see e.g. Gilbert and Wilson 2009, Kahneman et al. 2004, Wilson et al. 2000). For instance, helping out an elderly relative with their tax return next month may be imagined as an act of love, while doing it later today is imagined as hours of painstaking sorting through receipts. In a recent theoretical contribution, Gabaix and Laibson (2017) formalize the idea of *as-if discounting*, which results from a perfectly patient decision-maker who simulates future utility by combining priors with noisy, unbiased signals about future utility. The model implies that interventions which improve forecasting ability (or forecasting efforts) will lead to more patient behavior, without changing the underlying preference parameters. Models of quasi-hyperbolic discounting (Laibson 1997 and O'Donoghue and Rabin 1999) are based on a differential weight given to the present. Despite fundamental differences, both models predict that an increased focus on the future will be associated with fewer dynamic preference reversals.[19]

The theoretical predictions are matched by empirical evidence: in a randomized educational intervention in Turkish primary schools, Alan and Ertac (2018) find that eight weekly classes on "imagining future selves" result in the children making more patient decisions in incentivized choice tasks *three years* after the intervention. Research in psychology had previously established

[19]Both "as-if discounting" à la Gabaix and Laibson (2017) and quasi-hyperbolic discounting generate dynamic preference reversals, but only quasi-hyperbolic discounting generates self-control problems and commitment demand. Since there is no commitment in our study, the models are observationally equivalent for our purposes.

that participants exposed to vivid imagery of their future selves increased hypothetical retirement savings in laboratory and online studies (e.g. Hershfield et al. 2011). Further evidence on the effects of future visualization techniques comes from the literatures on "best possible selves" (Markus and Nurius 1986; Sheldon and Lyubomirsky 2006; King 2001), "episodic future thinking" (Daniel, Stanton, and Epstein 2013), and "mental contrasting" (though mental contrasting is commonly coupled with planning elements, such as implementation intentions, see e.g., Oettingen, Pak, and Schnetter 2001, Oettingen, Mayer, and Thorpe 2010, and Duckworth et al. 2011).

The core of our intervention was to encourage participants to a) connect their present behavior to outcomes in the future, b) visualize alternative realizations of the future, depending on their current behavior, and c) put themselves in the shoes of their future selves, imagine how they feel, and 'talk' to them. The approach was deliberately visual and emotional, with participants asked to close their eyes repeatedly for several minutes, and to imagine their future selves in as much graphic detail as possible. We hypothesize that increased vividness of the future may affect behavior through a change in preference parameters, a change in forecasting skills, or a stronger mapping between present behavior and future outcomes.

In the first intervention session, participants were given an interactive lecture on thinking about the future. Participants were asked to think about the effect of small everyday behaviors (such as spending their leftover budget on snacks) on future outcomes. The intervention carefully avoided changing participants' beliefs about which present behavior would entail which future outcome – it merely encouraged them to make the connection themselves. It moved on to several silent visualization exercises, with prompts including "Close your eyes for one minute. Imagine the person you will be in one year. Imagine your family in one year." and" Imagine that your future self can now talk to you. How does she feel? What does *she* want you to do?" In the second part of the session, participants listened to a story of a woman whose daily life is full of tasks and worries, and who consequently focuses only on what is necessary right now. Using examples of water chlorination, antenatal care visits, handwashing, and saving, she learns over time that thinking about the future in her everyday actions helps her and her family to have a better life. The story was followed by a group discussion on behaviors from the participants' own lives, and visualizing alternative realizations of their own future, depending on present behavior. The session finished by asking participants to draw one these "future realizations", and make notes on the corresponding present behavior. Unlike the Planning intervention, it focused on high-level behaviors and outcomes without implementation details. The first intervention session concluded with the information module described below. In the second Visualization intervention session, participants largely revisited the first session, albeit with increased emphasis on how future visualizations can be used to deal with temptations in the present.

3.2 Treatment 2: Planning + Information module ("P+INF")

The Planning intervention aimed to help participants to undertake activities that they had identified as important but were struggling to do regularly, by making concrete, specific plans. We drew on two different bodies of work in psychology. One approach to treating mild depression, known as behavioral activation, draws on literature on instrumental reinforcement and motivation (Lejuez et al. 2011).[20] Depressed people often reduce how often they undertake activities and avoid even basic tasks, which contributes to negative mood To break this cycle, behavioral activation teaches participants simple, structured skills: to set short-term goals, make plans, reduce avoidance, and problem-solve (Ekers et al. 2014; Weobong et al. 2017).

A second literature on "implementation intentions" (Gollwitzer and Sheeran 2006; Locke and Latham 2002) suggests that people are more likely to implement plans if they make plans concrete and detailed and take time to anticipate potential obstacles and strategies to overcome them. Such interventions have improved classroom behavior and test scores with school students (Duckworth et al. 2013; Morisano et al. 2010), increased reported exercise (Stadler, Oettingen, and Gollwitzer 2009) and voter turnout (Nickerson and Rogers 2010), and improved take-up of vaccination and preventive screening(Milkman et al. 2011; Milkman et al. 2013). Completing a particular action plan (in this case, to exercise) is also associated with an increase in self-efficacy (Lorig et al. 2014).

The first session began with a short lecture about how all people can get stuck in cycles of inactivity, where they lack motivation to get things done and lose out on the sense of achievement of accomplishing what they intended to do. Participants then listened to a story of a woman very similar to them who was tired and struggling to complete her chores, including fetching water and chlorinating it. Once she fell out of the habit of doing routine tasks, she found it very hard to find the motivation to start again. This had negative consequences for herself and her family. Participants could then share experiences of similar situations from their lives

In the second part of the first session, participants were asked to identify activities in their daily lives that they felt were important, but where they were struggling to "get going". In contrast to the Visualization intervention, they did not set long-term goals, but focused on activities they could do in the next week. With a partner, they completed a simple worksheet with two lists of activities: one set that they enjoyed doing, and one set that were necessary and important but they might not enjoy. They ranked the tasks on each list from most to least difficult. In the final part of the session, participants picked the easiest one or two activities from each of the "necessary" and "enjoyable" lists of activities, and scheduled them in a weekly diary. They then broke the task down into steps, visualized what they would need to do to complete the activity, anticipated potential obstacles, and made plans to overcome them. The first intervention session concluded

[20]Importantly, we do not screen for or target people with depression symptoms, or attempt to provide any treatment for depression. We use manuals by Richards and Whyte (2011) and Lejuez et al. (2011).

with the information module described below. In the second intervention session, they worked in the same group with the same partner where possible. They crossed off completed plans and circled uncompleted ones, discussed barriers they had faced to undertaking the activities, and brainstormed ways to overcome these barriers in the future.

3.3 Treatment 3: Active Control exercise + Information Module ("AC+INF") and Pure Control ("PC")

The "active control" intervention controls for any effects of simply attending a session and interacting with women from neighboring villages. The sessions followed the format of the two treatment interventions. The content of these sessions centered on the birds and plants of Kenya, topics chosen intentionally to be psychologically inactive. In the first session, participants listened to a short lecture on different kinds of birds that live in Kenya, followed by a short story about the daily routine of a woman similar to them. Participants discussed the birds they see in their village, and any birds they were particularly interested in. They wrote a list of all the birds they could think of, and then made some drawings of birds. The second session followed an identical structure, except that it did not include another short story about a woman, in line with the second sessions of the other treatment groups. The second session centered on plants in Kenya. In addition, participants in this group also received the same information module as the Visualization and Planning treatment groups described above.

All three intervention groups concluded with an information module about the benefits of chlorination. Participants were read information on chlorination, as well as on antenatal and postnatal care (ANC/PNC). The text on chlorination read: : "Not all "drinking water" is safe, even when it is sold as "treated". Only water properly treated with chlorine or boiled water is safe to drink. If you drink unsafe water you and your family may get dangerous diseases like diarrhea, typhoid and cholera. Children are the most seriously affected by diarrhea: children can quickly become dehydrated and very ill, and having diarrhea a lot can stop children from growing at a normal rate. But you can avoid one out of three times you or your children gets diarrhea simply by chlorinating your water with a product like WaterGuard, or chlorine from a dispenser. Chlorinating in this way takes 30 minutes and makes it unnecessary to boil the water. One capful of WaterGuard makes 20 l of water clean. If you keep the water covered and in a closed container with a lid, and don't dip dirty cups back into the water, the drinking water can't get re-contaminated. The smell of chlorine is not harmful and reduces over time. Chlorine is much cheaper than firewood for boiling. At some water points it is even available for free from a dispenser."

The pure control group were recruited at the same time as other groups. They completed a brief demographic questionnaire household recruitment but received no contact prior to endline.

4. Outcome measures

4.1 Behavioral measures

Our main pre-specified outcome is an objective measure of whether households chlorinate their drinking water. Enumerators made unannounced visits to participants' homes to test the household's stored drinking water for the presence of chlorine to minimize experimenter demand effects in the survey (de Quidt, Haushofer, and Roth 2018).We pre-specified Total Chlorine Residual (TCR) as our main chlorination outcome measure of interest. . TCR indicates the presence of any chlorine in the water; i.e., that the household had at some point added *some amount* of chlorine to the drinking water, which is suitable if one's primary interest is whether households attempt to chlorinate their water at all. [21] We note that our measure is a lower bound on actual usage. Chlorine decays over time after being added to water, and may be undetectable in water after 24 hours (Null and Lantagne 2012). Households who store water for two to three days may thus appear not to have treated water.

We follow the same chlorine testing procedure as Null et al. (2018) and Kremer et al. (2011a). To conduct the tests, enumerators filled vials with a sample of stored household drinking water and added DPD chlorine reagent powder, separately for total chlorine and free chlorine. Using color comparator boxes and DPD color discs, enumerators recorded the level of chlorine present in the water sample, between 0 mg/l and 3.4 mg/l. The TCR variable is 1 if any total chlorine is present in the sample, and 0 otherwise. The FCR variable is 1 if any free chlorine is present in the sample, and zero otherwise (CDC 2010).

While chlorination is our primary outcome of interest, our interventions were intended to be domain-general. Visualizing future benefits, as well as planning skills, are relevant for many everyday behaviors, and in particular for future investments like savings, education, and agricultural investments. During the endline survey, participants therefore completed several modules on economic and health behaviors. We pre-specified secondary outcomes as the extensive margins for savings and labor supply, as well as an index of investment into children's education. We also specified exploratory outcomes related to health (diarrhea, vaccinations, and ante-natal care visits).

4.2 Psychological measures

Planning: First, we measure whether people choose to make plans and follow through on them in their everyday lives, with a short form of the Behavioral Activation for Depression Scale (Kanter et al. 2007; Manos, Kanter, and Luo 2011). Participants are asked to identify how much statements

[21] This is also the outcome of interest in Kremer et al. (2011a) and Dupas et al. (2016). Null et al. (2018) and Ashraf, Berry, and Shapiro (2010) focus on Free Chlorine Residual (FCR), which indicates that not only has chlorine been added, but that there is still enough unreacted chlorine in the water to keep it sanitized; i.e., that the household added *sufficient* chlorine to the drinking water.

were true for them in the past week, including both positive (e.g. "I was an active person and accomplished the goals I set out to do") and negative items (e.g. "There were certain things I needed to do that I didn't do"). Responses range from "not at all" (1) to "completely" (7). Some items are reversed before summing to generate a composite score. Second, we use a version of the *Tower of London task* (TOL; also known as the "Stockings of Cambridge" task when implemented electronically), which measures a participant's higher-order cognitive ability to plan ahead in sequential strategies (Shallice 1982; Phillips et al. 2001). Appendix D.1 provides further details.[22]

Time Preferences over Effort: Following recent innovations in the elicitation of time preferences (Andreoni and Sprenger 2012; Augenblick, Niederle, and Sprenger 2015), and noting that water chlorination is an effortful task, we estimate time preferences in the effort domain. We use the methodology of Augenblick (2017), and implement it with a newly developed effort task, adapted to a field setting without computer or smartphone access in a developing country: participants choose how many units of an effort task they want to complete at a time t for a piece rate w, where t is 0, 1, 7, or 8 days from today, and the piece rate w is KES 2, 6, or 10. Each effort task consisted of sending a 30-digit random number string by SMS to a toll-free number.[23] Participants thus decided, for each combination of t and w, how many such strings they wanted to send by SMS. One time and one piece rate are randomly implemented at the end, and enforced using a KES 100 completion bonus. Following Augenblick (2017) as well as DellaVigna and Pope (2017), we structurally identify a present bias parameter (β) and an impatience parameter (δ) assuming quasi-linear utility and a power cost of effort function. Appendix D.2 provides full details on the estimation.

In addition to the effort discounting task, we included a conventional Multiple Price List (MPL) (see Andreoni and Sprenger 2012 for a discussion), to measure monetary discounting. Payments ranged from KES 100 to KES 300, and delays included today, four weeks, and eight weeks. See Appendix D.3 for details.

Risk Preferences: We use a modified Eckel-Grossman measure which offered a choice between one of three 50/50 lotteries, represented as bets on a coin flip (Charness, Gneezy, and Imas 2013). We construct an ordinal measure of risk aversion with three levels, based on the expected payout the participant is willing to forgo for an increase in certainty of payout.

[22]Some of the psychological measures – the Tower of London task, Monetary Price Lists for time preferences, and the General Self-Efficacy Scale – were validated in a previous study (Esopo et al. 2018). This study involved translation and back-translation of the measures into Kiswahili, cognitive interviewing to establish cultural acceptability of the measures, and tests of internal consistency, test-retest reliability, and construct validity.

[23]Although we did not screen on phone access, all participants in our sample have access to a mobile phone: 71 percent own one, 96 percent have one in their household, and the remainder shares the phone of friends or relatives. Since phones are often used by multiple individuals, phone access should be understood as continuous rather than binary.

Salience Task: We measure the salience of three future-oriented behaviors (chlorination, savings, and farm investment) compared to non-future oriented behaviors. Enumerators read out three lists of nine words each to every participant, and asked her to recall as many words as possible directly after reading each list (participants were paid KES 5 for every word they remembered). Each list contained three categories of future-related words (chlorine, savings, and farm investment), as well as non-future related filler words (see Table A.7 for the list of words). While the recall of words is clearly driven by memory, the recall of words *conditional* on the total number of words remembered captures whether a concept is "top-of-mind". We thus test whether our treatments differentially affect the probability to recall chlorine words, conditional on the total number of words remembered. If our treatments differentially affect the salience of chlorine, we further test whether this is due to increased salience of future-oriented behaviors in general, by asking whether the differential treatment effect also holds for two other future-oriented behaviors (saving and farm investment), which were not emphasized in the sessions (see Appendix D for the empirical specification).

Beliefs and Knowledge: We measure participants' beliefs about the effectiveness of chlorination in preventing disease. In particular, we measure beliefs across treatment groups about the proportion of pediatric diarrhea cases which can be prevented by water chlorination. We also measure knowledge about how to properly use chlorine with two multiple-choice questions: 1. how much chlorine to add to water, and 2. the amount of time that needs to pass after adding chlorine for water to be safe to drink. We score each question as a binary measure of whether the participant answered correctly and create a composite value that ranges from 0 to 2, which is then z-scored.[24] The answer to all three questions was given to all active treatment groups at baseline in the Information treatment.

5. Econometric approach

5.1 Experimental integrity

We test for balance across treatment groups in (1) demographic variables and key outcomes, (2) timing of the surveys relative to the intervention, (3) participation in the endline survey and the chlorination test at home, and (4) compliance with the assigned treatment (i.e., participation in the intervention sessions). We regress baseline demographics available for the entire recruited sample (age, years of education, marital status, and village, see Section 2.1) on indicators for all treatment groups. The reference group is either the active control (AC+INF) or the pure

[24]We similarly check for differential knowledge of WHO-recommended antenatal and postnatal care. These questions were included to pilot them for a future study.

control group (PC).The specification is identical to that used for the estimation of treatment effects (described in Sections 5.2 and 5.3), leaving out controls $\mathbf{X}_i$and lags y_{i0}.

We further test for differences in the timing of the endline survey relative to the baseline survey and first intervention date, as well as the timing of the chlorine test relative to the baseline survey and first intervention date. The outcome variables are the number of days between the first intervention and the endline survey, and between the first intervention and the chlorine test. For participants in the pure control group, and those in the treatment groups who did not attend the interventions, we use predictive mean matching to simulate a proxy intervention date, based on the actual intervention dates of other participants from their village of residence.

We also test for selective attrition in attending the endline survey and the chlorination test at home, for both the active control group comparison and the pure control comparison. We define attrition in two ways: firstly, failure to attend endline conditional upon being recruited to the study; and secondly, failure to attend endline conditional upon completing the baseline and first intervention. The specification is identical to that used for the estimation of treatment effects. Additional checks assess whether attriting individuals are different from non-attriting individuals in terms of observed demographics.

Finally, although recruited participants did not know their treatment assignment prior to arriving for the first intervention session, we test for differential compliance across treatment groups – i.e., the decision to participate in the first and second intervention session. Again, the specification is identical to that used for the estimation of treatment effects, except that the outcome variable is an indicator for session attendance.

5.2 Main specification: Planning and Visualization treatments compared to Active Control

We employ the following main specification:

$$y_{i1} = \alpha_0 + \alpha_1 T_{1i} + \alpha_2 T_{2i} + \mathbf{\Phi}'\mathbf{X}_i + \delta y_{i0} + \gamma_v + \theta_w + \eta_i \tag{1}$$

Here, y_{i1} is the outcome of interest for respondent i at time of endline. T_{1i} and T_{2i} are dummies equal to one if the respondent is assigned to the "Visualization" or "Planning" group, respectively. $\mathbf{X}$ is a vector of participant controls (year of birth, employment status, marital status, education level). γ_v are village fixed effects, and θ_w is an indicator for household wealth greater than the sample median (which was used in stratification). Standard errors are clustered by intervention cohort (five participants) to account for within-group dynamics. Psychological scales and behavioral tasks were the only outcomes collected at baseline. For these variables we include y_{i0}, the same outcome variable at time of baseline, in the regressors. Where only some baseline

observations of a variable are missing, we replace the missing values with zero and add a dummy variable indicating such cases, following Jones (1996). We remove outliers by winsorizing outcome variables which have no theoretical lower and upper bounds at the 1st and 99th percentile.

The sample excludes the PC group. Thus, the AC+INF group is the reference category. As pre-specified, the sample includes all participants who completed both the endline survey and at least the first intervention session. Note that selection into the sample based on treatment is not possible because the nature of the intervention was not revealed before the first intervention session.

The comparison between the two "psychological" treatment groups and the AC+INF group isolates the effects of the psychologically active elements of these treatments. It tests the effect of interventions specifically targeting the ability to visualize the future or to make and execute plans, respectively, over and above the effect of the AC+INF intervention (which provided information about the benefits of chlorination and got participants to gather as a group).

5.3 Comparison with pure control group

We also report results from comparing the three active (V+INF, P+INF, and AC+INF) to a pure control group (PC). As pre-specified, the sample includes all recruited participants who completed the endline survey, including "non-compliers" who were assigned to the active treatment groups, but chose not to participate in the baseline survey or the interventions. The comparison between the three "active" treatment groups and the pure control group gives the policy-relevant effect: the total effect on targeted behaviors of providing interventions such as ours in other, similar settings.

The specification is identical to that in equation 1, except that there is a third treatment indicator T_{3i} for the active control group, and the pure control group is used as the reference category. Further, since the pure control group was not surveyed at baseline, the estimation does not control for the baseline outcome y_{i0}. Finally, because the pure control completed the endline survey on average two days earlier than the active treatment groups, the specification includes fixed effects for i) the week of endline survey and ii) the day of the week of endline survey.[25]

5.4 WASH Benefits cross-randomization

To study differential effects of our treatments in villages which do and do not have free and convenient access to chlorine through dispensers, we separately estimate the same pure control specification for dispenser and control villages using seemingly unrelated regression (SUR). This approach allows coefficients on control variables to vary by village treatment status, and at the

[25]The inclusion of these fixed effects was not pre-specified. We report results in Table A.20 and Table A.21 to demonstrate that our results are robust to excluding the fixed effects.

same time allows us to compare treatment coefficients across models. The primary outcome of interest is an indicator for objective chlorination (TCR).

Due to a coding error in sampling (see Section 2.1), some participants were drawn from treatment arms of the WASH Benefits study other than "Water Quality" or "Passive Comparison." As noted above, we include these participants in the main analysis of the cross-randomization, but exclude them in a robustness check in Appendix F.

5.5 Treatment-on-the-Treated (TOT) Effects

As pre-specified, we also report "treatment-on-the-treated" effects from an instrumental variables specification in which treatment status is instrumented with treatment assignment. For this purpose, we consider all those in any "active" group, including the active control group, who attended at least the initial baseline session to be complying with treatment assignment, even if they didn't complete the second session.[26] For the comparison of the V+INF and PF+INF to the AC+INF group, the sample is already restricted to compliers, so that the TOT effect is the same as the Average Treatment Effect (ATE). We therefore report the TOT only for the comparison of the three active treatments to the pure control group.

5.6 Multiple hypothesis testing (MHT) correction

We use a step-down procedure to adjust p-values for the false discovery rate (FDR) among groups of outcomes, and report the resulting "q-values" (Benjamini, Krieger, and Yekutieli 2006). We adjust for multiple hypothesis testing within outcome groups (psychological mechanisms and behaviors) and hierarchical categories (primary and secondary), but not across them. Similarly, we consider the effects of our two active interventions to be theoretically distinct and therefore do not correct across them. Indices are constructed following Anderson (2008).

6. Results

6.1 Experimental integrity

Table 1 provides results on baseline balance on demographic variables, timing of the endline surveys relative to the intervention, differential endline participation, and compliance with treatment. To test for baseline balance, we estimate a version of equation 1 with baseline demographics as the outcome variables. Each row shows baseline balance for one demographic variable. Columns (1)–(5)

[26] All substantive content was covered in the first session: the second session was merely a reinforcement and repetition of the first session content.

show the comparison of the Visualization (V+INF) and Planning (P+INF) treatments to the active control (AC+INF) treatment, and Columns (6)–(10) show the comparison of the V+INF, P+INF, and AC+INF groups to the pure control group. Columns (1) and (6) show the mean and standard deviation of the respective comparison groups. Columns (2) and (3) show the treatment effects for the V+INF and P+INF treatments, respectively, relative to the AC+INF treatment. Column (4) is a test of equality between these two coefficients, and Column (5) shows the sample size, which varies slightly across rows because some respondents did not answer a small number of questions, some questions are restricted to certain respondent groups, e.g. those with children. Our demographic variables are balanced across treatments on the whole, with only two out of 30 coefficients on pairwise comparisons reaching statistical significance, at the 10 percent level.

Table A.2 presents results from a test of equality of means of outcome measures collected at baseline between active treatment groups, including self-reported chlorination practices and knowledge of how to use chlorine. We see no statistically significant differences in outcome scores between the active treatment groups prior to the interventions. 65 percent of respondents reported adding chlorine to water at some point during the last month, and the variable is balanced across arms.[27]

The second panel in Table 1 shows results across treatment groups in terms of the number of days between the date of baseline and first treatment and the date of endline, and then between the date of baseline and first treatment to the date of the chlorine test at the household. Column (1) shows that the average delay between the beginning of the interventions and endline was 68 days, i.e. ten weeks, and the average delay to the chlorine test was 78 days. Relative to the AC+INF group, the P group on average completed the endline survey 2 days later, significant at the 10 percent level. Relative to the pure control group, we see that both the V+INF group and the P+INF group on average completed the endline survey 2 days later. Although this difference is small in magnitude, we include a fixed effect for i) the week of endline survey and ii) the day of the week of the endline survey to the specification described in Section 5.3 for all survey outcome measures.

The third panel in Table 1 shows results on participation in the endline survey as well as in the chlorination measure, for two alternate definitions of attrition. Where attrition is defined as failure to complete the endline among all those initially censused, average attrition in the AC+INF and pure control groups was 8 and 24 percent, respectively. This more conservative definition captures the extent to which our sample deviates from being representative of our population of interest. Average attrition from the chlorine measurement, conducted at people's houses, was 12 percent

[27]In contrast, only 22 percent of the pure control group at endline have chlorine present in their water. Self-reported chlorination may be inflated by social desirability bias. As discussed in Section , in some households chlorine may have been added but decayed. Alternatively, households may chlorinate, but irregularly, reflecting the exact behavioral barriers we seek to target. Dupas et al. (2016) also find that 3 months after a study providing free chlorine or vouchers, 73 percent of respondents had non-empty chlorine bottles in their homes, but only 34 percent had verified chlorine in their stored drinking water.

in the AC+INF group and 26 percent in the pure control group. Where attrition is defined as failure to complete the endline among those recruited into the study who attended baseline (the standard definition of attrition for most experiments), average attrition in the AC+INF group was 8 percent; it is 0 percent by definition for the pure control group. Under this definition, average attrition from the chlorine measurement was 8 percent in the AC+INF group and 0 percent by definition for the pure control group.

Using the more conservative definition, in comparison to the AC+INF group, we find no differential attrition from either the endline survey, or the chlorine measurement. However, in comparison to the pure control group, the V+INF and AC+INF treatment groups have a small but statistically significantly smaller likelihood of attriting (-6 percentage points).[28] In addition, both the V+INF and AC+INF groups are 4 percentage points less likely to attrit from the chlorine measure than the pure control group, although this effect is only significant at the 10 and 5 percent levels.

Importantly, Appendix Tables A.4 and A.5 show that we find little evidence that this differential attrition led to differences in sample composition that would complicate inference. Columns (3) and (4) show that demographic variables do predict attrition from either endline measurement, once treatment status is controlled for. However, the interaction terms between demographic variables and treatment status in Columns (5) and (6) show that participants with particular characteristics are no more or less likely to drop out of the study in any one of the treatment groups compared to the pure control group. This result suggests that the composition of the sample is similar in all treatment groups, including in the pure control group, and that differences in sample composition are unlikely to be responsible for observed differences between treatment groups.[29]

To account for attrition, we conduct two additional analyses of our results. First, in Appendix Tables A.14 and A.15, we re-estimate our main results while restricting the sample in the active treatment groups to those who attended the baseline.[30] Second, to account for the differential attrition, we include controls selected by LASSO in the specification, following Belloni, Chernozhukov, and Hansen (2014). We present these results in Appendix A.16 and A.17. Finally, we present results in Appendix A.18 and A.19 where we both limit the sample to those who attended baseline and include controls selected by LASSO. All of these robustness checks lead to very similar results as our main specification.

[28] This is unsurprising: in the active arms, we contact people again fairly soon after recruitment. In the pure control arms, we contact them for the endline for the first time roughly twelve weeks after recruitment.

[29] Relative to the AC+INF group, we do see that married women in the V+INF group are slightly less likely to attrit, significant at the 10 percent level. However, given the number of comparisons made in our attrition analysis, without correction for multiple hypothesis testing, we are not concerned about this result.

[30] Under this definition, where the sample includes only those who attend baseline, there is also no differential attrition across the active treatment arms. Attrition from endline conditional on attending baseline is, of course, higher in the active treatment groups than the pure control group, which by definition has 0 percent attrition since the pure control was not invited to baseline.

The final panel in Table 1 shows compliance rates across the treatment groups. After the census, all respondents in the treatment groups were invited to the baseline and first intervention session, which were held at the same time. 78 percent of respondents attended the baseline and completed the first session. Only respondents who attended the first intervention session were invited to the second session. 74 percent of respondents completed both sessions, while 4 percent did not complete the second session. Compliance is balanced across treatment groups.

6.2 Behavioral outcomes

We present three main sets of results: first, a comparison of the V+INF and P+INF treatments to the AC+INF group; second, a comparison of these three groups to the pure control group; and finally, separate analyses in villages with chlorine dispensers vs. villages without dispensers, and the corresponding interaction terms with our treatment groups. The raw means and standard deviations of all outcome measures that were subsequently z-scored are reported in Appendix Table A.8. Table 2 shows results on behavioral outcome variables. The different panels show impacts on different families of outcomes, namely health, savings, labor, and other outcomes. The table panels indicate the variables which were pre-specified as primary, secondary or exploratory in the pre-analysis plan.

Both the Visualization and the Planning interventions lead to significant increases in our primary behavioral outcome measure, the presence of chlorine in household drinking water, as measured with an objective chlorine test during an unannounced household visit. In comparison to the pure control group, the V+INF group shows a 6 percentage point increase in the presence of total chlorine, significant at the 1 percent level. The P+INF group shows a 4 percentage point increase in the presence of total chlorine relative to the pure control group, significant at the 10 percent level.[31] The treatment effects in the V+INF and P+INF groups relative to pure control correspond to 27 and 18 percent increases relative to the pure control group mean of 22 percent. The treatment effect in the AC+INF group relative to the pure control group is small (3 percentage points) and not statistically significant. In line with this result, chlorination is significantly higher in the V+INF group (5 percentage points, significant at the 1 percent level) relative to the AC+INF group; the effect is smaller and not significant in the P+INF group relative to the AC+INF group (2 percentage points).

The results for free chlorine (FCR, see Section 4.1), which we had not pre-specified as a primary outcome, are similar in magnitude and statistical significance.

Note that our objective chlorination measures, obtained during unannounced household visits, leave little room for experimenter demand effects, such as social desirability bias in responding

[31] A self-report question on whether households treated their water in any way to make it safe to drink generated 99 percent affirmative responses in all groups, likely owing to experimenter demand effects; we therefore do not show this outcome in the tables. However, this result suggests that all treatments were equally informative about the objectives of the experimenter (de Quidt, Haushofer, and Roth 2018), leading to similar experimenter demand effects across all groups.

to survey questions. It is possible that households talked among themselves and anticipated our visits. However, the visits were concentrated in time within each village: 68 percent of chlorine tests were conducted on the first day that our team visited a given village (the remaining tests were spread out, with a median within-village range of 7 days). Appendix Table A.12 shows that the estimated treatment effects on chlorine are unaffected by including (i) a dummy for being tested on the first day within one's village, and (ii) the number of days elapsed since the first chlorination tests were conducted in the village.

Importantly, given our goal of improving health outcomes through psychological interventions, the V+INF and P+INF treatments both generate large and statistically significant reductions in the incidence of diarrhea among children. We find significant reductions in diarrhea episodes in both of these groups relative to the pure control group, with a 35 percent (8 percentage point) reduction for V+INF and a 26 percent (6 percentage point) reduction for P+INF. The effects of the V+INF and P+INF interventions are also significant relative to the AC+INF group (V+INF: 38 percent (9 percentage point) reduction, significant at the 1 percent level using both naïve and FDR-corrected p-values; P+INF: 25 percent (6 percentage point) reduction, significant at the 5 percent level using conventional p-values, but not significant using FDR-correction). We cannot reject that the V+INF and P+INF treatments are equally effective in reducing diarrhea. We find no strong effects on other health outcomes, with the exception of a small change in the number of children under the age of 15 who visited a healthcare provider in the last 3 months, potentially due to reduced diarrhea incidence.

The treatment effects of our interventions on diarrhea are rather large in relation to the effects on chlorination; with an effect on diarrhea of -35 percent and an effect on the presence of any chlorine of $+27$ percent, a back-of-the-envelope calculation suggests an instrumental variable estimate of -1.3 percent, i.e. a 1 percent increase in chlorination leads to a 1.3 percent reduction in diarrhea. This estimate compares favorable to that of a recent meta-analysis, which finds across six randomized controlled trials that a 1 percent increase of detectable free chlorine results in a 1 percent reduction in child diarrhea (Arnold and Colford Jr 2007). There are a few possible reasons for our slightly larger effects. First, we measure increases in total chlorine rather than the presence of residual chlorine. Second, our interventions are domain-general trainings that may affect multiple behaviors related to child diarrhea, such as handwashing or open defecation, while dispensers are likely to only affect chlorination. Third, there may be dependencies with the treatment effects on non-health outcomes: for instance, the observed increases in savings (described below) may allow for more consumption smoothing and thereby more advantageous health outcomes.

The second panel of Table 2 shows effects on savings-related outcomes. The V+INF treatment leads to a significant increase in our main savings-related outcome variable, the amount of money saved regularly. This effect corresponds to a 26 percent (KES 25) increase relative to the active control group, significant at the 5 percent level; relative to the pure control group, it is smaller

(20 percent; KES 18) and not statistically significant. In addition, we find a large and highly significant effect on savings on the extensive margin, with the share of respondents who save regularly increasing by 13 percentage points (36 percent) in the V+INF group relative to AC+INF, and 12 percentage points (38 percent) relative to the pure control group. Similarly, we find increases in the V+INF treatment on an indicator for whether the respondent saves for productive investments, 11 and 9 percentage points (65 and 53 percent) relative to the AC+INF and pure control groups, respectively. All of these effects are significant at the 1 percent level even after correcting for multiple comparisons. V+INF participants are also more likely to have joined a new ROSCA relative to AC+INF participants. The P+INF and AC+INF interventions did not show significant effects on savings-related outcomes, and most of the effects in the V+INF group described above are significantly larger than in those groups. Together, these results show that the V+INF treatment strongly affected savings-related behaviors.

This finding has two important implications. First, it shows that our V+INF intervention, while unsuccessful in affecting time preferences measured with laboratory-like discounting tasks (see Section 7.1), nevertheless strongly affected future-oriented behaviors. A possible explanation comes from the theoretical work of Gabaix and Laibson (2017): If the visualization intervention helped decision-makers to generate less noisy signals about the utility they will derive from future events, then more patient behavior would result even in absence of changes in the underlying discounting parameters.[32] Second, it provides evidence against an experimenter demand-effect account of the impacts of our interventions: while the interventions mentioned chlorination and health-related topics, savings were discussed only to the extent that participants brought them up themselves in the discussions. The treatment effects we report here are thus more likely to result from changes in the underlying preferences rather than from a simple desire to please the experimenters.

The third panel of Table 2 reports effects on labor-related outcomes. Somewhat surprisingly, we find a reduction in the total number of hours worked in the last 3 months in the P+INF group, with a magnitude of 24 hours (22 percent) relative both to the AC+INF group and to the pure control group. These effects are significant at the 1 percent and 5 percent level, respectively. A similar effect is observed in the total number of days worked, which is reduced by 4 days or 18 percent in the P+INF group, relative both to the AC+INF group and to the pure control group, significant at the 5 percent level, although the result does not survive correction for multiple hypothesis testing. One possible reason for this reduction is again that participants in the P+INF

[32]The V+INF intervention prompted participants to visualize alternative realizations of their future in one year – a time horizon where future utility is likely to feel vague and distant. Intuitively speaking, time preferences are estimated from participants' relative willingness to exert effort in 0 versus 1 days (for $\beta\delta$) and 7 versus 8 days (for δ, see Section 4.2). If the disutility from effort within the next 8 days is already salient and tangible at baseline, then the estimated β and δ parameters would not capture an "improved forecasting ability" in the V+INF group. In contrast, an improved simulation of utility in one year may have motivated participants to increase their savings in the present.

group may have improved their planning and execution ability and therefore be more efficient in accomplishing tasks. In support of this hypothesis, we find no significant change in average monthly earnings, despite the shortened work hours. The V+INF and AC+INF treatments do not show significant effects on labor outcomes.

We next compare these results to those obtained using a treatment-on-the-treated estimator. Results are shown in Table A.10. Note that, as explained above, the TOT estimator is defined only for the comparison to the pure control group because the sample in the AC+INF comparison is restricted to compliers at baseline and intervention. Again, the point estimates and levels of significance are similar, with the exception of the effects of P+INF on our secondary chlorination measure and diarrhea outcomes, where the TOT results do not survive correction for multiple hypothesis testing. Finally, in Appendix Tables A.14, A.16, and A.18, we re-estimate our results to account for differential endline participation, as described in Section 6.1. In Appendix Table A.20 we present our results excluding the endline week and day of week fixed effects described in Section 5.3, since the inclusion of these fixed effects was not pre-specified. We find no major differences in these results, except a stronger effect of the P+INF intervention relative to the pure control on our secondary chlorine measure when we exclude those in active treatment groups who did not attend the baseline and interventions.

6.3 Are psychological treatments more effective when cost barriers are removed?

We next ask whether our treatment effects differ by whether or not the village in which the interventions took place was randomly treated with a chlorine dispenser in the WASH Benefits study that preceded ours. Appendix Table A.3 presents results showing that, at baseline, participants from villages with chlorine dispensers were no different in observables from participants from villages without chlorine dispensers, except that participants from villages with dispensers have a slightly higher level of education than those in villages without dispensers. Table 3 shows results from the two main estimating equations, focusing on chlorination-related outcomes, separately for WASH control villages (Columns (1)–(5)) and WASH treatment villages (Columns (6)–(10)). The interaction terms on our two treatment groups with the WASH treatment (i.e. chlorine dispensers) are shown in Columns (11)–(13) for V+INF, P+INF, and AC+INF, respectively.

We find weak evidence of heterogeneity in the effects of the treatments on outcomes related to chlorination by village type. Specifically, the active treatment groups P+INF, V+INF, and AC+INF show significant treatment effects on water chlorination in the dispenser villages, but not in the villages without dispensers. The interaction term is significant at the 5 percent level for the AC+INF treatment, suggesting that information about the benefits of chlorination is

more effective in villages with a dispenser compared to those without. The other interaction terms are not statistically significant.[33]Together, these results present suggestive evidence that our psychological interventions, in particular the V+INF treatment, had somewhat larger effects in dispenser compared to non-dispenser villages. However, these effects are small and require replication before being taken seriously.

7. Mechanisms

7.1 Results for psychological variables

To elucidate potential psychological mechanisms which may account for our behavioral effects, we now turn to the results on our pre-specified psychological outcomes. Table 4 shows results on the psychological outcome variables, estimated using equation 1. The top panels show results on outcomes related to planning, time preferences, and self-efficacy, subdivided into primary and secondary outcomes according to our pre-analysis plan. Adjustment of p-values for multiple comparisons using FDR is done separately for primary and secondary outcomes in each family of variables. The final panel shows variables measuring beliefs and knowledge.

For potential mechanisms, the main comparison of interest is that of the "psychologically active" treatment groups to the AC+INF control group. The information conveyed by the treatments, as well as exposure to field staff and other participants, is held constant across these treatments. The exclusion of the non-compliers at baseline and the first intervention session implies that each individual in the sample received at least one treatment session. This comparison is also more precisely estimated than the comparison to the pure control group due to the inclusion of the outcome at baseline as a control variable on the right-hand side for some outcome variables. Finally, the comparison to active control is not confounded by a possible practice effect from having completed the tasks at baseline.

We find no statistically significant effects of any treatment on planning compared to the Active Control, although the Planning treatment leads to a non-significant 0.04 SD increase in the Behavioral Activation Score relative to the AC+INF mean, and a non-significant 0.04 SD reduction in the number of moves required to complete the Tower of London task. In comparison to the pure control group we observe a 0.11 SD increase in the Behavioral Activation Score,

[33]In Appendix Table A.13, we test the robustness of these results when restricting the villages to those from the Water Quality and Passive Comparison arms of the WASH Benefits study, as discussed in Appendix Section B. This analysis permits a cleaner distinction between dispensers and control, in the absence of any other interventions, at the expense of a somewhat smaller sample. We find that in this analysis, the interaction term for both chlorination measures is significant for the V+INF intervention, suggesting that this intervention had significantly larger effects on chlorination in dispenser villages compared to non-dispenser villages.

significant at the 10 percent level.[34]

The next panel reports results on outcomes related to time preferences. Our main outcome is the β parameter from our effort task, measuring present bias in the quasi-hyperbolic model of Laibson (1997); additional outcomes are the δ parameter from the same task, and corresponding parameters from the monetary discounting task. We find no statistically significant effects that survive multiple hypothesis testing for any of these outcomes, except a very small increase in δ^{money} in the V+INF and AC+INF groups, significant at the 10 percent level. Somewhat surprisingly, we find a reduction of the δ^{Effort} parameter in the P+INF group, driven by an increased willingness to supply effort 7 and 8 days in the future, at constant willingness to supply effort today or tomorrow (see Figure A.3). This result is consistent with the P intervention helping participants to "plan" future effort, considering aspects like phone access or child care. However, the effect is economically small (-0.002), and does not survive multiple inference correction.

The third panel of Table 4 shows the effect of our interventions on the General Self-Efficacy (GSE) scale, a measure of self-efficacy. Both the V+INF and the P+INF interventions generate statistically significant 0.16 SD and 0.12 SD increases, respectively, relative to the AC+INF group. This finding suggests that our interventions affected "inward-looking" beliefs about one's ability to achieve desired outcomes. We find similarly large (0.18 SD) and significant differences between the V+INF and P+INF groups and the pure control group on this outcome. The comparison between the AC+INF treatment and the pure control group reveals a very small difference that is not statistically significant.

We again compare these results to those obtained using a treatment-on-the-treated estimator. Results are shown in Appendix Table A.11. We find similar point estimates, and the same significance levels, with the exception of even larger and significant impacts of the V+INF and P+INF treatments on self-efficacy relative to the pure control group. We find no major differences when we re-estimate results accounting for attrition (as described in Section 6.1) or excluding endline week and day of week fixed effects (described in Section 5.3), except a stronger effect of the V+INF and P+INF treatments (0.21 SD and 0.19 SD, respectively) on self-efficacy when we exclude those in active treatment groups who did not attend the baseline and interventions. Results are shown in Section F.

7.2 Results for beliefs and knowledge

The last panel of Table 4 suggests that the information treatment had a similar effect across the active treatment groups: V+INF, P+INF, and AC+INF all significantly increase beliefs in the

[34]Performance on the Tower of London task is significantly improved in all active treatment groups relative to the pure control group, but this result is likely due to the fact that all active groups had performed this task previously at baseline, while the pure control group had not.

efficacy of chlorine in averting diarrhea relative to the pure control group, with similar effect sizes across the three interventions (0.13 SD, 0.16 SD, and 0.12 SD). Effects on knowledge about how to correctly chlorinate water are also statistically indistinguishable across the three active treatment groups. Compared to the pure control group mean, the V+INF group shows a significant increase in knowledge about chlorination of 0.12 SD. Effects are similar in magnitude for the other groups (0.07 SD for P+INF and 0.09 SD for AC+INF), but do not reach statistical significance, and we cannot reject that effects in all groups are of the same magnitude.[35] Importantly, we see effects on beliefs and knowledge in all three active arms, but we only see statistically significant effects on chlorination and diarrhea in the V+INF and P+INF arms. Thus, the Visualization and Planning treatments have additional effects on behavior compared to the effect of information on its own. These are likely to be due to changes in psychological variables, in particular self-efficacy (discussed above) and salience (discussed below), which are affected in both the V+INF and P+INF groups.

Finally, as expected and outlined in our PAP, we find no effects on risk aversion, suggesting that any behavioral effects are unlikely to result from changes in risk preferences induced by our treatments.

7.3 Results for salience

Section 4.2 explains the design of a test for increased salience of chlorination, and Appendix D explains the econometric specification. Column (1) of Table 5 shows that participants who had received the V or P intervention found it easier to remember chlorine-related words, conditional on the total number of words remembered. This effect was specific to chlorine-related words, rather than to future-oriented behaviors in general: Column (3) interacts treatment indicators with chlorine word indicators, and shows that the salience of chlorination increases differentially to the salience of other future-oriented words (the base category is farm investment). Finally, Column (4) suggests that neither treatment group had an effect on total words remembered, and thus on participants' memory in general.

Thus, we find an increase in the salience of chlorine-related words, but not savings-related words, as a result of treatment. Taken together, it is thus plausible that salience was a contributing factor to the observed treatment effect on chlorination. However, our treatment effect on savings, as well as on various other non-chlorine measures, is difficult to explain through salience, as the salience of savings was unaffected by treatment. Salience effects thus do not provide a consistent explanation across our findings, unless the mapping from salience to behavior is both non-linear

[35] We also find that all three interventions are effective in increasing knowledge of the correct schedule of antenatal and postnatal care visits a woman should attend before and after giving birth. We do not see an associated increase in the number of ANC visits attended by pregnant women in our sample, but our sample size for this analysis is less than 300 individuals. These questions were included for piloting purposes for a future study.

and differential across domains.

8. Conclusion

In this randomized field experiment, we study the effect of two light-touch interventions on behavioral and psychological outcomes among young women in Kenya. Specifically, we ask whether a "Visualization" and a "Planning" intervention increase take-up of drinking water chlorination. Both interventions increase chlorination of drinking water, and reduce the number of diarrhea episodes in children. Possible mechanisms include an increase in self-efficacy, i.e. the beliefs that one has the capacity to achieve desirable outcomes; and increased salience of chlorination of drinking water. We rule out mechanisms related to increased information or changed beliefs about the effectiveness of chlorination. We also find little evidence of changes in laboratory measures of time preferences or present bias, or planning. The fact that we found no effects on deeper underlying preferences and cognitive functions suggests that the latter may not respond strongly to light-touch interventions. In contrast, attention and self-efficacy appear more malleable.

We estimate high cost-effectiveness of the interventions; in terms of DALYs saved, our interventions are comparable to highly effective interventions such as the provision of insecticide-treated bed nets and deworming. Importantly, we find impacts even in villages which have free chlorine dispensers at water sources, suggesting that these interventions are effective over and above existing infrastructural interventions. Future work may attempt to replicate these effects, and extend the use of our interventions to other settings and behaviors of interest.

References

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy* 126 (5): 1865–1911.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: a Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

Andreoni, James, and Charles Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *American Economic Review* 102 (7): 3333–3356.

Arnold, Benjamin, and John Colford Jr. 2007. "Treating Water with Chlorine at Point-Of-Use to Improve Water Quality and Reduce Child Diarrhea in Developing Countries: A Systematic Review and Meta-Analysis." *The American Journal of Tropical Medicine and Hygiene* 76 (2): 354–364.

Ashraf, Nava, James Berry, and Jesse M Shapiro. 2010. "Can higher prices stimulate product use? Evidence from a field experiment in Zambia." *American Economic Review* 100 (5): 2383–2413.

Ashraf, Nava, Kelsey Jack, and Emir Kamenica. 2010. "Information and Subsidies: Complements or Substitutes?" *Journal of Economic Behavior and Organization* 88:133–139.

Augenblick, Ned. 2017. "Short-Term Time Discounting of Unpleasant Tasks." *Working Paper, UC Berkeley.*

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working Over Time: Dynamic Inconsistency in Real Effort Tasks." *The Quarterly Journal of Economics* 130 (3): 1067–1115.

Bai, Liang, Benjamin Handel, Edward Miguel, and Gautam Rao. 2017. "Self-Control and Demand for Preventive Health: Evidence from Hypertension in India." *NBER Working Paper No. 23727*.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2016. "Worms at Work: Long-Run Impacts of a Child Health Investment." *The Quarterly Journal of Economics* 131 (4): 1637–1680.

Bandura, Albert. 1977. "Self-Efficacy: Toward a Unifying Theory of Behavioral Change." *Psychological Review* 84 (2): 191–215.

Banerjee, Abhijit, Eliana la Ferrara, and Victor Orozco. 2018. "The Entertaining Way to Behavioral Change: Fighting HIV with MTV." *Working Paper, Massachusetts Institute of Technology*.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2017. "Mental Health and Women's Choices. Experimental Evidence from a Randomized Control Trial." *Working Paper, University of Melbourne*.

Beaman, Lori, Jeremy Magruder, and Jonathan Robinson. 2014. "Minding Small Change Among Small Firms in Kenya." *Journal of Development Economics* 108:69–86.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29–50.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli. 2006. "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate." *Biometrika* 93 (3): 491–507.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, and Alemayehu Seyoum Taffesse. 2018. "The Future in Mind: Short and Long-Run Impact of an Aspirations Intervention in Rural Ethiopia." *Working Paper, University of Oxford*.

Blattman, Christopher, Julian Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206.

Bleakley, Hoyt. 2010. "Malaria Eradication in the Americas: A Retrospective Analysis of Childhood Exposure." *American Economic Journal: Applied Economics* 2 (2): 1–45.

Bobonis, Gustavo, Edward Miguel, and Charu Puri-Sharma. 2006. "Anemia and School Participation." *Journal of Human Resources* 41 (4): 692–721.

CDC. 2010. *Chlorine Residual Testing Fact Sheet*. Centers for Disease Control and Prevention SWS Project.

Charness, Gary, Uri Gneezy, and Alex Imas. 2013. "Experimental Methods: Eliciting Risk Preferences." *Journal of Economic Behavior & Organization* 87:43–51.

Cohen, Jessica, and Pascaline Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." *The Quarterly Journal of Economics* 125 (1): 1–45.

Daniel, Tinuke, Christina Stanton, and Leonard Epstein. 2013. "The Future Is Now: Reducing Impulsivity And Energy Intake Using Episodic Future Thinking." *Psychological Science* 24 (11): 2339–2342.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3): 694–719.

DellaVigna, Stefano, and Devin Pope. 2017. "What Motivates Effort? Evidence and Expert Forecasts." *The Review of Economic Studies* 85 (2): 1029–1069.

de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2018. "Measuring and Bounding Experimenter Demand." *American Economic Review* 108 (11): 3266–3302.

Dillon, Andrew, Jed Friedman, and Pieter Serneels. 2014. "Health Information, Treatment, and Worker Productivity: Experimental Evidence from Malaria Testing and Treatment among Nigerian Sugarcane Cutters." *Policy Research Working Paper Series 7120, The World Bank*.

Dinkelman, Taryn, and Claudia Martínez. 2014. "Investing in Schooling in Chile: The Role of Information about Financial Aid for Higher Education." *Review of Economics and Statistics* 96 (2): 244–257.

Duckworth, Angela, Heidi Grant, Benjamin Loew, Gabriele Oettingen, and Peter Gollwitzer. 2011. "Self-Regulation Strategies Improve Self-Discipline in Adolescents: Benefits of Mental Contrasting and Implementation Intentions." *Educational Psychology* 31 (1): 17–26.

Duckworth, Angela, Teri Kirby, Anton Gollwitzer, and Gabriele Oettingen. 2013. "From Fantasy to Action: Mental Contrasting with Implementation Intentions (MCII) Improves Academic Performance in Children." *Social Psychological and Personality Science* 4 (6): 745–753.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." *American Economic Review* 105 (9): 2257–97.

Dupas, Pascaline. 2009. "What Matters (and What Does Not) in Households' Decision to Invest in Malaria Prevention?" *American Economic Review* 99 (2): 224–230.

———. 2011. "Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya." *American Economic Journal: Applied Economics* 3 (1): 1–36.

———. 2014a. "Getting Essential Health Products to Their End Users: Subsidize, but How Much?" *Science* 345 (6202): 1279–1281.

———. 2014b. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica* 82 (1): 197–228.

Dupas, Pascaline, Vivian Hoffmann, Michael Kremer, and Alix Peterson Zwane. 2016. "Targeting Health Subsidies through a Nonprice Mechanism: A Randomized Controlled Trial in Kenya." *Science* 353 (6302): 889–895.

Ekers, David, Lisa Webster, Annemieke Van Straten, Pim Cuijpers, David Richards, and Simon Gilbody. 2014. "Behavioural Activation for Depression; an Update of Meta-Analysis of Effectiveness and Sub Group Analysis." *PloS one* 9, no. 6.

Esopo, Kristina, Daniel Mellow, Catherine Thomas, Hannah Uckat, Justin Abraham, Prachi Jain, Chaning Jang, Nicholas Otis, Michala Riis-Vestergaard, Amanda Starcev, Kate Orkin, and Johannes Haushofer. 2018. "Measuring Self-Efficacy, Executive Function, and Temporal Discounting in Kenya." *Behaviour Research and Therapy* 101:30–45.

Fink, Günther, and Felix Masiye. 2015. "Health and Agricultural Productivity: Evidence from Zambia." *Journal of Health Economics* 42:151–164.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–178.

Gabaix, Xavier, and David Laibson. 2017. "Myopia and Discounting." *NBER Working Paper No. 23254.*

Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy. 2016. "Sex Workers, Self-Image and Stigma: Evidence from Kolkata Brothels." *Working Paper, University of Glasgow.*

Gilbert, Daniel, and Timothy Wilson. 2009. "Why the Brain Talks to Itself: Sources of Error in Emotional Prediction." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364 (1521): 1335–1341.

Giné, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics* 2 (4): 213–35.

Gollwitzer, Peter M, and Paschal Sheeran. 2006. "Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes." *Advances in Experimental Social Psychology* 38:69–119.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning Through Noticing: Theory and Evidence From a Field Experiment." *The Quarterly Journal of Economics* 129 (3): 1311–1353.

Heller, Sara, Anuj Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold Pollack. 2017. "Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago." *The Quarterly Journal of Economics* 132 (1): 1–54.

Hershfield, Hal, Daniel Goldstein, William Sharpe, Jesse Fox, Leo Yeykelis, Laura Carstensen, and Jeremy Bailenson. 2011. "Increasing Saving Behavior Through Age-Progressed Renderings of the Future Self." *Journal of Marketing Research* 48 (SPL): S23–S37.

Jalan, Jyotsna, and E. Somanathan. 2008. "The importance of being informed: experimental evidence on demand for environmental quality." *Journal of Development Economics* 87 (1): 14–28.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *The Quarterly Journal of Economics* 125 (2): 515–548.

———. 2012. "Do Labor Market Opportunities Affect Young Women's Work and Family Decisions? Experimental Evidence from India." *The Quarterly Journal of Economics* 127 (2): 753–792.

Jones, Michael. 1996. "Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression." *Journal of the American Statistical Association* 91 (433): 222–230.

Kahneman, Daniel, Alan Krueger, David Schkade, Norbert Schwarz, and Arthur Stone. 2004. "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method." *Science* 306 (5702): 1776–1780.

Kanter, Jonathan, Patrick Mulick, Andrew Busch, Kristoffer Berlin, and Christopher Martell. 2007. "The Behavioral Activation for Depression Scale (BADS): Psychometric Properties and Factor Structure." *Journal of Psychopathology and Behavioral Assessment* 29 (3): 191.

King, Laura. 2001. "The Health Benefits of Writing About Life Goals." *Personality and Social Psychology Bulletin* 27 (7): 798–807.

Kremer, Michael, and Edward Miguel. 2007. "The Illusion of Sustainability." *The Quarterly Journal of Economics* 122 (3): 1007–65.

Kremer, Michael, Edward Miguel, Sendhil Mullainathan, Clair Null, and Alix Peterson Zwane. 2011a. "Social Engineering: Evidence from a Suite of Take-Up Experiments in Kenya." *Working Paper, Harvard University.*

Kremer, Michael, Edward Miguel, Clair Null, and Alix Peterson Zwane. 2011b. "Sustainability of Long-Term Take-Up at Point-Of-Collection Chlorine Dispensers Provided Free of Charge in Rural Western Kenya." *Proceedings of the Water Environment Federation* 2011 (3): 249–250.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics* 112 (2): 443–478.

Lantagne, D, K Preston, E Blanton, N Kotlarz, H Gezagehn, E Van Dusen, J Berens, and K Jellison. 2010. "Hypochlorite solution expiration and stability in household water treatment in developing countries." *Journal of Environmental Engineering* 137 (2): 131–136.

Lejuez, Carl, Derek Hopko, Ron Acierno, Stacey Daughters, and Sherry Pagoto. 2011. "Ten Year Revision of the Brief Behavioral Activation Treatment for Depression: Revised Treatment Manual." *Behavior Modification* 35 (2): 111–161.

Lester, Richard, Paul Ritvo, Edward Mills, Antony Kariri, Sarah Karanja, Michael Chung, William Jack, James Habyarimana, Mohsen Sadatsafavi, Mehdi Najafzadeh, Carlo Marra, Benson Estambale, and Elizab Ngugi. 2010. "Effects of a Mobile Phone Short Message Service on Antiretroviral Treatment adherence in Kenya (WelTel Kenya1): A Randomised Trial." *The Lancet* 376 (9755): 1838–1845 (November).

Locke, Edwin, and Gary Latham. 2002. "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-year Odyssey." *American Psychologist* 57 (9): 705.

Lorig, Kate, Diana D Laurent, Kathryn Plant, Eswar Krishnan, and Philip L Ritter. 2014. "The Components of Action Planning and Their Associations With Behavior and Health Outcomes." *Chronic Illness* 10 (1): 50–59.

Luoto, Jill, David Levine, Jeff Albert, and Stephen Luby. 2014. "Nudging to Use: Achieving Safe Water Behaviors in Kenya and Bangladesh." *Journal of Development Economics* 110:13–21.

Luoto, Jill, Nusrat Najnin, Minhaj Mahmud, Jeff Albert, M Sirajul Islam, Stephen Luby, Leanne Unicomb, and David I Levine. 2011. "What Point-of-Use Water Treatment Products Do Consumers Use? Evidence from a Randomized Controlled Trial among the Urban Poor in Bangladesh." *PloS one* 6 (10): e26132.

Madajewicz, Malgosia, Alexander Pfaff, Alexander Van Geen, Joseph Graziano, Iftikhar Hussein, Hasina Momotaj, Roksana Sylvi, and Habibul Ahsan. 2007. "Can Information Alone Change Behavior? Response to Arsenic Contamination of Groundwater in Bangladesh." *Journal of Development Economics* 84 (2): 731–754.

Manos, Rachel, Jonathan Kanter, and Wen Luo. 2011. "The Behavioral Activation for Depression Scale–Short Form: Development and Validation." *Behavior Therapy* 42 (4): 726–739.

Markus, Hazel, and Paula Nurius. 1986. "Possible Selves." *American Psychologist* 41 (9): 954.

McKelway, Madeline. 2018. "Women's Self-Efficacy and Women's Employment: Experimental Evidence from India." *Working Paper, Massachusetts Institute of Technology.*

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159–217.

Milkman, Katherine, John Beshears, James Choi, David Laibson, and Brigitte Madrian. 2011. "Using Implementation Intentions Prompts to Enhance Influenza Vaccination Rates." *Proceedings of the National Academy of Sciences* 108 (26): 10415–10420.

———. 2013. "Planning Prompts as a Means of Increasing Preventive Screening Rates." *Preventive Medicine* 56 (1): 92–93.

Morisano, Dominique, Jacob Hirsh, Jordan Peterson, Robert Pihl, and Bruce Shore. 2010. "Setting, Elaborating, and Reflecting on Personal Goals Improves Academic Performance." *Journal of Applied Psychology* 95 (2): 255–264.

Nickerson, David, and Todd Rogers. 2010. "Do You Have a Voting Plan? Implementation Intentions, Voter Turnout, and Organic Plan Making." *Psychological Science* 21 (2): 194–199.

Null, Clair, and Daniele Lantagne. 2012. "Microbiological quality of chlorinated water after storage in ceramic pots." *Journal of Water, Sanitation and Hygiene for Development* 2 (4): 250–253.

Null, Clair, Christine Stewart, Amy Pickering, Holly Dentz, Benjamin Arnold, Charles Arnold, Jade Benjamin-Chung, Thomas Clasen, Kathryn Dewey, Lia Fernald, et al. 2018. "Effects of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on Diarrhoea and Child Growth in Rural Kenya: A Cluster-Randomised Controlled Trial." *The Lancet Global Health* 6 (3): e316–e329.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now Or Later." *American Economic Review* 89 (1): 103–124.

Oettingen, Gabriele, Doris Mayer, and Jennifer Thorpe. 2010. "Self-Regulation of Commitment to Reduce Cigarette Consumption: Mental Contrasting of Future With Reality." *Psychology and Health* 25 (8): 961–977.

Oettingen, Gabriele, Hyeon-ju Pak, and Karoline Schnetter. 2001. "Self-Regulation of Goal-Setting: Turning Free Fantasies About the Future Into Binding Goals." *Journal of Personality and Social Psychology* 80 (5): 736.

Ozier, Owen. 2018. "Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming." *American Economic Journal: Applied Economics* 10 (3): 235–62.

Phillips, Louise, Victor Wynn, Selwyn-Lloyd McPherson, and Ken Gilhooly. 2001. "Mental Planning and the Tower of London Task." *The Quarterly Journal of Experimental Psychology Section A* 54 (2): 579–597.

Pop-Eleches, Cristian, Harsha Thirumurthy, James Habyarimana, Joshua Graff Zivin, Markus Goldstein, Damien de Walque, Leslie Mackeen, Jessica Haberer, John Sidle, and Duncan Ngare. 2011. "Mobile Phone Technologies Improve Adherence to Antiretroviral Treatment in Resource-Limited Settings: A Randomized Controlled Trial of Text Message Reminders." *AIDS* 25 (6) (6): 825–34.

Richard, Stephanie, Robert Black, Robert Gilman, Richard Guerrant, Gagandeep Kang, Claudio Lanata, Kåre Mølbak, Zeba Rasmussen, R. Bradley Sack, Palle Valentiner-Branth, et al. 2013. "Diarrhea in Early Childhood: Short-Term Association with Weight and Long-Term Association with Length." *American Journal of Epidemiology* 178 (7): 1129–1138.

Richards, David, and Mark Whyte. 2011. *Reach Out: National Programme Student Materials to Support the Delivery of Training for Psychological Wellbeing Practitioners Delivering Low Intensity Interventions.* Rethink Mental Illness.

Royer, Heather, Mark Stehr, and Justin Sydnor. 2015. "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company." *American Economic Journal: Applied Economics* 7 (3): 51–84.

Schilbach, Frank. 2018. "Alcohol and Self-Control: A Field Experiment in India." *American Economic Review (Forthcoming).*

Shallice, Tim. 1982. "Specific Impairments of Planning." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 298 (1089): 199–209.

Sheldon, Kennon, and Sonja Lyubomirsky. 2006. "How to Increase and Sustain Positive Emotion: The Effects of Expressing Gratitude and Visualizing Best Possible Selves." *The Journal of Positive Psychology* 1 (2): 73–82.

Stadler, Gertraud, Gabriele Oettingen, and Peter Gollwitzer. 2009. "Physical Activity in Women: Effects of a Self-Regulation Intervention." *American Journal of Preventive Medicine* 36 (1): 29–34.

Stein, Jeffrey, Allison Tegge, Jamie Turner, and Warren Bickel. 2018. "Episodic Future Thinking Reduces Delay Discounting and Cigarette Demand: An Investigation of the Good-Subject Effect." *Journal of Behavioral Medicine* 41 (2): 269–276.

Troeger, Christopher, Danny V Colombara, Puja C Rao, Ibrahim A Khalil, Alexandria Brown, Thomas G Brewer, Richard L Guerrant, Eric R Houpt, Karen L Kotloff, Kavita Misra, William A Petri, and Jam Platts-Mills. 2018. "Global Disability-Adjusted Life-Year Estimates of Long-Term Health Burden and Undernutrition Attributable to Diarrhoeal Diseases in Children Younger than 5 Years." *The Lancet Global Health* 6 (3): e255–e269 (March).

Walker, Christa, Igor Rudan, Li Liu, Harish Nair, Evropi Theodoratou, Zulfiqar Bhutta, Katherine O'Brien, Harry Campbell, and Robert Black. 2013. "Global Burden of Childhood Pneumonia and Diarrhoea." *The Lancet* 381 (9875): 1405–1416.

Wang, Haidong, Mohsen Naghavi, Christine Allen, Ryan Barber, Zulfiqar Bhutta, Austin Carter, Daniel Casey, Fiona Charlson, Alan Chen, Matthew Coates, et al. 2016. "Global, Regional, and National Life Expectancy, All-Cause Mortality, and Cause-Specific Mortality for 249 Causes of Death, 1980–2015: A Systematic Analysis for the Global Burden of Disease Study 2015." *The Lancet* 388 (10053): 1459–1544.

Weobong, Benedict, Helen A Weiss, David McDaid, Daisy R Singla, Steven D Hollon, Abhijit Nadkarni, A-La Park, Bhargav Bhat, Basavraj Katti, Arpita Anand, et al. 2017. "Sustained Effectiveness and Cost-Effectiveness of the Healthy Activity Programme, a Brief Psychological Treatment for Depression Delivered by Lay Counsellors in Primary Care: 12-month Follow-Up of a Randomised Controlled Trial." *PLoS Medicine* 14, no. 9.

Wilson, Timothy, Thalia Wheatley, Jonathan Meyers, Daniel Gilbert, and Danny Axsom. 2000. "Focalism: A Source of Durability Bias in Affective Forecasting." *Journal of Personality and Social Psychology* 78 (5): 821–836.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean Karlan, Richard Hornbeck, Xavier Giné, et al. 2011. "Being Surveyed Can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences* 108 (5): 1821–1826.

Table 1: Experimental integrity

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Baseline balance										
Age	26.73	−0.06	−0.01	0.82	2116	26.88	−0.36	−0.43	−0.33	2984
	(4.50)	(0.23)	(0.24)			(4.63)	(0.25)	(0.26)*	(0.25)	
Married or cohabiting	0.89	0.00	0.00	0.88	2116	0.91	−0.02	−0.02	−0.03	2984
	(0.31)	(0.02)	(0.02)			(0.28)	(0.02)	(0.02)	(0.02)*	
Education level	5.85	−0.01	0.07	0.19	2116	5.92	−0.06	0.03	−0.03	2984
	(1.20)	(0.06)	(0.06)			(1.10)	(0.06)	(0.06)	(0.06)	
High wealth index	0.55	−0.01	−0.02	0.75	2116	0.53	0.01	−0.00	0.02	2984
	(0.50)	(0.03)	(0.03)			(0.50)	(0.03)	(0.03)	(0.03)	
Village of residence	89.08	−0.06	0.50	0.93	2116	85.46	1.49	1.73	1.16	2984
	(54.74)	(6.30)	(6.30)			(57.21)	(4.79)	(4.84)	(4.55)	
Delay variables										
Days between endline and baseline	67.73	0.62	1.80	0.23	2116	68.73	1.82	2.12	1.23	2984
	(20.63)	(1.00)	(0.93)*			(24.07)	(1.06)*	(1.01)**	(1.00)	
Days between chlorine test and baseline	78.11	0.57	2.52	0.16	1868	81.12	0.70	1.71	0.08	2646
	(26.65)	(1.37)	(1.31)*			(27.54)	(1.26)	(1.21)	(1.19)	
Chlorine test was conducted on the first day of village's tests	0.68	0.01	−0.00	0.69	2009	0.67	−0.01	−0.01	−0.01	2839
	(0.47)	(0.02)	(0.02)			(0.47)	(0.02)	(0.02)	(0.02)	
Endline participation										
Did not participate in endline	0.08	0.02	0.03	0.39	2337	0.24	−0.06	−0.04	−0.06	3750
	(0.27)	(0.02)	(0.02)*			(0.43)	(0.02)***	(0.02)*	(0.02)***	
Did not participate in endline, conditional on attending baseline	0.08	0.02	0.03	0.39	2337	0.00	0.07	0.08	0.06	3750
	(0.27)	(0.02)	(0.02)*			(0.05)	(0.01)***	(0.01)***	(0.01)***	
Did not participate in chlorine test	0.12	0.01	0.03	0.33	2337	0.26	−0.04	−0.02	−0.04	3750
	(0.33)	(0.02)	(0.02)			(0.44)	(0.02)*	(0.02)	(0.02)**	
Did not participate in chlorine test, conditional on attending baseline	0.12	0.01	0.03	0.33	2337	0.00	0.10	0.12	0.09	3750
	(0.33)	(0.02)	(0.02)			(0.05)	(0.01)***	(0.01)***	(0.01)***	
Compliance										
Completed both first and second intervention	0.74	0.01	−0.02	0.35	2975	–	–	–	–	-
	(0.44)	(0.02)	(0.02)			–	–	–	–	-
Completed first intervention	0.78	0.01	0.01	0.85	2975	–	–	–	–	-
	(0.41)	(0.02)	(0.02)			–	–	–	–	-
Completed no intervention	0.22	−0.01	−0.01	0.85	2975	–	–	–	–	-
	(0.41)	(0.02)	(0.02)			–	–	–	–	-

Notes: OLS estimates of balance across treatment groups. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Delay, attrition, and compliance specifications control for a vector of observed characteristics; baseline balance specifications do not. All specifications cluster standard errors at the level of intervention cohort. The "Baseline balance" and "Delay variables" samples are restricted to those who attended endline, and for the AC+INF comparison additionally restricted to those who completed baseline and the first intervention session, since these are the samples used for the primary analyses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Baseline balance variables were collected at point of recruitment into the study, approximately four weeks before the first intervention. 'High wealth index' denotes participants who scored above the median in a measure of the value of their holdings of a limited list of common household assets. Delay variables measure the number of days elapsed from when an individual participated in the first intervention session until i) the endline measures in the laboratory, and ii) the test in their home for the presence of chlorine in stored drinking water. The third delay variable indicates the proportion of study participants who had their drinking water tested for chlorine on the first day that the testing team came to their village, rather than during a subsequent visit. Endline participation variables show the proportion of the relevant sample who did not participate in i) the endline in the laboratory, and ii) the chlorine test at home. For the AC+INF comparison, this sample is restricted to those who completed baseline and the first intervention session, since this is the sample used for the primary analysis. Participants only received a chlorine test once they had completed endline in the laboratory, and so attrition from endline results in attrition from the chlorine test. Compliance variables show the proportion of the sample in active treatment groups who i) completed both of the two intervention sessions, ii) completed only the first and did not return for the second, iii) failed to complete any of the interventions. A comparison is not shown relative to the pure control group since that group was not invited to intervention sessions.

Table 2: Behavioral outcomes

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Health outcomes										
Pre-specified primary										
Objective measure: water has been treated with chlorine (TCR)	0.23	0.05	0.02	0.15	2012	0.22	0.06	0.04	0.03	2839
	(0.42)	(0.02)**	(0.02)			(0.42)	(0.02)***	(0.02)*	(0.02)	
Pre-specified exploratory										
Objective measure: water has sufficient chlorine to be safe (FCR)	0.21	0.04	0.01	0.16	2012	0.18	0.06	0.03	0.03	2839
	(0.40)	(0.02)**	(0.02)			(0.39)	(0.02)***	(0.02)	(0.02)	
		[0.07]*	[1.00]				[0.01]**	[0.81]	[1.00]	
Number of diarrhea incidences per child u15 in last 3 months	0.24	−0.09	−0.06	0.30	2012	0.23	−0.08	−0.06	0.00	2823
	(0.56)	(0.03)***	(0.03)**			(0.58)	(0.03)**	(0.03)**	(0.03)	
		[0.00]***	[0.15]				[0.02]**	[0.20]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.00	−0.02	0.23	1999	0.22	0.02	−0.02	0.01	2800
	(0.35)	(0.02)	(0.02)			(0.36)	(0.02)	(0.02)	(0.02)	
		[0.58]	[1.00]				[0.65]	[1.00]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.26	−0.22	0.04	0.45	200	1.19	−0.30	−0.07	0.15	272
	(1.19)	(0.41)	(0.35)			(1.17)	(0.40)	(0.42)	(0.45)	
		[0.50]	[1.00]				[0.65]	[1.00]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.21	−0.04	−0.03	0.43	2004	0.17	0.01	0.02	0.04	2806
	(0.34)	(0.02)**	(0.02)			(0.31)	(0.02)	(0.02)	(0.02)**	
		[0.06]*	[0.77]				[0.91]	[1.00]	[0.35]	
Savings outcomes										
Pre-specified secondary										
Amount saved regularly (per week, KES)	93.96	24.89	3.28	0.10	2108	88.76	17.57	7.38	3.87	2972
	(230.26)	(12.37)**	(12.51)			(228.12)	(12.67)	(12.86)	(11.90)	
		[0.16]	[1.00]				[0.42]	[0.23]	[0.99]	
Pre-specified exploratory										
Indicator: Amount saved regularly is positive	0.36	0.13	−0.02	0.00***	2108	0.32	0.12	0.01	0.04	2972
	(0.48)	(0.03)***	(0.03)			(0.47)	(0.03)***	(0.03)	(0.03)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Number of new ROSCAs joined in last 3 months	0.17	0.04	0.01	0.17	2108	0.21	0.02	−0.01	−0.02	2972
	(0.44)	(0.03)*	(0.02)			(0.46)	(0.03)	(0.03)	(0.03)	
		[0.10]	[1.00]				[0.65]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.11	−0.01	0.00***	2108	0.17	0.09	0.00	0.01	2972
	(0.38)	(0.02)***	(0.02)			(0.38)	(0.02)***	(0.02)	(0.02)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Labor outcomes										
Pre-specified secondary										
Total hours of work in last 3 months	106.11	−6.79	−23.83	0.06*	2108	108.78	−2.33	−23.59	−7.14	2972
	(174.61)	(9.54)	(8.99)***			(182.99)	(10.13)	(9.46)**	(9.64)	
		[0.90]	[0.03]**				[0.42]	[0.04]**	[0.85]	
Pre-specified exploratory										
Total days of work in last 3 months	21.22	−0.59	−3.91	0.04**	2108	21.73	−0.52	−3.94	−1.41	2972
	(30.09)	(1.64)	(1.60)**			(30.45)	(1.70)	(1.61)**	(1.64)	
		[0.56]	[0.15]				[0.91]	[0.17]	[1.00]	
Average monthly earnings in last 3 months	1094.50	3.11	−1.23	0.97	2108	1167.22	−43.58	−105.44	−142.78	2972
	(2865.35)	(147.23)	(163.19)			(3155.76)	(166.41)	(171.46)	(174.39)	
		[0.65]	[1.00]				[0.91]	[1.00]	[1.00]	
Other behavioral outcomes										
Pre-specified secondary										
Index of investment in children's education (z-score)	0.00	−0.02	0.01	0.60	1420	0.00	0.09	0.12	0.12	1967
	(1.00)	(0.06)	(0.07)			(1.00)	(0.07)	(0.07)*	(0.07)*	
		[0.90]	[1.00]				[0.42]	[0.10]	[0.36]	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session (all variables except the objective chlorine measures). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Objective measures of chlorine come from data collected in participants' homes, where stored household water was tested for the presence of Total and Free Chlorine Residual, which indicate whether chlorine has been added to water (TCR), and whether sufficient chlorine is in the water to make it safe to drink (FCR). The number of diarrhea incidences reports how many independent episodes of diarrhea each child under-15 had on average in the last three months. The index of investment in children's education consists of measures of money spent on children's education and average school days attended in the last 3 months, and is constructed only for those women with school-age children.

Table 3: Chlorine-related outcomes in dispenser vs. non-dispenser villages

	Village has no chlorine dispenser					Village has chlorine dispenser					Comparison		
	(1) Pure Control Mean (SD)	(2) V+INF Treatment Effect	(3) P+INF Treatment Effect	(4) AC+INF Treatment Effect	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N	(11) V+INF Interaction p-value	(12) P+INF Interaction p-value	(13) AC+INF Interaction p-value
Objective measure: water has been treated with chlorine (TCR)	0.23	0.04	0.02	−0.01	1534	0.21	0.08	0.05	0.07	1305	[0.31]	[0.42]	[0.04]**
	(0.42)	(0.03)	(0.03)	(0.03)		(0.41)	(0.03)**	(0.03)*	(0.03)**				
Objective measure: water has sufficient chlorine to be safe (FCR)	0.17	0.05	0.02	0.00	1534	0.20	0.06	0.03	0.05	1305	[0.81]	[0.80]	[0.24]
	(0.37)	(0.03)*	(0.03)	(0.03)		(0.40)	(0.03)*	(0.03)	(0.03)				
Number of diarrhea incidences per child u15 in last 3 months	0.23	−0.07	−0.06	−0.02	1493	0.17	−0.05	−0.03	0.04	1313	[0.65]	[0.68]	[0.39]
	(0.56)	(0.04)*	(0.04)	(0.04)		(0.45)	(0.03)	(0.04)	(0.04)				
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.16	0.17	0.08	1573	0.00	0.08	0.13	0.16	1382	[0.37]	[0.60]	[0.56]
	(1.00)	(0.07)**	(0.07)**	(0.08)		(1.00)	(0.08)	(0.08)	(0.08)*				
Chlorine knowledge score (z-score)	0.00	0.16	0.06	0.05	1573	0.00	0.05	0.06	0.12	1382	[0.29]	[1.00]	[0.57]
	(1.00)	(0.08)**	(0.08)	(0.08)		(1.00)	(0.08)	(0.08)	(0.08)				

Notes: OLS estimates of treatment effects. The outcome variables repeat those in Tables 1–3, but the table reports the analysis separately for villages which have at least one chlorine dispenser maintained at a village water source, and for villages which do not have chlorine dispensers. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. The specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session (all variables except the objective chlorine measures). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Columns (11)–(13) report the p-values on the differential effect of the treatments in villages with vs. without chlorine dispensers using SUR.

Table 4: Psychological outcomes

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Planning										
Pre-specified primary										
Behavioral Activation Score (BADS) (z-score)	0.00	−0.01	0.04	0.31	2103	0.00	0.03	0.11	0.04	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)	(0.06)*	(0.06)	
		[0.83]	[0.91]				[1.00]	[0.13]	[1.00]	
Pre-specified secondary										
Tower of London: Total Moves (z-score)	0.00	0.01	−0.04	0.28	2103	0.00	−0.38	−0.43	−0.37	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.05)***	(0.05)***	
		[0.85]	[0.50]				[0.00]***	[0.00]***	[0.00]***	
Time preferences										
Pre-specified primary										
β^{Effort}	0.982	0.007	0.005	0.33	2068	0.953	0.007	0.012	0.009	2906
	(0.005)	(0.006)	(0.007)			(0.020)	(0.018)	(0.018)	(0.018)	
		[0.83]	[0.91]				[1.00]	[0.32]	[1.00]	
Pre-specified secondary										
β^{MPL}	1.05	−0.02	0.02	0.21	2103	1.02	−0.01	0.02	0.01	2955
	(0.46)	(0.02)	(0.03)			(0.43)	(0.03)	(0.03)	(0.02)	
		[0.41]	[0.53]				[0.50]	[0.25]	[0.45]	
δ^{MPL}	0.98	−0.00	−0.00	0.86	2103	0.98	0.00	0.00	0.00	2955
	(0.02)	(0.00)	(0.00)			(0.02)	(0.00)*	(0.00)	(0.00)**	
		[0.41]	[0.40]				[0.08]*	[0.14]	[0.07]*	
δ^{Effort}	0.999	−0.001	−0.002	0.16	2068	0.995	−0.001	−0.002	0.003	2906
	(0.001)	(0.001)	(0.001)**			(0.002)	(0.002)	(0.002)	(0.002)	
		[0.26]	[0.06]*				[0.50]	[0.25]	[0.23]	
Risk Aversion Measure (z-score)	0.00	−0.03	−0.07	0.49	1926	0.00	0.01	−0.01	0.03	2735
	(1.00)	(0.06)	(0.06)			(1.00)	(0.06)	(0.06)	(0.06)	
Self-efficacy										
Pre-specified secondary										
General Self-Efficacy Score (GSE) (z-score)	0.00	0.16	0.12	0.38	2103	0.00	0.18	0.18	0.02	2955
	(1.00)	(0.05)***	(0.05)**			(1.00)	(0.06)***	(0.06)***	(0.06)	
		[0.00]***	[0.06]*				[0.01]**	[0.00]***	[0.45]	
Beliefs and knowledge										
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.08	0.06	0.61	2103	0.00	0.13	0.16	0.12	2955
	(1.00)	(0.05)	(0.06)			(1.00)	(0.05)**	(0.05)***	(0.06)**	
Chlorine knowledge score (z-score)	0.00	0.06	−0.01	0.20	2103	0.00	0.12	0.07	0.09	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)**	(0.06)	(0.06)	
ANC/PNC knowledge score (z-score)	0.00	0.04	−0.05	0.08*	2103	0.00	0.36	0.28	0.35	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)***	(0.05)***	(0.05)***	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. The BADS score measures a participant's ability and motivation to follow through on plans they make in their life. Tower of London is a lab game that measures a participant's ability to plan ahead. Time preferences over effort are estimated structurally from a real effort task involving data entry by SMSs. Time preferences over money are measured using Multiple Price Lists (MPL). The risk aversion measure is derived from responses to a coin flip for various monetary rewards. The General Self-Efficacy score measures a participant's belief in their own ability to achieve the outcomes they desire. The mechanisms variables check if the interventions differentially affected i) a participant's belief in the efficacy of chlorine to prevent diarrhea, and ii) their knowledge of how to correctly chlorinate water, to assess whether changes in beliefs or knowledge could have been the cause of changes in chlorination behavior.

Table 5: Salience & memory test outcomes

	(1) Chlorine word remembered	(2) Saving word remembered	(3) Future word remembered	(4) Total words remembered
Visualization	0.08	0.00	-0.01	-0.37
	(0.01)***	(0.01)	(0.01)	(0.23)
Planning	0.05	0.02	-0.01	-0.03
	(0.01)***	(0.01)	(0.01)	(0.24)
Active Control	0.02	0.01	-0.03	-0.12
	(0.01)	(0.01)	(0.01)*	(0.22)
V x Chlorine Word Interaction			0.09	
			(0.02)***	
P x Chlorine Word Interaction			0.06	
			(0.02)**	
AC x Chlorine Word Interaction			0.05	
			(0.02)*	
V x Saving Word Interaction			0.02	
			(0.02)	
P x Saving Word Interaction			0.03	
			(0.02)	
AC x Saving Word Interaction			0.04	
			(0.02)*	
Constant	5.15	2.27	3.36	-31.30
	(2.24)*	(2.09)	(1.09)**	(34.04)

Notes: The table reports the probability in the salience test of remembering a chlorine-related word, or a savings-related word. The OLS specifications control for the total number of words the participant remembered in each list and include a 'chlorine word' fixed effect. The coefficients of the interaction terms show the differential probability of remembering a 'future-related' word if that word is related to chlorine or saving. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Appendix

A. Cost-effectiveness per disability-adjusted life year saved

To compute the disability-adjusted life years saved with our interventions, we proceed as follows. First, the cost of our interventions was USD 4 per household, or USD 1.33 per child because households have three children on average. Troeger et al. (2018) estimate the DALYs lost per child in Kenya due to diarrhea in 2016, the year our study was conducted, to be 0.127. Our region is one of the poorer ones in Kenya and has worth health outcomes than others, so using this figure will lead to a conservative estimate. We next assume that the treatment effects of our interventions on diarrhea relative to the pure control group (-35 percent for visualization and -26 percent for planning) were never higher during the study period of 3 months than what we measured at the 3-month endline, and then immediately went to zero. This, too, is a conservative assumption and will lead to a lower-bound estimate. Together, these figures suggest that DALYs saved per child due to the visualization intervention are $0.127 \times 35\% \times \frac{3}{12} = 0.0111125$, and of the planning intervention 0.008255. Combined with an intervention cost of USD 1.33 per child for both interventions, this implies that the cost of one DALY saved is $\frac{\text{USD } 1.33}{0.0111125} = \text{USD}120$ for the visualization intervention and $\frac{\text{USD}1.33}{0.008255} = \text{USD}161$ for the planning intervention.

B. Description of WASH Benefits Kenya

The WASH Benefits Kenya study is a cluster-randomized controlled trial testing the effects of six "water, sanitation and handwashing" (WASH) interventions on childhood development (Null et al. 2018). Villages were eligible if they were rural, most of the population relied on communal water sources and had unimproved sanitation facilities, and there were no other ongoing water, sanitation, handwashing, or nutrition programs. Between 2012 and 2014, 8,246 pregnant women were enrolled from three counties in western Kenya: Bungoma, Kakamega and Vihiga. The six interventions were (W) improved water quality ("Water Quality"); (S) improved sanitation; (H) handwashing with soap; (WSH) combined water, sanitation and handwashing; (N) improved nutrition; (WSH+N) combined water, sanitation, handwashing and nutrition. The study also included two control arms: (A) an active control arm, who received monthly visits to measure children's arm circumferences; and (P) a passive comparison arm, who received no visits. The villages in our sample primarily came from either (W) improved water quality or (P) passive comparison arm.

In all Villages in Arm W, "Water Quality", chlorine dispensers were installed at public water sources used by study participants. All community members were able to use the dispensers free

of charge. After filling water carrying containers, usually a 20 l jerrycan, users turn the knob on the dispenser to add 3 ml of 1.25 percent sodium hypochlorite (chlorine), which yields 2.5 ml/l of free chlorine residual after 30 mins for 20 l of water (Kremer et al. 2011a). This is sufficient chlorine to sanitize the water for drinking. Community promoters encouraged use of the chlorine dispensers, which to this date are monitored and maintained by *Evidence Action's* "Dispensers for Safe Water" program. Sample households additionally received a six-month supply of bottled chlorine every six months, to be used for sanitizing water at home, in case the household drank harvested rainwater or chose not to use the water source with the installed dispenser.

About 20 percent of our sample was recruited uniformly from 71 villages in the remaining six WASH Benefits treatment arms in Mumias district, Kakamega county. The inclusion of participants from villages treated by the remaining six arms of the WASH intervention was not planned but occurred as a result of a coding error at the sampling stage. During the WASH Benefits study, 48 of these 71 villages received one of sanitation, handwashing or nutrition interventions, but no dispensers. We group these 48 villages with the passive comparison villages in our analysis. First, these additional interventions all took place at the household level. We exclude households which participated directly in the original WASH study from participation in the present study. Second, all interventions other than dispensers finished three to four years before our study. The other 23 villages received dispensers, as well as combinations of sanitation, handwashing or nutrition interventions. We group these villages with dispenser villages, as *Evidence Action* continued to maintain dispensers in these villages.

We include these villages in our main estimation of treatment effects, grouping treatment arms by whether or not they received the water quality interventions. As a robustness check, we exclude these 71 villages from the heterogeneity analysis by "Water Quality" assignment in Appendix F. We find little difference in the sign or magnitude of coefficients, and small differences in statistical significance, likely driven by reduced power from the smaller sample size.

C. Experimental integrity

Table A.1: Summary of participation

	Pure Control	Active Control	Visualization	Planning	Total
Recruited	775	992	992	991	3,750
Completed baseline	–	777	783	777	2,337
Completed baseline & Intervention 1	–	777	783	777	2,337
Completed baseline & Intervention 1 & Intervention 2	–	736	734	711	2,181
Completed baseline & endline	-	716	708	692	2,116
Completed endline	588	811	800	785	2,984
Completed baseline & endline & chlorine test	-	662	647	641	1,950
Completed endline & chlorine test	568	771	754	738	2,831
Completed chlorine test	571	771	756	741	2,839

Table A.2: Baseline balance: main outcomes

	Comparison with active control (AC+INF)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N
Baseline score					
Tower of London: Total Moves	23.32	−0.48	−0.30	0.70	2197
	(6.83)	(0.46)	(0.42)		
		[0.66]	[1.00]		
General Self-Efficacy Score (GSE)	42.84	−0.07	0.90	0.19	2175
	(11.59)	(0.68)	(0.74)		
		[1.00]	[1.00]		
β^{MPL}	0.99	0.01	0.00	0.64	2142
	(0.37)	(0.02)	(0.02)		
		[0.95]	[1.00]		
δ^{MPL}	0.98	0.00	0.00	0.52	2142
	(0.02)	(0.00)	(0.00)		
		[0.95]	[1.00]		
Self report: added chlorine in last month	0.65	−0.04	−0.04	0.82	2184
	(0.48)	(0.03)	(0.03)		
		[0.45]	[1.00]		
Chlorine knowledge score (z-score)	0.00	0.09	0.01	0.18	2337
	(1.00)	(0.05)	(0.05)		
		[0.45]	[1.00]		

Notes: OLS estimates of baseline values of main outcomes. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. The Tower of London is a lab game that measures a participant's ability to plan ahead. The General Self-Efficacy score measures a participant's belief in their own ability to achieve the outcomes they desire. Time preference parameters β and δ measured over money are derived from responses to Multiple Price Lists (MPL).

Table A.3: Baseline balance: dispenser vs. non-dispenser villages

	(1) Village without Chlorine Dispenser Mean (SD)	(2) Village with Chlorine Dispenser Difference	(3) N
Observables			
Age	26.25	0.21	3750
	(4.68)	(0.15)	
Married/ Cohabiting	0.89	−0.01	3750
	(0.31)	(0.01)	
Education Level	5.84	0.10	3750
	(1.18)	(0.04)***	
High Wealth Index	0.51	0.01	3750
	(0.50)	(0.02)	

Notes: OLS estimates of baseline balance on observed characteristics for villages with and without chlorine dispensers. For each variable, we report the mean of villages without a chlorine dispenser, with the standard deviation in parentheses. Column 2 reports the difference for villages with a chlorine dispenser, with standard errors in parentheses. All standard errors are clustered at the level of intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.4: Attrition analysis: treatments vs. active control

	(1) Attrited from endline	(2) Attrited from chlorine test	(3) Attrited from endline	(4) Attrited from chlorine test	(5) Attrited from endline	(6) Attrited from chlorine test
Visualization	0.02	0.01	0.02	0.01	0.02	−0.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.12)	(0.14)
Planning	0.03	0.03	0.03	0.03	0.14	0.15
	(0.02)*	(0.02)	(0.02)*	(0.02)	(0.14)	(0.15)
Age			−0.01	−0.01	−0.01	−0.01
			(0.00)***	(0.00)***	(0.00)***	(0.00)**
Married or cohabiting			−0.03	−0.02	0.00	0.01
			(0.02)	(0.02)	(0.03)	(0.04)
Education level			−0.00	−0.01	−0.00	−0.01
			(0.01)	(0.01)	(0.01)	(0.01)
V x Age Interaction					0.00	0.00
					(0.00)	(0.00)
P x Age Interaction					−0.00	−0.00
					(0.00)	(0.00)
V x Married Interaction					−0.09	−0.08
					(0.05)*	(0.05)
P x Married Interaction					−0.01	−0.01
					(0.06)	(0.06)
V x Education Interaction					0.01	0.00
					(0.01)	(0.01)
P x Education Interaction					−0.01	−0.01
					(0.01)	(0.02)
V x Wealth Interaction					−0.01	0.00
					(0.03)	(0.04)
P x Wealth Interaction					−0.03	−0.01
					(0.03)	(0.04)
Constant	0.09	0.15	0.31	0.39	0.27	0.37
	(0.01)***	(0.02)***	(0.06)***	(0.07)***	(0.08)***	(0.10)***

Notes: OLS estimates of the probability of attriting relative to the AC+INF control group. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Each column represents a different specification, with or without controls and interaction terms to assess whether i) there was differential attrition for groups with certain observed characterstics (Columns (3) and (4)) and ii) there was any differential effect of an observed characteristic on the probability of attriting for any treatment group (Columns (5) and (6)). All standard errors are clustered at the level of intervention cohort.
* denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.5: Attrition analysis: active treatments vs. pure control

	(1) Attrited from endline	(2) Attrited from chlorine test	(3) Attrited from endline	(4) Attrited from chlorine test	(5) Attrited from endline	(6) Attrited from chlorine test
V+INF	−0.05	−0.04	−0.06	−0.04	−0.02	0.01
	(0.02)***	(0.02)*	(0.02)***	(0.02)*	(0.16)	(0.17)
P+INF	−0.04	−0.02	−0.04	−0.02	−0.04	0.01
	(0.02)*	(0.02)	(0.02)**	(0.02)	(0.18)	(0.18)
AC+INF	−0.06	−0.04	−0.06	−0.04	−0.16	−0.09
	(0.02)***	(0.02)*	(0.02)***	(0.02)**	(0.17)	(0.17)
Age			−0.01	−0.01	−0.01	−0.01
			(0.00)***	(0.00)***	(0.00)**	(0.00)*
Married or cohabiting			−0.08	−0.08	−0.10	−0.13
			(0.02)***	(0.02)***	(0.06)*	(0.06)**
Education level			−0.01	−0.01	−0.01	−0.01
			(0.01)**	(0.01)**	(0.01)	(0.01)
V+INF x Age Interaction					−0.00	−0.00
					(0.00)	(0.00)
P+INF x Age Interaction					0.00	0.00
					(0.00)	(0.00)
AC+INF x Age Interaction					0.00	−0.00
					(0.00)	(0.00)
V+INF x Married Interaction					0.01	0.03
					(0.07)	(0.08)
P+INF x Married Interaction					0.04	0.08
					(0.08)	(0.08)
AC+INF x Married Interaction					0.05	0.09
					(0.07)	(0.08)
V+INF x Education Interaction					0.00	−0.01
					(0.02)	(0.02)
P+INF x Education Interaction					−0.01	−0.02
					(0.02)	(0.02)
AC+INF x Education Interaction					0.00	−0.00
					(0.02)	(0.02)
V+INF x Wealth Interaction					0.00	−0.00
					(0.04)	(0.04)
P+INF x Wealth Interaction					−0.02	−0.02
					(0.04)	(0.04)
AC+INF x Wealth Interaction					0.02	0.00
					(0.04)	(0.04)
Constant	0.26	0.29	0.59	0.62	0.61	0.61
	(0.02)***	(0.02)***	(0.06)***	(0.06)***	(0.13)***	(0.14)***

Notes: OLS estimates of the probability of attriting relative to the pure control group. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Each column represents a different specification, with or without controls and interaction terms to assess whether i) there was differential attrition for groups with certain observed characterstics (Columns (3) and (4)) and ii) there was any differential effect of an observed characteristic on the probability of attriting for any treatment group (Columns (5) and (6)). All standard errors are clustered at the level of intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

D. Detailed descriptions of outcome measures

D.1 Tower of London planning task

In our computerized version of the task, participants see a screen with two parts: on the left side is the word "start" with a picture of three "pegs" and various shapes positioned on the pegs; on the right side is the word "goal" with a similar picture of three "pegs" and the same shapes positioned differently on the pegs. To complete the task, participants must reposition the shapes underneath the "start" on the left to match the "goal" position on the right. They are instructed to complete each round in as few moves as possible, with the minimum number of moves shown as a number on the screen. In addition to a practice round, participants attempt four rounds of increasing complexity, beginning with one shape requiring only one move, and concluding with three shapes in a pattern that necessitates at least four moves. In all rounds, participants are limited to a maximum of 10 moves. If this occurs, the round ends and the participant is required to contact a staff member to ensure she understands the task before continuing to the next round. Therefore, the distribution of scores is censored at both ends. Performance on the Tower of London task is computed as the total number of moves used across the four rounds. An example of the participant's screen is shown in Figure A.1.

D.2 Effort discounting task

Following recent innovations in the elicitation of time preferences (Andreoni and Sprenger 2012; Augenblick, Niederle, and Sprenger 2015), we estimate time preferences in the effort domain, using the methodology of Augenblick (2017): participants choose how many units of an effort task they want to complete at a time t for a piece rate w, where t is 0, 1, 7, or 8 days from today, and the piece rate w is KES 2, 6, or 10. Variation in time identifies the discount rate, while variation in piece rates identifies the curvature of the utility function. One time and one piece rate are randomly implemented at the end (described below). Figure A.2 provides an example of the participant interface for the task.[36] In contrast to Augenblick (2017), we hold the time of decision constant and vary the time of effort provision, which requires us to control for weekday effects. All questions required a minimum effort allocation of one task at each time to control for the fixed costs of starting, and allow a maximum of 50 tasks.

Developing an effort task that is adapted to a field setting in a developing country, with low levels of literacy, was challenging: the required variation in timing meant that effort could not be completed in the laboratory. We needed to monitor and enforce *when* participants supply

[36] To consider the possibility that respondents feel obligated to carry out some effort regardless of the wage, a subsample of participants was also asked how many units of effort they would supply for a piece rate of KES 0 (but still receiving the KES 100 completion bonus explained below).

effort, and *how much*, while they are in their homes, and do not have access to a computer. We thus developed a new effort task that is adapted to our setting: participants completed data entry tasks by SMS, using toll-free numbers administered by the Busara Center.[37] Each SMS was supposed to contain a 30-digit random number string, which takes approximately two minutes to type. Participants were given a sheet with 50 such strings, including a counter to keep track. To ensure comprehension, participants completed one practice SMS during the survey. At the end of the survey, one decision (out of 12) was randomly selected to be the "decision that counts": at the selected piece rate and time horizon, participants had to send the exact number of SMS they chose. If they did, they received the full piece rate payment plus a KES 100 completion bonus. If they failed to implement the decision they made, they lost both the payment for this task and the completion bonus (see Augenblick 2017 for a full description of this method).[38] Earnings from this task were paid 14 days from the survey date, regardless of the selected effort time horizon.

We estimate time preferences over effort following the approach of Augenblick (2017) by assuming quasi-linear utility (linear in money, convex in effort) and a power cost of effort function. We additionally assume quasi-hyperbolic discounting. Following DellaVigna and Pope (2017), we allow for a non-monetary reward s, which participants receive for each task in addition to the piece rate. The non-monetary reward captures a range of motives, from norm or sense of duty, to reciprocity towards the employer (for the flat payment), to intrinsic motivation and personal competitiveness. It was motivated by the observation that participants supply non-zero amounts of effort even for low piece rates (DellaVigna and Pope 2017). The optimal level of effort is thus given by

$$e^* = argmax\,(s + D_m(14)\cdot\phi\cdot w)\cdot e - \beta^{I(t>0)}\cdot\delta^t\cdot(\frac{1}{\gamma}e^{\gamma} + d_w\cdot e) \tag{2}$$

where β and δ capture (hyperbolic) temporal discounting of effort, w is the piece rate, $D_m(14)$ captures monetary discounting of the payment in 14 days (this is constant for all questions, and thus allowed to differ from effort discounting), t is the time of effort provision, $\gamma > 1$ captures convex costs of effort, ϕ is a slope parameter, and d_w are weekday indicators which allow the opportunity cost of time to vary across weekdays. Within the non-linear objective function above, we estimate additive treatment effects of V, P, and AC on the parameters β, δ, s, and γ.[39]

[37] Although we did not screen on phone access, all participants in our sample have access to a mobile phone: 71 percent own one, 96 percent have one in their household, and the remainder shares the phone of friends or relatives. Since phones are often used by multiple individuals, phone access should be understood as continuous rather than binary.

[38] The field setting with SMS required some tolerance: while a laboratory computer can confirm correct and incorrect entries, and display the number of tasks still to complete, we relied on participants to do this themselves. We thus allowed for 75 percent accuracy in entering the number strings, and a tolerance of 10 in the number of completed SMS (subject to positive completion). The permissible time window started on the calendar day before the task was due, and ended with the calendar day of the task itself. Participants were told that there would be some tolerance for mistakes, but not how much.

[39] To control for the effect of introducing a 0 KES piece rate for a subsample of participants (Footnote 36), we

66 percent of participants identifiably sent at least one SMS (that was not a practice SMS during the session), 60 percent sent the correct number of SMS during the correct time window, and 41 percent additionally satisfied the required accuracy threshold (see footnote 38) and got paid. The key challenge for the verification of the effort task was matching SMS to participants: despite various safety provisions (including name and subject ID in each SMS, asking participants to report all phone numbers they might use), 59049 SMS from 3144 phone numbers could not be matched to any of our 2983 participants. This challenge arises from a field setting where individuals commonly share multiple phones within or across households (see footnote 37).

To test for difficulties in access to phones, we included a small module in the endline survey in which participants were asked about their access to a mobile phone, particularly at the times necessary to complete the SMS task. To alleviate the concern that respondents did not understand the payment scheme, we included three multiple-choice comprehension questions immediately before the task that asked participants to calculate the payout in different circumstances. Respondents could not participate in the task until they had answered the comprehension questions correctly.

Table A.9 shows phone access and task comprehension by treatment group. We find high rates of phone access and comprehension across all treatment groups, and no large differences across treatment groups. The exception is the pure control group, which showed lower comprehension at endline compared to the active treatment groups, presumably because it was their first time completing the task, while the other groups had already experienced it at baseline. We therefore interpret differences in time preferences between this group and the others with caution.

D.3 Money discounting task

In addition to the effort discounting task, we included a conventional Multiple Price List (MPL) task to measure monetary discounting. Participants were asked to make 10 choices between payments at earlier and later dates. The payment at the early date was always KES 100, while the payment at the later date increased gradually from KES 110 to KES 300, using gross interest rates 1.1, 1.25, 1.75, 2, and 3. Each decision was first made in a near time-frame (today vs. four weeks from today), and later in a future time-frame (four weeks vs. eight weeks from today). The list of decisions is presented in Table A.6. One decision was randomly selected to be paid out. As outcome measures from the MPL we estimate β and δ in the quasi-hyperbolic discounting model of Laibson (1997), assuming linearity of utility in money.

additionally allow treatment effects of being exposed to the zero rate on the parameters γ and s. These parameters measure the intercept and curvature of the effort supply function, and may thus be affected by variations in the set of wages. In contrast, time preferences are assumed to be orthogonal to wage variation effects.

D.4 Alternative mechanisms

D.4.1 Beliefs about effectiveness of chlorination

We assess differential beliefs across treatment groups about the proportion of pediatric diarrhea cases which can be prevented by water chlorination. At baseline, all participants in the active treatment groups ("V+INF", "P+INF", and "AC+INF") are told that water chlorination reduces childhood diarrhea by approximately one third. At endline they are asked this question in a multiple-choice format. We take the proportion of cases the participant believes chlorine can avert as a measure of belief about chlorine effectiveness.

D.4.2 Knowledge of how to use chlorine

We assess differential knowledge across treatment groups of how to use chlorine to sanitize water. We ask two multiple-choice questions at endline, to which all three active treatment groups were told the correct answer at baseline: i) how much chlorine to add to water; ii) the amount of time that needs to pass after chlorine is added for water to be safe to drink.

D.4.3 Risk Preferences

We include a modified Eckel-Grossman task to account for changes in risk preferences (Charness, Gneezy, and Imas 2013). Participants choose between one of three 50/50 lotteries, represented as bets on a coin flip. We construct an ordinal measure of risk aversion based on the expected payout the participant is willing to forgo for an increase in certainty of payout.

D.4.4 Demand Effects (Salience of Chlorination)

We test for the possibility that our treatments differentially increased the salience of water chlorination. During the endline survey, enumerators read out three lists of nine words each to every participant, and asked her to recall as many words as possible directly after reading each list. Each list contained three categories of future-related words (chlorine, savings, and farm investment), as well as non-future related filler words. The word lists are available in original Swahili and English translation in Table A.7. We estimate salience effects using equation 3:

$$w_{im} = a_0 + \sum_{j=1}^{3} a_j T_{ji} + \psi_0 X_{im} + \delta_m + \theta_{im} \tag{3}$$

where w_{im} is an indicator for participant i correctly recalling the word related to chlorine in list m; X_{im} refers to the number of words that the individual correctly recounted from that list; δ_m is

a fixed effect for list m; and T_j are treatment indicators. We test $H_0: \alpha_1 = \alpha_2 = \alpha_3$, with the null hypothesis corresponding to no differential salience of chlorine across (active) treatment groups.

In case our treatments differentially affected the salience of chlorine, we further test whether this is due to an increased salience of future-oriented behaviors in general - which may result from our main psychological mechanisms of interest. To this end, we estimate whether the differential treatment effect on chlorine words also holds for two other future-oriented behaviors (saving and farm investment), which were not emphasized in the sessions. We estimate

$$w_{imn} = a_0 + \sum_{j=1}^{3} a_j T_{ji} + \lambda chlorine_n + \psi X_{im} + \sum_{k=1}^{3} b_k T_{ki} \cdot chlorine_n + \delta_m + \theta_{imn} \quad (4)$$

where w_{imn} is an indicator for participant i correctly recalling the words in list m from future oriented behavior n (chlorination, savings or farm investment); and $chlorine_n$ is a dummy for the word being related to chlorine. The a_j coefficients capture increased future orientation due to treatment, while the b_j coefficients indicate that salience increased differentially for chlorination. We test $H_o: b_1 = b_2 = b_3$, with the null hypothesis corresponding a

Figure A.1: Tower of London Example Screen

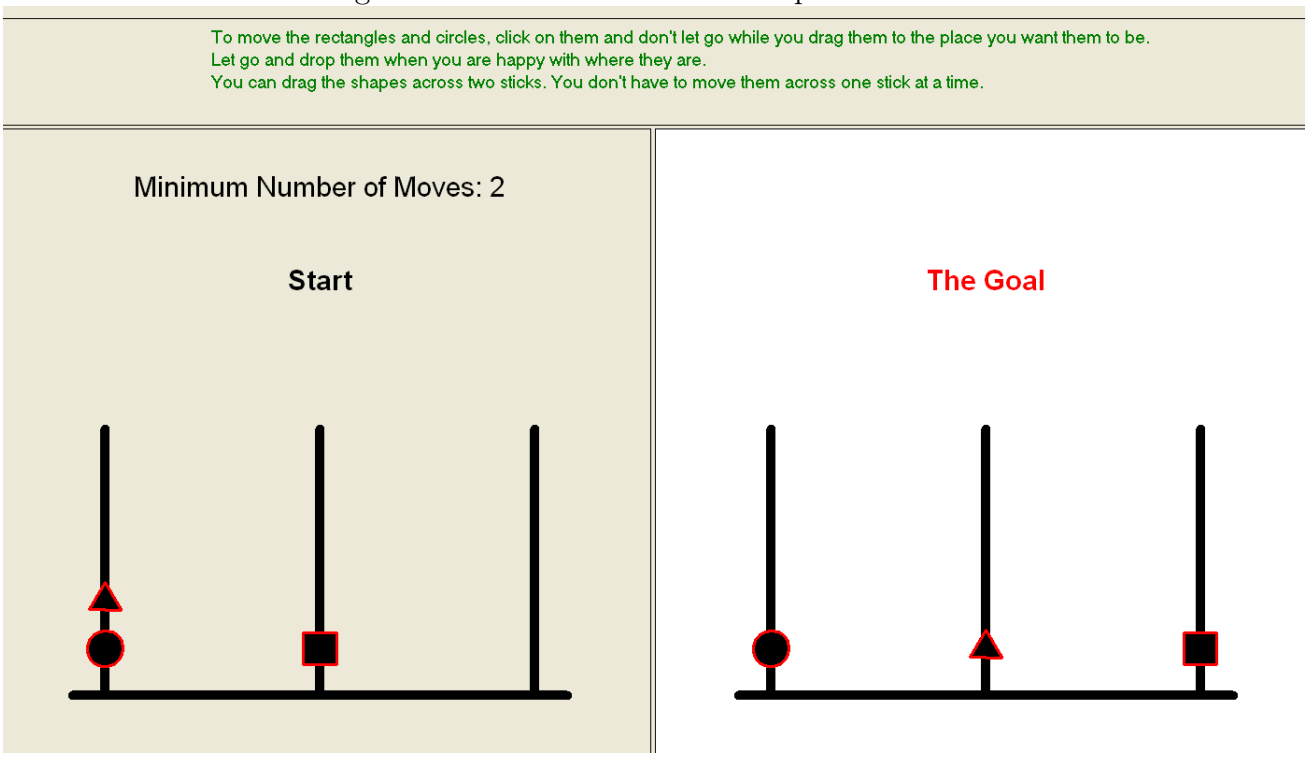


Table A.6: Monetary Discounting Decisions

Front-end delay (t)	Delay between payments (k)	Early (m)	Maximum Late ($m(1+r)$)	Implied interest rate ($1+r$)
Frame 1				
0	28	100	110	1.1
0	28	100	125	1.25
0	28	100	175	1.75
0	28	100	200	2
0	28	100	300	3
Frame 2				
28	28	100	110	1.1
28	28	100	125	1.25
28	28	100	175	1.75
28	28	100	200	2
28	28	100	300	3

Figure A.2: Effort discounting task example screen

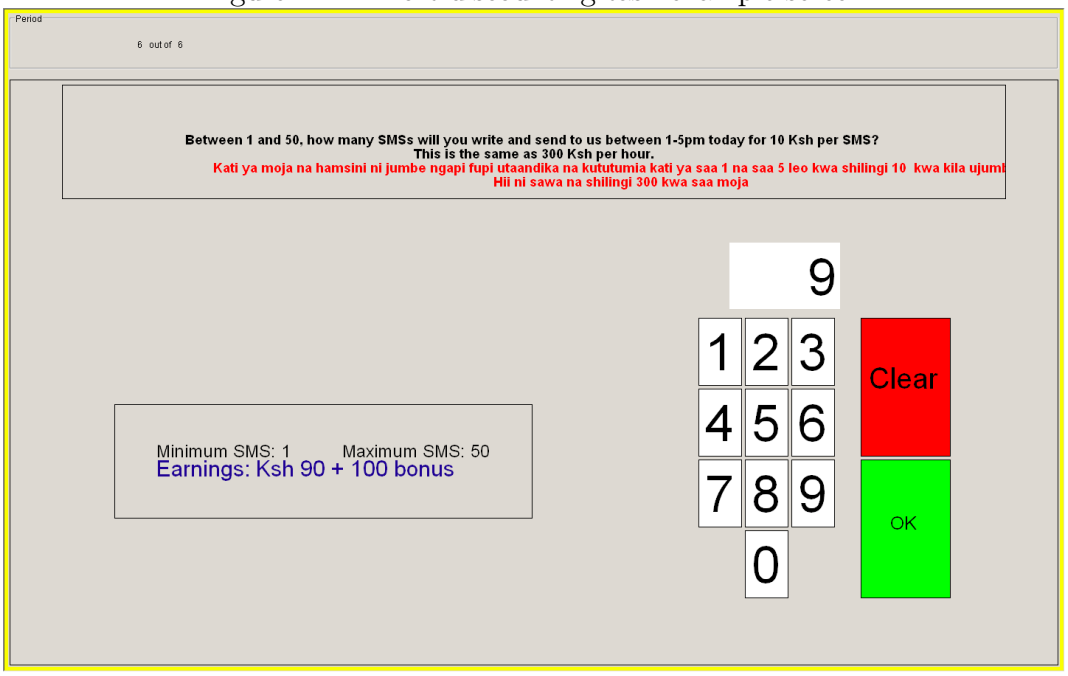
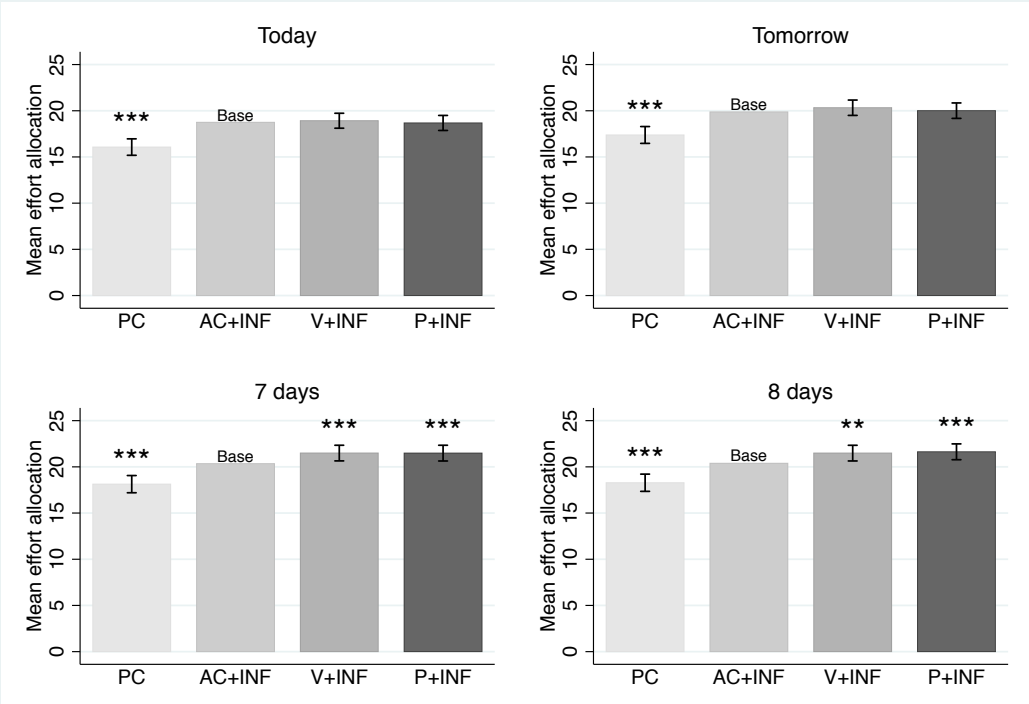


Figure A.3: Effort discounting: raw responses



* denotes significance at 10 pct., ** at 5 pct., *** at 1 pct. level

Table A.7: Word lists for salience test

List	Position	English Translation	Swahili	Group
A	1	Fence	Fence	Filler
A	2	Panadol	Panadol	Filler
A	3	WaterGuard	WaterGuard	Chlorine
A	4	Playing	Kucheza	Filler
A	5	Saving	Kuwekeza	Saving
A	6	Tarmac	Lami	Filler
A	7	Dairy Cow	Ng'ombe wa maziwa	Farm Investment
A	8	Safaricom	Safaricom	Filler
A	9	Resting	Kupumzika	Filler
B	1	Patterned Cloth	Kitenge	Filler
B	2	Thermos	Thermos	Filler
B	3	Savings Group	Chama	Savings
B	4	Baby Oil	Mafuta ya mtoto	Filler
B	5	Poultry Farming	Kilimo cha kuku	Farm investment
B	6	Petrol	Petroli	Filler
B	7	Chlorine	Klorini	Chlorine
B	8	Machete	Panga	Filler
B	9	Shoe Polish	Rangi ya viatu	Filler
C	1	Saucepan	Sufuria	Filler
C	2	Stool	Stool	Filler
C	3	Farm Lease	Kukodisha shamba	Farm investment
C	4	Transport	Transport	Filler
C	5	Dispenser	Dispensa	Chlorine
C	6	Photocopier	Photocopier	Filler
C	7	Piggybank	Benki ya nyumbani	Savings
C	8	Airtime	Airtime	Filler
C	9	Community Hall	Ukumbi wa jamii	Filler

Table A.8: Raw means of z-scored outcome variables

	(1) Active Control Group Mean (SD)	(2) Pure Control Mean (SD)
Behavioral Activation score (BADS) (9–63)	29.38	28.90
	(5.90)	(5.67)
Tower of London: Total moves (9–36)	20.89	23.86
	(6.58)	(6.79)
Risk aversion measure (1–3)	1.74	1.70
	(0.82)	(0.83)
General Self-Efficacy score (GSE) (0–72)	43.35	43.32
	(12.11)	(9.83)
Belief: Proportion of diarrhea incidences avoided through chlorination (0–1)	0.74	0.71
	(0.39)	(0.41)
Chlorine knowledge score (0–2)	1.14	1.07
	(0.68)	(0.69)
ANC/PNC knowledge score (0–2)	1.25	0.93
	(0.84)	(0.87)
Index of investment in children's education (-3–3)	0.03	−0.07
	(1.14)	(0.95)

Notes: Mean and standard deviation of the control group by which the z-scored outcome variables are standardized.

Table A.9: Phone access & task comprehension questions

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
SMS Task Checks										
Participant uses a phone she owns	0.71	0.01	−0.03	0.15	2108	0.70	0.00	−0.03	−0.00	2972
	(0.45)	(0.02)	(0.03)			(0.46)	(0.02)	(0.03)	(0.03)	
Participant uses a phone belonging to her household	0.96	0.01	−0.00	0.24	2108	0.96	0.02	0.00	0.01	2972
	(0.19)	(0.01)	(0.01)			(0.21)	(0.01)	(0.01)	(0.01)	
Proportion for whom accessing a phone for 30mins is very difficult or impossible	0.12	0.03	0.00	0.21	2107	0.13	0.02	−0.00	−0.01	2970
	(0.33)	(0.02)	(0.02)			(0.33)	(0.02)	(0.02)	(0.02)	
Proportion for whom accessing a phone for 1hr is very difficult or impossible	0.17	0.04	0.03	0.52	2107	0.19	0.02	0.00	−0.03	2970
	(0.37)	(0.02)*	(0.02)			(0.39)	(0.02)	(0.02)	(0.02)	
Proportion for whom accessing a phone for 4hrs is very difficult or impossible	0.32	0.02	0.01	0.54	2105	0.35	−0.02	−0.03	−0.04	2967
	(0.46)	(0.03)	(0.03)			(0.48)	(0.03)	(0.03)	(0.03)	
SMS Comprehension questions correct first time	0.81	−0.04	−0.01	0.29	2103	0.72	0.04	0.07	0.07	2955
	(0.39)	(0.03)	(0.03)			(0.45)	(0.03)	(0.03)***	(0.03)***	
Number of attempts at SMS comprehension questions	0.78	0.23	0.14	0.51	2103	1.22	−0.21	−0.27	−0.38	2955
	(1.89)	(0.12)*	(0.12)			(2.36)	(0.13)	(0.13)**	(0.13)***	

Notes: OLS estimates of responses to questions asked as validation checks for the SMS effort task. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. All columns include village-level fixed effects and a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

D.5 Schedule of tasks and treatments

Participants were invited to a 7:30AM or 12:30PM session at a village hall in their area. Sessions lasted between two and four hours. Participants received short breaks between each item on the agenda.

During zTree portions of the session, each participant sat in front of a Windows tablet computer, sufficiently spaced to prevent participants from seeing the answers of their neighbors. One enumerator read instructions and answer options aloud in Kiswahili from the center of the room, while several others were available to answer individual questions or assist with the technology.

During the SurveyCTO questionnaires at endline, five to eight enumerators went through questionnaires with participants individually, in the order that participants arrived.

Interventions were carried out in cohorts of approximately five, in a circle outside when weather permitted. Groups were physically separated to ensure participants could not be overheard. All participants received the same intervention on a given day.

Baseline Session 1:

At baseline, the demographic questionnaire, behavioral tasks and psychological measures were carried out on the zTree experimental interface.

1. Welcome, identification and screening
2. Consent
3. Demographics questionnaire
    (a) Marital status / household composition
    (b) Assets module
    (c) Water use
    (d) Chlorination behavior
    (e) Pregnancy health behaviors
4. Behavioral tasks and psychological scales
    (a) Tower of London
    (b) General Self-Efficacy scale
    (c) Effort discounting measure
    (d) Monetary discounting measure (Multiple Price List)
5. Administration of intervention: Part 1
6. Debrief
7. Payment

Baseline Session 2

1. Welcome, identification and screening
2. Administration of intervention: Part 2
3. Debrief

Endline

: At endline, the behavioral tasks and psychological measures were carried out on the zTree experimental interface, as at baseline. The salience task and individual survey were administered one-on-one with an enumerator.

1. Welcome, identification and screening
2. Consent
3. Salience task
4. Group tasks and measures
   (a) Behavioral Activation for Depression Scale (BADS)
   (b) General Self-Efficacy (GSE) scale
   (c) Tower of London
   (d) Risk measure
   (e) Effort discounting measure
   (f) Monetary discounting measure (Multiple Price Lists)
5. Individual Survey
   (a) Savings
   (b) Labor supply and search
   (c) Fertility & antenatal/postnatal care
   (d) Child education & health
   (e) Participant education
   (f) Phone access

E. Treatment effects on the treated

Table A.10: Behavioral outcomes (TOT)

	Comparison with pure control (PC)				
	(1) Pure Control Mean (SD)	(2) V+INF Treatment Effect	(3) P+INF Treatment Effect	(4) AC+INF Treatment Effect	(5) N
Health outcomes					
Pre-specified primary					
Objective measure: water has been treated with chlorine (TCR)	0.22	0.07	0.04	0.03	2839
	(0.42)	(0.02)***	(0.02)*	(0.02)	
Pre-specified exploratory					
Objective measure: water has sufficient chlorine to be safe (FCR)	0.18	0.07	0.03	0.03	2839
	(0.39)	(0.02)***	(0.02)	(0.02)	
		[0.01]***	[0.66]	[1.00]	
Number of diarrhea incidences per child u15 in last 3 months	0.23	−0.09	−0.07	0.00	2823
	(0.58)	(0.03)***	(0.03)**	(0.03)	
		[0.01]**	[0.16]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.02	−0.02	0.01	2800
	(0.36)	(0.02)	(0.02)	(0.02)	
		[0.61]	[0.89]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.19	−0.34	−0.07	0.19	272
	(1.17)	(0.29)	(0.28)	(0.35)	
		[0.42]	[1.00]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.17	0.01	0.02	0.05	2806
	(0.31)	(0.02)	(0.02)	(0.02)**	
		[0.89]	[0.89]	[0.26]	
Savings outcomes					
Pre-specified secondary					
Amount saved regularly (per week, KES)	88.76	19.80	8.24	4.29	2972
	(228.12)	(13.71)	(13.99)	(12.89)	
		[0.35]	[0.23]	[0.97]	
Pre-specified exploratory					
Indicator: Amount saved regularly is positive	0.32	0.14	0.01	0.04	2972
	(0.47)	(0.03)***	(0.03)	(0.03)	
		[0.00]***	[1.00]	[1.00]	
Number of new ROSCAs joined in last 3 months	0.21	0.02	−0.01	−0.02	2972
	(0.46)	(0.03)	(0.03)	(0.03)	
		[0.61]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.10	0.00	0.01	2972
	(0.38)	(0.02)***	(0.02)	(0.02)	
		[0.00]***	[1.00]	[1.00]	
Labor outcomes					
Pre-specified secondary					
Total hours of work in last 3 months	108.78	−2.65	−26.73	−8.01	2972
	(182.99)	(10.98)	(10.28)***	(10.46)	
		[0.37]	[0.03]**	[0.80]	
Pre-specified exploratory					
Total days of work in last 3 months	21.73	−0.59	−4.46	−1.58	2972
	(30.45)	(1.85)	(1.75)**	(1.77)	
		[0.89]	[0.12]	[1.00]	
Average monthly earnings in last 3 months	1167.22	−48.98	−119.15	−160.94	2972
	(3155.76)	(180.37)	(186.27)	(189.09)	
		[0.89]	[1.00]	[1.00]	
Other behavioral outcomes					
Pre-specified secondary					
Index of investment in children's education (z-score)	0.00	0.09	0.14	0.14	1967
	(1.00)	(0.07)	(0.08)*	(0.07)*	
		[0.35]	[0.08]*	[0.26]	

Notes: Treatment-on-the-treated estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. The specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session (all variables except the objective chlorine measures). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Objective measures of chlorine come from data collected in participants' homes, where stored household water was tested for the presence of Total and Free Chlorine Residual, which indicate whether chlorine has been added to water (TCR), and whether sufficient chlorine is in the water to make it safe to drink (FCR). The number of diarrhea incidences reports how many independent episodes of diarrhea each child under-15 had on average in the last three months. The index of investment in children's education consists of measures of money spent on children's education and average school days attended in the last 3 months, and is constructed only for those women with school-age children.

Table A.11: Psychological outcomes (TOT)

	Comparison with pure control (PC)				
	(1) Pure Control Mean (SD)	(2) V+INF Treatment Effect	(3) P+INF Treatment Effect	(4) AC+INF Treatment Effect	(5) N
Planning					
Pre-specified primary					
Behavioral Activation Score (BADS) (z-score)	0.00	0.03	0.12	0.04	2955
	(1.00)	(0.06)	(0.06)**	(0.06)	
		[1.00]	[0.10]	[1.00]	
Pre-specified secondary					
Tower of London: Total Moves (z-score)	0.00	−0.43	−0.48	−0.41	2955
	(1.00)	(0.06)***	(0.06)***	(0.06)***	
		[0.00]***	[0.00]***	[0.00]***	
Time preferences					
Pre-specified primary					
β^{Effort}	0.964	0.003	0.005	−0.004	2906
	(0.017)	(0.015)	(0.014)	(0.013)	
		[1.00]	[0.54]	[1.00]	
Pre-specified secondary					
β^{MPL}	1.02	−0.01	0.02	0.01	2955
	(0.43)	(0.03)	(0.03)	(0.02)	
		[0.60]	[0.26]	[0.40]	
δ^{MPL}	0.98	0.00	0.00	0.00	2955
	(0.02)	(0.00)*	(0.00)	(0.00)**	
		[0.07]*	[0.12]	[0.06]*	
δ^{Effort}	0.995	−0.000	0.000	0.004	2906
	(0.002)	(0.002)	(0.002)	(0.002)*	
		[0.60]	[0.67]	[0.09]*	
Risk Aversion Measure (z-score)	0.00	0.01	−0.01	0.04	2735
	(1.00)	(0.06)	(0.06)	(0.06)	
Self-efficacy					
Pre-specified secondary					
General Self-Efficacy Score (GSE) (z-score)	0.00	0.20	0.21	0.02	2955
	(1.00)	(0.07)***	(0.06)***	(0.07)	
		[0.01]***	[0.00]***	[0.40]	
Beliefs and knowledge					
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.14	0.18	0.13	2955
	(1.00)	(0.06)**	(0.06)***	(0.06)**	
Chlorine knowledge score (z-score)	0.00	0.13	0.08	0.10	2955
	(1.00)	(0.06)**	(0.06)	(0.06)*	
ANC/PNC knowledge score (z-score)	0.00	0.40	0.32	0.39	2955
	(1.00)	(0.06)***	(0.06)***	(0.06)***	

Notes: Treatment-on-the-treated estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. The specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. The BADS score measures a participant's ability and motivation to follow through on plans they make in their life. Tower of London is a lab game that measures a participant's ability to plan ahead. Time preferences over effort (β^{Effort} and δ^{Effort}) are estimated structurally, which makes a standard 2SLS approach infeasible. Instead, we report estimated treatment effects when the treatment indicator switches on for *compliance* with treatment, rather than assigned treatment. Due to the obvious selection issues, we prefer the estimates in Table 4. Time preferences over money are measured using Multiple Price Lists (MPL). The risk aversion measure is derived from responses to a coin flip for various monetary rewards. The General Self-Efficacy score measures a participant's belief in their own ability to achieve the outcomes they desire. The mechanisms variables check if the interventions differentially affected i) a participant's belief in the efficacy of chlorine to prevent diarrhea, and ii) their knowledge of how to correctly chlorinate water, to assess whether changes in beliefs or knowledge could have been the cause of changes in chlorination behavior.

F. Robustness checks

Table A.12: Robustness of chlorination effects to within-village testing order

	Comparison with active control (AC+INF)			Comparison with pure control (PC)		
	(1) Chlorine present in water	(2) Chlorine present in water	(3) Chlorine present in water	(4) Chlorine present in water	(5) Chlorine present in water	(6) Chlorine present in water
Visualization	0.05	0.05	0.05	0.06	0.06	0.06
	(0.02)**	(0.02)**	(0.02)**	(0.02)***	(0.02)***	(0.02)***
Planning	0.02	0.02	0.02	0.04	0.04	0.04
	(0.02)	(0.02)	(0.02)	(0.02)*	(0.02)*	(0.02)*
Active Control				0.03	0.03	0.03
				(0.02)	(0.02)	(0.02)
Days after first day of testing within village		−0.00			−0.00	
		(0.00)			(0.00)	
Dummy: test done on first day of testing within village			0.02			0.02
			(0.03)			(0.02)
Observations	2012	2012	2012	2839	2839	2839

Notes: * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.13: Chlorine-related outcomes in dispenser vs. non-dispenser villages: restricting to WASH villages in the Water Quality and Passive Comparison arms

	Village has no chlorine dispenser					Village has chlorine dispenser					Comparison		
	(1) Pure Control Mean (SD)	(2) V+INF Treatment Effect	(3) P+INF Treatment Effect	(4) AC+INF Treatment Effect	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N	(11) V+INF Interaction p-value	(12) P+INF Interaction p-value	(13) AC+INF Interaction p-value
Objective measure: water has been treated with chlorine (TCR)	0.26	−0.01	0.01	−0.04	1066	0.18	0.11	0.07	0.10	1091	[0.01]***	[0.15]	[0.00]***
	(0.44)	(0.03)	(0.03)	(0.03)		(0.38)	(0.03)***	(0.03)**	(0.03)***				
Objective measure: water has sufficient chlorine to be safe (FCR)	0.16	0.01	0.02	−0.01	1066	0.17	0.09	0.05	0.08	1091	[0.07]*	[0.58]	[0.04]**
	(0.37)	(0.03)	(0.03)	(0.03)		(0.38)	(0.04)**	(0.03)	(0.03)**				
Number of diarrhea incidences per child u15 in last 3 months	0.22	−0.05	−0.05	−0.00	1039	0.17	−0.05	−0.03	0.01	1099	[0.89]	[0.73]	[0.83]
	(0.56)	(0.05)	(0.05)	(0.05)		(0.46)	(0.04)	(0.04)	(0.04)				
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	−0.04	0.20	0.19	0.14	1092	0.01	0.07	0.09	0.11	1160	[0.32]	[0.48]	[0.87]
	(1.02)	(0.09)**	(0.09)**	(0.09)		(1.01)	(0.09)	(0.09)	(0.09)				
Chlorine knowledge score (z-score)	0.08	0.16	0.03	0.01	1092	0.02	0.01	0.04	0.10	1160	[0.22]	[0.90]	[0.47]
	(1.02)	(0.09)*	(0.09)	(0.09)		(1.01)	(0.09)	(0.09)	(0.09)				

Notes: Robustness check restricting the sample to individuals from villages in the Water Quality and Passive Comparison arms of the WASH Benefits study. The outcome variables repeat those in Tables 1–3, but the table reports the analysis separately for villages which have at least one chlorine dispenser maintained at a village water source, and for villages which do not have chlorine dispensers. OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. The specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session (all variables except the objective chlorine measures). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Columns (11)–(13) report the p-values on the differential effect of the treatments in villages with vs. without chlorine dispensers using SUR.

Table A.14: Behavioral outcomes for alternative attrition definition

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Health outcomes										
Pre-specified primary										
Objective measure: water has been treated with chlorine (TCR)	0.23	0.05	0.02	0.15	2012	0.22	0.06	0.04	0.02	2585
	(0.42)	(0.02)**	(0.02)			(0.42)	(0.02)***	(0.02)*	(0.02)	
Pre-specified exploratory										
Objective measure: water has sufficient chlorine to be safe (FCR)	0.21	0.04	0.01	0.16	2012	0.18	0.06	0.04	0.03	2585
	(0.40)	(0.02)**	(0.02)			(0.39)	(0.02)***	(0.02)*	(0.02)	
		[0.07]*	[1.00]				[0.01]***	[0.35]	[1.00]	
Number of diarrhea incidences per child u15 in last 3 months	0.24	−0.09	−0.06	0.30	2012	0.23	−0.08	−0.05	0.01	2569
	(0.56)	(0.03)***	(0.03)**			(0.58)	(0.03)***	(0.03)*	(0.03)	
		[0.00]***	[0.15]				[0.01]**	[0.35]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.00	−0.02	0.23	1999	0.22	0.01	−0.01	0.00	2550
	(0.35)	(0.02)	(0.02)			(0.36)	(0.02)	(0.02)	(0.02)	
		[0.58]	[1.00]				[0.88]	[1.00]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.26	−0.22	0.04	0.45	200	1.19	−0.50	−0.20	−0.08	242
	(1.19)	(0.41)	(0.35)			(1.17)	(0.42)	(0.46)	(0.51)	
		[0.50]	[1.00]				[0.39]	[1.00]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.21	−0.04	−0.03	0.43	2004	0.17	−0.00	0.01	0.04	2557
	(0.34)	(0.02)**	(0.02)			(0.31)	(0.02)	(0.02)	(0.02)**	
		[0.06]*	[0.77]				[0.99]	[1.00]	[0.65]	
Savings outcomes										
Pre-specified secondary										
Amount saved regularly (per week, KES)	93.96	24.89	3.28	0.10	2108	88.76	25.62	4.83	3.98	2693
	(230.26)	(12.37)**	(12.51)			(228.12)	(13.15)*	(13.19)	(12.07)	
		[0.16]	[1.00]				[0.18]	[0.31]	[1.00]	
Pre-specified exploratory										
Indicator: Amount saved regularly is positive	0.36	0.13	−0.02	0.00***	2108	0.32	0.16	0.01	0.04	2693
	(0.48)	(0.03)***	(0.03)			(0.47)	(0.03)***	(0.03)	(0.03)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Number of new ROSCAs joined in last 3 months	0.17	0.04	0.01	0.17	2108	0.21	0.03	−0.01	−0.02	2693
	(0.44)	(0.03)*	(0.02)			(0.46)	(0.03)	(0.03)	(0.03)	
		[0.10]	[1.00]				[0.51]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.11	−0.01	0.00***	2108	0.17	0.12	0.00	0.01	2693
	(0.38)	(0.02)***	(0.02)			(0.38)	(0.02)***	(0.02)	(0.02)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Labor outcomes										
Pre-specified secondary										
Total hours of work in last 3 months	105.87	−6.79	−23.66	0.06*	2108	108.54	−7.30	−23.35	−3.01	2693
	(173.63)	(9.48)	(8.95)***			(182.02)	(10.19)	(9.68)**	(10.00)	
		[0.90]	[0.03]**				[0.19]	[0.05]*	[1.00]	
Pre-specified exploratory										
Total days of work in last 3 months	21.22	−0.59	−3.91	0.04**	2108	21.73	−0.85	−3.80	−0.77	2693
	(30.09)	(1.64)	(1.60)**			(30.45)	(1.75)	(1.66)**	(1.70)	
		[0.56]	[0.15]				[0.88]	[0.29]	[1.00]	
Average monthly earnings in last 3 months	1094.50	3.11	−1.23	0.97	2108	1167.22	−106.79	−110.57	−110.87	2693
	(2865.35)	(147.23)	(163.19)			(3155.76)	(167.76)	(175.38)	(182.42)	
		[0.65]	[1.00]				[0.82]	[1.00]	[1.00]	
Other behavioral outcomes										
Pre-specified secondary										
Index of investment in children's education (z-score)	0.00	−0.02	0.01	0.60	1420	0.00	0.11	0.13	0.13	1825
	(1.00)	(0.06)	(0.07)			(1.00)	(0.07)*	(0.08)*	(0.07)*	
		[0.90]	[1.00]				[0.18]	[0.08]*	[0.33]	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session (all variables except the objective chlorine measures). For the comparison to the pure control, the sample is restricted in the active treatment groups to those who attended the baseline and first intervention. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Objective measures of chlorine come from data collected in participants' homes, where stored household water was tested for the presence of Total and Free Chlorine Residual, which indicate whether chlorine has been added to water (TCR), and whether sufficient chlorine is in the water to make it safe to drink (FCR). The number of diarrhea incidences reports how many independent episodes of diarrhea each child under-15 had on average in the last three months. The index of investment in children's education consists of measures of money spent on children's education and average school days attended in the last 3 months, and is constructed only for those women with school-age children.

Table A.15: Psychological outcomes for alternative attrition definition

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Planning										
Pre-specified primary										
Behavioral Activation Score (BADS) (z-score)	0.00	−0.01	0.04	0.31	2103	0.00	0.05	0.11	0.06	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)	(0.06)*	(0.06)	
		[0.83]	[0.91]				[1.00]	[0.15]	[1.00]	
Pre-specified secondary										
Tower of London: Total Moves (z-score)	0.00	0.01	−0.04	0.28	2103	0.00	−0.46	−0.48	−0.43	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.05)***	(0.05)***	
		[0.85]	[0.50]				[0.00]***	[0.00]***	[0.00]***	
Time preferences										
Pre-specified primary										
β^{Effort}	0.982	0.007	0.005	0.33	2068	0.953	0.007	0.012	0.009	2906
	(0.005)	(0.006)	(0.007)			(0.020)	(0.018)	(0.018)	(0.018)	
		[0.83]	[0.91]				[1.00]	[0.32]	[1.00]	
Pre-specified secondary										
β^{MPL}	1.05	−0.02	0.02	0.21	2103	1.02	−0.01	0.03	0.02	2686
	(0.46)	(0.02)	(0.03)			(0.43)	(0.03)	(0.03)	(0.02)	
		[0.41]	[0.53]				[0.46]	[0.17]	[0.40]	
δ^{MPL}	0.98	−0.00	−0.00	0.86	2103	0.98	0.00	0.00	0.00	2686
	(0.02)	(0.00)	(0.00)			(0.02)	(0.00)**	(0.00)	(0.00)**	
		[0.41]	[0.40]				[0.05]*	[0.17]	[0.02]**	
δ^{Effort}	0.999	−0.001	−0.002	0.16	2068	0.995	−0.001	−0.002	0.003	2906
	(0.001)	(0.001)	(0.001)**			(0.002)	(0.002)	(0.002)	(0.002)	
		[0.26]	[0.06]*				[0.46]	[0.27]	[0.21]	
Risk Aversion Measure (z-score)	0.00	−0.03	−0.07	0.49	1926	0.00	0.01	−0.02	0.04	2473
	(1.00)	(0.06)	(0.06)			(1.00)	(0.06)	(0.06)	(0.06)	
Self-efficacy										
Pre-specified secondary										
General Self-Efficacy Score (GSE) (z-score)	0.00	0.16	0.12	0.38	2103	0.00	0.21	0.19	0.01	2686
	(1.00)	(0.05)***	(0.05)**			(1.00)	(0.07)***	(0.06)***	(0.06)	
		[0.00]***	[0.06]*				[0.00]***	[0.01]***	[0.51]	
Beliefs and knowledge										
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.08	0.06	0.61	2103	0.00	0.17	0.16	0.10	2686
	(1.00)	(0.05)	(0.06)			(1.00)	(0.05)***	(0.06)***	(0.06)*	
Chlorine knowledge score (z-score)	0.00	0.06	−0.01	0.20	2103	0.00	0.18	0.10	0.12	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.06)*	(0.06)**	
ANC/PNC knowledge score (z-score)	0.00	0.04	−0.05	0.08*	2103	0.00	0.43	0.35	0.39	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)***	(0.05)***	(0.05)***	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session. For the comparison to the pure control, the sample is restricted in the active treatment groups to those who attended the baseline and first intervention. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. The BADS score measures a participant's ability and motivation to follow through on plans they make in their life. Tower of London is a lab game that measures a participant's ability to plan ahead. Time preferences over effort are estimated structurally from a real effort task involving data entry by SMSs. Time preferences over money are measured using Multiple Price Lists (MPL). The risk aversion measure is derived from responses to a coin flip for various monetary rewards. The General Self-Efficacy score measures a participant's belief in their own ability to achieve the outcomes they desire. The mechanisms variables check if the interventions differentially affected i) a participant's belief in the efficacy of chlorine to prevent diarrhea, and ii) their knowledge of how to correctly chlorinate water, to assess whether changes in beliefs or knowledge could have been the cause of changes in chlorination behavior.

Table A.16: LASSO specification: behavioral outcomes

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Health outcomes										
Pre-specified primary										
Objective measure: water has been treated with chlorine (TCR)	0.23	0.05	0.02	0.15	2012	0.22	0.06	0.04	0.03	2839
	(0.42)	(0.02)**	(0.02)			(0.42)	(0.02)***	(0.02)*	(0.02)	
Pre-specified exploratory										
Objective measure: water has sufficient chlorine to be safe (FCR)	0.21	0.04	0.01	0.16	2012	0.18	0.06	0.03	0.03	2839
	(0.40)	(0.02)**	(0.02)			(0.39)	(0.02)***	(0.02)	(0.02)	
		[0.07]*	[1.00]				[0.01]**	[0.81]	[1.00]	
Number of diarrhea incidences per child u15 in last 3 months	0.24	−0.09	−0.06	0.30	2012	0.23	−0.08	−0.06	0.00	2823
	(0.56)	(0.03)***	(0.03)**			(0.58)	(0.03)***	(0.03)**	(0.03)	
		[0.00]***	[0.15]				[0.02]**	[0.21]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.00	−0.02	0.23	1999	0.22	0.01	−0.02	0.01	2800
	(0.35)	(0.02)	(0.02)			(0.36)	(0.02)	(0.02)	(0.02)	
		[0.58]	[1.00]				[0.83]	[1.00]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.26	−0.22	0.04	0.45	200	1.19	−0.25	0.03	0.16	272
	(1.19)	(0.41)	(0.35)			(1.17)	(0.40)	(0.41)	(0.44)	
		[0.50]	[1.00]				[0.83]	[1.00]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.21	−0.04	−0.03	0.43	2004	0.17	0.01	0.02	0.04	2806
	(0.34)	(0.02)**	(0.02)			(0.31)	(0.02)	(0.02)	(0.02)**	
		[0.06]*	[0.77]				[0.91]	[1.00]	[0.33]	
Savings outcomes										
Pre-specified secondary										
Amount saved regularly (per week, KES)	93.96	24.89	3.28	0.10	2108	88.76	17.71	7.55	4.03	2972
	(230.26)	(12.37)**	(12.51)			(228.12)	(12.68)	(12.87)	(11.91)	
		[0.16]	[1.00]				[0.42]	[0.23]	[0.96]	
Pre-specified exploratory										
Indicator: Amount saved regularly is positive	0.36	0.13	−0.02	0.00***	2108	0.32	0.12	0.02	0.04	2972
	(0.48)	(0.03)***	(0.03)			(0.47)	(0.03)***	(0.03)	(0.03)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[0.89]	
Number of new ROSCAs joined in last 3 months	0.17	0.04	0.01	0.17	2108	0.21	0.02	−0.01	−0.02	2972
	(0.44)	(0.03)*	(0.02)			(0.46)	(0.03)	(0.03)	(0.03)	
		[0.10]	[1.00]				[0.83]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.11	−0.01	0.00***	2108	0.17	0.09	0.00	0.01	2972
	(0.38)	(0.02)***	(0.02)			(0.38)	(0.02)***	(0.02)	(0.02)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Labor outcomes										
Pre-specified secondary										
Total hours of work in last 3 months	106.11	−6.79	−23.83	0.06*	2108	108.78	−2.23	−23.47	−7.00	2972
	(174.61)	(9.54)	(8.99)***			(182.99)	(10.14)	(9.45)**	(9.64)	
		[0.90]	[0.03]**				[0.42]	[0.04]**	[0.88]	
Pre-specified exploratory										
Total days of work in last 3 months	21.22	−0.59	−3.91	0.04**	2108	21.73	−0.51	−4.01	−1.45	2972
	(30.09)	(1.64)	(1.60)**			(30.45)	(1.71)	(1.60)**	(1.64)	
		[0.56]	[0.15]				[0.91]	[0.14]	[1.00]	
Average monthly earnings in last 3 months	1094.50	3.11	−1.23	0.97	2108	1167.22	−43.58	−105.44	−142.78	2972
	(2865.35)	(147.23)	(163.19)			(3155.76)	(166.41)	(171.46)	(174.39)	
		[0.65]	[1.00]				[0.91]	[1.00]	[1.00]	
Other behavioral outcomes										
Pre-specified secondary										
Index of investment in children's education (z-score)	0.00	−0.02	0.01	0.60	1420	0.00	0.09	0.12	0.12	1967
	(1.00)	(0.06)	(0.07)			(1.00)	(0.07)	(0.07)*	(0.07)*	
		[0.90]	[1.00]				[0.42]	[0.10]	[0.36]	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session (all variables except the objective chlorine measures). The specification also includes additional controls selected by LASSO to account for differential attrition. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Objective measures of chlorine come from data collected in participants' homes, where stored household water was tested for the presence of Total and Free Chlorine Residual, which indicate whether chlorine has been added to water (TCR), and whether sufficient chlorine is in the water to make it safe to drink (FCR). The number of diarrhea incidences reports how many independent episodes of diarrhea each child under-15 had on average in the last three months. The index of investment in children's education consists of measures of money spent on children's education and average school days attended in the last 3 months, and is constructed only for those women with school-age children.

Table A.17: LASSO specification: psychological outcomes

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Planning										
Pre-specified primary										
Behavioral Activation Score (BADS) (z-score)	0.00	−0.01	0.04	0.31	2103	0.00	0.03	0.11	0.04	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)	(0.06)*	(0.06)	
		[0.83]	[0.91]				[1.00]	[0.13]	[1.00]	
Pre-specified secondary										
Tower of London: Total Moves (z-score)	0.00	0.01	−0.04	0.28	2103	0.00	−0.38	−0.43	−0.37	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.05)***	(0.05)***	
		[0.85]	[0.50]				[0.00]***	[0.00]***	[0.00]***	
Time preferences										
Pre-specified primary										
β^{Effort}	0.982	0.007	0.005	0.33	2068	0.953	0.007	0.012	0.009	2906
	(0.005)	(0.006)	(0.007)			(0.020)	(0.018)	(0.018)	(0.018)	
		[0.83]	[0.91]				[1.00]	[0.32]	[1.00]	
Pre-specified secondary										
β^{MPL}	1.05	−0.02	0.02	0.21	2103	1.02	−0.01	0.02	0.01	2955
	(0.46)	(0.02)	(0.03)			(0.43)	(0.03)	(0.03)	(0.02)	
		[0.41]	[0.53]				[0.50]	[0.25]	[0.40]	
δ^{MPL}	0.98	−0.00	−0.00	0.86	2103	0.98	0.00	0.00	0.00	2955
	(0.02)	(0.00)	(0.00)			(0.02)	(0.00)*	(0.00)	(0.00)**	
		[0.41]	[0.40]				[0.08]*	[0.14]	[0.07]*	
δ^{Effort}	0.999	−0.001	−0.002	0.16	2068	0.995	−0.001	−0.002	0.003	2906
	(0.001)	(0.001)	(0.001)**			(0.002)	(0.002)	(0.002)	(0.002)	
		[0.26]	[0.06]*				[0.50]	[0.25]	[0.21]	
Risk Aversion Measure (z-score)	0.00	−0.03	−0.07	0.49	1926	0.00	0.01	−0.01	0.03	2735
	(1.00)	(0.06)	(0.06)			(1.00)	(0.06)	(0.06)	(0.06)	
Self-efficacy										
Pre-specified secondary										
General Self-Efficacy Score (GSE) (z-score)	0.00	0.16	0.12	0.38	2103	0.00	0.18	0.18	0.02	2955
	(1.00)	(0.05)***	(0.05)**			(1.00)	(0.06)***	(0.06)***	(0.06)	
		[0.00]***	[0.06]*				[0.01]**	[0.00]***	[0.40]	
Beliefs and knowledge										
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.08	0.06	0.61	2103	0.00	0.13	0.16	0.12	2955
	(1.00)	(0.05)	(0.06)			(1.00)	(0.05)**	(0.05)***	(0.06)**	
Chlorine knowledge score (z-score)	0.00	0.06	−0.01	0.20	2103	0.00	0.12	0.07	0.09	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)**	(0.06)	(0.06)	
ANC/PNC knowledge score (z-score)	0.00	0.04	−0.05	0.08*	2103	0.00	0.36	0.28	0.35	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)***	(0.05)***	(0.05)***	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session. The specification also includes additional controls selected by LASSO to account for differential attrition. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. The BADS score measures a participant's ability and motivation to follow through on plans they make in their life. Tower of London is a lab game that measures a participant's ability to plan ahead. Time preferences over effort are estimated structurally from a real effort task involving data entry by SMSs. Time preferences over money are measured using Multiple Price Lists (MPL). The risk aversion measure is derived from responses to a coin flip for various monetary rewards. The General Self-Efficacy score measures a participant's belief in their own ability to achieve the outcomes they desire. The mechanisms variables check if the interventions differentially affected i) a participant's belief in the efficacy of chlorine to prevent diarrhea, and ii) their knowledge of how to correctly chlorinate water, to assess whether changes in beliefs or knowledge could have been the cause of changes in chlorination behavior.

Table A.18: LASSO specification: behavioral outcomes for alternative attrition definition

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Health outcomes										
Pre-specified primary										
Objective measure: water has been treated with chlorine (TCR)	0.23	0.05	0.02	0.15	2012	0.22	0.06	0.04	0.02	2585
	(0.42)	(0.02)**	(0.02)			(0.42)	(0.02)***	(0.02)*	(0.02)	
Pre-specified exploratory										
Objective measure: water has sufficient chlorine to be safe (FCR)	0.21	0.04	0.01	0.16	2012	0.18	0.06	0.04	0.03	2585
	(0.40)	(0.02)**	(0.02)			(0.39)	(0.02)***	(0.02)*	(0.02)	
		[0.07]*	[1.00]				[0.01]***	[0.37]	[1.00]	
Number of diarrhea incidences per child u15 in last 3 months	0.24	−0.09	−0.06	0.30	2012	0.23	−0.08	−0.05	0.01	2569
	(0.56)	(0.03)***	(0.03)**			(0.58)	(0.03)***	(0.03)*	(0.03)	
		[0.00]***	[0.15]				[0.01]**	[0.38]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.00	−0.02	0.23	1999	0.22	0.00	−0.01	0.00	2550
	(0.35)	(0.02)	(0.02)			(0.36)	(0.02)	(0.02)	(0.02)	
		[0.58]	[1.00]				[1.00]	[1.00]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.26	−0.22	0.04	0.45	200	1.19	−0.43	−0.14	0.00	242
	(1.19)	(0.41)	(0.35)			(1.17)	(0.42)	(0.44)	(0.50)	
		[0.50]	[1.00]				[0.51]	[1.00]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.21	−0.04	−0.03	0.43	2004	0.17	−0.00	0.01	0.04	2557
	(0.34)	(0.02)**	(0.02)			(0.31)	(0.02)	(0.02)	(0.02)**	
		[0.06]*	[0.77]				[1.00]	[1.00]	[0.50]	
Savings outcomes										
Pre-specified secondary										
Amount saved regularly (per week, KES)	93.96	24.89	3.28	0.10	2108	88.76	25.62	4.83	3.98	2693
	(230.26)	(12.37)**	(12.51)			(228.12)	(13.15)*	(13.19)	(12.07)	
		[0.16]	[1.00]				[0.18]	[0.31]	[1.00]	
Pre-specified exploratory										
Indicator: Amount saved regularly is positive	0.36	0.13	−0.02	0.00***	2108	0.32	0.16	0.02	0.04	2693
	(0.48)	(0.03)***	(0.03)			(0.47)	(0.03)***	(0.03)	(0.03)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Number of new ROSCAs joined in last 3 months	0.17	0.04	0.01	0.17	2108	0.21	0.03	−0.01	−0.02	2693
	(0.44)	(0.03)*	(0.02)			(0.46)	(0.03)	(0.03)	(0.03)	
		[0.10]	[1.00]				[0.51]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.11	−0.01	0.00***	2108	0.17	0.12	0.00	0.01	2693
	(0.38)	(0.02)***	(0.02)			(0.38)	(0.02)***	(0.02)	(0.02)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Labor outcomes										
Pre-specified secondary										
Total hours of work in last 3 months	105.87	−6.79	−23.66	0.06*	2108	108.54	−7.30	−23.35	−3.01	2693
	(173.63)	(9.48)	(8.95)***			(182.02)	(10.19)	(9.68)**	(10.00)	
		[0.90]	[0.03]**				[0.19]	[0.05]*	[1.00]	
Pre-specified exploratory										
Total days of work in last 3 months	21.22	−0.59	−3.91	0.04**	2108	21.73	−0.84	−3.92	−0.86	2693
	(30.09)	(1.64)	(1.60)**			(30.45)	(1.76)	(1.66)**	(1.71)	
		[0.56]	[0.15]				[0.90]	[0.22]	[1.00]	
Average monthly earnings in last 3 months	1094.50	3.11	−1.23	0.97	2108	1167.22	−106.79	−110.57	−110.87	2693
	(2865.35)	(147.23)	(163.19)			(3155.76)	(167.76)	(175.38)	(182.42)	
		[0.65]	[1.00]				[0.82]	[1.00]	[1.00]	
Other behavioral outcomes										
Pre-specified secondary										
Index of investment in children's education (z-score)	0.00	−0.02	0.01	0.60	1420	0.00	0.11	0.13	0.13	1825
	(1.00)	(0.06)	(0.07)			(1.00)	(0.07)*	(0.08)*	(0.07)*	
		[0.90]	[1.00]				[0.18]	[0.08]*	[0.33]	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session (all variables except the objective chlorine measures). For the comparison to the pure control, the sample is restricted in the active treatment groups to those who attended the baseline and first intervention. The specification also includes additional controls selected by LASSO to account for differential attrition. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Objective measures of chlorine come from data collected in participants' homes, where stored household water was tested for the presence of Total and Free Chlorine Residual, which indicate whether chlorine has been added to water (TCR), and whether sufficient chlorine is in the water to make it safe to drink (FCR). The number of diarrhea incidences reports how many independent episodes of diarrhea each child under-15 had on average in the last three months. The index of investment in children's education consists of measures of money spent on children's education and average school days attended in the last 3 months, and is constructed only for those women with school-age children.

Table A.19: LASSO specification: psychological outcomes for alternative attrition definition

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Planning										
Pre-specified primary										
Behavioral Activation Score (BADS) (z-score)	0.00	−0.01	0.04	0.31	2103	0.00	0.05	0.11	0.06	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)	(0.06)*	(0.06)	
		[0.83]	[0.91]				[1.00]	[0.15]	[1.00]	
Pre-specified secondary										
Tower of London: Total Moves (z-score)	0.00	0.01	−0.04	0.28	2103	0.00	−0.46	−0.48	−0.44	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.05)***	(0.05)***	
		[0.85]	[0.50]				[0.00]***	[0.00]***	[0.00]***	
Time preferences										
Pre-specified primary										
β^{Effort}	0.982	0.007	0.005	0.33	2068	0.953	0.007	0.012	0.009	2906
	(0.005)	(0.006)	(0.007)			(0.020)	(0.018)	(0.018)	(0.018)	
		[0.83]	[0.91]				[1.00]	[0.32]	[1.00]	
Pre-specified secondary										
β^{MPL}	1.05	−0.02	0.02	0.21	2103	1.02	−0.01	0.03	0.02	2686
	(0.46)	(0.02)	(0.03)			(0.43)	(0.03)	(0.03)	(0.02)	
		[0.41]	[0.53]				[0.46]	[0.17]	[0.40]	
δ^{MPL}	0.98	−0.00	−0.00	0.86	2103	0.98	0.00	0.00	0.00	2686
	(0.02)	(0.00)	(0.00)			(0.02)	(0.00)**	(0.00)	(0.00)**	
		[0.41]	[0.40]				[0.05]*	[0.17]	[0.02]**	
δ^{Effort}	0.999	−0.001	−0.002	0.16	2068	0.995	−0.001	−0.002	0.003	2906
	(0.001)	(0.001)	(0.001)**			(0.002)	(0.002)	(0.002)	(0.002)	
		[0.26]	[0.06]*				[0.46]	[0.27]	[0.21]	
Risk Aversion Measure (z-score)	0.00	−0.03	−0.07	0.49	1926	0.00	0.01	−0.02	0.04	2473
	(1.00)	(0.06)	(0.06)			(1.00)	(0.06)	(0.06)	(0.06)	
Self-efficacy										
Pre-specified secondary										
General Self-Efficacy Score (GSE) (z-score)	0.00	0.16	0.12	0.38	2103	0.00	0.21	0.19	0.01	2686
	(1.00)	(0.05)***	(0.05)**			(1.00)	(0.07)***	(0.06)***	(0.06)	
		[0.00]***	[0.06]*				[0.00]***	[0.01]***	[0.51]	
Beliefs and knowledge										
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.08	0.06	0.61	2103	0.00	0.17	0.16	0.10	2686
	(1.00)	(0.05)	(0.06)			(1.00)	(0.05)***	(0.06)***	(0.06)*	
Chlorine knowledge score (z-score)	0.00	0.06	−0.01	0.20	2103	0.00	0.18	0.11	0.13	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.06)*	(0.06)**	
ANC/PNC knowledge score (z-score)	0.00	0.04	−0.05	0.08*	2103	0.00	0.43	0.35	0.39	2686
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)***	(0.05)***	(0.05)***	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. For the comparison to the pure control group, the specification also includes fixed effects for i) the week and ii) the day of the week of the endline survey, for measures collected in the endline session. For the comparison to the pure control, the sample is restricted in the active treatment groups to those who attended the baseline and first intervention. The specification also includes additional controls selected by LASSO to account for differential attrition. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. The BADS score measures a participant's ability and motivation to follow through on plans they make in their life. Tower of London is a lab game that measures a participant's ability to plan ahead. Time preferences over effort are estimated structurally from a real effort task involving data entry by SMSs. Time preferences over money are measured using Multiple Price Lists (MPL). The risk aversion measure is derived from responses to a coin flip for various monetary rewards. The General Self-Efficacy score measures a participant's belief in their own ability to achieve the outcomes they desire. The mechanisms variables check if the interventions differentially affected i) a participant's belief in the efficacy of chlorine to prevent diarrhea, and ii) their knowledge of how to correctly chlorinate water, to assess whether changes in beliefs or knowledge could have been the cause of changes in chlorination behavior.

Table A.20: Behavioral outcomes (without endline date fixed effects)

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Health outcomes										
Pre-specified primary										
Objective measure: water has been treated with chlorine (TCR)	0.23	0.05	0.02	0.15	2012	0.22	0.06	0.04	0.03	2839
	(0.42)	(0.02)**	(0.02)			(0.42)	(0.02)***	(0.02)*	(0.02)	
Pre-specified exploratory										
Objective measure: water has sufficient chlorine to be safe (FCR)	0.21	0.04	0.01	0.16	2012	0.18	0.06	0.03	0.03	2839
	(0.40)	(0.02)**	(0.02)			(0.39)	(0.02)***	(0.02)	(0.02)	
		[0.07]*	[1.00]				[0.01]**	[0.81]	[1.00]	
Number of diarrhea incidences per child u15 in last 3 months	0.24	−0.09	−0.06	0.30	2012	0.23	−0.08	−0.06	0.01	2823
	(0.56)	(0.03)***	(0.03)**			(0.58)	(0.03)**	(0.03)**	(0.03)	
		[0.00]***	[0.15]				[0.02]**	[0.26]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.00	−0.02	0.23	1999	0.22	0.01	−0.02	0.00	2800
	(0.35)	(0.02)	(0.02)			(0.36)	(0.02)	(0.02)	(0.02)	
		[0.58]	[1.00]				[0.90]	[0.82]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.26	−0.22	0.04	0.45	200	1.19	−0.27	−0.06	0.23	272
	(1.19)	(0.41)	(0.35)			(1.17)	(0.40)	(0.42)	(0.43)	
		[0.50]	[1.00]				[0.90]	[1.00]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.21	−0.04	−0.03	0.43	2004	0.17	0.00	0.01	0.03	2806
	(0.34)	(0.02)**	(0.02)			(0.31)	(0.02)	(0.02)	(0.02)*	
		[0.06]*	[0.77]				[1.00]	[1.00]	[1.00]	
Savings outcomes										
Pre-specified secondary										
Amount saved regularly (per week, KES)	93.96	24.89	3.28	0.10	2108	88.76	14.15	4.77	0.11	2972
	(230.26)	(12.37)**	(12.51)			(228.12)	(12.69)	(12.93)	(12.07)	
		[0.16]	[1.00]				[0.66]	[0.31]	[1.00]	
Pre-specified exploratory										
Indicator: Amount saved regularly is positive	0.36	0.13	−0.02	0.00***	2108	0.32	0.12	0.01	0.03	2972
	(0.48)	(0.03)***	(0.03)			(0.47)	(0.03)***	(0.03)	(0.03)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Number of new ROSCAs joined in last 3 months	0.17	0.04	0.01	0.17	2108	0.21	0.02	−0.01	−0.02	2972
	(0.44)	(0.03)*	(0.02)			(0.46)	(0.03)	(0.03)	(0.03)	
		[0.10]	[1.00]				[0.90]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.11	−0.01	0.00***	2108	0.17	0.09	0.00	0.01	2972
	(0.38)	(0.02)***	(0.02)			(0.38)	(0.02)***	(0.02)	(0.02)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Labor outcomes										
Pre-specified secondary										
Total hours of work in last 3 months	106.11	−6.79	−23.83	0.06*	2108	108.78	1.18	−21.17	−3.36	2972
	(174.61)	(9.54)	(8.99)***			(182.99)	(10.16)	(9.51)**	(9.70)	
		[0.90]	[0.03]**				[0.66]	[0.09]*	[1.00]	
Pre-specified exploratory										
Total days of work in last 3 months	21.22	−0.59	−3.91	0.04**	2108	21.73	0.25	−3.40	−0.58	2972
	(30.09)	(1.64)	(1.60)**			(30.45)	(1.71)	(1.63)**	(1.65)	
		[0.56]	[0.15]				[1.00]	[0.26]	[1.00]	
Average monthly earnings in last 3 months	1094.50	3.11	−1.23	0.97	2108	1167.22	−44.55	−113.88	−146.26	2972
	(2865.35)	(147.23)	(163.19)			(3155.76)	(166.31)	(171.26)	(174.92)	
		[0.65]	[1.00]				[1.00]	[1.00]	[1.00]	
Other behavioral outcomes										
Pre-specified secondary										
Index of investment in children's education (z-score)	0.00	−0.02	0.01	0.60	1420	0.00	0.08	0.12	0.12	1967
	(1.00)	(0.06)	(0.07)			(1.00)	(0.07)	(0.07)*	(0.07)*	
		[0.90]	[1.00]				[0.66]	[0.10]*	[0.38]	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. In contrast to Table 2, for the comparison to the pure control group, the specification does not include fixed effects for i) the week and ii) the day of the week of the endline survey. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. Objective measures of chlorine come from data collected in participants' homes, where stored household water was tested for the presence of Total and Free Chlorine Residual, which indicate whether chlorine has been added to water (TCR), and whether sufficient chlorine is in the water to make it safe to drink (FCR). The number of diarrhea incidences reports how many independent episodes of diarrhea each child under-15 had on average in the last three months. The index of investment in children's education consists of measures of money spent on children's education and average school days attended in the last 3 months, and is constructed only for those women with school-age children.

Table A.21: Psychological outcomes (without endline date fixed effects)

	Comparison with active control (AC+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) N
Planning										
Pre-specified primary										
Behavioral Activation Score (BADS) (z-score)	0.00	−0.01	0.04	0.31	2103	0.00	0.02	0.11	0.04	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)	(0.06)*	(0.06)	
		[0.83]	[0.91]				[1.00]	[0.13]	[1.00]	
Pre-specified secondary										
Tower of London: Total Moves (z-score)	0.00	0.01	−0.04	0.28	2103	0.00	−0.37	−0.42	−0.36	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.05)***	(0.05)***	
		[0.85]	[0.50]				[0.00]***	[0.00]***	[0.00]***	
Time preferences										
Pre-specified primary										
β^{Effort}	0.982	0.007	0.005	0.33	2068	0.953	0.007	0.012	0.009	2906
	(0.005)	(0.006)	(0.007)			(0.020)	(0.018)	(0.018)	(0.018)	
		[0.83]	[0.91]				[1.00]	[0.32]	[1.00]	
Pre-specified secondary										
β^{MPL}	1.05	−0.02	0.02	0.21	2103	1.02	0.00	0.03	0.02	2955
	(0.46)	(0.02)	(0.03)			(0.43)	(0.03)	(0.03)	(0.02)	
		[0.41]	[0.53]				[0.60]	[0.27]	[0.45]	
δ^{MPL}	0.98	−0.00	−0.00	0.86	2103	0.98	0.00	0.00	0.00	2955
	(0.02)	(0.00)	(0.00)			(0.02)	(0.00)	(0.00)	(0.00)*	
		[0.41]	[0.40]				[0.20]	[0.27]	[0.21]	
δ^{Effort}	0.999	−0.001	−0.002	0.16	2068	0.995	−0.001	−0.002	0.003	2906
	(0.001)	(0.001)	(0.001)**			(0.002)	(0.002)	(0.002)	(0.002)	
		[0.26]	[0.06]*				[0.60]	[0.27]	[0.28]	
Risk Aversion Measure (z-score)	0.00	−0.03	−0.07	0.49	1926	0.00	0.02	−0.00	0.04	2735
	(1.00)	(0.06)	(0.06)			(1.00)	(0.06)	(0.06)	(0.06)	
Self-efficacy										
Pre-specified secondary										
General Self-Efficacy Score (GSE) (z-score)	0.00	0.16	0.12	0.38	2103	0.00	0.17	0.17	0.01	2955
	(1.00)	(0.05)***	(0.05)**			(1.00)	(0.06)***	(0.06)***	(0.06)	
		[0.00]***	[0.06]*				[0.02]**	[0.01]***	[0.57]	
Beliefs and knowledge										
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.08	0.06	0.61	2103	0.00	0.13	0.15	0.11	2955
	(1.00)	(0.05)	(0.06)			(1.00)	(0.05)**	(0.05)***	(0.06)**	
Chlorine knowledge score (z-score)	0.00	0.06	−0.01	0.20	2103	0.00	0.11	0.07	0.08	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)**	(0.06)	(0.06)	
ANC/PNC knowledge score (z-score)	0.00	0.04	−0.05	0.08*	2103	0.00	0.35	0.28	0.34	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)***	(0.05)***	(0.05)***	

Notes: OLS estimates of treatment effects. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, control for a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. For the comparison to the AC+INF group, the specification also controls for the baseline value of the outcome, for those that were collected at baseline. In contrast to Table 4, for the comparison to the pure control group, the specification does not include fixed effects for i) the week and ii) the day of the week of the endline survey. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Outcome measures are listed on the left, and are described in detail in Section 4. The BADS score measures a participant's ability and motivation to follow through on plans they make in their life. Tower of London is a lab game that measures a participant's ability to plan ahead. Time preferences over effort are estimated structurally from a real effort task involving data entry by SMSs. Time preferences over money are measured using Multiple Price Lists (MPL). The risk aversion measure is derived from responses to a coin flip for various monetary rewards. The General Self-Efficacy score measures a participant's belief in their own ability to achieve the outcomes they desire. The mechanisms variables check if the interventions differentially affected i) a participant's belief in the efficacy of chlorine to prevent diarrhea, and ii) their knowledge of how to correctly chlorinate water, to assess whether changes in beliefs or knowledge could have been the cause of changes in chlorination behavior.