

Can Simple Psychological Interventions Increase Preventive Health Investment?*

Anett John[†] and Kate Orkin[‡]

August 2021

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. After 12 weeks, visualization increases objectively measured chlorination, reduces diarrhea episodes among children, and increases savings. Effects on chlorination and savings persist after almost three years. Effects of the planning intervention are weaker and largely insignificant. Analysis of mechanisms suggests both interventions increase self-efficacy – beliefs about one's ability to achieve desired outcomes. Visualization also increases participants' skill in forecasting their future utility (Gabaix and Laibson 2017). The interventions do not differentially affect beliefs and knowledge about chlorination. Results suggest simple psychological interventions can increase future-oriented behaviors, including use of preventive health technologies.

Keywords: preventive health; water chlorination; visualization; planning; randomized controlled trial. JEL codes: O12, D91, I12

*We thank Johannes Haushofer, Clair Null, Michael Kremer, and the WASH Benefits Kenya team for collaboration and advice; Jane Dougherty, Magdalena Larrebourg, Daniel Mellow, Winnie Mughogho, Moritz Poll, Clemence Poug  -Biyong, and the Busara Center for Behavioral Economics for excellent research assistance; and Nava Ashraf, Pascaline Dupas, David Laibson, Xavier D'Haultfoeuille, Alessandro Iaria, and Pedro Rey-Biel for comments. Two pre-analysis plans for this study are registered at <https://www.socialscienceregistry.org/trials/2850>. Funding is gratefully acknowledged from NIH (grant NIH UH2 NR016378 to Johannes Haushofer), the CSAE Research Fund and UKRI Global Challenges Research Fund (to Kate Orkin), and Labex Ecodec Investissements d'Avenir (grant ANR-11-LABX-0047 to Anett John). The project was approved by the Princeton (#7376) and KEMRI (#536) IRBs.

[†]Department of Economics, University of Birmingham. A.John@bham.ac.uk.

[‡]Blavatnik School of Government, Centre for Study of African Economies, University of Oxford. kate.orkin@bsg.ox.ac.uk.

1. Introduction

Individuals often fail to invest in preventive healthcare, even when such investments cost little and individuals are aware of their benefits.¹ 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 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 25 Kenyan shillings (KES, USD 0.25). Interventions which provide chlorine for free, often in combination with information or marketing campaigns, increase usage to between 23 and 60 percent, but take-up remains far below complete (Dupas et al. 2016; Null et al. 2018; Luoto et al. 2014; Dupas et al. 2020).

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 for present bias in health decisions (Schilbach 2019; Bai et al. 2017; Giné, Karlan, and Zinman 2010; DellaVigna and Malmendier 2006). The success of planning interventions suggests that planning skills and limited attention may be a factor (Stadler, Oettingen, and Gollwitzer 2009; Milkman et al. 2011). Most recently, economists have explored the role of beliefs about oneself – from self-image to agency and sense of control over one’s life – in determining health behavior (Ghosal et al. 2020).

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 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 increase patient behavior by making future outcomes more vivid and tangible in participants’ minds, thus increasing their perceived value relative to the more immediate costs required to attain them. The second group received a two-session intervention that aimed to improve planning skills, helping participants to undertake activities that they were struggling to do regularly by making specific plans, and establishing routines (“Planning” or “P”). Both interventions required participants to think about how their current

¹For instance, insecticide-treated bednets reduce incidence of malaria, which increases farmworker productivity (Dillon, Friedman, and Serneels 2020), farm yields (Fink and Masiye 2015) and children’s earnings in adulthood (Bleakley 2010). Similarly, treating intestinal parasites improves child health and school attendance (Miguel and Kremer 2004), test scores (Ozier 2018), and adult wages (Baird et al. 2016). Yet demand for bednets and deworming pills is low and price-sensitive (Cohen and Dupas 2010; Dupas 2014; Kremer and Miguel 2007).

behavior affects their future outcomes. A third, active control group (“Active Control” or “AC”) also gathered as a group for workshops of the same format, but to discuss topics likely to be psychologically inactive (birds in Kenya).. 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. 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 two psychological treatment groups and the AC+INF group, on which we focus, tests the effect of interventions targeting the ability to visualize the future or to make and execute plans, respectively, over and above those targeting lack of information. This comparison is of primary academic interest, as it best isolates the mechanisms at work. The comparison to the pure control group gives the total effect of providing interventions such as ours in similar settings. We measure health, economic, and psychological effects of our interventions using in-person surveys and choice tasks after ten weeks, and also conduct chemical tests of chlorine levels in water in unannounced household visits after twelve weeks. We follow up after 30-36 months using phone surveys, given pandemic-related safety concerns in 2020.

First, we report economically large and statistically significant effects of Visualization on our main outcome of interest, objectively measured chlorination of household drinking water. Specifically, we find a significant increase of 22 percent (5 percentage points) in the share of households whose drinking water contains chlorine, relative to the active control group, twelve weeks after the interventions. Self-reports confirm this finding. We show suggestive evidence that this effect persists over time: After almost three years, when asked an unprompted question about how they treat their water, participants in the Visualization group are 5 percentage points more likely to state that they chlorinate it. The number of diarrhea episodes per child under 15 decreases by 46 percent (0.12 episodes) relative to the active control after ten weeks. This effect does not persist over time, although we show this may be due to differences in season between rounds. In contrast, the Planning intervention has positive but largely insignificant effects on health outcomes: Chlorination does not increase in either round. Diarrhea episodes decrease somewhat (potentially driven by an increase in boiling water), but neither effect persists over time.

The Visualization intervention is highly cost-effective, even under conservative assumptions. Delivery of the interventions costs approximately USD 4 per household. Considering only benefits to children under 5, the cost per Disability-Adjusted Life Year (DALY) saved is USD 248.² The WHO classifies an intervention as “cost-effective” for a cost per DALY saved below USD 4525, and “highly cost-effective” below USD 1508.

Second, while our primary interest is in preventive health investments, our interventions are domain-general in nature, and may thus reduce barriers to forward-looking

²We conservatively assume that treatment effects drop to zero immediately after three months. If, instead, effects persisted for one year, the cost per DALY would be reduced to USD 62 per child under 5. We also do not account for benefits for children older than 5. Extrapolation to children under 15 yields a cost per DALY of USD 108 for the three-month effects. See Appendix B for detailed calculations.

behavior in general. Indeed, we observe that the Visualization intervention led to significant increases in all four pre-specified savings indicators (amount saved per week, likelihood of saving at all, whether participants joined new rotating savings and credit associations (ROSCAs), and likelihood of saving for productive investments). These effects persist after almost three years, allowing the Visualization group to accumulate 41 percent higher savings balances. The long-term effects are observed during the 2020 global pandemic, which decreased real income in rural Kenya by 7.9 percent (Nechifor et al. 2020), suggesting that participants who learned visualization techniques in 2017 benefited from an increased financial buffer in 2020.

Third, we test if our psychological interventions have larger effects in villages where infrastructure lowers the monetary 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 have been maintained since the trial. Our treatment effects on chlorination and diarrhea are very similar in both types of villages. Our results suggest that relieving psychological constraints is effective, whether or not cost and access barriers remain.

Fourth, we use lab-in-the-field methods and psychological questionnaires to study effects on pre-specified mechanisms which were targeted by the interventions. We hypothesized that Visualization may affect time preferences and self-efficacy, while Planning may affect planning skills and self-efficacy.³ We find that conventional laboratory measures of time preferences – both an incentivized real-effort task and multiple price lists over money – do not respond to our interventions. The interventions also have no measurable impact on self-reported planning behavior, nor underlying cognitive functions linked to planning. In contrast, we show large, significant effects of both the Visualization and Planning treatments on self-efficacy. After ten weeks, self-efficacy increases by 0.15 and 0.11 standard deviations, respectively. In the long-term follow-up, participants in the Visualization group maintain significantly higher levels of self-efficacy than those in the Planning group, consistent with the differential long-term effects on chlorination.

While conventional measures of time preferences did not respond to our Visualization intervention, we find suggestive evidence of improved *utility forecasting*, a conceptualization of intertemporal choice recently proposed by Gabaix and Laibson (2017). In this framework, decision-makers simulate future utility by combining priors with noisy, unbiased signals. Imperfect foresight of future utility results in choices that are biased towards the present, with choice patterns resembling those predicted by hyperbolic discounting models. The framework implies that interventions which reduce forecasting noise will

³Self-efficacy is a psychological concept which captures beliefs about one’s ability to achieve desirable outcomes and overcome obstacles (Bandura 1977; Reich and Infurna 2016). We hypothesized that both interventions might affect self-efficacy, as both required a strong mapping between current behavior and future outcomes. Specifically, visualizing alternative realizations of the future depending on current behavior requires the belief that current behavior can *affect* future outcomes. Similarly, making plans both requires and enforces the belief that desired outcomes are within the individual’s reach. Self-efficacy is closely related to the economic concept of believed returns to effort.

lead to more patient behavior, without changing the underlying preference parameters. Consistent with this prediction, we find that participants in the Visualization group report significantly more clear and vivid mental forecasts of the future after almost three years.⁴ There is no such effect in the Planning group, which may explain the differential effects on both chlorination and savings. Summarizing our results on mechanisms, our results suggest that deeper underlying preferences and cognitive functions may not respond strongly to light-touch interventions. In contrast, beliefs about the self (such as self-efficacy) as well as the ability to imagine the future seem to be both malleable in the short and long term, and potentially powerful drivers of human behavior.

Finally, we test a range of alternative plausible psychological channels. First, we show 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. Second, our interventions do not affect lab measures of risk preferences. Third, using a “salience” task, we find that participants’ attention is more focused on chlorination in both psychological treatment groups than in the information treatment alone, even 10 weeks after the interventions. In the Visualization group, this effect appears to grow *larger* after almost three years. We cannot disentangle whether increased salience causes or is caused by increased use of chlorine. However, given the long-term nature of the salience effects, and the fact that the salience of savings does not increase but savings do increase, the evidence is more consistent with reverse causality than with salience as a driver of treatment effects.

Although we cannot completely rule out that participants were influenced by social desirability bias in answering questions, 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 after 12 weeks. In the long-run survey, where we only measure self-reported chlorination, we also conduct explicit experimenter demand treatments (following de Quidt, Haushofer, and Roth 2018, and orthogonal to our psychological treatments). We find that explicitly telling the respondents the experimenter’s hypothesis has no effect on self-reported chlorination, suggesting social desirability bias is limited.

Our paper contributes to existing work in behavioral and development economics on psychological factors which affect investment decisions. First, and beyond the health domain, to the best of our knowledge, our paper is the first to show the effects of a visualization-style intervention on real-world behavior of adults, in multiple domains. We understand visualization as the simulation of *expected* future utility, and thus as distinct from goal-setting (Gollwitzer and Sheeran 2006). Psychologists have long been interested in interventions which fall into this category. Typically, interventions are administered *while* the outcome behavior is observed in a laboratory setting: In “Episodic Future

⁴We discuss possible reasons for the divergence of our time preference measures in Section 7.

Thinking (EFT),” participants’ mental focus is shifted to unrelated future events, such as attending a birthday party next month (Daniel, Stanton, and Epstein 2013). Vividness interventions aim to make the future self more vivid, typically by showing age-progressed photos or avatars of the decisionmaker (Hershfield, John, and Reiff 2018; Hershfield et al. 2011). Sample sizes are usually around or below 50. Alan and Ertac (2018) take visualization techniques to the field, and show that an eight-session intervention in Turkish primary schools leads children to make more patient decisions in incentivized choice tasks almost three years later. We innovate on this literature by showing that even light-touch visualization interventions can affect real-world behavior of adults, twelve weeks and three years after the intervention. We also elicit comprehensive measures of the mechanisms through which the intervention may work. Our work is also distinct from studies testing light-touch interventions to increase aspirations (Bernard et al. 2019; Baranov, Haushofer, and Jang 2020; Orkin et al. 2020). A key difference is that such interventions assume decision-makers with reference-dependent preferences, and aim to enhance effort and investment by raising people’s reference points (Genicot and Ray 2020). Instead, we present suggestive evidence that visualization increases patient behavior both by strengthening the belief that current behaviors affect future outcomes (self-efficacy),⁵ and by improving participants’ mental forecasts of the future (Gabaix and Laibson 2017).⁶

Second, we compare our Visualization intervention against another intervention aimed at increasing future-oriented behavior: Planning. This comparison of two interventions which target different psychological mechanisms is novel, and allows us to explore which psychological targets are the most malleable to intervention, and which most affect behavior. The Planning intervention was adapted from a psychotherapy approach called Behavioral Activation (BA), which teaches people simple, structured planning skills to help them undertake activities that they have identified as important but are struggling to do regularly (Lejuez et al. 2011; Ekers et al. 2014). To the best of our knowledge, the effects of behavioral activation on economic behavior or preventive health choices have not been shown.⁷ Economists have examined effects of a related but distinct therapy, Cognitive Behavioral Therapy (CBT) (Blattman, Jamison, and Sheridan 2017; Baranov et al. 2020). Relative to this work, we work with a general population, rather than those suffering from depression or who are criminally engaged. We also study a different therapy: CBT interventions are typically bundled, containing multiple modules teaching different skills and cognitions. They are usually longer (typically 7-8 sessions) and require skilled personnel, while BA is shorter and facilitators require little training (Richards et al. 2016; Patel et al. 2017). Our Planning intervention is also distinct from “implementation intentions” interventions (Gollwitzer and Sheeran 2006; Duckworth et al. 2013), as well as from planning prompts (Milkman et al. 2011; Stadler, Oettingen, and Gollwitzer 2009). These identify a specific domain of behavior

⁵See McKelway (2018) for the effect of self-efficacy on women’s labor supply.

⁶Future research is needed to confirm and better understand this mechanism, such as the ongoing work on visualization-based business skills training by Ashraf et al. (2020).

⁷One ongoing study examines the effects of a four-week behavioral activation treatment targeted at depressed people on labor market outcomes four years later (de Quidt et al. 2020).

where participants are encouraged to make changes, while in the Planning intervention, participants identified areas where they aimed to change behavior themselves.

Third, we contribute to the question of how to increase preventive health investment in low-income countries. Within this literature, to the best of our knowledge, we are the first to study a domain-general psychological intervention which is light-touch and easy to scale. A large literature has studied the impact of information (Madajewicz et al. 2007; Jalan and Somanathan 2008), price (Ashraf, Berry, and Shapiro 2010; Kremer et al. 2011b), combinations of information and subsidies (Ashraf, Jack, and Kamenica 2010; Null et al. 2018; Kremer et al. 2011a; Luoto et al. 2011), and non-monetary costs (Dupas et al. 2016; Dupas et al. 2020) on the demand for preventive healthcare.⁸ Following the hypothesis that the remaining barriers to adoption may be behavioral, some studies use “nudges” to complement information or free provision. These have included verbal commitments (Kremer and Miguel 2007; Dupas 2009) or marketing messages (Kremer et al. 2011a; Dupas 2009). They have also shown limited effectiveness, except when they are highly personalized (e.g. Luoto et al. 2014 prints posters using photos of the respondent) and thus potentially difficult to scale. Nudges are also targeted to specific, momentary behaviors, and do not attempt to change fundamental preferences or constraints. We contribute to this literature by showing that domain-general psychological interventions can be a low-cost and scalable tool to increase preventive health investments. In combination with subsidy or information campaigns, they may be used to deliver clean water to the 2 billion people who currently use drinking water contaminated with feces.⁹

The paper proceeds 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), a cluster-randomized controlled trial conducted from 2012 to 2014 in 1,226 villages (Null et al. 2018). Villages were randomized to eight arms testing household-level water, sanitation, handwashing, and nutrition interventions, either in isolation or in combination.¹⁰

We sampled 205 villages from the WASH study and recruited 3,750 women aged 18–35 between October 2017 and January 2018 in these villages. Women are primarily responsible for collecting water, and thus for chlorination. We recruit women aged 18–35 as they are the

⁸Given the size of the literature, our citations here focus on interventions pertaining to water quality and chlorination. For comprehensive reviews of the preventive health literature, see the Dupas and Miguel (2017) and Kremer and Glennerster (2011).

⁹<https://www.who.int/news-room/fact-sheets/detail/drinking-water>.

¹⁰For more information on the WASH Benefits study, see Appendix C.

most likely to have small children, who are the most vulnerable to water-borne illnesses. Enumerators visited all households in each village and conducted a census to determine household eligibility. Screening criteria included: 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 and have on average 6 years of education. 89 percent are married or cohabiting. We split our sample into three “active” treatment groups and one pure control group. We randomly assign 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 stratify the randomization on village of residence and a wealth index.¹¹

We run our study in WASH sample villages because there is random variation between WASH treatment arms in levels of access to chlorine dispensers, and hence in the monetary and time costs households face to chlorinate. We cross-cut this village-level variation with our household-level randomization to psychological treatments. We analyze the effects of psychological interventions on take-up of preventive health products when take-up is more or less costly.

We randomly selected 90 villages from the WASH arms that receive chlorine dispensers (henceforth “dispenser villages”). First, we randomly selected 67 villages from the “Water Quality” treatment arm of the WASH study. In this treatment, chlorine dispensers were installed at an average of five community water points per village cluster and community promoters encouraged their use. The NGO *Evidence Action* still maintains these dispensers, ensures they are filled with chlorine, and retains a local promoter in each community. In these villages, the cost of chlorinating is lower. Second, we sample 23 villages who received the “Water Quality” intervention in combination with other interventions, detailed in Appendix C. All other interventions (sanitation, handwashing, nutrition, or a combination thereof) took place at the household level, and finished three to four years before our study. Our sampling excludes households which participated directly in the original WASH study. As a result, we simply classify all 90 villages as dispenser villages.

We also randomly selected 105 “non-dispenser” WASH villages. 67 of these villages were in the WASH “Passive Comparison” arm, which received no interventions. The other 48 villages were in WASH treatment arms which received no dispensers, but did receive other household-level interventions. As above, former WASH beneficiary households are excluded from our study. In the 105 non-dispenser villages, the monetary and time cost of chlorinating is higher. The randomization has remained intact, so no dispensers have been provided by *Evidence Action* in non-dispenser villages.

¹¹The wealth index consisted of the total value of a limited set of assets (bicycles, cellphones, gas stoves, all livestock, radios, sofas and televisions).

2.2 Design and timeline

Participants in the three active treatment groups participated in four sessions: a combined baseline and first intervention session, a second intervention session one week later, an endline 10 weeks after the first session, and a long-term follow-up after 30 to 36 months. Participants in the pure control group did not participate in the baseline or the interventions. Households in all four groups received an unannounced visit 12 weeks after the first session, where enumerators collected a drinking water sample. Figure 1 shows the study timeline.

We conducted the baseline survey, treatment interventions, and 10-week endline survey in “mobile labs” operated by the Busara Center. The baseline and endline surveys lasted about two hours each. Behavioral tasks and psychological questionnaires were self-administered using touch screen computers and color-coded response buttons in zTree (Fischbacher 2007). Enumerators read out instructions in Kiswahili to maximize comprehension, though all participants were literate. Economic and health questions were administered one-on-one by enumerators.

For the intervention, participants were split into the same cohorts of five for both sessions. No participant was invited for the second session without having already participated in the first session. Each intervention session lasted about two hours. Interventions were run by locally-trained female facilitators.

Between July and December 2020, an average of 33 months after the interventions, we conducted a long-term follow-up survey. Given public health restrictions in 2020, this survey was conducted by phone, and lasted about 30 minutes. To minimize attrition, enumerators went back to the villages of all respondents whose phone numbers were no longer operational (about 22 percent) during an intensive tracking stage, to collect new phone numbers. To ensure consistency of surveying and respondent safety, surveys were conducted by phone even in these cases.

Participants received participation fees of KES 250 (USD 2.50) for the baseline and each intervention session, KES 350 for the 10-week endline survey, and KES 100 for the long-term follow-up.¹² Participants were reimbursed for their transport costs to the mobile laboratory (typically 30 minutes travel time). They also received payments from the experimental tasks, as described in Section 4. Typical total study payments (excluding transport) were around KES 1700 in the active treatment groups, and around KES 850 in the pure control group.¹³ All recruited participants were invited to attend the endline and follow-up surveys, regardless of whether they attended the baseline or intervention sessions. Table A.9 shows the number of participants at each stage of the study.

¹²USD 1 was equivalent to approximately KES 100 at the time of the study.

¹³Participants were informed of the participation fees 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.

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 sample at baseline, diarrhea prevalence was 16 percent in the past two weeks among children under age 3. 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 and store it in jerry cans or pottery jars. Water can be (re)contaminated easily if storage containers are left open, or when water is removed from storage pots.

Chlorinating water kills many of the pathogens that cause diarrhea.¹⁴ Absent point-of-collection chlorine dispensers, households can purchase dilute chlorine. The main brand 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) and comes with instructions in Kiswahili and in pictures. There is widespread awareness of this product: even at baseline of the WASH study in 2012, 89 percent of the sample had heard of *WaterGuard*, and 29 percent had used it at least once (Null et al. 2018). Households also believe that chlorination reduces diarrhea.¹⁵

However, take-up of chlorination is low. In our sample during the unannounced home visits, 18 percent of pure control households had detectable free chlorine in their water, and 22 percent had detectable total chlorine.¹⁶ 66 percent of households report chlorination as one of their main methods of treating water (which may be true even if their current water is not chlorinated, or the chlorine present has decayed, see Section 4.1).

Chlorination may have monetary, time, attention, and disutility costs, in both dispenser and non-dispenser villages. Dispensers are not present at every water source, so people might need to plan to go to the correct water source. Once at the water source, they may face an attention cost, e.g. when chatting with fellow villagers while fetching water. They might need to make alternative arrangements if the dispenser is not filled, or remind children fetching water to chlorinate it. In non-dispenser villages, women need to purchase *WaterGuard*, paying both time and monetary costs, and remember to add it to each batch of water.

Finally, taste has been mentioned as a potential barrier to chlorination (Dupas et al. 2016; Ashraf, Berry, and Shapiro 2010; Dupas et al. 2020). One can taste if water has been chlori-

¹⁴A meta-analysis finds point-of-use water chlorination methods reduced relative risk of any episode of child diarrhea in the last two weeks by 34 percent, across 9 studies (Arnold and Colford Jr 2007).

¹⁵In the 10-week survey, participants are asked, "If three children are sick with diarrhea, how many of these cases could have been prevented with chlorination?". 61 percent say chlorine would prevent diarrhea in all three children; only 20 percent say it would not prevent diarrhea in any children. The true number is slightly less than one case.

¹⁶See 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.

nated: in a blind taste test among staff of the Busara Center, participants correctly identified chlorinated drinking water 75 percent of the time. We find no statistically significant differences in the taste rating participants gave to chlorinated vs. unchlorinated water.¹⁷ However, habit formation plays a large role in taste (Atkin 2013; Loewenstein, Price, and Volpp 2016). Taste habituation could be modeled as a standard investment choice, where people need to pay a fixed disutility cost to get used to the unfamiliar taste of chlorinated water. In this sense, taste could be seen as a cost barrier, rather than as an innate, fixed preference.

3. Interventions and Theory of Change

Here, we discuss the structure of interventions and the mechanisms they target. 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 exercises and activities. In all three active arms, the first intervention session concluded with the information module described in Section 3.3.

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

The core of our Visualization 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.

In the first intervention session, participants were given an interactive lecture on thinking about the future. Participants were asked to think of examples of small everyday behaviors (such as spending their leftover budget on snacks) which could affect 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. Instead of prescribing a list of everyday behaviors, facilitators largely relied on examples brought up by the participants. The session then 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 and antenatal care visits, 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. The session concluded with participants visualizing and

¹⁷Unchlorinated water mean score = 6.21/10, chlorinated water mean score = 6.06/10, $p = 0.66$, $n = 90$.

drawing an alternative realization of their own future, depending on present behavior.¹⁸ In the second session, participants revisited all concepts from the first session and repeated the visualization exercises, applied to different situations in their lives. Particular emphasis was put on how they could use future visualizations in their everyday life in order to overcome any temptations they may encounter. Unlike the Planning intervention, Visualization focused on high-level behaviors and outcomes without implementation details.

From a theoretical perspective, the Visualization 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 (Gilbert and Wilson 2009; Kahneman et al. 2004). 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 forms beliefs about the utility from a future reward $u_t \sim \mathcal{N}(\mu, \sigma_u^2)$ by combining her prior μ with mental simulations of the reward s_t . Simulations are noisy, unbiased signals $s_t = u_t + \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0, \sigma_{\epsilon_t}^2)$. Importantly, the simulation noise $var(\epsilon_t) = \sigma_{\epsilon_t}^2$ increases in the time horizon t . Assuming for simplicity that $\mu = 0$ (this can be relaxed), the average posterior belief of u_t becomes $D(t)u_t$, where $D(t) = 1/(1 + \sigma_{\epsilon_t}^2/\sigma_u^2)$ is the *as-if* discounting function. $D(t)$ decreases in t , and takes a hyperbolic shape when the simulation noise is linear in the time horizon ($\sigma_{\epsilon_t}^2 = \sigma_{\epsilon}^2 t$), thus generating present-biased behavior with preference reversals. The model implies that interventions which improve forecasting ability (and thus reduce simulation noise $\sigma_{\epsilon_t}^2$) will lead to more patient behavior. Gabaix and Laibson (2017) themselves note that forecasting precision is likely to improve with more time spent thinking about a trade-off, higher intelligence or imaginative capacity, and more effort spent forecasting.¹⁹ These predictions are matched by empirical evidence that patience in discounting tasks can be increased by introducing a deliberation period before subjects can make a choice (Imas, Kuhn, and Mironova 2020), and by letting subjects interact with age-progressed computer renderings of themselves while they choose (Hershfield et al. 2011).

We hypothesize that this intervention works by teaching participants to generate more precise utility forecasts when making intertemporal trade-offs in their daily life, and thus act more patiently.²⁰ In addition, we hypothesize that visualization exercises will strengthen the mapping between present behavior and future outcomes (self-efficacy). Applied (but not restricted) to the context of chlorination, this may mean that participants connect chlorinating in the present more to health benefits in the future, have a more

¹⁸Behaviors participants wanted to change included avoiding impulse purchases, saving, searching more for work, taking a loan to start a business, treating water, purchasing a water tank to get cleaner water, eating better, or encouraging children to wash their hands after going to the bathroom.

¹⁹One way to microfound this is to think of forecasting effort as the number of simulations s_t that are generated. With n simulations, forecasting variance decreases by $1/n$.

²⁰It is possible to obtain a similar prediction using conventional models of present-biased preferences (Laibson 1997 and O'Donoghue and Rabin 1999), which place a differential weight $1/\beta$ on the present. Using these models, the key question for our intervention is whether the fundamental preference parameters β and δ can be changed persistently by increasing the focus on the future (see also Becker and Mulligan 1997). See Section 7.1 for more discussion on similarities and differences between the models' predictions.

precise image of these benefits, and thus weigh them more heavily against small present costs in terms of money, attention, or taste habituation.

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

The Planning intervention taught participants skills to help them undertake activities that they had identified as important but were struggling to do regularly. It is based on planning skills taught during a psychotherapy called “Behavioral Activation” (BA) (Ekers et al. 2014; Patel et al. 2017). Our exercises drew on an existing manual for implementing this approach (Richards and Whyte 2011).

BA draws on literature on instrumental reinforcement and motivation (Lejuez et al. 2011), which finds that people often avoid necessary tasks, and even rewarding activities when mildly depressed or feeling demotivated. They thus lose out on the sense of achievement of accomplishing what they intended to do. Avoidance exacerbates negative moods, which further increases avoidance and rumination, creating a cycle of inactivity. Behavioral activation teaches simple, structured skills to help people “get going” and re-engage with meaningful activities. Through simple diaries and planning exercises, they learn to set short-term goals, establish routines and reduce avoidance. Compared to other planning interventions, BA is distinctive in how it teaches people to structure plans to make them very easy to implement.: For example, people are taught to start with the easiest tasks, so they quickly earn a sense of accomplishment, which improves mood and increases the likelihood they persist with the task. People are also taught to reward themselves (e.g. plan a fun activity) upon accomplishing a task.

The first session began with an interactive lecture of how anyone can get stuck in cycles of inactivity. Participants then listened to a story of a woman similar to them who was overwhelmed by basic chores, fell out of the habit of doing routine tasks, and struggled to find the motivation to start again. This had negative consequences for herself and her family. Fetching water and chlorinating it were used as examples. Participants could then share their experiences of similar situations.

Participants then did a series of worksheets. They identified activities in their daily lives that they felt were important, but where they were struggling to “get going.” They wrote two lists of activities they could do in the next week: one set they enjoyed doing, and one set that were necessary. They ranked the tasks on each list from most to least difficult, picked the easiest activities on each list, broke the tasks down into steps, and scheduled them in a weekly diary. In the second session, participants discussed barriers they had faced and brainstormed ways to overcome these barriers. Similar to the Visualization intervention, facilitators avoided prescribing what activities participants should do, but merely aimed to teach skills to help them achieve their priorities.

We hypothesize that the planning intervention helps participants to make and implement plans for daily necessary tasks (including but not restricted to chlorination). We also hypothesize that concrete experiences of executing one’s plans and achieving a desired

goal – “mastery experiences” – will increase self-efficacy, consistent with the psychological literature (Bandura 1997; Lorig et al. 2014).

3.3 Treatment 3: Active Control exercise + Information Module (“AC+INF”)

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 which did not capture any of the psychological elements contained in the Visualization or Planning modules. 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. They wrote a list of all the birds they could think of, and then made some drawings of birds. The second session, centered on plants in Kenya, 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 treatments.

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 chlorination message stated that only water properly treated with chlorine or boiled water is safe to drink, while unsafe water can cause illness. It noted that children are the most affected by diarrhea, which causes dehydration and can impede growth. It highlighted that people could avoid one out of three episodes of diarrhea by chlorinating with products like WaterGuard or dispenser chlorine. The message gave instructions on chlorination dosage and the importance of covering water containers. It highlighted that the smell of chlorine is not harmful and reduces over time, and that chlorinating is cheaper than the costs of firewood for boiling.

4. Outcome measures

We have two families of outcomes, behaviors and psychological mechanisms. Within each family, we pre-specified primary, secondary or exploratory hypotheses and the variables used to test each hypothesis, outlined in Appendix Table A.8. We use this ranking in correcting for multiple testing, as described in Section 5.

4.1 Health-related outcomes

Our primary hypothesis is that the interventions affect whether households chlorinate water. In the short-term, we conduct objective tests of the household’s drinking water in unannounced household visits, as well as self-reported measures of how households treat water in the 10-week survey. In the 33-month survey, we repeat the self-reported measure.

In the short-term, our primary outcome is an indicator for whether the Total Chlorine Residual (TCR) is above 0.2 mg/l in household drinking water. This captures whether households have chlorinated water in the past 24 to 72 hours. The measure is a lower bound

on whether households have chlorinated this water at any point: TCR decays and disappears, at different rates depending on the type of container and how much organic material is present in the water, and thus may not be present in water which has been chlorinated more than 24 hours ago (Murphy et al. 2016; Null and Lantagne 2012).²¹ We additionally measure whether the Free Chlorine Residual (FCR) level is above 0.2 mg/l. Put simply, FCR is an alternative, slightly more stringent measure for water safety (WHO 2011). Like some other studies in economics (Kremer et al. 2011a; Dupas et al. 2016), we focus on TCR. In the health literature, TCR is increasingly preferred to FCR because it is more strongly correlated with the presence of *E. coli* (Murphy et al. 2016), the bacteria which causes diarrhea.

In surveys, we asked households whether and how they generally treat water, and recorded up to two treatments. An indicator for answering “chlorine” is our primary outcome at 33 months. We minimize demand effects through question administration. In the baseline and 10-week survey, the module is self-administered on a tablet, so enumerators do not see answers. Water treatment options are presented as buttons on a tablet screen, and include ineffective options like straining or letting it settle. In the 33-month survey, this question is asked in open format on the phone, without prompting any options. Enumerators coded participants’ answers. In all surveys, this was the first mention of water treatment.

We view the objective measure as a lower bound on the use of chlorine and the self-report as an upper bound. Importantly, they are not directly comparable: the objective measure captures whether current water is treated, while the self-report captures general behavior. Both measures have strengths and weaknesses, so many health studies measure both (Murphy et al. 2016). The objective measure minimizes experimenter demand effects, but may not capture households who have added chlorine some time ago and may still have safe water. The self-reported measure is better suited to capture broader chlorination behavior and also covers boiling of water. However, it is more susceptible to experimenter demand effects. We discuss such concerns further in Section 4.4.

We find TCR present in current drinking water for 22 percent of active control households and FCR present for 18 percent at the 12-week chlorine test. 73 percent of households report generally using chlorine. Other studies also find gaps between self-reports of whether water is *currently* chlorinated and water tests: Blanton et al. (2010) find self-reported chlorine use in current water of 39 percent and positive tests of 21 percent in Kenya; Gupta et al. (2007) find 81 and 50 percent in Indonesia.

We test whether interventions affect health outcomes related to chlorination and other preventive health behaviors. We examine self-reported child diarrhea as an exploratory outcome in the 10-week survey and a primary outcome in the 30-36 month survey. For each child in the household, we ask about the number of independent episodes of diarrhea over the last three months. We also measure other preventive health behaviors, including vaccinations and health check-ups.

²¹Murphy et al. (2016) tested water samples which were properly chlorinated and kept in normal conditions in rural Kenyan homes. 17 percent of samples kept in jerry cans had no measured TCR after 48 hours. In clay pots, used by most of our sample, TCR had disappeared in 51 percent of samples after 48 hours.

4.2 Savings, labor supply and other economic outcomes

The ability to visualize future benefits and planning skills are relevant for many behaviors. Our secondary hypothesis for behavioral variables is that our interventions (which were domain-general) affect other investment behaviors, in particular savings, education, and labor supply. During both short-run and long-run surveys, participants therefore completed several modules on economic behaviors. We pre-specified secondary outcomes measuring savings and labor supply in both rounds, as well as an index of investment into children's education in the 10-week follow-up.

4.3 Targeted psychological channels

We hypothesized three mechanisms for the effect of our interventions on behavior: the Visualization intervention targeted time preferences and self-efficacy, while the Planning intervention targeted planning skills and self-efficacy. We use incentivized choice tasks or validated psychological scales to measure these constructs.²²

Planning skills: Our first primary hypothesis in relation to mechanisms is that the intervention alters whether people make plans and follow through on them. The primary outcome is a short form of the Behavioral Activation for Depression Scale (Kanter et al. 2007; Manos, Kanter, and Luo 2011), capturing agreement with items about setting and following through on goals (e.g. "I was an active person and accomplished the goals I set out to do"). As a secondary outcome, we conduct an incentivized Tower of London task (ToL), which measures a participant's higher-order cognitive ability to plan ahead in sequential strategies (Shallice 1982; Phillips et al. 2001). Appendix G.1 provides further details.

Time preferences and utility forecasting: Our second primary hypothesis is that the intervention affects time preferences. In the 10-week survey, we estimate time preferences in the effort domain, 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 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. Each effort task consisted of sending a 30-digit random number string by SMS to a toll-free number.²³ Participants choose how many SMS they want to send 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. We structurally identify a present bias parameter (β), the primary outcome, as well as an impatience parameter (δ), assuming quasi-linear utility and a power cost of effort function (Augenblick 2017; DellaVigna and Pope 2017). In addition to the effort discounting task, we include monetary discounting parameters from a conventional Multiple Price List (MPL) (Andreoni and Sprenger 2012) as secondary outcomes. Delays included today, four weeks, and eight weeks. Appendices G.2 and G.3

²²We validated the Tower of London task, Monetary Price Lists for time preferences, and the self-efficacy scale in Kenya, testing internal consistency, test-retest reliability, and construct validity (Esopo et al. 2018).

²³All participants 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.

provide full details on the estimation. In the long-run survey, we administer a shorter form of a multiple price list using the “staircase procedure” and a qualitative question on patience from the Global Preference Survey (GPS) (Falk et al. 2018).

In the long-run follow-up, we designed a measure to capture how well people can form a precise mental image of their future, adapted from the Plymouth Sensory Imagery Questionnaire (Andrade et al. 2014). We asked respondents to form an image in their mind of themselves and their family in one year, and then rate the clarity and vividness of the image. The measure proxies for utility forecasting noise $\sigma_{\epsilon_t}^2$ in Gabaix and Laibson (2017) (see Section 3.1).²⁴ It is our primary psychological outcome in the long-run survey (along with self-efficacy, below). We discuss how our measures of intertemporal choice relate to each other in Section 7.

Self-Efficacy: Our final potential mechanism is self-efficacy, the belief that one is able to achieve desired outcomes (see footnote 3). We use a widely used psychological scale, the Generalized Self-Efficacy (GSE) scale (Schwarzer and Jerusalem 2010). Participants rate their agreement (on a scale from 1 to 5) with 12 statements such as “I always manage to solve difficult problems if I try hard enough” and “It is easy for me to stick to my aims and accomplish my goals”. Responses are aggregated into a summary score and z-scored.

4.4 Alternative Mechanisms

We also measure alternative mechanisms and potential confounds. First, we measure beliefs and knowledge about chlorination. The addition of the information module to all active arms aims to ensure these are held constant. We verify this using multiple-choice questions referring to the information module. We measure participants’ beliefs about the effectiveness of chlorination in preventing pediatric diarrhea. We also measure knowledge about how to properly use chlorine with questions about how much chlorine to add to water, and how much time needs to pass after adding chlorine until the water is safe to drink.²⁵

Second, we test whether our interventions differentially focus participants’ attention on chlorination. 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 then asked her to recall as many words as possible 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.15). The task captures whether a concept is “top-of-mind”, and is conceptually similar to the audio word search task in Lichand and Mani (2016). We test whether our treatments differentially affect the probability to recall chlorine words, *conditional* on the total number of words remembered. See Appendix G for full specification details.

²⁴In line with the Gabaix and Laibson (2017) prediction that forecasting ability increases with intelligence, this measure correlates with participants’ education level ($t = 3.16$).

²⁵We similarly check for differential knowledge of WHO-recommended antenatal and postnatal care. These questions were included to pilot them for a future study.

Third, we measure risk preferences using a modified Eckel-Grossman task, which offered a choice between one of three 50/50 lotteries, represented as bets on a coin flip (Charness, Gneezy, and Imas 2013). One participant per 25 people was paid their choice.

5. Econometric approach

5.1 Main specification

Our main specification compares the two “psychological” treatment groups, V+INF and P+INF, to the active control group, AC+INF. We employ the following specification, separately for data collected 10 weeks and 33 months after the first intervention session:

$$y_{i1} = \alpha_0 + \alpha_1 T_{1i} + \alpha_2 T_{2i} + \phi X_i + \delta y_{i0} + \gamma_v + \varphi_w + \rho_d + \theta_a + \eta_i \quad (1)$$

where 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. X_i is a vector of participant controls (year of birth, employment status, marital status, and education level). γ_v are village fixed effects. φ_w and ρ_d are fixed effects for the survey week and weekday.²⁶ θ_a is an indicator for household asset wealth greater than the sample median (used for stratification). Standard errors are clustered by intervention cohort (five participants) to account for within-group dynamics. For variables collected at baseline (Table A.10), we include y_{i0} , the outcome at baseline.²⁷ We winsorize outcome variables with no theoretical lower and upper bounds at the 1st and 99th percentile. As pre-specified, the sample includes all participants who completed both the relevant survey (after 10 weeks or 33 months) and the combined baseline and first intervention session, and thus excludes the pure control group. Selection into the sample based on treatment is not possible because the nature of the intervention was not revealed before the first intervention session.

5.2 Multiple hypothesis testing

We calculate two sets of p -values: “per-comparison” p -values, which are appropriate for readers interested in a specific outcome, and sharpened “ q -values” (Benjamini, Krieger, and Yekutieli 2006), which adjust p -values for the false discovery rate (FDR) among groups of outcomes. We adjust for multiple hypothesis testing within outcome groups (behaviors and psychological mechanisms) and hierarchical outcome categories (primary, secondary, and exploratory), but not across them. We consider the effects of our two active interventions to be theoretically distinct and therefore do not correct across them. Appendix Table A.8 shows the hypothesis under which a given variable falls in each survey round. Indices are constructed following Anderson (2008).

²⁶Week and day of week fixed effects were not pre-specified, but were added to account for a slight imbalance in survey delay (see Section 6.1). We exclude these in a robustness check.

²⁷Where only some baseline observations are missing, we replace the missing values with the sample mean at baseline and add a dummy variable indicating such cases.

5.3 Comparison with the pure control group

In Appendix E, 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 respective survey (10 weeks or 33 months), including “non-compliers” who were assigned to the active treatment groups, but chose not to participate in the baseline survey or the interventions. The specification is identical to that in equation 1, except that there is a third treatment indicator T_{3i} for the active control group, the pure control group is used as the reference category, and the estimation does not control for the baseline outcome y_{i0} . This comparison gives the total effect on targeted behaviors of providing interventions such as ours in other, similar settings. In contrast, the active control comparison isolates the effects of the psychologically active elements of Visualization and Planning. The information module, exposure to field staff and other participants, and experimental payments are held constant across the three active arms and the restriction to baseline survey participants holds constant any practice effects on tasks.

6. Results

6.1 Experimental integrity

Table 1 shows randomization balance, differential survey participation, intervals between surveys, and compliance with treatment. The first panel shows that demographic variables are balanced across treatments, with only one out of 25 coefficients reaching statistical significance, at the 10 percent level. Appendix Table A.10 shows that baseline outcomes are balanced prior to interventions.

The second panel shows participation in each round. Defining attrition as a failure to complete the 10-week or 33-month surveys among those who attended baseline (the standard definition of attrition for most experiments), average attrition in the active control group was 8 percent in the 10-week endline, 12 percent in the chlorine test, and 10 percent in the long-run follow-up. Attrition is balanced across the active control, Planning and Visualization groups. Demographic variables predict attrition from all survey rounds, but not in interaction with treatment status, suggesting that the composition of the sample remains similar between groups in all rounds (Appendix Tables A.12 and A.13).

We put more weight on the active control comparison, where attrition by treatment group is balanced, than the pure control comparison. In the pure control comparison, there is more drop-out and there are some differences in participation in the endline and the chlorine test across groups. In the pure control group, 24 percent of people *in the recruitment census* did not complete the endline, 26 percent did not complete the chlorine test, and 16 percent did not complete the follow-up. Non-participation in surveys is higher as we cannot condition on baseline attendance in this sample, and thus include all participants from the recruitment census. The difference is important: The recruitment census was conducted door-to-door in the villages, and includes all women who meet our sampling

criteria (Section 2.1). The baseline and 10-week survey were conducted in mobile labs, typically 30 minutes from participants' home. Thus, the 24 percent of the pure control group who fail to participate in the endline are not 'attriters' in the conventional sense that they choose to drop out of the study – they are individuals who choose not to participate in the first place (perhaps because the mobile lab is too far, or they do not have childcare).

The third panel in Table 1 shows some small imbalances in the delay between the date of the baseline (and first intervention) to the subsequent rounds. Relative to the active control group, the Planning group completed the endline survey and chlorine test about 2 days later. We include a fixed effect for the survey week and survey weekday, for all outcomes in the short- and long-run survey (see equation 1). We present estimates without these fixed effects in Tables A.20 and A.21 and the results are robust.

The final panel in Table 1 shows compliance rates, which are balanced across the treatment groups. After the census, all respondents in the active arms were invited to the baseline and first intervention session, which were held on the same day. 78 percent of respondents attended the baseline and completed the first session, and 74 percent of completed both sessions.

6.2 Health-related outcomes

Table 2 shows results on behavioral outcomes, with health outcomes shown in the top panel. The column "Multiple Hypothesis Test (MHT) level" indicates whether variables were pre-specified as primary (1), secondary (2), or exploratory outcomes (3), or were not pre-specified (np). The first number relates to the outcome pre-specification in the 10-week survey; the second to the outcome pre-specification in the 30-36 month survey. The ranking of variables is given in Table A.8.

Chlorination: After 12 weeks, the Visualization intervention led to a significant increase in our primary behavioral outcome measure, the presence of chlorine in household drinking water, measured during unannounced household visits. Among the Visualization group, there is a 5 percentage point (22 percent) increase in the presence of Total Chlorine Residual, significant at the 5 percent level, relative to an active control mean of 23 percent. There is a small, statistically insignificant increase on chlorination in the Planning group. We cannot reject that the Visualization and Planning interventions have the same effect ($p = 0.15$). The results for free chlorine (FCR) are similar in magnitude and statistical significance. Both results are indicative of increased safety of household drinking water.

We see a similar pattern of treatment effects on a self-reported measure of whether households chlorinate water, after 10 weeks and 30-36 months. We asked households whether and how they generally treat water and recorded up to two treatments (see Section 4.1). In the Visualization arm, the percentage of respondents who report generally chlorinating water increases by 7 percentage points after 10 weeks and by 5 percentage points after 30 to 36 months. Both effects are significant at the 1 percent level. We see no significant effects in the Planning arm in either round. We can reject that the Visualization and Planning

interventions have the same effect in the 10-week survey, but not in the long-run follow-up ($p = 0.15$). Our findings are consistent with studies of long-term effects of point-of-use chlorination, which find chlorine use becomes a habit once households have learned to chlorinate and, potentially, become accustomed to taste (Luby et al. 2006; Luby et al. 2004).

We also find that our treatments increase boiling of water after 10 weeks, which suggests that increased chlorination did not crowd out other effective methods of water treatment (boiling was mentioned in our information module, see Section 3.3).²⁸ We find little evidence that the 2020 pandemic affected water treatment practices: 97 percent of households report treating their water in the same way than they did before March 2020.

Other health outcomes: The Visualization and Planning treatments both generate large and statistically significant reductions in the incidence of diarrhea among children after 10 weeks. Relative to the active control group, we find a 46 percent (0.12 episodes) reduction in diarrhea episodes per child under 15 in the last 3 months for Visualization and a 23 percent (0.06 episodes) reduction for Planning.²⁹ However, only the effect of Visualization survives multiple test corrections. We can reject that the Visualization and Planning treatments are equally effective in reducing diarrhea ($p = 0.09$). We find similar effects on a non-prespecified outcome, diarrhea per child under 5, although only the Visualization effect is statistically significant. We find no effects on other health outcomes, including vaccinations and pre-natal care visits, with the exception of a small decrease in the number of children under 15 who visited a healthcare provider in the last 3 months in the Visualization group, potentially due to reduced diarrhea incidence.

After almost three years, there is little persistence in the effects on diarrhea. One possible reason is that the short-run effects on boiling do not persist in either group. We also find suggestive evidence that the difference in short and long-run diarrhea results is linked to the seasonality of diarrhea in Kenya (Chao et al. 2019), and the fact that the surveys are in different seasons. Water chlorination is more likely to affect diarrhea in the rainy season than in the dry season.³⁰ The 10-week survey was conducted in February and March 2018, during rainy season. The 33-month survey ran from July–December 2020, with the majority of surveys conducted during dry season. Table A.1 shows treatment effect heterogeneity by whether the survey is conducted during rainy season (November–December 2020),

²⁸We see some increases over time in self-reported water treatment. In the short-run (long-run) follow-up, in the active control, 73 (85) percent report they generally chlorinate, 35 (63) percent generally boil water, and 27 (54) percent report using both methods. Only 20 (6) percent report no or only ineffective methods, which suggests that effective water treatment is relatively widespread in our study area, though it may not be used consistently. Time trends in water treatment may be due to changing social norms, or increased public health awareness during the Covid-19 pandemic, including public health campaigns. However, we cannot disentangle time effects in underlying behavior from the effect of adapting questions to a phone survey format (see Section 4.1). Estimates of treatment effects as a percentage of the mean can be compared between rounds.

²⁹We pre-specified diarrhea in children under 15 (as opposed to younger age groups) to maximize sample size, and because children in a wide age range suffer negative health effects from ongoing enteric dysfunction.

³⁰In the two rainy seasons (February–May and November–December), one main cause of diarrhea is rain washing faeces into water sources, contaminating water. In the dry season (June to September), rotavirus, a major cause of diarrhea, becomes more prevalent. It is transmitted through air and human contact. See Luby et al. (2006) and Arnold and Colford Jr (2007) on the importance of seasonality in interpreting the effects of chlorination interventions on diarrhea.

and by whether the household's water source is protected.³¹ We find no effects of our interventions during the dry season, but some effects during rainy season. During the rainy season, Visualization decreases the probability that any child under 15 in the household had diarrhea in the last 7 days by 10 percentage points, though we lack power to detect significance. This effect increases to 34 percentage points ($p < 0.05$) for those whose water source is not protected.

The short-run treatment effects of our interventions on diarrhea are rather large in relation to the effects on chlorination; with an effect on diarrhea of -46 percent and an effect on the presence of any chlorine of $+22$ percent in the Visualization arm. A back-of-the-envelope calculation suggests an instrumental variable estimate of -2.09 percent, i.e. a 1 percent increase in chlorination leads to a 2.09 percent reduction in diarrhea episodes. This estimate is within the confidence interval of a recent meta-analysis, which considers effects of dispensers or distribution of free chlorine (Arnold and Colford Jr 2007).³² There are a few possible reasons for our relatively large effects. First, our interventions affect both chlorination and boiling of water. Second, they are domain-general trainings that may affect other behaviors related to child diarrhea, such as handwashing, washing fruit and vegetables, open defecation and general hygiene. Third, there could be dependencies with the treatment effects on non-health outcomes, such as the observed increases in savings (described below).

Concerns about demand effects in chlorination measures: Our objective chlorination measures, obtained during unannounced household visits two weeks after the 10 week survey, leave little room for experimenter demand effects, such as social desirability bias in responding to survey questions. It was also difficult for households to anticipate our visits for chlorine tests. 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.19 shows that the estimated treatment effects on chlorine are unaffected by including a dummy for being tested on the first day within one's village, or the number of days elapsed since the first chlorination tests were conducted in the village. When we include interactions, we find that participants in the Visualization group are (insignificantly) *more* likely to have chlorine in their water if tested on the first day in their village – the opposite of what would be expected if knowledge about chlorination visits spreads. Finally, we see effects on child diarrhea in the 10-week endline survey, two weeks before the chlorine test, suggesting increases in chlorination had already occurred. For the same reason, it is unlikely that our treatment effects are driven by the 10-week survey itself acting as a reminder to chlorinate by the 12-week test. First, we see effects on diarrhea and self-reported chlorination in the 10-week survey. Second, questions about water treatment are included in surveys to all four

³¹We elicited 13 types of water sources following WHO guidelines in the 33-month survey, so this is potentially endogenous. We classify as unprotected: surface water, rainwater, unprotected springs and wells.

³²Across six randomized controlled trials which measure both outcomes, a 1 percent increase of detectable free chlorine is correlated with a 1 percent reduction in child diarrhea incidence in the past two weeks, with confidence intervals from a 2.3 percent decrease to a 0.5 percent increase.

groups. Third, we do not see an effect on the active control group, suggesting the survey does not remind participants of our information module on chlorination (Table A.5).

Concerns about demand effects may be more plausible in our self-reported measures of chlorination. Treatment effects across the objective and self-reported measures are strikingly consistent, which is suggestive evidence that households are reporting truthful answers. To further investigate the possibility of experimenter demand effects in self-reports, we include randomized demand treatments (de Quidt, Haushofer, and Roth 2018) in the follow-up survey (at the end, orthogonal to our psychological treatments), which reveal the objective of the experimenter to the participants.³³ This has a precisely estimated zero effect on self-reported chlorination behavior (difference of means -0.006 SD, $p = 0.92$, see Table A.18). Thus, while we do not rule out demand effects in self-reports, we do not find any evidence for them.

6.3 Savings, labor supply and other economic outcomes

The second panel of Table 2 shows effects on savings-related outcomes. After 10 weeks, the Visualization treatment leads to a significant increase in our main savings-related outcome variable, the amount of money saved regularly (converted to savings contributions per week). This effect corresponds to a 26 percent (KES 24) increase relative to the active control group, significant at the 5 percent level. After three years, the treatment effect on weekly savings is 15 percent (KES 60).³⁴

In addition, we find large and significant effects on all pre-specified savings outcomes in the Visualization group, relative to the active control group. Savings on the extensive margin increase, with the share of respondents who save regularly increasing by 13 percentage points (36 percent) after 10 weeks and 5 percentage points (7 percent) after three years.³⁵ Similarly, we find increases in the Visualization group on an indicator for whether the respondent saves for productive investments (education, business, agriculture) after 10 weeks, although this effect does not persist. Visualization participants have joined more ROSCAs in both rounds. Their weekly savings contributions to ROSCAs increase by 32

³³ Respondents are assigned to a group A (B), and told “We hypothesize that people who participated in this study and received the same treatment as you will give higher (lower) responses to these questions than others.” They are then asked how often they added chlorine to water collected from their primary source in the last 7 days. Following de Quidt, Haushofer, and Roth (2018), the responses can be used to obtain bounds $a^+(\zeta)$ and $a^-(\zeta)$ for the impact of experimenter demand effects on self-reports. We do not reject the equality $a^+(\zeta) = a^-(\zeta)$, neither for the overall sample ($p = 0.92$) nor within any of the treatment groups (Table A.18).

³⁴ We made pre-specified adjustments to how the savings variable is elicited between rounds, which likely explains the sizable increase in means (94 KES in the 10-week endline vs. 408 KES in the long-run follow-up). The 10-week survey asked respondents how often they put money aside for a given savings *purpose*, and how much each time. The variable “Amount saved regularly (per week, KES) is obtained by converting to weekly frequency, and summing across savings purposes. In the long-run survey, we ask how often they put money aside in a given savings *place* (safe hiding place, bank account, ROSCA...), and how much each time. This will yield a more comprehensive measure if people also saved for unspecified purposes. We suspect that participants included ROSCA savings when asked by place in the long-run survey, but not when asked by purpose after 10 weeks: Given the inflexibility of ROSCA schedules to individual needs, ROSCA savings are more likely to be for unspecified purposes. We also report ROSCA savings separately in both rounds. If our interpretation is correct, total weekly savings are 300 KES in the short-run, and 408 KES in the long-run.

³⁵ As before, the difference in means is likely explained by a pre-specified change in the elicitation method.

KES (16 percent) after 10 weeks and 41 KES (19 percent) after three years (see footnote 34 on how this relates to overall savings).

After almost three years, the Visualization group has accumulated 41 percent higher savings balances. The long-term survey occurred during the 2020 global pandemic, an economic shock which reduced real income in rural Kenya by 7.9 percent (Nechifor et al. 2020). Our results may suggest that participants who learned visualization techniques in 2017 benefited from an increased financial buffer in 2020.³⁶ There are no significant effects of the Planning treatment on savings. Together, these results show that the Visualization treatment strongly affected savings behavior and effects persist over time.

The third panel of Table 2 reports effects on labor-related outcomes. Somewhat surprisingly, we find a reduction of 23 hours (22 percent) in the total number of hours worked in the last 3 months in the Planning group, and a similar effect on the number of days worked. The Planning group may have improved their daily scheduling of activities and been more efficient. Respondents work mainly in subsistence farming or self-employment, and thus their income is determined largely by their output rather than their work hours. Indeed, we find no significant change in monthly earnings, despite the shortened work hours. However, the effect is temporary, and completely disappears by the three year follow-up. In contrast, we find a highly significant positive effect of Visualization on hours worked in the long-run, of 2.3 hours worked (18 percent) in the last 7 days. The Visualization intervention focused on the future in one year, and some participants may have used this to develop longer-term career visions, including applying for salaried jobs. But these longer term visions may not have affected labor supply in the short term. We see no effects on an index of children's education investment after 10 weeks, and did not measure it in the long-term follow-up.

Our findings on non-health behaviors have two important implications. First, our Visualization intervention strongly affected future-oriented behaviors across different domains, in both short- and long-run. We discuss evidence on potential mechanisms in Section 7. Second, it provides further evidence against experimenter demand effects: while the interventions mentioned chlorination and health-related topics, savings and work 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 an increased valuation of the future rather than from a simple desire to please the experimenters.

6.4 Treatment effects when the costs of chlorination vary

We examine whether our treatment effects differ by whether or not the village in which the interventions took place was randomly allocated chlorine dispensers in the WASH Benefits study.³⁷ Dispensers reduce the monetary and non-monetary costs of chlorination, by pro-

³⁶Savings behavior has changed since the pandemic. About 11 percent of respondents report they reduced their regular savings since the start of public health restrictions in Kenya (March 2020), with an average conditional reduction of 73 percent in contributions. Treatment effects in the absence of the pandemic may have been different.

³⁷We verify the WASH randomization remains intact. 70 percent of active control participants in dispenser villages in the 33-month survey report they have access to a working dispenser at their primary or secondary

viding free chlorine at specific water sources, and by reminding people to chlorinate when collecting water. Chlorination rates are slightly higher in dispenser than in non-dispenser villages, both in objective and self-reported data, and use of boiling is slightly lower (Table 3). For young children, rates of diarrhea are slightly higher in non-dispenser than dispenser villages.³⁸

Table 3 shows results from the main estimating equation, focusing on chlorination-related outcomes, separately for non-dispenser villages (columns (1)–(4)) and dispenser villages (columns (5)–(8)). The interaction terms on our two treatment groups with the randomized dispenser treatment are shown in columns (9) and (10). We find little evidence of heterogeneity in treatment effects across dispenser and non-dispenser villages, in either the short or long term. The Visualization treatment leads to a 5 percentage point increase in positive chlorine tests after 12 weeks in non-dispenser villages and a 4 percentage point increase in dispenser villages. Visualization increases self-reported use of chlorination and boiling to a similar extent in dispenser and non-dispenser villages. Consistent with increases in water treatment, Visualization decreases diarrhea incidents among children under 15 and under 5, with few differences across dispenser and non-dispenser villages. After 33 months, effects on self-reported chlorination persist; effects on diarrhea do not, as occurs in the average effects (potentially relating to seasonality, see Section 6.2). In the 10-week survey, Planning has smaller and mostly insignificant effects on chlorination and leads to a small increase in boiling decreases in diarrhea, regardless of village dispenser status. No effects persist in the 33-month follow-up.

The similarity in effects across dispenser and non-dispenser villages suggests that monetary costs may not be the binding constraint to chlorination for the households who respond to our interventions. If costs were a strong barrier, psychological interventions might have *smaller* effects in non-dispenser villages, where using chlorine is somewhat more costly. Instead, we find effects on chlorination are the same in absolute terms in non-dispenser and dispenser villages. In relative terms, they are even slightly *larger* in non-dispenser villages, as fewer households chlorinate in non-dispenser villages to begin with.

In search of further evidence, we explore where households obtained chlorine at the bottom of Table 3. Households in non-dispenser villages are still fairly easily able to access bottled chlorine, which is cheap and widely available: 47 percent reported adding bottled chlorine to their current water, compared to 31 percent in dispenser villages. Households in non-dispenser villages are less likely to report using dispenser chlorine (21 percent compared to 54 percent in dispenser villages). We see suggestive evidence that increases in chlorination are driven somewhat more by increases in use of bottled chlorine, in both types of villages. This is plausible: there are more barriers to use of bottled chlorine – although it is cheap, households must remember to buy and add it. However, we are

water source, and 82 percent have a dispenser within 30min walking distance. This proportion is lower, 34 and 49 percent respectively, in non-dispenser villages. These figures suggest that some non-dispenser villagers can still access dispensers, potentially because they live at the edge of dispenser villages.

³⁸Appendix Table A.11 shows that participants in dispenser and non-dispenser villages were comparable on baseline characteristics, although in dispenser villages they had a slightly higher level of education.

cautious about this conclusion due to concerns about power.

7. Mechanisms

Our interventions targeted three psychological mechanisms, registered in our first pre-analysis plan: First, the Visualization intervention encouraged participants to connect everyday behaviors to distant future outcomes, and to build their ability to make these future outcomes vivid and tangible in their minds. It thus aimed to increase patient behavior, both through valuation of the future (time preferences), and through strengthening the mapping between current behavior and future outcomes (captured in self-efficacy, or domain-general returns to effort, see footnote 3). Second, our Planning intervention aimed to build planning skills by teaching participants to structure tasks, break them down into individual steps, and approach them in a way that induces self-reinforcing motivation cycles. The intervention may also build self-efficacy by strengthening participants' beliefs that desired outcomes are within their reach.

7.1 Targeted psychological mechanisms

Table 4 shows results on pre-specified psychological mechanisms, estimated using equation 1. All outcomes are described in Section 4.3. The raw means and standard deviations of all outcome measures that were z-scored are reported in Appendix Table A.16.

Planning skills: In the first panel of Table 4, we find no statistically significant effects of any treatment on planning skills, measured by the Behavioral Activation score (BADs) and the Tower of London task (ToL), compared to the active control group after 10 weeks. However, the coefficients in the Planning treatment have the expected signs (positive for BADs, and negative for the number of moves required to complete the ToL task).

Time preferences and utility forecasting: The second panel shows that the Visualization intervention has few effects on conventional measures of time preferences, but appears to improve utility forecasting. Our primary outcome in the 10-week follow-up is β^{Effort} , which captures present bias in our effort task (see Section 4.3 and Appendix G.2 for details of the structural estimation). Secondary outcomes are the δ^{Effort} parameter from the same task, and corresponding parameters from the monetary discounting task. We find no statistically significant effects of the Visualization treatment on the estimated preference parameters.³⁹

In the long-run survey, we measure outcomes more finetuned to the Gabaix-Laibson (2017) model of intertemporal choice (Section 3.1), namely participants' ability to forecast future utility. Our primary pre-specified psychological outcome for this survey round captures how vivid participants' images of themselves and their family are in one year's time,

³⁹Somewhat surprisingly, we find an (economically small -0.002) reduction of the δ^{Effort} parameter in the Planning 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.4). This result is consistent with the Planning intervention helping participants to plan the logistics of future effort provision (SMS data entry), such as phone access or child care (which is possible for 7 or 8 days, but perhaps too short notice for 0 or 1 days).

and is our proxy for forecasting noise $\sigma_{\epsilon_t}^2$ in the Gabaix-Laibson model. We find that participants in the Visualization group report significantly more clear and vivid utility simulations after almost three years. There is no effect in the Planning group. A secondary outcome asks participants whether they typically mentally imagine the consequences of decisions when making everyday choices (the “practice” of utility forecasting). This measure relates to both the Gabaix-Laibson model (regular practice of forecasting may improve forecasting ability and reduce forecasting noise), as well as to the concept of self-efficacy (the belief that one’s behavior affects future outcomes). We find significant increases in both Visualization and Planning treatments. A consistent interpretation is that effects on behavior are driven by forecasting ability, and that the regular practice of utility forecasting by itself (without an associated increase in forecasting skill) is not sufficient to generate behavioral change.

Given our utility forecasting results, it may seem puzzling that we do not observe an effect on conventional measures of time preferences: An increased weight of the future relative to the present should be captured in our estimates of β^{effort} . However, there are important differences between the Gabaix-Laibson model and conventional models of quasi-hyperbolic discounting (“ $\beta\delta$ –models”): While both models can generate hyperbolic discounting patterns, $\beta\delta$ –models assume the existence of fundamental and stable parameters (β and δ). In contrast, the Gabaix-Laibson model stipulates that patience is inherently unstable, as the forecasting noise $\sigma_{\epsilon_t}^2$ can vary with cognitive load, time spent thinking about a problem, and how cognitively well-simulated the relevant future period is. Thus, details of the choice frame and setting matter.

Importantly, the 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. In contrast, 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 4.3). It is plausible that the intervention made far-future utility from abstract rewards (e.g. having healthy children in one year) more tangible and salient, while the disutility from near-future effort (SMS data entry within the next 8 days) was already tangible at baseline. This could explain the observed changes in chlorination and savings behavior without affecting our lab measures of β and δ .

We investigate the importance of the time horizon further in the long-run follow-up by including a hypothetical monetary discounting task (Falk et al. 2018), which estimates the discount factor between today and 12 months from now. We find no effects on this measure, but note that the informational value is compromised by the fact that only 34 percent of respondents had one-year discount factors in the measurable range $[0.46, 1]$,⁴⁰ as well as by standard concerns about hypothetical and monetary discounting measures (Andreoni and Sprenger 2012). However, we do find that the Visualization intervention leads to a 0.09 SD increase in the GPS qualitative time preference measure (reported willingness to give up current benefits to obtain future rewards). The effect is significantly larger than in the

⁴⁰We use the official stakes for Kenya from the GPS, (Falk et al. 2018).

Planning group ($p < 0.01$). We conclude there is limited evidence of effects on conventional time preference measures.

Self-efficacy: The third panel of Table 4 shows the effect of our interventions on self-efficacy. After 10 weeks, both the Visualization and the Planning interventions generate statistically significant 0.15 SD and 0.11 SD increases, respectively. The effect in the Visualization group is 36 percent larger than that in the Planning group, though the difference is not statistically significant. We find evidence that the effect of Visualization on self-efficacy persists in the long-run: After almost three years, this group has 0.14 SD higher self-efficacy than the Planning group ($p < 0.01$), and 0.07 SD higher self-efficacy than the active control ($p = 0.10$). The result is robust to multiple hypothesis test correction.⁴¹

Finally, we present simple predictive regressions of key chlorine and savings outcomes on our psychological measures in Tables A.2 and A.3. These do not have a causal interpretation, as many of the regressors are endogenous to treatment. We find that our chlorination measures are robustly and significantly linked with self-efficacy across all survey rounds. There is a positive but insignificant association with utility forecasting. In contrast, savings measures are strongly predicted by both utility forecasting and self-efficacy. For both chlorination and savings, other psychological mechanisms have little predictive power. We conclude that both self-efficacy and utility forecasting are likely drivers of our treatment effects, with self-efficacy playing a larger role for chlorination, and utility forecasting playing a relatively larger role for savings.

In short, we find plausible evidence of which psychological mechanisms – utility forecasting and self-efficacy – are activated by our treatments, in particular by the Visualization treatment. These mechanisms are correlated with chlorination and savings in the absence of treatment. There is persistence in effects, despite the light-touch nature of the interventions.

7.2 Beliefs, knowledge, and risk preferences

We test a range of alternative psychological channels, with results shown in Table 5. The first panel shows results on beliefs and knowledge about chlorination. Beliefs in the efficacy of chlorine in averting diarrhea, as well as knowledge about how to correctly chlorinate water, are similar in the Visualization, Planning and active control groups. This is consistent with the fact that all active arms received the information module. Importantly, beliefs and knowledge do increase in all active arms, relative to the pure control group (see Table A.7). Thus, the information module was somewhat effective. Correlations of beliefs and knowledge with TCR after 12 weeks are shown in Table A.4.

We see effects on beliefs and knowledge in all three active arms, but only Visualization has statistically significant effects on chlorination and diarrhea. Thus, the Visualization treatment has additional effects on behavior compared to the effect of information on its own. These are likely due to changes in psychological outcomes discussed above. Finally, we find no effects on risk aversion, suggesting that any behavioral effects are unlikely to

⁴¹The sharpened q values are less than unadjusted p -values. This can occur when many hypotheses in the outcome group are rejected (Anderson 2008).

result from changes in risk preferences induced by our treatments.

7.3 Saliency

Section 4.4 explains the design of a test for increased saliency of chlorination, and Appendix G explains the econometric specification. The bottom panel of Table 5 shows that, in the 10-week endline, participants who had received the Visualization or Planning intervention found it easier to remember chlorine-related words, *conditional* on the total number of words remembered. In the long-run survey, this effect persists only for Visualization.⁴²

There are three possibilities for the role of saliency in our study. First, the interventions may cause participants to pay more attention to chlorination, and thus chlorinate more (i.e., saliency is one mechanism which accounts for our treatment effects). Although they were domain-general, Visualization and Planning interventions use water treatment as an example behavior in engaging stories or exercises. Second, participants may chlorinate more, and this makes chlorination more salient to them (i.e., reverse causality). Third, the interventions may change the saliency of chlorination, but this may not relate to behavior change (i.e., those for whom saliency responds are not those who chlorinate).

We find evidence consistent with the second and third explanations, but little evidence consistent with the first explanation. Consistent with the second explanation, effects on both chlorination and the saliency of chlorination persist only in Visualization. However, any chlorine references in the Visualization and Planning group were identical. Furthermore, saliency effects in Visualization appear to *increase* over time, suggesting they are not a direct response to the intervention. Consistent with the third explanation, there is no correlation between the saliency of chlorination and chlorination behavior (Table A.4). Moreover, we observe changes in behavior without a change in saliency: despite persistent effects on savings measures, the saliency of savings is unaffected. We tentatively conclude that saliency is less likely as a mechanism than other psychological mechanisms we have identified.

8. Conclusion

Behavioral constraints may explain low demand for preventive health products, as well as other behaviors requiring participants to incur current costs to secure future benefits. Potential constraints may include people's ability to visualize future benefits of investments, or their ability to make concrete plans. We study whether two light-touch interventions, one targeting each of these constraints, affect behavioral and psychological outcomes among young women in Kenya. We conduct an in-person survey and incentivized choice tasks after 10 weeks, and test drinking water for the presence of chlorine in unannounced household visits after 12 weeks; we follow up by phone after 30 to 36 months.

⁴²The interventions are unlikely to have affected participants' memory in general, proxied by the total number of words remembered. There is a small increase of 0.2-0.3 words remembered in the long-run survey, but this likely reflects that participants remembered chlorine words *in addition* rather than *instead* of filler words. All saliency regressions condition on the total number of words remembered.

The Visualization intervention is more effective than the Planning intervention in the short and particularly in the longer term. The Visualization intervention increases chlorination in a water test after 12 weeks and self-reported chlorination after both 10 weeks and 30 to 36 months, relative to an active control group. We observe reductions in child diarrhea in the short but not in the long term. Visualization also increases savings in the short and long term and labor supply in the long term. The Planning intervention, in contrast, has few significant effects. We estimate the Visualization intervention is highly cost-effective relative to WHO benchmarks. Possible mechanisms for behavioral effects include self-efficacy (participants' belief that their current behavior affects future outcomes) and utility forecasting (participants' ability to mentally simulate future utility (Gabaix and Laibson 2017)).

Our results suggest that visualization-based interventions might be effective in increasing take-up of other preventive health products and be a valuable focus of government policy. Our modules were administered by field workers with no specific training. It would thus be straightforward to incorporate them into the curricula that community health workers (CHWs) administer during household visits or community education sessions to promote preventive health products. Instead of domain-general content like ours, practitioners may choose to further focus the exercises on specific target behaviors. The regularity of the work of CHWs might also overcome the fact that our effects on diarrhea fade out after three years – effects may persist if the intervention is refreshed periodically. In the context of recent work raising concerns about the efficacy of CHW programs in improving chlorination take-up (Dupas et al. 2020), our results suggest that improving the content of preventive health promotion interventions might enhance their effectiveness at a low cost.

The observed effects on savings outcomes and labor supply suggest that visualization-based interventions might be used in other domains where policy-makers wish to encourage forward-looking behavior. Such interventions could form part of the training for members of ROSCAs or Village Savings and Loan Associations (VSLAs). They could also be used to enhance repayment discipline among borrowers of microfinance institutions. Furthermore, in ongoing work by Ashraf et al. (2020), visualization is incorporated in an intensive business skills training for entrepreneurs in Colombia.

The experiment suggests a number of areas for future research. A key one is to better understand and measure the mechanisms behind visualization-based interventions. Task-based measures of utility forecasting ability might be particularly valuable. Conducting water testing in longer-term follow-ups, outside a pandemic context, may further increase the credibility and external validity of the intervention. A second potential area of research is to examine if “topping-up” the intervention with repeat sessions ensures that effects on health behaviors persist. In other trials, reminders improved adherence to existing health programs (Pop-Eleches et al. 2011). A third potential area of future research may be to explore whether domain-general planning interventions can be better adapted to target populations, or whether planning skills do not constitute a binding constraint to future-oriented behavior.

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.
- Andrade, Jackie, Jon May, Catherine Deeproose, Sarah-Jane Baugh, and Giorgio Ganis. 2014. "Assessing Vividness of Mental Imagery: the Plymouth Sensory Imagery Questionnaire." *British Journal of Psychology* 105 (4): 547–563.
- 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, Gharad Bryan, Alexia Delfino, Emily Holmes, Leonardo Iacovono, and Ashley Pople. 2020. "Learning to See the World's Opportunities: The Impact of Visualisation on Entrepreneurial Success." *Pre-Analysis Plan, AEARCTR-0004695, October 2020*.
- Ashraf, Nava, Kelsey Jack, and Emir Kamenica. 2010. "Information and Subsidies: Complements or Substitutes?" *Journal of Economic Behavior and Organization* 88:133–139.
- Atkin, David. 2013. "Trade, tastes, and nutrition in India." *American Economic Review* 103 (5): 1629–63.
- 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.
- . 1997. *Self-Efficacy: The Exercise of Control*. New York: W.H. Freeman.
- Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2020. "Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial." *American Economic Review* 110 (3): 824–59.
- Baranov, Victoria, Johannes Haushofer, and Channing Jang. 2020. "Can Positive Psychology Improve Psychological Well-Being and Economic Decision Making? Experimental Evidence from Kenya." *Economic Development and Cultural Change* 68 (4): 000–000.
- Becker, Gary S, and Casey B Mulligan. 1997. "The Endogenous Determination of Time Preference." *The Quarterly Journal of Economics* 112 (3): 729–758.
- 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. 2019. "The Future in Mind: Short and Long-Run Impact of an Aspirations Intervention in Rural Ethiopia." *Working Paper, University of Oxford*.
- Blanton, Elizabeth, Sam Ombeki, Gordon Otieno Oluoch, Alex Mwaki, Kathleen Wannemuehler, and Rob Quick. 2010. "Evaluation of the Role of School Children in the Promotion of Point-of-Use Water Treatment and Handwashing in Schools and Households: Nyanza Province, Western Kenya, 2007." *American Journal of Tropical Medicine and Hygiene* 82 (4): 664–671.
- 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.
- Chao, Dennis, Anna Roose, Min Roh, Karen Kotloff, and Joshua Proctor. 2019. "The Seasonality of Diarrheal Pathogens: A Retrospective Study of Seven Sites over Three Years." *PLOS Neglected Tropical Diseases* 13 (8): 1–20 (08).

- 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.
- de Quidt, Jonathan, Johannes Haushofer, Frank Schilbach, Gautam Rao, and Pierre-Luc Vautrey. 2020. "The Economics Consequences of Depression: Evidence from Goa." *Pre-Analysis Plan, AEARCTR-0003823, October 2020*.
- Dillon, Andrew, Jed Friedman, and Pieter Serneels. 2020. "Health Information, Treatment, and Worker Productivity." *Journal of the European Economic Association*, 06. jvaa024.
- 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.
- Dupas, Pascaline. 2009. "What Matters (and What Does Not) in Households' Decision to Invest in Malaria Prevention?" *American Economic Review* 99 (2): 224–230.
- . 2014. "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.
- Dupas, Pascaline, and Edward Miguel. 2017. "Impacts and Determinants of Health Levels in Low-Income Countries." In *Handbook of Economic Field Experiments*, Volume 2, 3–93. Elsevier.
- Dupas, Pascaline, Basimenye Nhlema, Zachary Wagner, Aaron Wolf, and Emily Wroe. 2020. "Expanding Access to Clean Water for the Rural Poor: Experimental Evidence from Malawi." *NBER Working Paper No. 27570*.
- 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 (6): e100100.
- Esopo, Kristina, Daniel Mellow, Catherine Thomas, Hannah Uckat, Justin Abraham, Prachi Jain, Channing 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.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–1692.
- 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*.
- Genicot, Garance, and Debraj Ray. 2020. "Aspirations and Economic Behavior." *Annual Review of Economics* 12 (1): 715–746.
- Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy. 2020. "Sex Workers, Stigma and Self-Image: Evidence from Kolkata Brothels." *Review of Economics and Statistics*, vol. Advance.
- 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.

- Gupta, Sundeep, Astrid Suantio, Alicia Gray, Endang Widyastuti, Neena Jain, Reineke Rolos, Robert Hoekstra, and Rob Quick. 2007. "Factors Associated with E. Coli Contamination of Household Drinking Water Among Tsunami and Earthquake Survivors, Indonesia." *The American Journal of Tropical Medicine and Hygiene* 76 (6): 1158–1162.
- 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.
- Hershfield, Hal, Elicia John, and Joseph S Reiff. 2018. "Using Vividness Interventions to Improve Financial Decision-Making." *Policy Insights from the Behavioral and Brain Sciences* 5 (2): 209–215.
- Imas, Alex, Michael Kuhn, and Vera Mironova. 2020. "Waiting to choose." *Working Paper*.
- 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.
- 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): 191–202.
- Kremer, Michael, and Rachel Glennerster. 2011. "Improving Health in Developing Countries: Evidence from Randomized Evaluations." In *Handbook of Health Economics*, Volume 2, 201–315. Elsevier.
- 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.
- 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.
- Lichand, Guilherme, and Anandi Mani. 2016. "Cognitive Droughts." *Working Paper, Harvard University*.
- Loewenstein, George, Joseph Price, and Kevin Volpp. 2016. "Habit Formation in Children: Evidence from Incentives for Healthy Eating." *Journal of Health Economics* 45:47–54.
- Lorig, Kate, Diana Laurent, Kathryn Plant, Eswar Krishnan, and Philip Ritter. 2014. "The Components of Action Planning and Their Associations With Behavior and Health Outcomes." *Chronic Illness* 10 (1): 50–59.
- Luby, Stephen, Mubina Agboatwalla, Robert Hoekstra, Mohammad Rahbar, Ward Billhimer, and Bruce Keswick. 2004. "Delayed Effectiveness of Home-Based Interventions in Reducing Childhood Diarrhea, Karachi, Pakistan." *American Journal of Tropical Medicine and Hygiene* 71 (4): 420–427.
- Luby, Stephen, Mubina Agboatwalla, John Painter, Arshad Altaf, Ward Billhimer, Bruce Keswick, and Robert Hoekstra. 2006. "Combining Drinking Water Treatment and Hand Washing for Diarrhoea Prevention: A Cluster Randomised Controlled Trial." *Tropical Medicine & International Health* 11 (4): 479–489.
- 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 Unicom, 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.
- 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.
- Murphy, Jennifer, Tracy Ayers, Jacqueline Knee, Jared Oremo, Aloyce Odhiambo, Sitnah Faith, Ronald Nyagol, Christine Stauber, Daniele Lantagne, and Robert Quick. 2016. "Evaluating Four Measures of Water Quality in Clay Pots and Plastic Safe Storage Containers in Kenya." *Water Research* 104:312–319.
- Nechifor, Victor, Emanuele Ferrari, Evelyne Kihiu, Joshua Laichena, Daniel Omany, Rogers Musamali, and Benson Kiriga. 2020. "COVID-19 Impacts and Short-Term Economic Recovery in Kenya." *Publications Office of the European Union, Luxembourg*.
- 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.
- Orkin, Kate, Rob Garlick, Mahreen Mahmud, Richard Sedlmayr, Johannes Haushofer, and Stefan Dercon. 2020. "Assets, Aspirations and Anti-Poverty Policies." *Working Paper, University of Oxford*.
- Ozier, Owen. 2018. "Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming." *American Economic Journal: Applied Economics* 10 (3): 235–62.
- Patel, Vikram, Benedict Weobong, Helen Weiss, Arpita Anand, Bhargav Bhat, Basavraj Katti, Sona Dimidjian, Ricardo Araya, Steve Hollon, Michael King, et al. 2017. "The Healthy Activity Program (HAP), a Lay Counsellor-Delivered Brief Psychological Treatment for Severe Depression, in Primary Care in India: A Randomised Controlled Trial." *The Lancet* 389 (10065): 176–185.
- 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.
- Reich, John, and Frank Infurna. 2016. *Perceived Control: Theory, Research, and Practice in the First 50 Years*. Oxford University Press.
- 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.
- Richards, David A, David Ekers, Dean McMillan, Rod S Taylor, Sarah Byford, Fiona C Warren, Barbara Barrett, Paul A Farrand, Simon Gilbody, Willem Kuyken, et al. 2016. "Cost and Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial." *The Lancet* 388 (10047): 871–880.
- Schilbach, Frank. 2019. "Alcohol and Self-Control: A Field Experiment in India." *American Economic Review* 109 (4): 1290–1322.
- Schwarzer, Ralf, and Matthias Jerusalem. 2010. "The General Self-Efficacy Scale (GSE)." *Anxiety, Stress, and Coping* 12:329–345.
- Shallice, Tim. 1982. "Specific Impairments of Planning." *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 298 (1089): 199–209.
- 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.
- Troeger, Christopher, Danny Colombara, Puja Rao, Ibrahim Khalil, Alexandria Brown, Thomas Brewer, Richard Guerrant, Eric Houpt, Karen Kotloff, Kavita Misra, William Petri, James Platts-Mills, Mark Ridde, Scott Swartz, Mohammad Forouzanfar, Robert Reiner, Simon Hay, and Ali Mokdad. 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.

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.

WHO. 2011. *Guidelines for Drinking-Water Quality*. World Health Organisation.

Wisniewski, Janna, Angela Acosta, Jan Kolaczinski, Hannah Koenker, and Joshua Yukich. 2020. "Systematic Review and Meta-analysis of the Cost and Cost-effectiveness of Distributing Insecticide-treated Nets for the Prevention of Malaria." *Acta Tropica* 202:105229.

Figure 1: Timeline of the Study

	Recruitment	Baseline + Intervention Session 1	Intervention Session 2	Endline	Chlorine Test	Long-run Follow-up
When?	(Week -2)	Week 0	Week 1	Week 10	Week 12	Week 140 (33 months)
Where?	Home	Mobile Lab	Mobile Lab	Mobile Lab	Home	Phone
Who?	V, P, AC, PC	V, P, AC	V, P, AC	V, P, AC, PC	V, P, AC, PC	V, P, AC, PC

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) Col. 2 vs. Col. 3 <i>p</i> -value	(5) <i>N</i>	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) <i>N</i>
<i>Balance on recruitment census variables</i>										
Age	26.57 (4.52)	-0.02 (0.22)	0.00 (0.22)	0.93	2337	26.62 (4.69)	-0.42 (0.22)*	-0.36 (0.22)	-0.31 (0.22)	3750
Married or cohabiting	0.89 (0.31)	-0.00 (0.02)	0.00 (0.02)	0.76	2337	0.90 (0.30)	-0.02 (0.02)	-0.01 (0.01)	-0.02 (0.02)	3750
Education level	5.85 (1.22)	0.01 (0.06)	0.08 (0.06)	0.19	2337	5.93 (1.08)	-0.08 (0.05)	0.00 (0.05)	-0.05 (0.05)	3750
High wealth index	0.55 (0.50)	-0.02 (0.02)	-0.03 (0.02)	0.67	2337	0.52 (0.50)	0.00 (0.02)	-0.01 (0.02)	0.02 (0.02)	3750
Village of residence	87.59 (54.51)	0.46 (6.11)	0.42 (6.05)	0.99	2337	83.31 (56.43)	0.80 (4.18)	-0.24 (4.14)	-0.07 (3.97)	3750
<i>Survey participation</i>										
Did not participate in endline (10w)	0.08 (0.27)	0.02 (0.02)	0.03 (0.02)*	0.39	2337	0.24 (0.43)	-0.06 (0.02)***	-0.04 (0.02)*	-0.06 (0.02)***	3750
Did not participate in chlorine test (12w)	0.12 (0.33)	0.01 (0.02)	0.03 (0.02)	0.33	2337	0.26 (0.44)	-0.04 (0.02)*	-0.02 (0.02)	-0.04 (0.02)**	3750
Did not participate in follow-up (33m)	0.10 (0.31)	0.02 (0.02)	0.02 (0.02)	0.85	2337	0.17 (0.37)	-0.01 (0.02)	-0.02 (0.02)	-0.01 (0.02)	3750
<i>Delay variables</i>										
Days between endline and baseline	67.73 (20.63)	0.62 (1.00)	1.80 (0.93)*	0.23	2116	68.73 (24.07)	1.82 (1.06)*	2.12 (1.01)**	1.23 (1.00)	2984
Days between chlorine test and baseline	79.19 (26.42)	0.73 (1.35)	2.62 (1.31)**	0.16	2007	81.58 (27.27)	0.75 (1.21)	1.57 (1.17)	-0.01 (1.16)	2832
Chlorine test was conducted on first day in village	0.68 (0.47)	0.01 (0.02)	-0.00 (0.02)	0.69	2009	0.67 (0.47)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	2839
Days between follow-up and baseline	980.47 (33.29)	2.76 (1.98)	2.02 (1.86)	0.70	2066	987.06 (36.04)	-1.16 (1.91)	-2.35 (1.82)	-4.36 (1.84)**	2828
<i>Compliance</i>										
Completed baseline & both interventions	0.74 (0.44)	0.01 (0.02)	-0.02 (0.02)	0.35	2975	-	-	-	-	-
Completed baseline & first intervention	0.78 (0.41)	0.01 (0.02)	0.01 (0.02)	0.85	2975	-	-	-	-	-

Notes: OLS estimates of balance across treatment groups. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Survey participation, delay, and compliance specifications control for a vector of observed characteristics; randomization balance specifications do not. All specifications cluster standard errors at the level of the intervention cohort. In the active control comparison, the sample is restricted to those who completed baseline and the first intervention session, since these are the samples used for the primary analyses. The exception are the "compliance" regressions, where baseline and intervention attendance are the outcomes of interest. Survey participation can only be interpreted as attrition in the active control comparison: In the pure control comparison, we cannot condition on baseline participation, and thus include all participants from the recruitment census. Balance variables were collected during the recruitment census in the villages. 'High wealth index' denotes participants with an above-median value of assets from a limited list of common household assets.

Table 2: Behavioral outcomes

	MHT Level	Endline (10-12 weeks)					Follow-Up (30-36 months)				
		(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Col. 2 vs. Col. 3 p-value	(5) N	(6) Active Control Group Mean (SD)	(7) Visualization Treatment Effect	(8) Planning Treatment Effect	(9) Col. 2 vs. Col. 3 p-value	(10) N
HEALTH OUTCOMES											
Objective measure: chlorine present in water (TCR)	1/-	0.23 (0.42)	0.05 (0.02)**	0.02 (0.02)	0.15	2012					
Objective measure: chlorine sufficient to be safe (FCR)	3/-	0.21 (0.40)	0.04 (0.02)**	0.01 (0.02)	0.16	2012					
Main Treatment: Chlorine (self-report)	np/1	0.73 (0.45)	0.07 (0.02)**	-0.00 (0.02)	0.00***	2116	0.85 (0.35)	0.05 (0.02)**	0.02 (0.02)	0.15	2073
Main Treatment: Boil (self-report)	np/np	0.35 (0.48)	0.07 (0.03)**	0.05 (0.03)*	0.46	2116	0.63 (0.48)	0.03 (0.03)	-0.01 (0.03)	0.19	2073
Diarrhea incidences per child u15, last 3 months	3/1	0.26 (0.69)	-0.12 (0.03)**	-0.06 (0.03)*	0.09*	2004	0.27 (0.73)	0.00 (0.04)	-0.04 (0.03)	0.26	2045
Diarrhea incidences per child u5, last 3 months	np/3	0.34 (0.86)	-0.16 (0.05)**	-0.06 (0.05)	0.02**	1682	0.39 (1.04)	0.01 (0.07)	-0.01 (0.06)	0.70	1612
Proportion of children taken for healthcare check-up	3/3	0.21 (0.34)	-0.04 (0.02)**	-0.02 (0.02)	0.34	1995	0.36 (0.37)	-0.02 (0.02)	-0.01 (0.02)	0.88	2045
Proportion of children u15 vaccinated, last 3 months	3/-	0.22 (0.35)	0.00 (0.02)	-0.01 (0.02)	0.49	1990					
Number of ANC visits, last 3 months (if pregnant)	3/-	1.26 (1.19)	-0.10 (0.49)	0.21 (0.36)	0.45	200					
SAVINGS OUTCOMES											
Amount saved regularly (per week, KES)	2/2	93.96 (230.26)	24.37 (12.38)**	3.58 (12.55)	0.12	2108	407.50 (605.28)	56.90 (33.78)*	23.66 (33.29)	0.34	2073
Indicator: Amount saved regularly is positive	3/3	0.36 (0.48)	0.13 (0.03)**	-0.02 (0.03)	0.00***	2108	0.78 (0.42)	0.05 (0.02)**	0.02 (0.02)	0.13	2073
Number of ROSCAs [joined in last 3 months/total]	3/3	0.17 (0.44)	0.04 (0.03)	0.01 (0.02)	0.24	2108	1.08 (1.05)	0.12 (0.06)*	0.10 (0.06)	0.75	2073
Weekly ROSCA savings	np/np	205.93 (304.72)	32.44 (15.73)**	11.22 (16.35)	0.17	2108	246.07 (363.71)	41.36 (22.13)*	33.23 (21.21)	0.72	2073
Indicator: Saves for productive investments	3/3	0.17 (0.38)	0.11 (0.02)**	-0.01 (0.02)	0.00***	2108	0.62 (0.49)	0.02 (0.02)	0.01 (0.03)	0.84	2073
Total Savings Balance (KES)	-/3						2542.85 (6378.76)	1066.69 (428.62)**	39.07 (404.17)	0.01***	2073
LABOR OUTCOMES											
Total hours of work [last 3 months/ last 7 days]	2/2	106.11 (174.61)	-5.38 (9.54)	-23.39 (8.93)**	0.05*	2108	13.25 (15.41)	2.36 (0.88)**	0.89 (0.87)	0.09*	2073
Total days of work, last 3 months	3/-	21.22 (30.09)	-0.31 (1.62)	-3.70 (1.58)**	0.04**	2108					
Earnings, cash and in-kind [monthly/last 7 days]	3/3	1094.50 (2865.35)	0.17 (147.14)	10.16 (163.57)	0.94	2108	677.26 (1343.83)	104.62 (77.27)	101.91 (80.34)	0.97	2073
OTHER BEHAVIOURAL OUTCOMES											
Index of investment in children's education (z-score)	2/-	0.00 (1.00)	-0.02 (0.06)	0.00 (0.07)	0.68	1420					

Notes: OLS estimates of treatment effects. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. The column "MHT level" indicates the multiple hypothesis test level: whether variables were pre-specified as primary (1), secondary (2), or exploratory outcomes (3), or were not pre-specified (np). The first number relates to the outcome pre-specification in the 10-week survey; the second to the outcome pre-specification in the 30-36 month survey. All columns include village-level fixed effects, a vector of individual characteristics, fixed effects for the week and the day of the week of the relevant survey, and standard errors which are clustered at the level of the intervention cohort. The sample in all regressions is restricted to participants in active treatment groups who attended the baseline survey. Where available, we control for the baseline outcome of the dependent variable. Outcome measures are listed on the left, and are described in detail in Section 4.

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

	Village has no chlorine dispenser				Village has chlorine dispenser				Comparison	
	(1) Active Control Mean (SD)	(2) V+INF Treatment Effect	(3) P+INF Treatment Effect	(4) N	(5) Active Control Mean (SD)	(6) V+INF Treatment Effect	(7) P+INF Treatment Effect	(8) N	(9) V+INF Interaction <i>p</i> -value	(10) P+INF Interaction <i>p</i> -value
ENDLINE: 10 - 12 WEEKS										
Objective measure: chlorine present in water (TCR)	0.19 (0.39)	0.05 (0.03)*	0.04 (0.03)	1082	0.29 (0.45)	0.04 (0.03)	-0.00 (0.03)	930	[0.87]	[0.34]
Objective measure: chlorine sufficient to be safe (FCR)	0.16 (0.36)	0.05 (0.03)**	0.03 (0.03)	1082	0.27 (0.44)	0.03 (0.03)	-0.01 (0.03)	930	[0.60]	[0.32]
Main Treatment: Chlorine (self-report)	0.65 (0.48)	0.09 (0.03)***	0.02 (0.03)	1129	0.82 (0.39)	0.06 (0.03)*	-0.01 (0.03)	987	[0.34]	[0.49]
Main Treatment: Boiling (self-report)	0.37 (0.48)	0.06 (0.04)	0.04 (0.04)	1129	0.33 (0.47)	0.09 (0.04)**	0.07 (0.04)*	987	[0.52]	[0.56]
Diarrhea incidences per child u15, last 3 months	0.26 (0.67)	-0.09 (0.05)**	-0.04 (0.05)	1066	0.25 (0.71)	-0.13 (0.05)**	-0.09 (0.05)*	938	[0.59]	[0.46]
Diarrhea incidences per child u5, last 3 months	0.37 (0.88)	-0.14 (0.06)**	-0.06 (0.07)	908	0.30 (0.83)	-0.15 (0.06)**	-0.06 (0.07)	774	[0.97]	[0.96]
FOLLOW-UP: 30 - 36 MONTHS										
Main Treatment: Chlorine (self-report)	0.81 (0.39)	0.05 (0.03)*	0.02 (0.03)	1103	0.90 (0.30)	0.04 (0.02)**	0.03 (0.02)	970	[0.78]	[0.78]
Main Treatment: Boiling (self-report)	0.63 (0.48)	0.04 (0.04)	0.02 (0.04)	1103	0.62 (0.49)	0.02 (0.04)	-0.03 (0.04)	970	[0.67]	[0.37]
Diarrhea incidences per child u15, last 3 months	0.29 (0.71)	0.01 (0.06)	-0.10 (0.05)**	1088	0.25 (0.76)	-0.04 (0.06)	-0.02 (0.05)	957	[0.48]	[0.23]
Diarrhea incidences per child u5, last 3 months	0.46 (1.12)	-0.02 (0.09)	-0.09 (0.09)	867	0.32 (0.93)	0.02 (0.09)	0.06 (0.08)	745	[0.74]	[0.15]
Chlorine source: Bottle	0.47 (0.50)	0.06 (0.04)	0.04 (0.04)	1103	0.31 (0.46)	0.02 (0.04)	0.06 (0.04)	970	[0.48]	[0.77]
Chlorine source: Dispenser	0.21 (0.41)	-0.04 (0.03)	-0.04 (0.03)	1103	0.54 (0.50)	-0.01 (0.04)	-0.05 (0.04)	970	[0.57]	[0.93]
Reports working dispenser at main source	0.34 (0.47)	-0.03 (0.04)	0.01 (0.04)	1103	0.70 (0.46)	0.03 (0.03)	-0.04 (0.03)	970	[0.17]	[0.33]
Reports working dispenser within 30min walk	0.49 (0.50)	-0.01 (0.04)	-0.02 (0.04)	1103	0.82 (0.38)	0.01 (0.03)	-0.01 (0.03)	970	[0.57]	[0.77]

Notes: OLS estimates of treatment effects. Columns (1) and (5) report the mean and standard deviation of the control group. Columns (2)-(3) and (6)-(7) report the coefficients of interest and standard errors in parentheses. The analysis repeats that in Table 2, separately for villages which randomly received at least one chlorine dispenser in the WASH Benefits study (and which was subsequently maintained by Evidence Action), and for villages which did not receive chlorine dispensers. We use seemingly unrelated regression (SUR) to allow us to compare treatment coefficients across models. For each variable, we report the mean of the comparison group, the coefficients of interest, and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All columns include village-level fixed effects, a vector of individual characteristics, fixed effects for the week and the day of the week of the relevant survey, and standard errors which are clustered at the level of the intervention cohort. The sample in all regressions is restricted to participants in active treatment groups who attended the baseline survey. Where available, we control for the baseline outcome of the dependent variable. Outcome measures are listed on the left, and are described in detail in Section 4. Columns (9)-(10) report the *p*-values on the differential effect of the treatments in villages with vs. without chlorine dispensers using SUR.

Table 4: Psychological outcomes

	MHT Level	Endline (10-12 weeks)					Follow-Up (30-36 months)				
		(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 <i>p</i> -value	(5) <i>N</i>	(6) Active Control Group Mean (SD)	(7) Visualization Treatment Effect	(8) Planning Treatment Effect	(9) Column 2 vs. Column 3 <i>p</i> -value	(10) <i>N</i>
PLANNING SKILLS											
BADS score (z-score)	1/-	0.00 (1.00)	-0.03 (0.05) [0.83]	0.05 (0.05) [0.91]	0.21	2103					
Tower of London (z-score)	2/-	0.00 (1.00)	0.02 (0.05) [0.49]	-0.04 (0.05) [0.39]	0.25	2103					
TIME PREFERENCES											
β^{Effort}	1/-	0.982 (0.005)	0.007 (0.006) [0.83]	0.005 (0.007) [0.91]	0.33	2068					
δ^{Effort}	2/-	0.999 (0.001)	-0.001 (0.001) [0.26]	-0.002 (0.001)** [0.08]*	0.16	2068					
Utility Forecasting: Vividness (z-score)	-/1						0.00 (1.00)	0.12 (0.05)** [0.04]**	-0.04 (0.06) [0.72]	0.00***	2073
Utility Forecasting: Practice (z-score)	-/2						-0.00 (1.00)	0.10 (0.05)* [0.17]	0.12 (0.05)** [0.09]*	0.68	2073
β^{MPL}	1/-	1.05 (0.46)	-0.01 (0.02) [0.49]	0.02 (0.03) [0.39]	0.27	2103					
δ^{MPL}	2/-	0.98 (0.02)	-0.00 (0.00) [0.34]	-0.00 (0.00) [0.20]	0.83	2103					
Time Preferences, qualitative (z-score)	-/2						-0.00 (1.00)	0.09 (0.05)* [0.17]	-0.06 (0.06) [0.47]	0.01***	2073
One-year discount factor (GPS staircase)	-/2						0.52 (0.13)	0.00 (0.01) [0.94]	-0.01 (0.01) [0.47]	0.21	2073
SELF-EFFICACY											
General Self-Efficacy Score (GSE) (z-score)	2/1	0.00 (1.00)	0.15 (0.05)*** [0.01]***	0.11 (0.05)** [0.08]*	0.45	2103	-0.00 (1.00)	0.07 (0.05) [0.09]*	-0.07 (0.05) [0.72]	0.00***	2073

Notes: OLS estimates of treatment effects. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Square brackets contain additional *p*-values corrected for multiple hypothesis testing using the false discovery rate. The column "MHT level" indicates the multiple hypothesis test level: whether variables were pre-specified as primary (1), secondary (2), or exploratory outcomes (3), or were not pre-specified (np). The first number relates to the outcome pre-specification in the 10-week survey; the second to the outcome pre-specification in the 30-36 month survey. All columns include village-level fixed effects, a vector of individual characteristics, fixed effects for the week and the day of the week of the relevant survey, and standard errors which are clustered at the level of the intervention cohort. The sample in all regressions is restricted to participants in active treatment groups who attended the baseline survey. Where available, we control for the baseline outcome of the dependent variable. Outcome measures are listed on the left, and are described in detail in Section 4.

Table 5: Alternative Mechanisms

	Endline (10-12 weeks)					Follow-Up (30-36 months)				
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 <i>p</i> -value	(5) <i>N</i>	(6) Active Control Group Mean (SD)	(7) Visualization Treatment Effect	(8) Planning Treatment Effect	(9) Column 2 vs. Column 3 <i>p</i> -value	(10) <i>N</i>
BELIEFS AND KNOWLEDGE										
Belief: Diarrhea avoided through chlorination (z-score)	0.00 (1.00)	0.08 (0.05)	0.05 (0.05)	0.56	2103					
Chlorine knowledge score (z-score)	0.00 (1.00)	0.07 (0.05)	0.00 (0.05)	0.22	2103					
ANC/PNC knowledge score (z-score)	0.00 (1.00)	0.04 (0.05)	-0.05 (0.05)	0.09*	2103					
Risk Aversion Measure (z-score)	0.00 (1.00)	-0.03 (0.06)	-0.07 (0.06)	0.49	1926					
SALIENCE TASK										
Chlorine word remembered	0.40 (0.49)	0.06 (0.01)***	0.03 (0.01)**	0.095*	6336	0.42 (0.49)	0.10 (0.03)**	0.04 (0.03)	0.028**	2073
Savings word remembered	0.45 (0.50)	-0.02 (0.01)	0.00 (0.01)	0.124	6336	0.45 (0.50)	0.04 (0.02)	-0.01 (0.02)	0.069*	2073
Total words remembered	4.23 (1.64)	-0.09 (0.07)	0.04 (0.07)	0.105	6336	4.45 (1.61)	0.28 (0.09)**	0.20 (0.09)*	0.336	2073

Notes: OLS estimates of treatment effects. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. The column "MHT level" indicates the multiple hypothesis test level: whether variables were pre-specified as primary (1), secondary (2), or exploratory outcomes (3), or were not pre-specified (np). The first number relates to the outcome pre-specification in the 10-week survey; the second to the outcome pre-specification in the 30-36 month survey. All columns include village-level fixed effects, a vector of individual characteristics, fixed effects for the week and the day of the week of the relevant survey, and standard errors which are clustered at the level of the intervention cohort. The sample in all regressions is restricted to participants in active treatment groups who attended the baseline survey. Where available, we control for the baseline outcome of the dependent variable. Outcome measures are listed on the left, and are described in detail in Section 4. The bottom panel of the table reports the probability of remembering a chlorine-related word, or a savings-related word, on a given word list in the salience task. In the endline survey, participants were read three word lists, resulting in three observations per individual. In the long-run follow-up, participants were read one word list (randomly selected out of three lists). Salience regressions additionally control for the total number of words the participant remembered on that word list.

Online Appendix

A. Overview of Interventions

Figure A.1: Overview of the Interventions

	VISUALIZATION	PLANNING	ACTIVE CONTROL
Target problem	Impatient behavior caused by inability to imagine the future	Inactivity cycles caused by avoidance and negative mood	
Key conceptual reference	Gabaix and Laibson (2017)	Lejuez et al (2011)	
Key Content	<ol style="list-style-type: none"> 1. Connect present behavior to future outcomes 2. Visualize alternative realizations of the future depending on current behavior 3. Put yourself in the shoes of your future selves, imagine how they feel, and 'talk' to them. 	<ol style="list-style-type: none"> 1. Write lists of necessary, routine, and pleasurable tasks 2. Rate tasks from most to least difficult 3. Schedule in diary, starting systematically with easiest tasks 4. Break tasks into steps, anticipate obstacles 	Placebo session: Lectures, exercises and drawings on birds and plants of Kenya
Psychological Targets	Time Preferences Self-Efficacy	Planning Skills Self-Efficacy	Not psychologically active

Sessions for all treatments included interactive lectures, case stories, exercises and drawings

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

To compute the cost per disability-adjusted life years saved with our interventions, we proceed as follows. The cost of our interventions was USD 4 per household. This is USD 3.70 per child under 5 or USD 1.58 per child under 15 because participants have 1.08 children under 5 and 2.53 children under 15 on average. This includes the costs relevant to a potential scale-up, i.e. the cost of running the sessions including overheads, but excludes our incentivized surveys, as well as the cost of sampling/targeting 18-35 year old women. Troeger et al. (2018) estimate the DALYs lost per child under 5 in Kenya due to diarrhea in 2016, the year before our study was conducted, to be 0.127. This includes both acute, immediate effects of illness and/or death and the longer-term burden of disease associated with growth impairment due to diarrhea. Our region is one of the poorer ones in Kenya and likely has worse health outcomes and a higher burden of disease from diarrhea than the Kenya-wide average, so using this figure will lead to a conservative estimate.

We focus on children under 5, as these are the focus of the estimates in Troeger et al. (2018). We assume that the treatment effects of our interventions on diarrhea relative to the active control group were never higher during the study period of three 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. For children under 5, Table 2 shows reductions in under-5 diarrhea of 47 percent in the Visualization group, and no significant effect in the Planning group. DALYs saved due to the Visualization intervention are $0.127 \times 47\% \times \frac{3}{12} = 0.0149225$, if one accounts for both acute and long term effects of diarrhea. The cost of one DALY saved is $\frac{\text{USD } 3.70}{0.0149225} = \text{USD } 248$.

For children under 15, the effect of our interventions on diarrhea relative to the active control group is 46 percent for Visualization and 23 percent for Planning. We could not find published estimates for Kenya of DALYs lost per child under 15 annually to diarrhea, so we use the same figure of 0.127 as for under 5 children. This likely overstates the disease burden on children under 15, as older children are less likely to die. However, older children do benefit from a reduction in ongoing enteric dysfunction, which may cause adverse effects like stunting and impaired cognition in a wide age range. Thus, a restriction to children under 5 would be too narrow. With this assumption, DALYs saved per child due to the Visualization intervention are 0.014605, and of the Planning intervention 0.0073025. Combined with an intervention cost of USD 1.58 per child for both interventions, this implies that the cost of one DALY saved is $\frac{\text{USD } 1.58}{0.014605} = \text{USD } 108$ for the Visualization intervention and $\frac{\text{USD } 1.58}{0.0073025} = \text{USD } 216$ for the Planning intervention.

Extrapolating to other contexts, these numbers will be sensitive to the number of children per treated participant. Estimates are also sensitive to how long the effect on diarrhea lasts: If, instead of three months, the effects lasted for one year and then faded out, the estimated cost per DALY saved for the Visualization intervention would be reduced to USD 27 (when considering children under 15) and USD 62 (when considering only children under 5). All our estimates remain highly cost-effective by WHO standards. The WHO classifies an intervention as “cost-effective” for a cost per DALY saved below USD 4525, and “highly cost-effective” below USD 1508 (<https://www.who.int/bulletin/volumes/93/2/14-138206/en/>).

Policymakers may be interested in the comparison of our interventions to other cost-effective preventive health interventions, such as insecticide-treated nets. Cost effectiveness estimates for malaria bednets range from USD 29 to 100/DALY (Wisniewski et al. 2020), depending on assumptions. The WHO estimates a cost per DALY of USD 29 to 34 (https://www.who.int/choice/results/mal_afrd/en/). The higher estimate of USD 100 is from GiveWell (<https://forum.effectivealtruism.org/posts/HbunzTyFPRwcYihg6/long-lasting-insecticide-treated-nets-usd3-340-per-life>). A recent meta-analysis of studies in Africa estimates a cost per DALY between 42 and 80 USD (Wisniewski et al. 2020).

C. 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 liter 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 minutes for 20 liters of water (Kremer et al. 2011a). This is sufficient chlorine to sanitize a jerrycan of 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. WASH participant households (excluded in our study) 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.

We compare villages which received the dispenser intervention to villages which do not. Our “dispenser villages” include 67 villages from the “Water Quality” (henceforth “dispenser”) treatment arm. We also include 23 villages which received dispensers, as well as combinations of sanitation, handwashing or nutrition interventions at household level. Our “non-dispenser villages” include 67 villages from the “Passive Comparison” arm, which received no interventions. We also include 48 villages who received one of sanitation, handwashing or nutrition interventions, but no dispensers. These additional interventions all took place at the household level, and we do not sample households which participated directly in the original WASH study. Even in WASH households, Null et al. (2018) find no effect of these interventions on whether stored drinking water has detectable free chlorine. In addition, all interventions other than dispensers finished three to four years before our study. We had intended to select only villages from the Water Quality and Passive Comparison arms. The 23 villages with dispensers and other WASH interventions, and the 48 villages with no dispensers and WASH interventions were selected due to a

coding error in the sampling in Mumias district, Kakamega county.

D. Supplementary Evidence on Mechanisms

Table A.1: Treatment Effect Heterogeneity on Diarrhea by Season and Water Source

	Any diarrhea in u15 children, last 7 days (33 months)		
	(1)	(2)	(3)
Visualization	-0.0148 (0.0162)	-0.0150 (0.0144)	-0.0156 (0.0165)
Planning	-0.0125 (0.0170)	-0.00896 (0.0147)	-0.0129 (0.0172)
Unprotected water source (33m)	0.00994 (0.0259)		-0.00912 (0.0266)
Surveyed during rainy season (33m)		0.0500 (0.0778)	-0.102*** (0.0390)
Surveyed during rainy season × Unprotected water source (33m)			0.359** (0.163)
V × Unprotected water source (33m)	-0.0143 (0.0338)		0.00237 (0.0350)
P × Unprotected water source (33m)	0.00988 (0.0343)		0.0175 (0.0348)
V × Surveyed during rainy season (33m)		-0.0842 (0.0824)	0.0731 (0.0499)
P × Unprotected water source (33m)		-0.0387 (0.100)	0.0583 (0.0741)
V × Surveyed during rainy season × Unprotected water source (33m)			-0.399** (0.173)
P × Surveyed during rainy season × Unprotected water source (33m)			-0.0702 (0.246)
Observations	2045	2045	2045
Control Mean	0.07	0.07	0.07

Notes: The table reports OLS estimates of heterogeneous treatment effects on diarrhea in the 33-month survey, by season and type of water source. The sample is restricted to individuals who attended the baseline survey. *Surveyed during rainy season* is an indicator for whether the respondent was surveyed after November 1, 2020 for the follow-up survey. *Unprotected water source* equals 1 if the household's primary water source is not protected (unprotected well, unprotected spring, rainwater, surface water), and 0 otherwise (private or public tap, borehole or tubewell, protected well or spring). We report standard errors in parentheses. All columns include village-level fixed effects, control for diarrhea at baseline and a vector of individual characteristics, and cluster standard errors at the level of the intervention cohort. As regressors are potentially endogenous to treatment, all regressions should be interpreted as correlational evidence. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.2: Correlates of Chlorination

	Main Treatment: Chlorine (baseline)	Chlorine present in water (TCR) (12 weeks)		Main Treatment: Chlorine (10 weeks)		Main Treatment: Chlorine (33 months)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Visualization		0.0426*	0.0593***	0.0752***	0.0852***	0.0577***	-0.00542
		(0.0221)	(0.0214)	(0.0223)	(0.0254)	(0.0191)	(0.0180)
Planning		0.0154	0.0318	-0.0252	0.0285	0.0304	-0.0268
		(0.0257)	(0.0216)	(0.0255)	(0.0265)	(0.0199)	(0.0180)
Active Control			0.0230		0.0380		-0.0479**
			(0.0213)		(0.0255)		(0.0189)
Tower of London (baseline)	-0.00172	-0.0000		-0.00411***			
	(0.00164)	(0.00141)		(0.00159)			
β^{Effort} (baseline)	-0.00598	0.00272		0.00355			
	(0.00467)	(0.00364)		(0.00520)			
δ^{Effort} (baseline)	-0.00797	-0.00794		0.0104			
	(0.0138)	(0.00847)		(0.0126)			
General Self-Efficacy Score (GSE) (z-score) (baseline)	0.0212**	0.000179		0.0400***			
	(0.0104)	(0.00943)		(0.00965)			
Tower of London (10w)			-0.000982		-0.00595***	-0.00117	-0.000768
			(0.00114)		(0.00132)	(0.00125)	(0.000966)
β^{Effort} (10w)			-0.00106		-0.00340	-0.00797	-0.00112
			(0.00132)		(0.00210)	(0.00909)	(0.00123)
δ^{Effort} (10w)			-0.0196*		0.00855	-0.0210	-0.0106
			(0.0101)		(0.0142)	(0.0163)	(0.0109)
General Self-Efficacy Score (GSE) (z-score) (10w)			0.0108*		0.0268***	0.0110*	0.00778
			(0.00627)		(0.00711)	(0.00644)	(0.00555)
Utility Forecasting: Vividness (z-score) (33m)						0.00463	0.00311
						(0.00858)	(0.00689)
Utility Forecasting: Practice (z-score) (33m)						0.00850	0.0132*
						(0.00940)	(0.00779)
Time Preferences, qualitative (z-score) (33m)						-0.00979	-0.00373
						(0.00888)	(0.00710)
One-year discount factor (GPS staircase) (33m)						0.0984	0.0859
						(0.0601)	(0.0529)
Main Treatment: Chlorine(baseline)				0.213***		0.0816***	
				(0.0235)		(0.0183)	
High wealth index (census)	-0.00431	-0.00139	0.00778	0.0138	0.0254	0.00574	0.00525
	(0.0230)	(0.0212)	(0.0168)	(0.0212)	(0.0178)	(0.0170)	(0.0137)
Completed secondary education (census)	0.0396	0.0337	0.0264	0.105***	0.0670***	-0.0221	-0.0237
	(0.0256)	(0.0230)	(0.0181)	(0.0220)	(0.0197)	(0.0189)	(0.0156)
Birth Year (census)	-0.00231	-0.00442**	-0.00391**	-0.000805	-0.000908	0.00287	0.00213
	(0.00245)	(0.00207)	(0.00164)	(0.00237)	(0.00194)	(0.00189)	(0.00148)
Employed (census)	-0.0402	0.0195	0.00593	0.0283	0.0259	-0.00229	-0.00405
	(0.0288)	(0.0262)	(0.0205)	(0.0263)	(0.0212)	(0.0217)	(0.0177)
Married or cohabiting (census)	-0.000306	-0.0170	-0.0155	-0.00590	0.0113	0.0246	0.0156
	(0.0354)	(0.0334)	(0.0277)	(0.0340)	(0.0293)	(0.0281)	(0.0228)
BASELINE SAMPLE	YES	YES	NO	YES	NO	YES	NO
Observations	2.129	1.821	2.782	1.918	2.934	1.894	2.638
R-squared	0.181	0.354	0.338	0.231	0.139	0.152	0.134

Notes: Predictive OLS regressions of chlorination measures on psychological mechanisms, as well as on demographics. All regressors except those measured at baseline are potentially endogenous to treatment, and thus provide only correlational evidence. All regressions include village-level fixed effects, the demographic controls listed, and standard errors which are clustered at the level of the intervention cohort. Where available, we control for the baseline outcome of the dependent variable. Covariates are listed on the left, and are described in detail in Section 4. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.3: Correlates of Savings

	Amount Saved Regularly (per week, KES)(10 weeks)		Amount Saved Regularly (per week, KES)(33 months)	
	(1)	(2)	(3)	(4)
Visualization	10.22 (37.93)	44.46 (29.31)	39.41 (45.73)	25.31 (43.43)
Planning	-70.81 (46.71)	-19.57 (23.18)	20.31 (41.64)	16.76 (41.97)
Active Control		45.60 (34.24)		-16.94 (40.91)
Tower of London (baseline)	-2.990* (1.697)			
β^{Effort} (baseline)	-3.112 (2.509)			
δ^{Effort} (baseline)	17.71 (30.80)			
General Self-Efficacy Score (GSE) (z-score) (baseline)	21.75 (14.41)			
Tower of London (10w)		-0.882 (1.665)	3.333 (2.914)	2.083 (2.164)
β^{Effort} (10w)		-0.916 (1.033)	-17.51 (24.16)	-4.349 (2.810)
δ^{Effort} (10w)		-17.65 (14.41)	1.577 (31.70)	1.090 (22.23)
General Self-Efficacy Score (GSE) (z-score) (10w)		23.11*** (8.341)	42.09*** (13.48)	42.21*** (11.76)
Utility Forecasting: Vividness (z-score) (33m)			53.00*** (20.08)	34.73** (16.57)
Utility Forecasting: Practice (z-score) (33m)			56.18*** (21.11)	36.65** (17.45)
Time Preferences, qualitative (z-score) (33m)			1.533 (17.31)	-1.490 (13.91)
One-year discount factor (GPS staircase) (33m)			186.7 (152.7)	102.4 (123.8)
High wealth index (census)	-6.586 (37.05)	-18.98 (33.24)	50.51 (37.80)	42.59 (29.60)
Completed secondary education (census)	-5.398 (47.60)	13.84 (34.25)	116.1** (48.84)	109.4*** (40.63)
Birth Year (census)	-2.408 (4.435)	-2.539 (3.163)	0.208 (4.119)	0.122 (3.185)
Employed (census)	80.75** (37.20)	84.03** (35.19)	53.64 (46.13)	107.2*** (37.29)
Married or cohabiting (census)	-32.91 (76.21)	-0.689 (61.38)	18.22 (54.24)	1.447 (46.36)
BASELINE SAMPLE	YES	NO	YES	NO
Observations	1.910	2.924	1.894	2.638
R-squared	0.136	0.070	0.167	0.135

Notes: Predictive OLS regressions of savings outcomes on psychological mechanisms, as well as on demographics. All regressors except those measured at baseline are potentially endogenous to treatment, and thus provide only correlational evidence. All regressions include village-level fixed effects, the demographic controls listed, and standard errors which are clustered at the level of the intervention cohort. Covariates are listed on the left, and are described in detail in Section 4. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.4: Correlates of Chlorination and Savings – Alternative Mechanisms

	Chlorine present in water (TCR) (12 weeks)		Amount Saved Regularly (per week, KES)(10 weeks)	
	(1)	(2)	(3)	(4)
Visualization	0.0491** (0.0216)	0.0571*** (0.0211)	12.42 (37.85)	45.08 (30.02)
Planning	0.0234 (0.0228)	0.0351* (0.0212)	-51.50 (39.85)	-16.11 (22.11)
Active Control		0.0227 (0.0212)		48.20 (33.57)
Belief: Diarrhea avoided through chlorination (z score) (10w)	0.00544 (0.00904)	0.00670 (0.00750)	-11.58 (24.94)	-2.917 (21.28)
Chlorine knowledge score (z-score) (10w)	0.0203** (0.00910)	0.0152** (0.00720)	-0.196 (14.19)	8.097 (10.61)
Risk Aversion Measure (z-score) (10w)	0.00966 (0.00927)	0.00795 (0.00769)	-20.38 (12.59)	-22.17 (13.79)
Total chlorine words remembered (10w)	-0.00181 (0.0110)	0.00365 (0.00927)	16.80 (27.30)	5.103 (20.80)
Total saving words remembered (10w)	-0.00445 (0.0127)	0.00117 (0.0106)	55.86 (35.86)	23.74 (29.34)
Total words remembered (10w)	0.000847 (0.00299)	0.000754 (0.00256)	-8.796 (6.637)	-3.922 (6.178)
High wealth index (census)	-0.00894 (0.0203)	0.00727 (0.0167)	-14.28 (35.61)	-23.37 (31.67)
Completed secondary education (census)	0.0339 (0.0219)	0.0311* (0.0180)	20.17 (41.54)	27.22 (32.32)
Birth Year (census)	-0.00467** (0.00190)	-0.00419*** (0.00162)	-2.637 (3.824)	-1.859 (3.134)
Employed (census)	0.0140 (0.0248)	0.00756 (0.0208)	86.34** (34.19)	86.11** (34.92)
Married or cohabiting (census)	-0.00295 (0.0317)	-0.0142 (0.0276)	-49.00 (74.34)	2.180 (61.79)
BASELINE SAMPLE	YES	NO	YES	NO
Observations	1.998	2.798	2.095	2.938
R-squared	0.367	0.336	0.123	0.072

Notes: Predictive OLS regressions of chlorination and savings measures on alternative mechanisms, as well as on demographics. All regressors except those measured at baseline are potentially endogenous to treatment, and thus provide only correlational evidence. All regressions include village-level fixed effects, the demographic controls listed, and standard errors which are clustered at the level of the intervention cohort. Covariates are listed on the left, and are described in detail in Section 4. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

E. Pure Control Comparison

Table A.5: Behavioral outcomes (comparison with pure control group)

	MHT Level	Endline (10-12 weeks)				(5)	Follow-Up (30-36 months)				(10)
		(1) Pure Control Mean(SD)	(2) V+INF Treatment Effect	(3) P+INF Treatment Effect	(4) AC+INF Treatment Effect		(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	
HEALTH OUTCOMES											
Objective measure: chlorine present in water (TCR)	1/-	0.22 (0.42)	0.06 (0.02)***	0.04 (0.02)*	0.03 (0.02)	2839					
Objective measure: chlorine sufficient to be safe (FCR)	3/-	0.18 (0.39)	0.06 (0.02)***	0.03 (0.02)	0.03 (0.02 0	2839					
Main Treatment; Chlorine (self-report)	np/1	0.66 (0.47)	0.10 (0.03)***	0.04 (0.03)*	0.05 (0.03)**	2984	0.88 (0.33)	-0.01 (0.02)	-0.01 (0.02)	-0.03 (0.02)*	3194
Main Treatment: Boil (self-report)	np/np	0.36 (0.48)	0.05 (0.03)*	0.02 (0.03)	-0.01 (0.03)	2984	0.58 (0.49)	0.05 (0.03)*	0.02 (0.03)	0.05 (0.03)**	3194
Diarrhea incidences per child u15, last 3 months	3/1	0.24 (0.73)	-0.10 (0.04)***	-0.08 (0.04)**	-0.00 (0.04)	2815	0.29 (0.76)	-0.04 (0.04)	-0.07 (0.04)*	-0.02 (0.04)	3130
Diarrhea incidences per child u5, last 3 months	np/3	0.28 (0.83)	-0.11 (0.05)**	-0.03 (0.05)	0.03 (0.05)	2357	0.44 (1.14)	-0.05 (0.07)	-0.03 (0.07)	-0.03 (0.07)	2482
Proportion of children taken for healthcare check-up	3/3	0.16 (0.31)	0.01 (0.02)	0.02 (0.02)	0.05 (0.02)**	2796	0.36 (0.37)	0.00 (0.02)	0.00 (0.02)	0.02 (0.02)	3130
Proportion of children u15 vaccinated, last 3 months	3/-	0.22 (0.35)	[0.46] (0.02)	[1.00] (0.02)	[0.25] (0.02)	2790		[1.00]	[1.00]	[1.00]	
Number of ANC visits, last 3 months (if pregnant)	3/-	1.19 (1.17)	-0.22 (0.43)	0.05 (0.44)	0.26 (0.48)	272					
			[0.50]	[1.00]	[1.00]						
SAVINGS OUTCOMES											
Amount saved regularly (per week, KES)	2/2	88.76 (228.12)	16.62 (12.79)	7.67 (13.17)	3.09 (12.06)	2972	424.48 (618.21)	37.29 (32.36)	22.76 (33.39)	-18.28 (32.01)	3194
Indicator: Amount saved regularly is positive	3/3	0.32 (0.47)	0.12 (0.03)***	0.01 (0.03)	0.04 (0.03)	2972	0.79 (0.41)	0.03 (0.02)	0.02 (0.02)	-0.01 (0.02)	3194
Number of ROSCAs [joined in last 3 months/total]	3/3	0.21 (0.46)	0.02 (0.03)	-0.00 (0.03)	-0.02 (0.03)	2972	1.05 (1.06)	0.12 (0.06)**	0.09 (0.06)*	0.01 (0.05)	3191
Weekly ROSCA savings	np/np	202.18 (315.21)	36.60 (16.85)**	16.03 (17.11)	5.50 (16.68)	2972	257.93 (382.92)	28.77 (20.90)	19.66 (20.58)	-12.71 (19.80)	3194
Indicator: Saves for productive investments	3/3	0.17 (0.38)	0.10 (0.02)***	0.01 (0.02)	0.02 (0.02)	2972	0.61 (0.49)	0.04 (0.02)	0.03 (0.02)	0.01 (0.02)	3194
Total Savings Balance (KES)	-/3		[0.00]***	[1.00]	[1.00]		2984.17 (7912.65)	556.07 (424.38)	-263.13 (391.31)	-522.20 (388.07)	3194
								[0.98]	[1.00]	[1.00]	
LABOR OUTCOMES											
Total hours of work [last 3 months/last 7 days]	2/2	108.32 (183.05)	0.97 (10.12)	-21.77 (9.52)**	-4.77 (9.78)	2972	14.45 (17.25)	0.98 (0.88)	-0.10 (0.87)	-1.35 (0.86)	3194
Total days of work, last 3 months	3/-	21.73 (30.45)	-0.07 (1.72)	-3.63 (1.62)**	-1.17 (1.66)	2972		[0.46]	[1.00]	[0.36]	
Earnings, cash and in-kind [monthly/last 7 days]	3/3	1167.22 (3155.76)	6.11 (168.76)	-55.35 (174.31)	-105.20 (177.88)	2972	828.16 (1571.16)	-82.87 (76.14)	-77.48 (77.91)	-204.74 (77.89)***	3194
			[0.71]	[1.00]	[1.00]			[1.00]	[1.00]	[0.14]	
OTHER BEHAVIORAL OUTCOMES											
Index of investment in children's education (z-score)	2/-	0.00 (1.00)	0.06 (0.07)	0.11 (0.07)	0.11 (0.07)	1967					
			[1.00]	[0.14]	[0.64]						

Notes: OLS estimates of treatment effects, relative to the pure control group. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects; a vector of individual characteristics, fixed effects for the week and the day of the week of the relevant survey, and standard errors which are clustered at the level of the intervention cohort. Outcome measures are listed on the left, and are described in detail in Section 4.

Table A.6: Psychological outcomes (comparison with pure control group)

	MHT Level	Endline (10-12 weeks)				(5) N	Follow-Up (30-36 months)				(10) N
		(1) Pure Control Mean(SD)	(2) V+INF Treatment Effect	(3) P+INF Treatment Effect	(4) AC+INF Treatment Effect		(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	
<i>PLANNING SKILLS</i>											
BADS score (z-score)	1/-	0.00 (1.00)	0.03 (0.06) [1.00]	0.12 (0.06)** [0.08]*	0.05 (0.06) [1.00]	2955					
Tower of London (z-score)	2/-	0.00 (1.00)	-0.36 (0.05)*** [0.00]***	-0.40 (0.05)*** [0.00]***	-0.34 (0.05)*** [0.00]***	2955					
<i>TIME PREFERENCES</i>											
β^{Effort}	1/-	0.953 (0.020)	0.007 (0.018) [1.00]	0.012 (0.018) [0.32]	0.009 (0.018) [1.00]	2906					
δ^{Effort}	2/-	0.995 (0.002)	-0.001 (0.002) [0.46]	-0.002 (0.002) [0.22]	0.003 (0.002) [0.23]	2906					
Utility Forecasting: Vividness (z-score)	-/1						-0.00 (1.00)	0.06 (0.05) [0.74]	-0.07 (0.05) [0.12]	-0.02 (0.05) [1.00]	3191
Utility Forecasting: Practice (z-score)	-/2						0.00 (1.00)	0.01 (0.05) [1.00]	0.04 (0.05) [1.00]	-0.08 (0.05) [0.35]	3191
β^{MPL}	1/-	1.02 (0.43)	-0.01 (0.03) [0.46]	0.01 (0.03) [0.29]	0.00 (0.02) [0.50]	2955					
δ^{MPL}	2/-	0.98 (0.02)	0.00 (0.00)* [0.09]*	0.00 (0.00)* [0.11]	0.00 (0.00)** [0.05]**	2955					
Time Preferences, qualitative (GPS)	-/2						0.00 (1.00)	0.12 (0.05)** [0.09]*	-0.01 (0.05) [1.00]	0.09 (0.05)* [0.35]	3191
One-year discount factor (GPS staircase)	-/2						0.52 (0.12)	0.00 (0.01) [1.00]	-0.01 (0.01) [1.00]	0.00 (0.01) [0.42]	3191
<i>SELF-EFFICACY</i>											
General Self-Efficacy Score (GSE) (z-score)	2/1	0.00 (1.00)	0.17 (0.07)*** [0.02]**	0.17 (0.06)*** [0.01]***	0.02 (0.06) [0.50]	2955	-0.00 (1.00)	-0.01 (0.05)	-0.13 (0.05)***	-0.05 (0.05)	3191

Notes: OLS estimates of treatment effects, relative to the pure control group. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Square brackets contain additional *p*-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, a vector of individual characteristics, fixed effects for the week and the day of the week of the relevant survey, and standard errors which are clustered at the level of the intervention cohort. Outcome measures are listed on the left, and are described in detail in Section 4.

Table A.7: Alternative Mechanisms (comparison with pure control group)

	Endline (10-12 weeks)					Follow-Up (30-36 months)				
	(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
BELIEFS AND KNOWLEDGE										
Belief: Diarrhea avoided through chlorination (z-score)	0.00 (1.00)	0.13 (0.06)**	0.15 (0.05)***	0.11 (0.06)**	2955					
Chlorine knowledge score (z-score)	0.00 (1.00)	0.11 (0.05)**	0.06 (0.06)	0.08 (0.06)	2955					
ANC/PNC knowledge score (z-score)	0.00 (1.00)	0.34 (0.06)**	0.28 (0.05)***	0.33 (0.05)***	2955					
Risk Aversion Measure (z-score)	0.00 (1.00)	0.01 (0.06)	-0.01 (0.06)	0.04 (0.06)	2735					
SALIENCE TASK										
Chlorine word remembered	0.38 (0.48)	0.08 (0.01)***	0.05 (0.01)***	0.02 (0.01)	8934	0.43 (0.50)	0.08 (0.03)**	0.03 (0.03)	0.00 (0.02)	3191
Savings word remembered	0.44 (0.50)	0.00 (0.01)	0.02 (0.01)	0.01 (0.01)	8934	0.48 (0.50)	0.01 (0.02)	-0.03 (0.02)	-0.03 (0.02)	3191
Total words remembered	4.23 (1.66)	-0.12 (0.08)	-0.01 (0.08)	-0.04 (0.07)	8934	4.50 (1.60)	0.19 (0.08)*	0.09 (0.08)	-0.06 (0.08)	3191

Notes: OLS estimates of treatment effects, relative to the pure control group. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All columns include village-level fixed effects, a vector of individual characteristics, fixed effects for the week and the day of the week of the relevant survey, and standard errors which are clustered at the level of the intervention cohort. Outcome measures are listed on the left, and are described in detail in Section 4. The bottom panel of the table reports the probability of remembering a chlorine-related word, or a savings-related word, on a given word list in the salience task. In the endline survey, participants were read three word lists, resulting in three observations per individual. In the long-run follow-up, participants were read one word list (randomly selected out of three lists). Salience regressions additionally control for the total number of words the participant remembered on that word list.

F. Test corrections and experimental integrity

Table A.8 lists our hypotheses and pre-specified outcome variables. We adjust for multiple hypothesis testing within outcome groups (behaviors and psychological mechanisms) and hierarchical outcome categories (primary, secondary, and exploratory), but not across them.

Behavioral outcomes are our main focus. Our primary hypothesis is that interventions affect water chlorination, measured with the primary outcome of objectively measured water chlorination (10 weeks) and self-reported chlorination (30-36 months). In the follow-up, we add another primary hypothesis, that the intervention affects chlorination-related health outcomes, and measure child diarrhea. After 30-36 months, we correct p -values across these two primary hypotheses.

Our secondary hypothesis tests if the interventions have domain-general effects on future investments. We consider one pre-specified outcome measuring savings behavior, labor supply, and education investment (10 weeks) and savings behavior and labor supply (30-36 months). We adjust p -values across this group of outcomes in each round.

For analysis on psychological outcomes, after 10 weeks, we test three main hypotheses, namely that interventions affect planning, time preferences and self-efficacy, with one primary variable to capture each concept. We correct p -values over the three hypotheses. After 30-36 months, we only examine time preferences and self-efficacy, as we found few short term effects on planning measures. We also run exploratory analysis on some pre-specified and some non-specified variables. We correct across all the exploratory tests we run on behaviors, and separately, on psychological outcomes.

Table A.8: Primary and secondary hypotheses by round

	10-week	30-36 month
BEHAVIOR		
Hypothesis 1: the intervention affects water chlorination		
Presence of any chlorine in household drinking water	Primary	
General water treatment: Indicator for chlorine	Not pre-specified	Primary
Hypothesis 1a: the intervention affects health outcomes		
Number of diarrhea episodes per child under 15 in last 3 months	Exploratory	Primary
Hypothesis 2: the intervention affects future investments		
Amount saved regularly (weekly, KES)	Secondary	Secondary
Total hours of work ¹	Secondary	Secondary
Education investment index ²	Secondary	Not measured
PSYCHOLOGICAL MECHANISMS		
Hypothesis 1: the intervention affects planning ability		
Behavioral Activation for Depression Scale - Short Form (BADSD-SF)	Primary	Not measured
Tower of London task: total moves across all four rounds	Secondary	Not measured
Hypothesis 2: the intervention affects time preferences		
β^{Effort} (estimated from the effort discounting task)	Primary	Not measured
Utility forecasting: Vividness rating	Not measured	Primary
Hypothesis 3: the intervention affects self-efficacy		
Generalized Self-Efficacy (GSE) scale	Secondary	Primary

¹ Hours of work is measured over the last 3 months in the 10-week survey and last 7 days in 30-36 month survey.

² This is made up of any school days missed across all children, expenditure per child on schooling in last three months.

Table A.9: 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
Completed follow-up survey	649	849	848	849	3195
Completed baseline & follow-up survey	-	697	692	684	2073
Completed baseline & endline & follow-up survey	-	652	637	627	1916

Table A.10: Baseline balance: main outcomes

	Comparison with active control (AC+INF)				(5) N
	(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Column 2 vs. Column 3 p-value	
<i>Baseline score</i>					
Tower of London: Total Moves	23.32 (6.83)	-0.48 (0.46) [0.98]	-0.30 (0.42) [1.00]	0.70	2197
General Self-Efficacy Score (GSE)	42.84 (11.59)	-0.07 (0.68) [1.00]	0.90 (0.74) [1.00]	0.19	2175
β^{MPL}	0.99 (0.37)	0.01 (0.02) [1.00]	0.00 (0.02) [1.00]	0.64	2142
δ^{MPL}	0.98 (0.02)	0.00 (0.00) [1.00]	0.00 (0.00) [1.00]	0.52	2142
Self report: added chlorine in last month	0.65 (0.48)	-0.04 (0.03) [0.56]	-0.04 (0.03) [1.00]	0.82	2184
Chlorine knowledge score (z-score)	0.00 (1.00)	0.09 (0.05) [0.56]	0.01 (0.05) [1.00]	0.18	2337
Prevalence of diarrhea among u15 children, last 2 weeks	0.09 (0.24)	0.01 (0.01) [0.41]	0.00 (0.01) [0.82]	0.53	2061

Notes: OLS estimates of baseline values of main outcomes. For each variable, column (1) reports the mean and standard deviation of the control group. Columns (2)-(3) report 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.11: Baseline balance: dispenser vs. non-dispenser villages

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

Notes: OLS estimates of baseline balance on observed characteristics for villages with and without WASH 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 the intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.12: Attrition analysis: treatments vs. active control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Not in endline	Not in chlorine test	Not in follow-up	Not in endline	Not in chlorine test	Not in follow-up	Not in endline	Not in chlorine test	Not in follow-up
Visualization	0.02 (0.02)	0.01 (0.02)	0.01 (0.01)	0.02 (0.02)	0.01 (0.02)	0.01 (0.01)	0.02 (0.12)	-0.05 (0.14)	0.12 (0.12)
Planning	0.03 (0.02)*	0.03 (0.02)	0.02 (0.01)	0.03 (0.02)*	0.03 (0.02)	0.02 (0.01)	0.14 (0.14)	0.15 (0.15)	0.01 (0.12)
Age				-0.01 (0.00)***	-0.01 (0.00)***	-0.01 (0.00)***	-0.01 (0.00)***	-0.01 (0.00)**	-0.01 (0.00)***
Married or cohabiting				-0.03 (0.02)	-0.02 (0.02)	0.00 (0.03)	0.00 (0.03)	0.01 (0.04)	-0.07 (0.04)
Education level				-0.00 (0.01)	-0.01 (0.01)	-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)	-0.00 (0.00)
P x Age Interaction							-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
V x Married Interaction							-0.09 (0.05)*	-0.08 (0.05)	0.10 (0.06)*
P x Married Interaction							-0.01 (0.06)	-0.01 (0.06)	0.12 (0.06)**
V x Education Interaction							0.01 (0.01)	0.00 (0.01)	-0.02 (0.01)
P x Education Interaction							-0.01 (0.01)	-0.01 (0.02)	-0.01 (0.01)
V x Wealth Interaction							-0.01 (0.03)	0.00 (0.04)	0.03 (0.03)
P x Wealth Interaction							-0.03 (0.03)	-0.01 (0.04)	0.02 (0.03)
Constant	0.09 (0.01)***	0.15 (0.02)***	0.13 (0.01)***	0.31 (0.06)***	0.39 (0.07)***	0.41 (0.05)***	0.27 (0.08)***	0.37 (0.10)***	0.38 (0.08)***
Observations	2337	2337	2337	2337	2337	2337	2337	2337	2337

Notes: OLS estimates of the probability of attriting relative to the active control group. For each variable, we report 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 characteristics (columns (4)–(6)) and ii) there was any differential effect of an observed characteristic on the probability of attriting for any treatment group (columns (7)–(9)). All standard errors are clustered at the level of the intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.13: Attrition analysis: active treatments vs. pure control

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Not in endline	Not in chlorine test	Not in follow-up	Not in endline	Not in chlorine test	Not in follow-up	Not in endline	Not in chlorine test	Not in follow-up
V+INF	-0.05 (0.02)***	-0.04 (0.02)*	-0.01 (0.02)	-0.06 (0.02)***	-0.04 (0.02)*	-0.02 (0.02)	-0.02 (0.16)	0.01 (0.17)	0.21 (0.13)*
P+INF	-0.04 (0.02)*	-0.02 (0.02)	-0.01 (0.02)	-0.04 (0.02)**	-0.02 (0.02)	-0.02 (0.02)	-0.04 (0.18)	0.01 (0.18)	0.14 (0.14)
AC+INF	-0.06 (0.02)***	-0.04 (0.02)*	-0.01 (0.02)	-0.06 (0.02)***	-0.04 (0.02)**	-0.01 (0.02)	-0.16 (0.17)	-0.09 (0.17)	0.14 (0.13)
Age				-0.01 (0.00)***	-0.01 (0.00)***	-0.01 (0.00)***	-0.01 (0.00)**	-0.01 (0.00)*	-0.01 (0.00)**
Married or cohabiting				-0.08 (0.02)***	-0.08 (0.02)***	0.02 (0.02)	-0.10 (0.06)*	-0.13 (0.06)**	0.04 (0.05)
Education level				-0.01 (0.01)**	-0.01 (0.01)**	-0.00 (0.00)	-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)	-0.00 (0.00)
P+INF x Age Interaction							0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)
AC+INF x Age Interaction							0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
V+INF x Married Interaction							0.01 (0.07)	0.03 (0.08)	-0.00 (0.06)
P+INF x Married Interaction							0.04 (0.08)	0.08 (0.08)	-0.01 (0.06)
AC+INF x Married Interaction							0.05 (0.07)	0.09 (0.08)	-0.08 (0.06)
V+INF x Education Interaction							0.00 (0.02)	-0.01 (0.02)	-0.02 (0.01)
P+INF x Education Interaction							-0.01 (0.02)	-0.02 (0.02)	-0.02 (0.01)
AC+INF x Education Interaction							0.00 (0.02)	-0.00 (0.02)	-0.01 (0.01)
V+INF x Wealth Interaction							0.00 (0.04)	-0.00 (0.04)	0.01 (0.04)
P+INF x Wealth Interaction							-0.02 (0.04)	-0.02 (0.04)	0.01 (0.04)
AC+INF x Wealth Interaction							0.02 (0.04)	0.00 (0.04)	-0.02 (0.04)
Constant	0.26 (0.02)***	0.29 (0.02)***	0.19 (0.02)***	0.59 (0.06)***	0.62 (0.06)***	0.44 (0.05)***	0.61 (0.13)***	0.61 (0.14)***	0.29 (0.10)***
Observations	3750	3750	3750	3750	3750	3750	3750	3750	3750

Notes: OLS estimates of the probability of attriting relative to the pure control group. For each variable, we report 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 characteristics (columns (4)–(6)) and ii) there was any differential effect of an observed characteristic on the probability of attriting for any treatment group (columns (7)–(9)). All standard errors are clustered at the level of the intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

G. Detailed descriptions of outcome measures

G.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.2. Payment was based on a randomly selected round, with a payment of KES 250 for completing the task with the minimum number of moves, and a KES 50 deduction for each additional move.

G.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.3 provides an example of the participant interface for the task.⁴³ 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 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:

⁴³To 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).

participants completed data entry tasks by SMS, using toll-free numbers administered by the Busara Center.⁴⁴ 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).⁴⁵ 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

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

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 .⁴⁶

66 percent of participants identifiably sent at least one SMS (that was not a practice

⁴⁴Although 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.

⁴⁵The 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.

⁴⁶To control for the effect of introducing a 0 KES piece rate for a subsample of participants (Footnote 43), we 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.

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 45) 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 44).

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.17 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.

G.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.14. 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.

G.4 Alternative mechanisms

G.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 (“Visualization”, “Planning”, 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.

G.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.

G.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.

G.4.4 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. 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. The word lists are available in original Swahili and English translation in Table A.15. We estimate salience effects using equation 3:

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

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 $H_0 : b_1 = b_2 = b_3$, with the null hypothesis corresponding to no differential salience of chlorine across (active) treatments.

G.5 Schedule of tasks and treatments

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. 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.

Figure A.2: Tower of London Example Screen

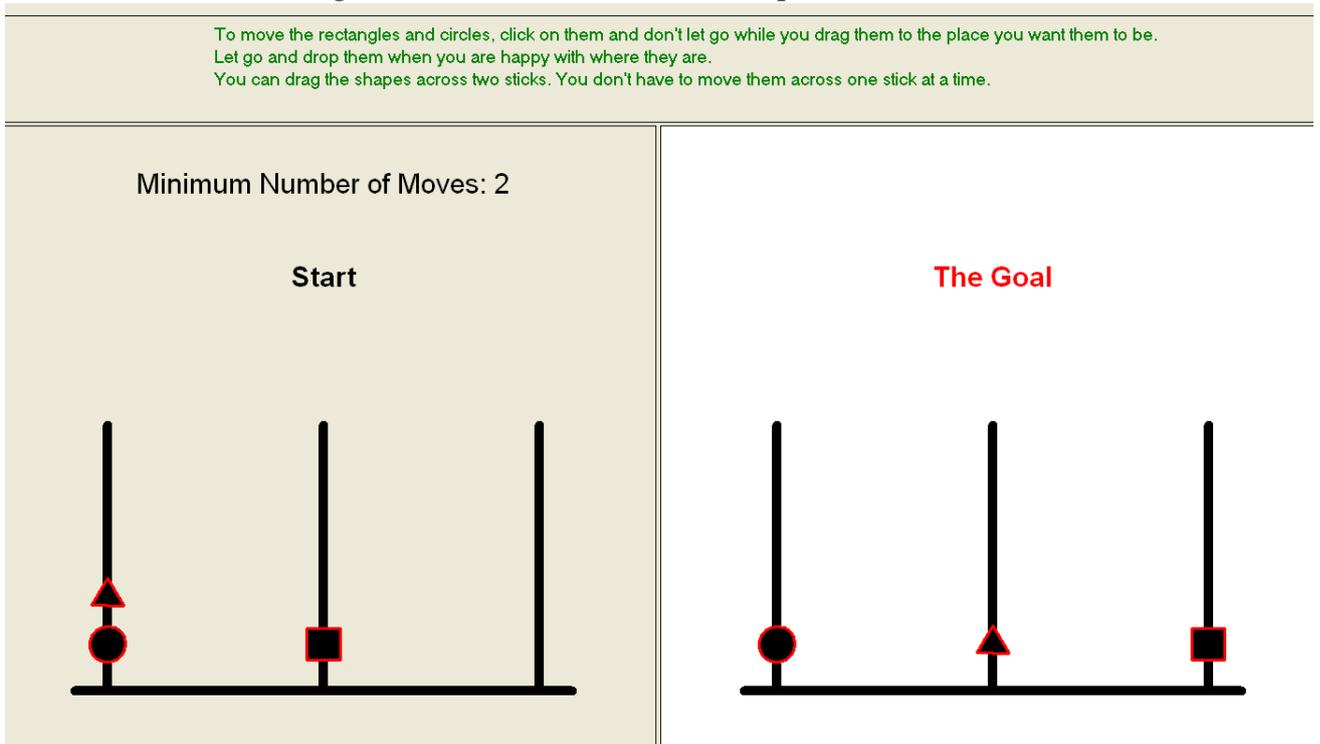


Figure A.3: Effort discounting task example screen

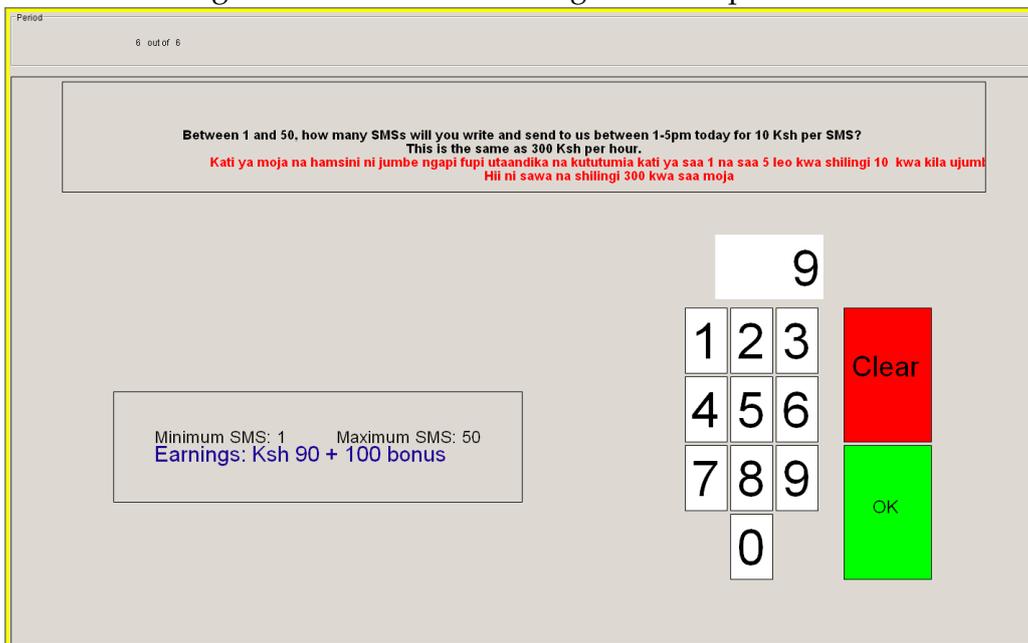


Figure A.4: Effort discounting: raw responses

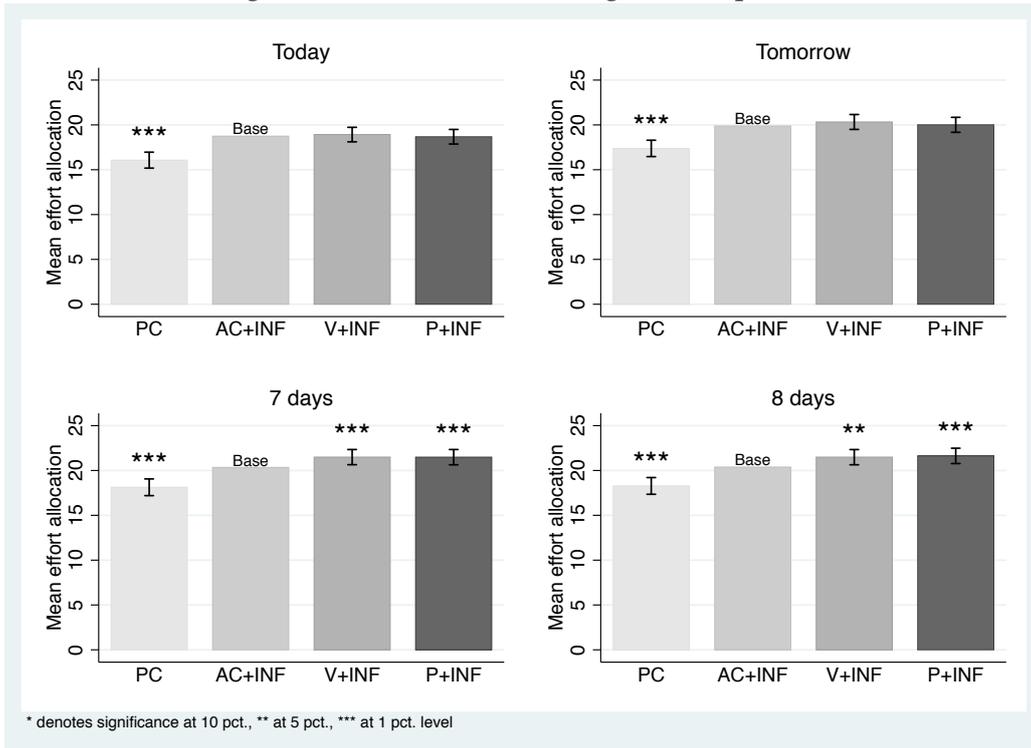


Table A.14: Monetary Discounting Decisions

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

Table A.15: 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.16: 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 (5.90)	28.90 (5.67)
Tower of London: Total moves (9–36)	20.89 (6.58)	23.86 (6.79)
Risk aversion measure (1–3)	1.74 (0.82)	1.70 (0.83)
General Self-Efficacy score (GSE) (0–72)	43.35 (12.11)	43.32 (9.83)
Belief: Proportion of diarrhea incidences avoided through chlorination (0–1)	0.74 (0.39)	0.71 (0.41)
Chlorine knowledge score (0–2)	1.14 (0.68)	1.07 (0.69)
ANC/PNC knowledge score (0–2)	1.25 (0.84)	0.93 (0.87)
Index of investment in children’s education (-3–3)	0.03 (1.14)	-0.07 (0.95)

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

Table A.17: 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 <i>p</i> -value	(5) <i>N</i>	(6) Pure Control Mean (SD)	(7) V+INF Treatment Effect	(8) P+INF Treatment Effect	(9) AC+INF Treatment Effect	(10) <i>N</i>
<i>SMS Task Checks</i>										
Participant uses a phone she owns	0.71 (0.45)	0.01 (0.02)	-0.03 (0.03)	0.15	2108	0.70 (0.46)	0.00 (0.02)	-0.03 (0.03)	-0.00 (0.03)	2972
Participant uses a phone belonging to her household	0.96 (0.19)	0.01 (0.01)	-0.00 (0.01)	0.24	2108	0.96 (0.21)	0.02 (0.01)	0.00 (0.01)	0.01 (0.01)	2972
Proportion for whom accessing a phone for 30mins is very difficult or impossible	0.12 (0.33)	0.03 (0.02)	0.00 (0.02)	0.21	2107	0.13 (0.33)	0.02 (0.02)	-0.00 (0.02)	-0.01 (0.02)	2970
Proportion for whom accessing a phone for 1hr is very difficult or impossible	0.17 (0.37)	0.04 (0.02)*	0.03 (0.02)	0.52	2107	0.19 (0.39)	0.02 (0.02)	0.00 (0.02)	-0.03 (0.02)	2970
Proportion for whom accessing a phone for 4hrs is very difficult or impossible	0.32 (0.46)	0.02 (0.03)	0.01 (0.03)	0.54	2105	0.35 (0.48)	-0.02 (0.03)	-0.03 (0.03)	-0.04 (0.03)	2967
SMS Comprehension questions correct first time	0.81 (0.39)	-0.04 (0.03)	-0.01 (0.03)	0.29	2103	0.72 (0.45)	0.04 (0.03)	0.07 (0.03)***	0.07 (0.03)***	2955
Number of attempts at SMS comprehension questions	0.78 (1.89)	0.23 (0.12)*	0.14 (0.12)	0.51	2103	1.22 (2.36)	-0.21 (0.13)	-0.27 (0.13)**	-0.38 (0.13)***	2955

Notes: OLS estimates of responses to questions asked as validation checks for the SMS effort task. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report 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 the intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

H. Robustness checks

Table A.18: Randomized Experimenter Demand Treatments (de Quidt, Haushofer, and Roth 2018)

	$\alpha^+(\zeta)$ Mean(SD)	$\alpha^-(\zeta)$ Mean (SD)	Test: $\alpha^-(\zeta) = \alpha^+(\zeta)$ p-value
Visualization	2.337 (1.773)	2.264 (1.695)	0.539
Planning	2.269 (2.375)	2.149 (1.649)	0.394
Active Control	2.087 (1.742)	2.291 (1.734)	0.089*
Pure Control	2.163 (1.644)	2.174 (1.595)	0.929
All Treatment Groups Combined	2.217 (1.924)	2.223 (1.674)	0.923
Obs	1556	1616	3,172

Notes: The table shows results from randomized experimenter demand treatments (de Quidt, Haushofer, and Roth 2018), which were included in the long-run follow-up survey. Respondents are assigned to a group A (B), and told “We hypothesize that people who participated in this study and received the same treatment as you will give higher (lower) responses to these questions than others.” They are then asked how often they added chlorine to water collected from their primary source in the last 7 days. Following de Quidt, Haushofer, and Roth (2018), the responses can be used to obtain bounds $a^+(\zeta)$ and $a^-(\zeta)$ for the impact of experimenter demand effects on self-reports. The table shows means and standard deviations for $a^+(\zeta)$ and $a^-(\zeta)$ in all treatment groups, as well as p-values from a test of equality $a^+(\zeta) = a^-(\zeta)$.

Table A.19: Robustness of chlorination effects to within-village testing order

	Comparison with active control (AC+INF)			Comparison with pure control (PC)		
	(1) Chlorine present (TCR)	(2) Chlorine present (TCR)	(3) Chlorine present (TCR)	(4) Chlorine present (TCR)	(5) Chlorine present (TCR)	(6) Chlorine present (TCR)
Visualization	0.05 (0.02)**	0.05 (0.02)**	0.01 (0.04)	0.06 (0.02)***	0.06 (0.02)***	0.02 (0.04)
Planning	0.02 (0.02)	0.01 (0.02)	0.02 (0.05)	0.04 (0.02)*	0.03 (0.02)	0.02 (0.04)
Active Control				0.03 (0.02)	0.02 (0.02)	0.04 (0.04)
Days after first day of testing within village		-0.00 (0.00)			-0.01 (0.00)	
Dummy: test done on first day of testing within village			0.00 (0.04)			0.00 (0.04)
V x with first day of testing			0.06 (0.05)			0.07 (0.05)
P x first day of testing			0.01 (0.05)			0.03 (0.05)
AC x with first day of testing						-0.02 (0.05)
V x with days after first test		-0.00 (0.00)			0.00 (0.00)	
P x with days after first test		0.00 (0.01)			0.00 (0.01)	
AC x with days after first test					0.01 (0.01)	
Observations	2012	2012	2012	2839	2839	2839

Notes: OLS estimates of treatment effects on chlorine in water (TCR) after 12 weeks, with additional controls for testing order within the village. For each variable, we report the coefficients of interest, and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. All columns include village-level fixed effects, a vector of individual characteristics, and standard errors which are clustered at the level of the intervention cohort. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

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

	MHT Level	Endline (10-12 weeks)					Follow-Up (30-36 months)				
		(1) Active Control Group Mean (SD)	(2) Visualization Treatment Effect	(3) Planning Treatment Effect	(4) Col. 2 vs. Col. 3 p-value	(5) N	(6) Active Control Group Mean (SD)	(7) Visualization Treatment Effect	(8) Planning Treatment Effect	(9) Col. 2 vs. Col. 3 p-value	(10) N
HEATH OUTCOMES											
Objective measure: chlorine present in water (TCR)	1/-	0.23 (0.42)	0.05 (0.02)**	0.02 (0.02)	0.15	2012					
Objective measure: chlorine sufficient to be safe (FCR)	3/-	0.21 (0.40)	0.04 (0.02)** [0.04]**	0.01 (0.02) [1.00]	0.16	2012					
Main Treatment; Chlorine (self-report)	np/1	0.73 (0.45)	0.07 (0.02)** [0.00]**	-0.00 (0.02) [1.00]	0.00***	2116	0.85 (0.35)	0.05 (0.02)** [0.03]**	0.02 (0.02) [0.38]	0.15	2073
Main Treatment: Boil (self-report)	np/np	0.35 (0.48)	0.07 (0.03)** [0.02]**	0.05 (0.03)* [0.38]	0.45	2116	0.63 (0.48)	0.03 (0.03) [1.00]	-0.01 (0.03) [1.00]	0.21	2073
Diarrhea incidences per child u15, last 3 months	3/1	0.26 (0.69)	-0.11 (0.03)** [0.00]**	-0.07 (0.03)* [0.38]	0.12	2004	0.27 (0.73)	-0.01 (0.04) [0.27]	-0.05 (0.03) [0.38]	0.27	2045
Diarrhea incidences per child u5, last 3 months	np/3	0.34 (0.86)	-0.16 (0.05)** [0.00]**	-0.06 (0.05) [0.76]	0.03**	1682	0.39 (1.04)	0.01 (0.06) [1.00]	-0.02 (0.06) [1.00]	0.73	1612
Proportion of children taken for healthcare check-up	3/3	0.21 (0.34)	-0.04 (0.02)** [0.03]**	-0.02 (0.02) [0.76]	0.41	1995	0.36 (0.37)	-0.01 (0.02) [1.00]	-0.01 (0.02) [1.00]	0.90	2045
Proportion of children u15 vaccinated, last 3 months	3/-	0.22 (0.35)	0.00 (0.02) [0.35]	-0.01 (0.02) [1.00]	0.35	1990					
Number of ANC visits, last 3 months (if pregnant)	3/-	1.26 (1.19)	-0.24 (0.49) [0.26]	0.06 (0.35) [1.00]	0.44	200					
SAVINGS OUTCOMES											
Amount saved regularly (per week, KES)	2/2	93.96 (230.26)	24.89 (12.37)** [0.16]	3.28 (12.51) [1.00]	0.10	2108	407.50 (605.28)	60.00 (33.65)* [0.10]*	26.67 (33.16) [0.51]	0.34	2073
Indicator: Amount saved regularly is positive	3/3	0.36 (0.48)	0.13 (0.03)** [0.00]**	-0.02 (0.03) [1.00]	0.00***	2108	0.78 (0.42)	0.05 (0.02)** [0.14]	0.02 (0.02) [1.00]	0.13	2073
Number of ROSCAs [joined in last 3 months/ total]	3/3	0.17 (0.44)	0.04 (0.03)* [0.06]*	0.01 (0.02) [1.00]	0.17	2108	1.08 (1.05)	0.13 (0.06)** [0.14]	0.10 (0.06)* [1.00]	0.59	2073
Weekly ROSCA savings	np/np	205.93 (304.72)	34.67 (15.81)** [0.03]**	10.83 (16.02) [1.00]	0.12	2108	246.07 (363.71)	45.90 (22.36)* [0.10]	33.84 (21.24) [0.42]	0.60	2073
Indicator: Saves for productive investments	3/3	0.17 (0.38)	0.11 (0.02)** [0.00]**	-0.01 (0.02) [1.00]	0.00***	2108	0.62 (0.49)	0.02 (0.02) [1.00]	0.01 (0.03) [1.00]	0.74	2073
Total Savings Balance (KES)	-/3						2542.85 (6378.76)	1046.08 (423.20)** [0.14]	45.72 (401.32) [1.00]	0.01***	2073
LABOR OUTCOMES											
Total hours of work [last 3 months/last 7 days]	2/2	106.11 (174.61)	-6.78 (9.56) [0.90]	-24.33 (9.03)** [0.02]**	0.06*	2108	13.25 (15.41)	2.42 (0.87)** [0.05]**	0.91 (0.87) [0.42]	0.09*	2073
Total days of work, last 3 months	3/-	21.22 (30.09)	-0.59 (1.64) [0.32]	-3.91 (1.60)** [0.27]	0.04**	2108					
Earnings, cash and in-kind [monthly/last 7 days]	3/3	1094.50 (2865.35)	3.11 (147.23) [0.39]	-1.23 (163.19) [1.00]	0.97	2108	677.26 (1343.83)	108.14 (77.42) [0.81]	95.12 (79.96) [1.00]	0.87	2073
OTHER BEHAVIORAL OUTCOMES											
Index of investment in children's education (z-score)	2/-	0.00 (1.00)	-0.02 (0.06) [0.90]	0.01 (0.07) [1.00]	0.60	1420					

Notes: OLS estimates of treatment effects, using the specification from the PAP, without survey date fixed effects. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, a vector of individual characteristics, and standard errors which are clustered at the level of the intervention cohort. The sample in all regressions is restricted to participants in active treatment groups who attended the baseline survey. Where available, we control for the baseline outcome of the dependent variable. Outcome measures are listed on the left, and are described in detail in Section 4.

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

	MHT Level	Endline (10-12 weeks)					Follow-Up (30-36 months)				
		(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) Active Control Group Mean (SD)	(7) Visualization Treatment Effect	(8) Planning Treatment Effect	(9) Column 2 vs. Column 3 p-value	(10) N
PLANNING SKILLS											
BADS score (z-score)	1/-	0.00 (1.00)	-0.01 (0.05) [0.83]	0.04 (0.05) [0.91]	0.31	2103					
Tower of London (z-score)	2/-	0.00 (1.00)	0.01 (0.05) [0.79]	-0.04 (0.05) [0.50]	0.28	2103					
TIME PREFERENCES											
β^{Effort}	1/-	0.982 (0.005)	0.007 (0.006) [0.83]	0.005 (0.007) [0.91]	0.33	2068					
δ^{Effort}	2/-	0.999 (0.001)	-0.001 (0.001) [0.26]	-0.002 (0.001)** [0.06]*	0.16	2068					
Utility Forecasting: Vividness (z-score)	-/1						0.00 (1.00)	0.11 (0.05)** [0.06]**	-0.03 (0.06) [0.71]	0.01***	2073
Utility Forecasting: Practice (z-score)	-/2						-0.00 (1.00)	0.10 (0.05)* [0.15]	0.11 (0.05)** [0.13]	0.77	2073
β^{MPL}	1/-	1.05 (0.46)	-0.02 (0.02) [0.41]	0.02 (0.03) [0.53]	0.21	2103					
δ^{MPL}	2/-	0.98 (0.02)	-0.00 (0.00) [0.36]	-0.00 (0.00) [0.31]	0.84	2103					
Time Preferences, qualitative (z-score)	-/2						-0.00 (1.00)	0.10 (0.05)* [0.15]	-0.06 (0.06) [0.36]	0.00***	2073
One-year discount factor (GPS staircase)	-/2						0.52 (0.13)	0.00 (0.01) [0.93]	-0.01 (0.01) [0.36]	0.17	2073
SELF-EFFICACY											
General Self-Efficacy Score (GSE) (z-score)	2/1	0.00 (1.00)	0.16 (0.05)*** [0.00]***	0.12 (0.05)** [0.06]*	0.38	2103	-0.00 (1.00)	0.07 (0.05) [0.08]*	-0.07 (0.05) [0.71]	0.00***	2073

Notes: OLS estimates of treatment effects, using the specification from the PAP, without survey date fixed effects. For each variable, columns (1) and (6) report the mean and standard deviation of the control group. Columns (2)-(3) and (7)-(8) report the coefficients of interest and standard errors in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Square brackets contain additional p-values corrected for multiple hypothesis testing using the false discovery rate. All columns include village-level fixed effects, a vector of individual characteristics, and standard errors which are clustered at the level of the intervention cohort. The sample in all regressions is restricted to participants in active treatment groups who attended the baseline survey. Where available, we control for the baseline outcome of the dependent variable. Outcome measures are listed on the left, and are described in detail in Section 4.