WHAT MOTIVATES HEALTH BEHAVIOR: PREFERENCES, CONSTRAINTS, OR BELIEFS? EVIDENCE FROM PSYCHOLOGICAL INTERVENTIONS IN KENYA*

Johannes Haushofer[†], Anett John[‡], Kate Orkin[§]

November 2018

Abstract

We test the effect of light-touch psychological interventions on water chlorination and related health and psychological outcomes using a randomized controlled trial among 3750 young women in rural Kenya. We randomly allocate participants to two light-touch psychological interventions: one targeting planning skills (executive functions); and one targeting present bias and impatience

*Fieldwork and Princeton and Busara research assistance was supported by grant NIH UH2 NR016378 from the National Institutes of Health to JH, which is part of the NIH Science of Behavior Change program. The project was approved by the Princeton (#7376) and KEMRI (#536) IRBs. Our pre-analysis plan is registered at https://www.socialscienceregistry.org/trials/2850/history/27566 (April 2, 2018). For more information on this study's role in the Science of Behavior Change program, please visit our Open Science Framework page: https://osf.io/twbu8. We are grateful to Jane Dougherty, Daniel Mellow, Moritz Poll, and the staff of the Busara Center for Behavioral Economics for excellent research assistance and data collection. We thank Clair Null, Michael Kremer, and the WASH Benefits Kenya team, for permitting us to cross-randomize in the WASH Benefits study villages and for advice on chlorine measurement protocols. We thank Xavier D'Haultfoeuille and Alessandro Iaria for useful comments. We are also grateful to our NIH Project Scientist, Dr. Rosalind King, for scientific oversight.

[†]Department of Psychology, Woodrow Wilson School of Public and International Affairs, and Department of Economics, Princeton University, Princeton, USA; National Bureau of Economic Research; and Busara Center for Behavioral Economics, Nairobi, Kenya. haushofer@princeton.edu.

[‡]Department of Economics, CREST Paris, France. Anett.JOHN@ensae.fr

[§]Blavatnik School of Government, Centre for Study of African Economies and Merton College, University of Oxford, Oxford, UK. kate.orkin@bsg.ox.ac.uk.

(time preferences). A third group received only information about the benefits of chlorination, and a pure control group received no intervention. Ten weeks after the interventions, the time preferences and executive function interventions led to significant 27 and 18 percent increases, respectively, in the share of households who chlorinated their drinking water, compared to the pure control group. This increase was accompanied by significant 25–26 percent (executive function) and 35–38 percent (time preferences) reductions in the number of child diarrhea episodes relative to both placebo and control. The time preferences intervention also significantly increased participants' savings per week by 26 percent. We further study the psychological channels through which effects occur. The executive function intervention did not affect lab measures of planning, and the time preferences intervention did not affect choices in lab discounting tasks. However, both interventions increased self-efficacy, i.e. beliefs about one's ability to achieve desired outcomes. We show effects are not driven by changes in information about the benefits of chlorination. Together, these results suggest that low self-efficacy may be a psychological barrier to health behavior.

Keywords: time preferences; executive function; self-efficacy; health behaviors; preventive health; randomized controlled trial

JEL codes: O12, D91, I12

1. Introduction

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

In this paper, we consider three potential psychological channels which may explain why some households fail to chlorinate and target them directly with psychological interventions. First, small costs in the present, such as buying and using chlorine, may outweigh distant benefits, such as fewer diarrhea episodes among children. This possibility represents an account in terms of (time) *preferences*.[2] Second, people may have deficits in executive function, i.e. the ability to plan or execute the actions required to

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

[2] Kremer and Glennerster (2011) argue that time preferences – in particular, present bias – and may explain why small but non-zero prices dramatically reduce the observed demand for preventive health products (but see Dupas and Miguel 2017 for a counterargument). Demand for commitment products is also consistent with present bias; see Schilbach (2015), Bai et al. (2017), and Giné, Karlan, and Zinman (2010) in developing countries, and DellaVigna and Malmendier (2006) and Royer, Stehr, and Sydnor (2015) in the United States.

implement their preferences.[3] This possibility represents an account in terms of (cognitive) *constraints*. Finally, individuals may have incorrect beliefs about the effectiveness of chlorination, or pessimistic beliefs about themselves and their ability to improve their family's health outcomes. Both of these mechanisms are accounts in terms of *beliefs*; beliefs about one's ability to achieve desired outcomes are referred to as self-efficacy in the psychology literature (Bandura 1977).[4]

In a randomized controlled trial in rural Kenya, we allocate 3750 young women to four treatment groups. The first group received a two-session intervention that aimed to reduce present bias and increase respondents' valuation of outcomes in the future, to test if changes in time preferences affect take-up of chlorination ("Time Preferences" or "TP" intervention). The second group received a two-session intervention that aimed to improve executive function, i.e. cognitive skills required to form goals, make specific plans, and carry them out ("Executive Function" or "EF" intervention). This intervention tests the role of this specific cognitive constraint in decision-making about chlorination. As both interventions presume and build upon participants' agency, both interventions might plausibly have effects on a third psychological channel, self-efficacy, i.e. people's beliefs about their ability to achieve desired outcomes.

To isolate the effects of the psychologically active elements of these treatments, a third, placebo group received all elements of the intervention except the psychologically active components ("Placebo" or "PLA" intervention): these participants also gathered as a group, but to discuss "psychologically inactive" topics. In addition, all three of these groups received a short information module about the benefits of chlorination ("Information" or "INF" intervention), in an attempt to hold constant beliefs about

[3]Executive functions are the cognitive processes required for forming goals, planning, and carrying out plans directed by goals (Lezak 1983; Miller and Cohen 2001). We target one process, planning: the ability to generate a strategy, including the sequencing of steps, to achieve intended goals, and implement the strategy (Carlin et al. 2000). Previous research suggests that improving planning with detailed implementation plans increases take-up of preventive screening and vaccination (Milkman et al. 2011; Milkman et al. 2013).

[4]Self-efficacy has been explored to a more limited extent as a determinant of preventive health investment. Interventions targeted at improving self-efficacy increase adherence to exercise regimens (Williams and French 2011; Barkley and Fahrenwald 2013; Seghers et al. 2014). A short personal growth course increased both measures of competence and health checkup attendance among Indian sex workers (Ghosal et al. 2016). Furthermore, interventions targeting self-efficacy or related measures, locus of control and aspirations, increase other future-oriented investments, such as education or use of agricultural technology (Beaman et al. 2012; Bernard et al. 2018), as well as women's labor market participation (McKelway 2018).

chlorination across groups.[5] Finally, we compare these treatments to a fourth, pure control group ("PC"), who were simply surveyed at endline. Thus, our groups are "TP+INF", "EF+INF", "PLA+INF", and "PC". The comparison between the three "active" treatment groups and the pure control group gives the policy-relevant effect: the total effect on targeted behaviors of providing interventions such as ours in other, similar settings. The comparison between the two "psychological" treatment groups and the PLA+INF group allows us to test the effect of interventions targeting time preferences, executive function, and self-efficacy over and above those targeting lack of information.

Somewhat surprisingly, we find no effects of our psychological interventions on some of the targeted psychological outcomes. Specifically, eleven weeks after the interventions, the Executive Function treatment showed no significant improvements in planning ability, measured by the "Tower of London" task, a lab measure of ability to plan or sequence activities, in which participants have to make and implement a plan to move a set of shapes on a screen into a new configuration. The Executive Function treatment also did not affect a self-reported measure of whether participants were making plans to do necessary tasks and following through on them, rather than avoiding them, the Behavioral Activation for Depression Scale (BADS), relative to the Placebo and Time Preferences groups (although it did result in a small increase relative to the pure control group, significant at the 10 percent level). Similarly, the Time Preferences treatment did not work exactly as theoretically predicted, in that it did not significantly affect its target, lab measures of present bias and discount rates, measured separately over money and over effort, relative to the Placebo and pure control groups.

However, we do find large and significant effects of both the Time Preferences and Executive Function treatments on a self-efficacy scale, which was pre-specified as a third psychological channel. Together, these results suggest that lab measures of planning ability and time preferences may not respond strongly to light-touch interventions, while beliefs about one's ability to achieve desired outcomes appear more malleable.

We next examine the effects of these interventions on water chlorination and related health outcomes, as well as economic outcomes. In the Time Preferences and Executive Function groups relative to the pure control group, we find statistically significant increases of 27 and 18 percent, respectively, in the share of households whose drinking

[5]Our intervention is similar in format to Jensen (2010), who gives information on returns to education.

water contains chlorine 13 weeks after the interventions. This measure is objective and collected in unannounced household visits, largely ruling out demand effects. The effect in the Placebo group, who only receive information and gather in groups and perform unrelated tasks, is smaller and not significant compared to the pure control group. It is also smaller than the effect of the Executive Function and Time Preferences interventions, although only the difference between the Time Preferences and Placebo groups is statistically significant. In line with these findings, the Time Preferences and Executive Function treatments significantly reduce the number of diarrhea episodes in children. This effect is statistically significant both in comparison to the pure control group (35 and 26 percent, respectively), and in comparison to the Placebo group (38 and 25 percent, respectively). In contrast, the effect in the Placebo group relative to the pure control group is small and not statistically significant. Importantly, all three "active" treatment groups which received the information treatment show statistically significant and similar increases in their belief that chlorination can prevent diarrhea, but the effects of these treatments on health outcomes are different.

These findings suggest that psychological interventions targeting executive function, time preferences, and self-efficacy significantly affected health behaviors and outcomes. These effects go beyond those achieved by simply providing information, even when information successfully changes beliefs about the effectiveness of engaging in health behaviors. The interventions may change behavior by affecting people's internal beliefs about their ability to realize desired outcomes: both "psychologically active" interventions increased self-efficacy, and did not affect other psychological targets. We thus conclude that psychological interventions targeting preferences and psychological constraints can complement traditional interventions targeting incorrect beliefs about facts, but they may take their effects by affecting beliefs about ability.

The effect of our interventions is not limited to the health domain, but also extended to economic behavior: while the Time Preferences intervention did not affect our laboratory measure of time preferences, it caused a statistically significant 36 percent increase in the share of individuals who save regularly, and a 26 percent increase in the amount saved per week. These effects were larger than those of the Executive Function intervention. Given the fact that self-efficacy increased in both groups, an alternative interpretation of the savings results is that our Time Preferences intervention increased patience and future orientation in a way that is not captured in our lab measure of time preferences.

The Time Preferences and Executive Function interventions were designed to be domain-general, but the scripts also mentioned chlorination. Mention of chlorination may focus participants' attention on this behavior (Zwane et al. 2011). We test for this possibility by measuring the salience of three future-oriented behaviors (chlorination, savings, and farm investment) compared to non-future oriented behaviors. We find indeed that the Time Preferences, Executive Function, and Placebo treatments all increased the salience of chlorination (but not of savings or farm investment), with a stronger effect for the Time Preferences and Executive Function groups than the Placebo group. This effect constitutes a possible explanation for our treatment effects on chlorination. However, our treatment effect on savings, as well as on various other non-chlorine measures, is difficult to explain through salience, as the salience of savings was unaffected by treatment. Thus, increases in salience do not provide a consistent explanation across our findings, unless the mapping from salience to behavior is both non-linear and differential across domains. Relatedly, treatment effects may reflect social desirability bias in answering questions related to chlorination. However, we observe increases in objectively measured chlorine content of household drinking water during unannounced household visits, suggesting that this possibility is unlikely.

Finally, our design also allows us to investigate the relative role of psychological factors and monetary and effort costs in determining chlorination. We cross-cut these four treatment groups with a previous randomized experiment, the "WASH Benefits" study, in which villages were randomly assigned to receive chlorine dispensers placed at the water source (Null et al. 2018). This feature allows us to test if our psychological interventions have larger effects in villages where access to chlorination is easier or more difficult. Our interventions may have larger effects in villages without dispensers if (i) people in villages with dispensers chlorinate already, and (ii) changes in psychological targets (such as time preferences, executive function, and self-efficacy) can compensate for higher costs of chlorination. Alternatively, if the cost of accessing chlorine without dispensers remains a barrier to chlorine use, then treatment effects may be higher in dispenser than in control villages. Indeed, our treatment effects on chlorination are somewhat larger in villages with dispensers, although differences are mostly statistically insignificant. Thus, when both psychological and cost/effort constraints are alleviated simultaneously, effects on behavior may be larger than when cost and access barriers remain, but we observe only weak evidence of this.

Our study builds on a small literature which uses light-touch interventions to affect

real-world behaviors through psychological mechanisms. Bernard et al. (2018) show that showing farmers videos of role models similar to them who have improved their economic position increases aspirations, savings, educational investment, and investment in productive technologies in Ethiopia. Similar to our setting, their interventions work through respondents' beliefs about their own ability, rather than through changes in preferences. Ghosal et al. (2016) show that a short course on personal growth for sex workers improved their self-worth and their beliefs about their own ability, as well as increasing savings and attendance at health checkups. McKelway (2018) shows that an intervention specifically targeted at increasing self-efficacy produces increases in employment. All of these results lend further credence to the hypothesis that beliefs about one's ability may be malleable and affect economic and health outcomes. At the same time, Alan and Ertac (2018) use an eight-session educational intervention in Turkish primary schools to increase patience, suggesting that changing preferences may be possible. Relatedly, more involved, multi-session interventions that resemble psychotherapy have been shown to improve both psycho-social and economic outcomes (Blattman, Jamison, and Sheridan 2017; Heller et al. 2017; Baranov et al. 2017). We build on this work by testing the relative effect of interventions specifically targeting preferences, beliefs, and constraints independently, rather than focusing on the effect of one psychological intervention. Second, by cross-cutting our intervention with one which provides chlorine dispensers at the water source, we can study how psychological interventions such as ours interact with others that have been shown to affect health behaviors by reducing cost or increasing ease of access to technologies.

Our work also builds on, but is distinct from, research demonstrating that limited information and attention may affect economic decisions. People are known to increase investment in high-return opportunities, such as health and education, when information about benefits and returns is provided (Jensen 2010; Jensen 2012; Dinkelman and Martínez 2014). This is also true in the context of drinking water, where information about unsafe arsenic levels can lead to substitution to different water sources (Madajewicz et al. 2007), and promotion of chlorination by community health workers can increase adoption (Kremer et al. 2011b; Kremer et al. 2011a).

Similarly, countering people's limited attention by pointing out low-productivity behaviors, such as farmers not noticing important factors in the growth of crops (Hanna, Mullainathan, and Schwartzstein 2014), or tradespeople not noticing the time lost looking for change (Beaman, Magruder, and Robinson 2014), can alter behavior in ways

which increase returns. We find that information has some effects on its own, but the effects from targeting psychological constraints, preferences, and beliefs about oneself go beyond them.[6]

Finally, our paper also contributes to the literature on demand for preventive health care, particularly take-up of chlorination (Kremer et al. 2011b; Dupas et al. 2016; Kremer et al. 2011a). A number of scholars have suggested that psychological constraints, in particular present bias and poor planning, may explain low demand for such products, and many empirical findings are consistent with the presence of such constraints (Dupas and Miguel 2017; Kremer and Glennerster 2011). Our study goes beyond previous work by targeting these psychological constraints directly to examine their causal effect on behavior.

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

2. Experimental design

2.1 Study site

Our trial areas are Bungoma and Kakamega counties in rural Western Kenya. These counties were included in the WASH Benefits study (henceforth WASH) of Null et al. (2018). In one treatment group of the WASH study, villages were provided with chlorine dispensers next to water sources, as well as community health promoters (see Section D). We focus on women because they are primarily responsible for household chores, including collecting water, and thus for water chlorination. We recruit women aged 18–35 as they are most likely to have small children, who in turn are the most vulnerable to water-borne illnesses. As shown in Table 1, the women in our sample are on average

[6]Indeed, our results suggest a possible reinterpretation of past findings: some "information" interventions in this literature may work by combining pure information with elements targeting constraints, preferences, or beliefs about oneself. For example, information about chlorination provided by promoters may also operate through psychological mechanisms (Kremer et al. 2011b; Kremer et al. 2011a); information on financial aid for university delivered through videos of role models might not only provide information, but also enhance self-efficacy (Dinkelman and Martínez 2014). In line with this view, messaging around free safe water products is particularly effective in increasing adoption when it is behaviorally inspired (Luoto et al. 2014).

26 years old, 89 percent are married or co-habiting, and they have on average 6 years of education.

2.2 Sampling

The WASH study was a cluster-randomized controlled trial, conducted from 2012 to 2014, which included 1,226 villages (Null et al. 2018). Villages were 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. Villages were grouped into clusters of one to three neighboring villages. Clusters were then randomised to eight treatment arms, six of which which tested household-level water, sanitation, handwashing, and nutrition interventions in isolation and different combinations of interventions.[7]

We sampled 205 villages from the full sample of 1,226 villages. We randomly selected 67 villages from the "Water Quality" (henceforth "dispenser") treatment arm and 67 villages from the "Passive Comparison" arm.[8] In the dispenser villages, chlorine dispensers were installed at an average of five community water points per village cluster, and refilled as needed. The "Dispensers for Safe Water" program operated by the NGO *Evidence Action* has since maintained these dispensers, ensured they are filled with chlorine, and retained a local promoter in each community. The randomization has remained intact. "Passive comparison" villages received no interventions. An additional 71 villages were selected uniformly from the remaining six WASH Benefits arms in Mumias constituency, Kakamega county. The inclusion of participants from villages treated by the remaining six arms of the WASH intervention was not planned in our study, but occurred as a result of a coding error at the sampling stage. During the WASH Benefits study, 48 of these 71 villages received one of sanitation, handwashing or nutrition interventions, but no dispensers. Howver, these additional interventions all took place at the household level, and we exclude households which participated directly in the original WASH study from participation in the present study. In addition, all interventions other than dispensers finished three to four years before our study. For these reasons, we group these 48 villages withthe passive comparison villages

[7]For more information on the WASH Benefits study, see Appendix D.

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

in our analysis.[9] The other 23 villages received dispensers, as well as combinations of sanitation, handwashing or nutrition interventions. We group these villages with dispenser villages, as *Evidence Action* continued to maintain dispensers in these villages. As a robustness check, we exclude these 71 villages from the heterogeneity analysis by "Water Quality" assignment in Appendix A.

In the 205 study villages, we recruited 3750 women aged 18–35 between October 2016 and January 2017. With the help of local guides, enumerators visited all households in each included village (see Section D) and conducted a census to determine household eligibility. Enumerators collected demographic information on women that met the screening criteria: i) aged 18–35 inclusive, ii) within this age range, the most senior woman in their household, and iii) their household did not participate in the WASH Benefits study. The target sample of 3750 women represents all eligible women in the 205 villages.

2.3 Randomization

We split our sampleinto three "active" treatment groups and one pure control group. We randomly assigned 992 participants to the "Time Preferences" (TP+INF) group, 991 to the "Executive Function" (EF+INF) group, 992 to the "Placebo" (PLA+INF) group, and 775 to the "Pure Control" group. We stratified the randomization on village of residence, and a wealth index collected during the census.[10] Redundant observations during stratification were equally distributed across treatment groups.

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

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

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

2.4 Design and timeline

An overview of the study timeline is shown in Figure 1. Participants in the three "active" treatment groups participated in three sessions: a combined baseline and first intervention session; a second intervention session one week later; and an endline 11 weeks after the first session. Participants in the pure control group did not participate in the baseline and the first and second intervention sessions, but participated in the endline 11 weeks after the beginning of the study in their village. In addition, households in all four groups received an unannounced visit 13 weeks after the beginning of the study, where enumerators collected a drinking water sampleto be assayed for the presence of chlorine.

The baseline and endline sessions lasted about two hours each and consisted of behavioral tasks, psychological questionnaires, and a socioeconomic survey. These measures were collected in "mobile labs" operated by the Busara Center for Behavioral Economics in Bungoma and Kakamega counties, which each accommodated up to 25 participants at a time. The behavioral tasks and psychological questionnaires were presented using touch screen computers and the zTree experimental interface (Fischbacher 2007). Enumerators read instructions to the respondents in Kiswahili to maximize comprehension.[11] The combination of touchscreen interfaces, audio instructions, and color-coded response buttons allowed both computer-illiterate and entirely illiterate individuals to participate. The socioeconomic survey was administered one-on-one by enumerators using tablets.

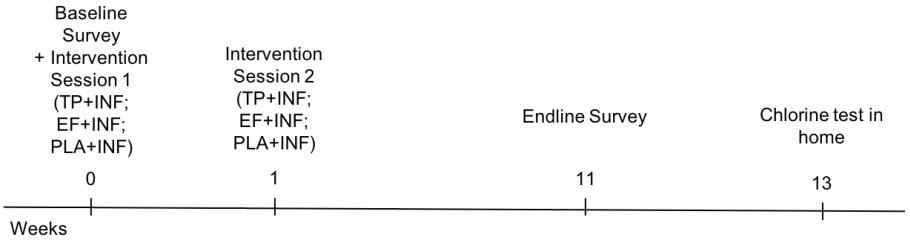
Each intervention session in the three "active" treatment groups lasted about two hours. The structure of each session was held constant across treatment groups: each included a short lecture, followed by a story of a woman like them, reflections of how the themes relate to participants' own lives, and finally drawing and list-writing exercises and activities. Participants were split into cohorts of five for each session, which were run by locally-trained female facilitators. Participants were reconvened in the same cohorts for the second session. No participant was invited for the second session without having already participated in the first session.

Participants received KES 200 (USD 2) for participating in the baseline survey and first intervention session, KES 200 for the second session, and KES 300 for participating

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

in the endline session.[12] They additionally received a KES 50 bonus for arriving on time for each appointment. Participants were reimbursed for their transport costs, using public transport rates from their village of residence to the mobile laboratory. Finally, participants received payments from the experimental tasks, as described in Section 4. Participants were informed of the fixed amounts that they would be paid during the phone call inviting them to a session, and told that they would have the opportunity to earn some extra money during the session itself. All participants recruited to the sample were invited to attend endline sessions, regardless of whether they attended the baseline and/or intervention sessions. A full overview of the number of participants at each stage of the study can be found in Table A.10.

Figure 1: Study timeline



2.5 Background on chlorination use

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

Chlorinating water kills many of the pathogens that cause diarrhea.[13] Absent point-of-collection chlorine dispensers (as installed by WASH), households can purchase dilute

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

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

chlorine. The main brand of dilute chlorine is *WaterGuard*, which has been distributed, heavily marketed, and quality-controlled 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 of household drinking water), and comes with instructions in Swahili and in pictures. There is widespread awareness of this product: even at baseline of the WASH study in 2012, 89 percent had heard of *WaterGuard,* and 29 percent had used it at least once. Households also believe chlorine is effective: at endline, the pure control group believe, on average, that 71 percent of incidences of diarrhea can be avoided by using chlorine.

However, take-up of chlorination is low. Even using chlorine from a dispenser requires repeated, proactive behavior. Using *WaterGuard* from a bottle requires, in addition, paying for the product and the time cost of repeated purchases. In our sample at endline, 27 percent of pure control households report having chlorinated their current drinking water supply, 18 percent had detectable free chlorine in their water, and 22 percent had detectable total chlorine.[14]

Chlorinating drinking water will only partially reduce diarrhea incidence, as it does not reduce other risk factors for diarrhea. For example, in the baseline data from Null et al. (2018), 77 and 14 percent of children aged 0–3 and 4–8, respectively, defecate in the open air. However, other risk factors are fairly contained in this setting: 75 percent of households in our study area had an improved drinking water source, 96 percent report using a latrine for defecation, and 82 percent own a latrine. Water chlorination thus may be an important margin in reducing high levels of diarrhea.

3. Interventions

3.1 Treatment 1: Time Preferences + Information module ("TP+INF")

The time preference 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,

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

focusing on abstract qualities rather than details of execution (see e.g. Gilbert and Wilson 2009, Kahneman et al. 2004, Wilson et al. 2000). For instance, helping out an elderly relative with their tax return next month may be imagined as an act of love, while doing it later today is imagined as hours of painstaking sorting through receipts. In a recent theoretical contribution, Gabaix and Laibson (2017) formalize the idea of *as-if discounting*, which results from a perfectly patient decision-maker that simulates future utility by combining priors with noisy, unbiased signals about future utility. The model implies that interventions which improve forecasting ability (or forecasting efforts) will lead to more patient behavior. Models of quasi-hyperbolic discounting (Laibson 1997 and O'Donoghue and Rabin 1999) are based on a differential weight (or salience) given to the present. Despite fundamental differences, both models predict that an increased salience of the future will be associated with fewer dynamic preference reversals.[15] This theoretical prediction is matched by empirical evidence: in a randomized educational intervention in Turkish primary schools, Alan and Ertac (2018) find that weekly classes and exercises on "imagining future selves" result in the children making more patient decisions in incentivized choice tasks *three years* after the intervention. Our intervention is conceptually similar to that used in Alan and Ertac (2018), though it is shorter (two two-hour sessions instead of eight two-hour sessions).

Through interactive lectures, case stories, exercises and drawings, we encouraged 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 being asked to close their eyes repeatedly for several minutes, to imagine 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. During the lecture, participants were asked to think about the effect of small everyday behaviors (such as smoking a cigarette, or spending their leftover budget on snacks) on future outcomes. The intervention carefully avoided changing participants' beliefs about which present behavior would entail which future outcome – it merely encouraged them to make a connection themselves. The lecture

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

further suggested that the future may feel vague and distant, and thus we may not connect it to our actions today. It moved on to several silent visualization exercises, with the following prompts: 1. "Close your eyes for one minute. Imagine the person you will be in one year. Imagine your family in one year. Use details." 2. "Now connect your present behaviors with your future self. If you behave as you behave in the present, which kind of future will you get?" 3. "Close your eyes again. Imagine that your future self can now talk to you. How does she feel? What does she think about your behavior in the present? What does *she* want you to do?"

In the second part of the first intervention session, participants listened to a story of a woman very similar to them. Her daily life is full of tasks and worries. As a result, she always focuses on what is necessary right now, and does not think too much about the future. Using examples of water chlorination, antenatal care visits, handwashing, and saving, she learns over time that thinking about the future in her everyday actions helps her and her family to have a better life.

The story was followed by a group discussion on alternative realizations of the protagonists' future. It then moved on to eliciting example behaviors from the participants' lives, and visualizing alternative realizations of their own future, depending on present behavior. The first session finished by asking participants to draw one these "future realizations", and make notes on the corresponding present behavior. To distinguish the Time Preference intervention from the Executive Function intervention, it minimized any planning features, and focused on high-level behaviors and outcomes without implementation details.

In the second TP intervention session, participants largely revisited the first session, albeit with increased emphasis on how future visualizations can be used to deal with temptations in the present.

Information module ("INF")

The intervention 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). These behaviors were used as real-world examples of important health behaviors in the Time Preferences and Executive Function treatments. The information provided on chlorination consisted of the following text, which was read out to the participants: "Not all "drinking water" is safe, even when it is sold as

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

3.2 Treatment 2: Executive Function + Information module ("EF+INF")

Executive functions are the cognitive processes required for forming goals, planning, and carrying out plans directed by goals (Lezak 1983; Miller and Cohen 2001). The Executive Function intervention aimed to increased respondents' ability to set goals, make plans to achieve them, and execute their plans, rather than avoiding them. We drew on an approach used with patients with mild depression known as behavioral activation.[16] One symptom of depression is that people reduce how often they undertake activities and avoid even basic tasks (Lejuez et al. 2011). We adapted a low-intensity treatment manual called "Reach Out" (Richards and Whyte 2011) from a clinical trial of depression management in the UK (Richards et al. 2008), and added elements from "A Brief Behavioral Activation Treatment for Depression" (Lejuez et al. 2011). In addition, we included goal-setting exercises, which involved mental contrasting (contrasting one's actual situation with the desired situation), describing implementation intentions (small manageable steps to achieve goals), and listing if-then strategies for overcoming obstacles (Duckworth et al. 2013; Morisano et al. 2010; Milkman et al. 2011).

In the first intervention session, participants first listened to a story of a woman very similar to them who was tired and struggling to complete her chores, including fetching

[16]Importantly, we do not screen for or target people with depression symptoms, or attempt to provide any treatment for depression.

water and chlorinating it. Participants could then volunteer to share experiences of similar situations from their lives. The goal of this exercise was to convey to participants that it is very common for people to become stuck in inactivity and avoid important tasks, especially if they are facing difficult events or adversity; and that they may feel a lack of energy and motivation to do things, or find it difficult to get going on tasks or achieve goals.

In the second part of the first intervention session, participants were asked to set simple, achievable goals. In contrast to the time preferences intervention, they did not set long-term goals, but merely sought to identify a few current activities in their daily lives where they were struggling to get going. Working in pairs on a simple worksheet, using drawing or writing, they made two lists of activities: one set they enjoyed doing, and one set that were necessary and important but they might not necessarily enjoy doing. They ranked the tasks on each list from most to least difficult.

In the final part of the first intervention session, participants made achievable plans towards a subset of these goals, and plans to overcome obstacles. Again in pairs, they picked the easiest one or two activities from each of the "necessary activities" and "enjoyable activities" lists, and scheduled them in a weekly diary. They then broke the task down into steps, visualized what they would need to do to complete the activity, anticipated potential obstacles, and made plans to overcome them. The first intervention session concluded with the same information module described above.

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

3.3 Treatment 3: Placebo exercise + Information Module ("PLA+INF")

The goal of the "placebo" intervention was to control 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, and hence included a lecture, discussion, some drawing and some list-writing. The content of these sessions centered on the birds and plants of Kenya, topics chosen intentionally to be psychologically inactive. In the first session, participants listened to a short lecture on different kinds of birds that live in Kenya, followed by a short story about the daily routine of a woman similar to

them. Participants discussed the birds they see in their village, and any birds they were particularly interested in. They wrote a list of all the birds they could think of, and then made some drawings of birds. The second session followed a identical structure, except that it did not include another short story about a woman, in line with the second sessions of the other treatment groups. The second session centered on plants in Kenya. In addition, participants in this group also received the same information module as the Time Preferences and Executive Function treatment groups described above.

3.4 Pure Control ("PC")

The pure control group received no contact prior to endline, except for the brief demographic questionnaire administered during household recruitment.

4. Outcome measures

4.1 Primary behavioral measure: Validated measures of chlorination

Enumerators made unannounced visits to participants' homes to test the household's stored drinking water for the presence of chlorine. These tests were conducted roughly two weeks after the endline survey, to minimize experimenter demand effects in the survey (de Quidt, Haushofer, and Roth 2018). We test both Total Chlorine Residual (TCR) and Free Chlorine Residual (FCR), using TCR as our main chlorination outcome measure of interest as specified in our pre-analysis plan. TCR indicates the presence of any chlorine in the water; i.e., that the household had at some point added *some amount* of chlorine to the drinking water. FCR indicates that not only has chlorine been added, but that there is still enough unreacted chlorine in the water to keep it sanitized; i.e., that the household added *sufficient* chlorine to the drinking water. Although correct usage of chlorine is an important outcome, we are primarily interested in whether households attempt to chlorinate their water at all, and thus focus on TCR.[17]

We follow the same chlorine testing procedure as Null et al. (2018) and Kremer et al.

[17]This is also the outcome of interest in Kremer et al. (2011a), who are interested in chlorine take-up. Null et al. (2018) focus on FCR.

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

4.2 Other behavioral measures

While chlorination is our primary outcome of interest, and was pre-specified as such, our interventions were not intended to be specific to chlorination. Time preferences and executive function are relevant for many everyday behaviors, and in particular for future investments like savings, education, and agricultural investments. During the endline survey, participants therefore completed several modules on economic and health behaviors. We pre-specified secondary and exploratory outcomes in the domains of health (diarrhea, vaccinations, and ante-natal care visits), savings (internal and external margin), labor supply, and educational investment.

4.3 Psychological measures

The majority of the psychological measures were selected for inclusion based on validation in a previous study (Esopo et al. 2018). This study involved translation and back-translation of the measures into Kiswahili, cognitive interviewing to establish cultural accceptability of the measures, and tests of internal consistency, test-retest reliability, and construct validity. The measures used in the present study, which were validated in this previous study, are Monetary Price Lists for time preferences, the Tower of London task, and the General Self-Efficacy Scale.

4.3.1 Time preferences

Typical tasks to measure time preferences involve choices over monetary payments available at different times. However, because water chlorination is an effortful task, rather than a monetary payment, these tasks may not adequately capture time preferences in this domain. Following recent innovations in the elicitation of time preferences

(Andreoni and Sprenger 2012; Augenblick, Niederle, and Sprenger 2015), we therefore estimate time preferences in the effort domain, using a newly developed real effort task from 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.[18] 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: participants completed data entry tasks by SMS, using toll-free numbers administered by the Busara Center.[19] 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).[20] Earnings from this task were paid 14 days from

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

[19]Although we did not screen on phone access, all participants in our sample have access to a mobile phone: 70 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.

[20]The field setting with SMS required some tolerance: while a laboratory computer can confirm

the survey date, regardless of the selected effort time horizon.

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

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

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 TP, EF, and PLA on the parameters β, δ, s, and γ.[21]

66 percent of participants identifiably sent at least one SMS (that was not a practice SMS during the session), 60 percent sent the correct number of SMS during the correct time window, and 41 percent additionally satisfied the required accuracy threshold (see footnote 20) 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

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.

[21] To control for the effect of introducing a 0 KES piece rate for a subsample of participants (Footnote 18), 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.

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 19).

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

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

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.11. Figure A.2 provides an example of the participant interface for the MPL. 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.

4.3.2 Executive Function

We measure two aspects of the planning component of executive function. First, to measure whether people choose to make plans and follow through on them, we employ a modified version of the Behavioral Activation for Depression Scale (Kanter et al.

2007). The full questionnaire contains 29 items. We use the short form of the scale, "(BADS-SF)", developed by Manos, Kanter, and Luo (2011), who carried out item reduction procedures until 9 items remained. In these questionnaires, participants are asked to identify how much statements about BA were true for them in the past week, including both positive (e.g. "I was an active person and accomplished the goals I set out to do") and negative items (e.g. "There were certain things I needed to do that I didn't do"). Responses range from "not at all" (1) to "completely" (7). Some items are reversed before summing to generate a composite score.

Second, to measure the higher-order cognitive skill of ability to plan, we use a common psychological measure, a version of the *Tower of London task* (TOL; also known as the Stockings of Cambridge task when implemented electronically), which is designed to measure a participant's ability to plan ahead in sequential strategies (Shallice 1982; Phillips et al. 2001). In our computerized version of the task, participants see a screen with two parts: on the left side is the word "start" with a picture of three "pegs" and various shapes positioned on the pegs; on the right side is the word "goal" with a similar picture of three "pegs" and the same shapes positioned differently on the pegs. To complete the task, participants must reposition the shapes underneath the "start" on the left to match the "goal" position on the right. They are instructed to complete each round in as few moves as possible, with the minimum number of moves shown as a number on the screen. In addition to a practice round, participants attempt four rounds of increasing complexity, beginning with one shape requiring only one move, and concluding with three shapes in a pattern that necessitates at least four moves. In all rounds, participants are limited to a maximum of 10 moves. If this occurs, the round ends and the participant is required to contact a staff member to ensure she understands the task before continuing to the next round. Therefore, the distribution of scores is censored at both ends. Performance on the Tower of London task is computed as the total number of moves used across the four rounds. An example of the participant's screen is shown in Figure A.1.

4.3.3 Self-Efficacy

We measure self-efficacy using the General Self-Efficacy (GSE) scale (Schwarzer and Jerusalem 2010). This self-report questionnaire measures individuals' general belief in their ability to cope with problems and perform novel or difficult tasks. Participants

are asked to rate the truthfulness of statements such as "I can always manage to solve difficult problems if I try hard enough" on a scale from "Never true" (1) to "Always true" (6). Our version contains 12 items: 10 from the generic version, and two which are repeated and reversed, to check for consistency. Higher scores indicate higher self-efficacy.

4.3.4 Salience effects

Our interventions were designed to target time preferences and executive function in general, rather than specifically to increase chlorination use. However, chlorination was mentioned in the intervention scripts. This raises the possibility of salience effects: it is possible that the mention of chlorination itself represents a nudge to chlorinate, by making chlorination salient to participants. We test for this possibility by measuring the salience of three future-oriented behaviors (chlorination, savings, and farm investment) compared to non-future oriented behaviors. 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 (see Table A.12 for the list of words). While the recall of words is clearly driven by memory, the recall of words *conditional* on the total number of words remembered captures whether a concept is "top-of-mind". We thus test whether our treatments differentially affect the probability to recall chlorine words, conditional on the total number of words remembered. If our treatments differentially affect the salience of chlorine, we further test whether this is due to increased salience of future-oriented behaviors in general by asking whether the differential treatment effect also holds for two other future-oriented behaviors (saving and farm investment), which were not emphasized in the sessions (see Appendix C for the empirical specification).

Alternative Mechanisms We collect a range of measures which allow us to test whether behavioral changes occurred through mechanisms other than time preferences, self-efficacy, or executive function (see Appendix C for a full description and empirical specification).

First, the treatments may change participants' beliefs about the effectiveness of chlorination in preventing disease. We test this hypothesis by assessing differential

beliefs across treatment groups about the proportion of pediatric diarrhea cases which can be prevented by water chlorination. At baseline, all participants (except the pure control) were told that water chlorination reduces childhood diarrhea by approximately one third. At endline participants are asked this question in a multiple choice format. We take the proportion of diarrhea cases the participant believes chlorine can avert as a measure of the strength of their beliefs about chlorine effectiveness.

Second, treatment may affect chlorination by providing information about how to properly use it, as was done explicitly in the INF module. We ask two multiple-choice questions at endline, to which all three groups were told the correct answer at baseline: 1. how much chlorine to add to water, and 2. the amount of time that needs to pass after adding chlorine for water to be safe to drink. We score each question as a binary measure of whether the participant answered correctly and create a composite value that ranges from 0 to 2, which is then z-scored.[22]

Finally, to test for the possibility that the treatments affect risk preferences, we include a modified Eckel-Grossman measure of risk preferences in the endline survey (Charness, Gneezy, and Imas 2013). This measure offered a choice between one of three 50/50 lotteries, represented as bets on a coin flip. We construct an ordinal measure of risk aversion with three levels, based on the expected payout the participant is willing to forgo for an increase in certainty of payout.

5. Econometric approach

5.1 Experimental integrity

To ensure experimental integrity, we test for balance across treatment groups in (1) demographic variables, (2) timing of the surveys relative to the intervention, (3) attrition in the endline survey and the chlorination test at home, and (4) compliance with the assigned treatment (i.e., participation in the intervention sessions).

To determine whether the randomization was balanced, we regress baseline demographics available for the entire recruited sample (age, years of education, marital sta-

[22]We similarly check for differential knowledge of WHO-recommended antenatal and postnatal care. These questions were included to pilot them for a future study. We ask participants how many antenatal and postnatal care visits a woman should make before and after giving birth, information that they were given at baseline. We score each qusetion as a binary measure of whether the participant answered correctly and create a composite which ranges from 0 to 2, which is then z-scored.

tus, and village, see Section 2.2) on indicators for all treatment groups. The reference group is either the placebo control (PLA+INF) or the pure control group (PC). The specification is identical to that used for the estimation of treatment effects (described in Sections 5.2 and 5.3), leaving out controls $\mathbf{X}_i$ and lags y_{i0}.

We further test for differences in the timing of the endline survey relative to the baseline survey and first intervention date (Figure 1), as well as the timing of the chlorine test relative to the baseline survey and first intervention date. For participants in the pure control group, and those in the treatment groups who did not attend the interventions, we use predictive mean matching to simulate a proxy intervention date, based on the actual intervention dates of other participants from their village of residence. The outcome variables are the number of days between the first intervention and the endline survey, and between the first intervention and the chlorine test. The specification is identical to that used for the estimation of treatment effects.

We also test for selective attrition in attending the endline survey and the chlorination test at home, for both the placebo group comparison and the pure control comparison. The specification is identical to that used for the estimation of treatment effects. Additional checks assess whether attriting individuals are different in terms of observed demographics.

Finally, although recruited participants did not know their treatment assignment prior to arriving for the first intervention session (see Figure 1), we test for differential compliance across treatment groups - i.e., the decision to participate in the first and second intervention session. Again the specification is identical to that used for the estimation of treatment effects, except that the outcome variable is an indicator for session attendance.

5.2 Main specification: Executive Function and Time Preferences treatments compared to Placebo

We employ the following main specification:

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

Here, y_{i1} is the outcome of interest for respondent i at time of endline, and y_{i0} is the outcome at baseline; the latter is only included where data exist. The sample

excludes the PC group. Thus, the PLA+INF group is the reference category, and T_{1i} and T_{2i} refer to the "Time Preferences" and "Executive Function" groups, respectively. $\mathbf{X}$ represents a vector of participant controls (year of birth, employment status, marital status, education level), γ_v are village fixed effects, and θ_w is an indicator for household wealth greater than the sample median (which was used in stratification). Standard errors are clustered by intervention cohort (five participants) to account for within-group dynamics.

Only the psychological scales and behavioral tasks were collected at baseline and endline. For these variables we include y_{i0}, the same outcome variable at time of baseline, in the regressors. Where only some baseline observations of a variable are missing, we replace the missing values with zero and add a dummy variable indicating such cases, following Jones (1996). We remove outliers by winsorizing outcome variables which have no theoretical lower and upper bounds at the 1st and 99th percentile. As pre-specified, the sample includes all participants who completed both the endline survey and at least the first intervention session. Note that this approach does introduce the potential for selection based on treatment because the nature of the intervention was not revealed before the first intervention session.

5.3 Comparison with pure control group

We also report results from comparing the three active (EF+INF, TP+INF, and PLA+INF) to a pure control group (PC). The specification is identical to that in equation 2, except that there is a third treatment indicator T_{3i} for the placebo group, and the pure control group is used as the reference category. Further, since the pure control group was not surveyed at baseline, the estimation does not control for the baseline outcome y_{i0}. Finally, because the pure control completed the endline survey on average two days earlier than the active treatment groups, the specification controls for the number of days between baseline and endline. The sample includes all recruited participants who completed the endline survey, including 'non-compliers' who were assigned to the active treatment groups, but chose not to participate in the baseline survey or the interventions.

5.4 WASH Benefits cross-randomization

To study differential effects of our treatments in villages which do and do not have free and convenient access to chlorine through dispensers, we separately estimate the same pure control specification for dispenser and control villages using seemingly unrelated regression (SUR). This approach allows coefficients on control variables to vary by village treatment status, and at the same time allows us to compare treatment coefficients across models. The primary outcome of interest is an indicator for objective chlorination (TCR).

Due to a coding error in sampling (see Section 2.2), some participants were drawn from treatment arms of the WASH Benefits study other than "Water Quality" or "Passive Comparison." As noted above, we include these participants in the main analysis of the cross-randomization, but exclude them in a robustness check in Appendix A.

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

As pre-specified, we also report "treatment-on-the-treated" effects from an instrumental variables specification in which treatment status is instrumented with treatment assignment. For this purpose, we consider all those in any non-pure control group, including the placebo group, who attended at least the initial baseline session to be complying with treatment assignment, even if they didn't complete the second session.[23] For the comparison of the TP+INF and EF+INF to the PLA+INF group, the sample is already restricted to compliers, so that the TOT effect is the same as the Average Treatment Effect (ATE). We therefore report the TOT only for the comparison of the three active treatments to the pure control group.

5.6 Multiple hypothesis testing (MHT) correction

We use a stepdown procedure to adjust p-values for the false discovery rate (FDR) among groups of outcomes, and report the resulting "q-values" (Benjamini, Krieger, and Yekutieli 2006). We adjust for multiple hypothesis testing within outcome groups (psychological mechanisms and behaviors) and hierarchical categories (main and additional), but not across them. Similarly, we consider the effects of our two active

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

interventions to be theoretically distinct and therefore do not correct across them. Indices are constructed following Anderson (2008).

6. Results

6.1 Experimental integrity

Table 1 provides results on baseline balance on demographic variables, timing of the endline surveys relative to the intervention, differential attrition, and compliance with treatment. To test for baseline balance, we estimate a version of equation 2 with baseline demographics as the outcome variables. Each row shows baseline balance for one demographic variable. Columns (1)–(5) show the comparison of the Time Preferences (TP+INF) and Executive Function (EF+INF) treatments to the placebo (PLA+INF) treatment, and Columns (6)–(10) show the comparison of the TP+INF, EF+INF, and PLA+INF groups to the pure control group. Columns (1) and (6) show the mean and standard deviation of the respective comparison groups. Columns (2) and (3) show the treatment effects for the TP+INF and EF+INF treatments, respectively, relative to the PLA+INF treatment. Column (4) is a test of equality between these two coefficients, and Column (5) shows the sample size, which varies slightly across rows because some respondents did not answer a small number of questions, some questions are restricted to certain respondent groups, e.g. those with children, and some observations are removed in trimming as described in 5.2. Our demographic variables are well-balanced across treatments on the whole, with only one out of 30 coefficients on pairwise comparisons reaching statistical significance, at the 10 percent level.

The second panel in Table 1 shows results across treatment groups in terms of the number of days between the date of baseline and first treatment and the date of endline, and then between the date of baseline and first treatment to the date of the chlorine test at the household. Column (1) shows that the average delay between the beginning of the interventions and endline was 69 days, i.e. ten weeks, and the average delay to the chlorine test was 79 days. There are no statistically significant differences in survey timing or chlorine testing relative to the PLA+INF group.We see that both the TP+INF group and the EF+INF group on average completed the endline survey 2 days later than the pure control group. Although this difference is small in magnitude, we include a control for the days elapsed between intervention and endline

to the specfication described in Section 5.3 for outcome measures collected at endline in the laboratory. This approach ensures that differences between groups are driven by treatment rather than by differences in the length of time elapsing between intervention and endline.

The third panel in Table 1 shows results on attrition in the endline survey as well as in the chlorination measure. In the endline survey, average attrition in the PLA+INF and pure control groups was 8 and 24 percent, respectively. Average attrition from the chlorine measurement, conducted at people's houses, was 12 percent in the PLA+INF group and 26 percent in the pure control group. In comparison to the PLA+INF group, we find no differential attrition from either the endline survey or the chlorine measurement in the TP+INF group, and a small increase in the likelihood of attrition in the EF+INF group, significant at the 10 percent level. In comparison to the pure control group, we find small but statistically significant differential attrition on the endline survey for the pure control group and the TP+INF and PLA+INF treatment groups (-6 percentage points). In addition, both the TP+INF and PLA+INF groups are 4 percentage points less likely to attrit from the chlorine measure than the pure control group, although this effect is only significant at the 10 percent level.

Importantly, Tables A.1 and A2 show that we find little evidence that this differential attrition led to differences in sample composition that would complicate inference. Columns (2) and (3) show that demographic variables do predict attrition from either endline measurement, once treatment status is controlled for. However, the interaction terms between demographic variables and treatment status in Columns (3) and (4) show that participants with particular characteristics are no more or less likely to drop out of the study in any one of the treatment groups compared to the pure control group. This result suggests that the composition of the sample is similar in all treatment groups, including in the pure control group, and that differences in sample composition are unlikely to be responsible for observed differences between treatment groups. Relative to the PLA+INF group, we do see that married women in the TP+INF group are slightly less likely to attrit, significant at the 10 percent level. However, given the number of comparisons made in our attrition analysis, without correction for multiple hypothesis testing, we are not overly concerned about this result.

The final panel in Table 1 shows compliance rates across the treatment groups. After the census, all respondents in the treatment groups were invited to the baseline and first intervention session, which were held at the same time. 78 percent of respondents

completed the first session. Only respondents who attended the first intervention session were invited to the second session. 74 percent of respondents completed both sessions, while 4 percent did not complete the second session. Compliance is balanced across treatment groups.

Finally, Table A.4 presents results from atest of equality of means of outcome measures collected at baseline between active treatment groups. We see no statistically significant differences in outcome scores between the active treatment groups prior to the interventions.

6.2 Results for psychological targets of our interventions

We now turn to the results on the psychological and behavioral outcomes of interest. We present three main sets of results: first, a comparison of the TP+INF and EF+INF treatments to the PLA+INF group; second, a comparison of these three groups to the pure control group; and finally, separate analyses in villages with chlorine dispensers vs. villages without dispensers, and the corresponding interaction terms with our treatment groups. The raw means and standard deviations of all outcome measures that were subsequently z-scored are reported in Table A.5.

Table 2 shows results on the psychological outcome variables, estimated using equation 2. The arrangement of columns is as described above. The top panels show results on outcomes related to executive function, time preferences, and self-efficacy, each of which is subdivided into main and additional outcomes according to our pre-analysis plan. Adjustment of p-values for multiple comparisons using FDR is done separately for main and additional outcomes in each family of variables. The final panel shows variables measuring mechanisms.

For outcomes related to executive function, time preferences, and self-efficacy, the main comparison of interest is that of the "psychologically active" treatment groups to the PLA+INF control group. The information conveyed by the treatments, as well as exposure to field staff and other participants, is held constant across these treatments, providing the cleanest identification of changes in outcomes related to constraints, preferences, and beliefs. The exclusion of the non-compliers at baseline and the first intervention session implies that each individual in the sample received at least one treatment session. This comparison is also more precisely estimated than the comparison to the pure control group due to the inclusion of the outcome at baseline as a control variable

on the right-hand side. Finally, the comparison to placebo is not confounded by a possible practice effect from having completed the tasks at baseline.

We find no statistically significant effects of the Executive Function treatment on executive function, with a non-significant 0.04 SD increase in the Behavioral Activation Score relative to a PLA+INF mean, and a non-significant 0.04 SD reduction in the number of moves required to complete the Tower of London task. Thus, the EF+INF treatment did not affect the psychological outcomes it was designed to move.

In comparison to the pure control group we observe a 0.10 SD increase in the Behavioral Activation Score, significant at the 10 percent level. Performance of the EF+INF group on the Tower of London task is significantly improved relative to the pure control group, but this result is likely due to the fact that this group had performed this task previously, while the pure control group had not.

Neither of the executive function outcomes are affected by the TP+INF treatment, which is expected given that the intervention is not designed for this purpose. We also find no significant effects of the TP+INF, and PLA+INF treatments on executive function outcomes relative to the pure control group, except increased performance in the Tower of London task. As above, this result is most likely due to the fact that the three active groups had completed the task before.

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

The third panel of Table 2 shows the effect of our interventions on the General Self-Efficacy (GSE) scale, our measure of self-efficacy. Both the TP+INF and the EF+INF interventions generate statistically significant 0.17 SD and 0.25 SD increases, respectively, in this measure of self-efficacy relative to the PLA+INF group. This finding suggests that interventions geared to affect (time) preferences and constraints

can also affect beliefs; in this case, "inward-looking" beliefs about one's ability to achieve desired outcomes. We find similarly large and significant differences between the Time Preferences and Executive Function groups and the pure control group on this outcome. The comparison between the PLA+INF treatment and the pure control group reveals a very small difference that is not statistically significant.

We next compare these results to those obtained using a treatment-on-the-treated estimator. Results are shown in Table A.6. Note that, as explained above, the TOT estimator is defined only for the comparison to the pure control group because the sample in the PLA+INF comparison is restricted to compliers at baseline and intervention. We find very similar point estimates, and the same significance levels, with the exception of an even more strongly significant impact of the TP+INF treatment on self-efficacy relative to the pure control group. Thus, the TOT results mirror very closely those of the Intention-to-Treat (ITT) estimation.

6.3 Results for beliefs and alternative psychological measures

We next discuss results on a range of outcomes that we measure to confirm that differences between the two psychological treatments and the placebo are largely due to differences in the effects of the psychologically active components. The first obvious concern is that all treatments could simplychange people's beliefs about the effectiviness of chlorination or their level of information. Our experimental design includes an information module in all three active treatment groups. This ensures that any differences between the EF+INF, TP+INF, and PLA+INF groups are unlikely to be driven by changes in people's beliefs about chlorination in response to receiving information, or from differential knowledge about how to chlorinate water. They are much more likely to be driven by the psychologically active components in the EF+INF or TP+INF treatments.[24]

[24]Strictly, the treatment effect of each psychological treatment compared to the pure control consists of the psychologically active element (e.g. elements targeting executive function), the content of the placebo (information and the effect of gathering in a group), and the interaction between the psychologically active element and the placebo (i.e. $EF + INF$ compared to pure control $= ENF + INF + EF \times INF$, and $TP + INF$ compared to pure control $= TP + INF + TP \times INF$). Thus, the treatment effect of each psychological treatment compared to the placebo consists of the psychologically active element and the interaction between the psychologically active element and the content of the placebo (e.g. EF compared to placebo $= EF + EF \times INF$). The interaction term (e.g. $EF \times INF$) would be positive if, for example, the psychologically active treatment made people process information better or if the psychologically active treatment made gathering in a group have an effect

Of course, this argument relies on the EF+INF, TP+INF, and PLA+INF groups actually affecting beliefs and knowledge in the same way. If, for example, the placebo group was very boring, such that people stopped paying attention and did not take on board the information component, the information module in this group may have been less effective. However, the last panel of Table 2 shows that the information treatment worked as anticipated: We find that all three interventions, TP+INF, EF+INF, and PLA+INF, significantly increase beliefs in the efficacy of chlorine in averting diarrhea relative to the pure control group, with similar effect sizes across the three interventions (0.12 SD, 0.15 SD, and 0.11 SD, respectively, all significant at the 5 percent level).

Effects on knowledge about how to correctly chlorinate water are also very similar across the three active treatment groups. Compared to the pure control group mean, the TP+INF group shows a significant increase in knowledge about chlorination of 0.11 SD. Effects are similar in magnitude for the other groups (0.07 SD for EF+INF and 0.08 for PLA+INF), although these effects do not reach statistical significance.[25]

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

In sum, neither the TP+INF nor the EF+INF intervention affected time preferences and executive function, as intended, but both interventions affected self-efficacy. This finding suggests that it may be difficult to affect executive function and time preferences over a short time horizon, but that it may be possible to change beliefs about one's ability to achieve desired outcomes.

6.4 Behavioral outcomes

We next turn to behavioral outcomes, which are shown in Table 3. The arrangement of columns is the same as in the previous table. The different panels show impacts on

on outcomes. We cannot separate the effect of the interaction from the main effect of the psychologically active treatments and consider them both as part of the effect of the psychologically active treatment. Furthermore, it is arguably unlikely that such interaction effects would be positive if the psychologically active treatment had no main effect, or that they would be large.

[25]We also find that all three interventions are effective in increasing knowledge of the correct schedule of antenatal and postnatal care visits a woman should attend before and after giving birth. Compared to the pure control group, all three treatment groups show increased knowledge of approximately 0.30 SD. We do not see an associated increase in the number of ANC visits attended by pregnant women in our sample, but our sample size for this analysis is less than 300 individuals. These questions were included for piloting purposes for a future study.

different families of outcomes, namely health, savings, labor, and other outcomes. Each of these families is again subdivided into main and additional outcomes.

Both the Time Preferences and the Executive Function interventions lead to significant increases in our primary behavioral outcome measure, the presence of chlorine in household drinking water, as measured with an objective chlorine test during an unannounced household visit. In comparison to the pure control group, the TP+INF group shows a 6 percentage point increase in the presence of total chlorine, significant at the 1 percent level. The EF+INF group shows a 4 percentage point increase in the presence of total chlorine relative to the pure control group, significant at the 10 percent level.[26] The treatment effects in the TP+INF and EF+INF groups relative to pure control correspond to 27 and 18 percent increases relative to the pure control group mean of 22 percent. The treatment effect in the PLA+INF group relative to the pure control group is small (3 percentage points) and not statistically significant. In line with this result, chlorination is significantly higher in the TP+INF group (5 percentage points, significant at the 1 percent level) relative to the PLA+INF group; the effect is smaller and not significant in the EF+INF group relative to the PLA+INF group (2 percentage points).

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

Importantly, given our goal of improving health outcomes through psychological interventions, the TP+INF and EF+INF treatments both generate large and statistically significant reductions in the incidence of diarrhea among children. We find significant reductions in diarrhea episodes in both of these groups relative to the pure control group, with a 35 percent reduction for TP+INF and a 26 percent reduction for EF+INF.[27] The effects of the TP+INF and EF+INF interventions are also signifcant relative to the PLA+INF group (TP+INF: 38 percent reduction, significant at the 1 percent level using both naïve and FDR-corrected p-values; EF+INF: 25 percent reduc-

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

[27] We cannot isolate whether these effects occur through increased use of chlorine or an increase in other behaviors which might reduce diarrhea, such as washing hands more frequently or discouraging defecation in the open. However, we did not directly promote changes in these behaviors in our interventions.

tion, significant at the 5 percent level using conventional p-values, but not significant using FDR-correction). We cannot reject that the TP+INF and EF+INF treatments are equally effective in reducing diarrhea.

We find no strong effects on other health outcomes, with the exception of a small changes in the number of children under the age of 15 who completed a healthcare check-up in the last 3 months. We have no good explanation for this finding.

The second panel of Table 3 shows effects on savings-related outcomes. The TP+INF treatment leads to a significant increase in our main savings-related outcome variable, the amount of money saved regularly. This effect corresponds to a 26 percent increase relative to the placebo group, significant at the 5 percent level; relative to the pure control group, it is smaller (17 percent) and not statistically significant. In addition, we find a large and highly significant effect on savings on the extensive margin, with the share of respondents who save regularly increasing by 13 percentage points (36 percent) in the TP+INF group relative to PLA+INF, and 12 percentage points (38 percent) relative to the pure control group. Similarly, we find increases in the TP+INF treatment on an indicator for whether the respondent saves for productive investments, 11 and 9 percentage points (65 and 53 percent) relative to the PLA+INF and pure control groups, respectively. All of these effects are significant at the 1 percent level even after correcting for multiple comparisons. TP+INF participants are also more likely to have joined a new ROSCA relative to PLA+INF participants. The EF+INF and PLA+INF interventions did not show significant effects on savings-related outcomes, and most of the effects in the TP+INF group described above are significantly larger than in those groups. Together, these results show that the TP+INF treatment strongly affected savings-related behaviors.

This finding has two important implications. First, it shows that our TP+INF intervention, while unsuccessful in affecting time preferences measured with laboratory-like discounting tasks, nevertheless strongly affected future-oriented behaviors. Second, it provides evidence against an experimenter demand-effect account of the impacts of our interventions: while the interventions mentioned chlorination and health-related topics, they did not prompt participants to engage in saving behaviors, and therefore the treatment effects we report here are likely to result from changes in the underlying preferences rather than a simple desire to please the experimenters.

The third panel of Table 3 reports effects on labor-related outcomes. Somewhat surprisingly, we find a reduction in the total number of hours worked in the last 3

months in the EF+INF group, with a magnitude of 24 hours (22 percent) relative to the PLA+INF group, and 21 hours (20 percent) relative to the pure control group. These effects are significant at the 1 percent and 5 percent level, respectively. A similar effect is observed in the total number of days worked, which is reduced by 4 days or 18 percent in the EF+INF relative to the PLA+INF group, and by 3 days or 16 percent relative to the pure control group, significant at the 5 percent and 10 percent level after FDR correction, respectively. One possible reason for this reduction is again that participants in the EF+INF group may have improved their planning and execution ability and therefore be more efficient in accomplishing tasks. In support of this hypothesis, we find no significant change in average monthly earnings, despite the shortened work hours. The TP+INF and PLA+INF treatments do not show significant effects on labor outcomes.

We again compare these results to those of the TOT estimation, shown in Table A.7. Again the point estimates and levels of significance are similar, with the exception of the effects of TP+INF and EF+INF on our secondary chlorination measure and diarrhea outcomes, where the TOT results do not survive correction for multiple hypothesis testing.

6.5 What explains changes in behavior: information, psychological targets, or both?

We next explore the extent to which these changes in behavior are due to changes in the psychological mechanisms targeted by our two interventions, or, alternatively, due to changes in beliefs, information, or some other aspect of being involved in the treatment (such as gathering as cohorts). We conclude that, while some of the effects of our intervention undoubtedly work by altering respondents' level of information about chlorination, the Time Preferences and Executive Function treatments have additional effects on behavior that are likely to be due to changes in psychological variables, in particular self-efficacy, which is affected in both main treatment groups. We present five pieces of evidence in line with this argument.

First, the Time Preferences and Executive Function interventions increase chlorination and reduce child diarrhea compared not only relative to the pure control group, but also the placebo group, which receives the same information about chlorination. This result suggests that information alone cannot account for the effect of the psychological

interventions.

This is not to say that the PLA+INF treatment has no effect on behavior. Indeed, the difference in the effect sizes and significance in the comparisons to PLA+INF and the comparisons to pure control suggests that a combination of group meetings and the provision of some information on the benefits of chlorination may work in conjunction with our interventions to affect chlorination behavior to some extent (recall that the information module was administered to all groups). But the effect is undoubtedly smaller than the effect of the psychological treatment.

Second, as shown above, there are very few differences in knowledge and beliefs between the two psychological treatments and the placebo group which might account for differences in behavior. If increased chlorination rates in the EF+INF and TP+INF group were a result of changes in knowledge or beliefs about the effectiveness of chlorination (for example, if the psychological treatments cause some differences in how information is received and taken up), we should also observe more accurate beliefs and greater knowledge in these groups, but this is not the case.

Third, we observe effects of the Time Preferences and Executive Function treatments on behaviors which are not mentioned at all in the intervention. The TP treatment affects savings, which makes sense as it attempts to make respondents consider and invest more in the future. The EF treatment potentially makes people somewhat more efficient, lowering their hours and days worked without affecting earnings. These impacts on more "distal" outcomes suggest that our interventions may have had domain-general psychological effects, which then caused behavioral changes in multiple domains.

Fourth, increased attention and salience of chlorination induced by the interventions cannot fully explain the results. Section 4.3.4 explains the design of a test for increased salience of chlorination, and Appendix C explains the econometric specification. Column (1) of Table 5 shows that participants who had received the TP or EF intervention indeed found it easier to remember chlorine-related words, conditional on the total number of words remembered. This effect was specific to chlorination-related words: Column (3) interacts treatment indicators with chlorine word indicators, and shows that the salience of chlorination increases differentially to the salience of other future-oriented words (the base category is farm investment). Finally, Column (4) suggests that neither treatment group had an effect on total words remembered, and thus on participants' memory in general.

Thus, we find an increase in the salience of chlorination-related words, but not

savings-related words, as a result of treatment. This pattern of results argues against the possibility that the effect of the interventions on chlorination behavior and savings are driven by increases in salience, since only the salience of chlorination is increased. At the same time, reverse causality is unlikely: the fact that we find increases in future-oriented behaviors such as savings without an accompanying increase in their salience suggests that engaging in a behavior has limited effects on the behavior's salience. We therefore conclude that the increase in salience of chlorination is likely to be incidental to the intervention.

Finally, experimenter demand effects, such as social desirability bias in responding to survey questions about chlorination, is an unlikely mechanism for our results because we observe treatment effects not only on self-reported outcomes, but also on objective measures of chlorine in household drinking water during unannounced home visits.

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

We next ask whether our treatment effects differ by whether or not the village in which the interventions took place was randomly treated with a chlorine dispenser in the WASH Benefits study that preceded ours. Table A.3 presents results showing that, at baseline, participants from villages with chlorine dispensers were no different in observables from participants from villages without chlorine dispensers, except that participants from villages with dispensers have a slightly higher level of education than those in villages without dispensers. Table 4 shows results from the two main estimating equations, focusing on chlorination-related outcomes, separately for WASH control villages (Columns (1)–(5)) and WASH treatment villages (Columns (6)–(10)). The interaction terms on our two treatment groups with the WASH treatment (i.e. chlorine dispensers) are shown in Columns (11)–(13) for TP+INF, EF+INF, and PLA+INF, respectively.

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

dispenser compared to those without. The other interaction terms are not statistically significant.

In Table A.8, we test the robustness of these results when restricting the villages to those from the Water Quality and Passive Comparison arms of the WASH Benefits study. This analysis permits a cleaner distinction between dispensers and control, in the absence of any other interventions, at the expense of a somewhat smaller sample. We find that in this analysis, the interaction term for both chlorination measures is significant for the TP+INF intervention, suggesting that this intervention had significantly larger effects on chlorination in dispenser villages compared to non-dispenser villages.

Together, these results present suggestive evidence that our psychological interventions, in particular the TP+INF treatment, had somewhat larger effects in dispenser compared to non-dispenser villages. However, these effects are small and require replication before being taken seriously.

7. Conclusion

In this randomized experiment, we study the effect of two light-touch interventions on psychological and behavioral outcomes among young women in Kenya. Specifically, we ask whether a "Time Preferences" intervention reduces present bias and/or increases patience, and whether an "Executive Function" intervention improves behavioral activation and executive function, and whether both interventions affect self-efficacy. We find no evidence for an effect of the interventions on present bias, patience, behavioral activation, or executive function. However, both interventions lead to an increase in self-efficacy, i.e. positive beliefs about one's abiltiy to achieve desired outcomes. Both interventions affect our primary behavioral outcome, chlorination of drinking water, relative to the pure control group, as well as the number of diarrhea episodes in children, and other outcomes. Together, these results suggest that low self-efficacy may be a psychological barrier to health behavior.

These results also suggest that light-touch psychological interventions have the potential to move intermediate psychological outcomes, as well as distal behavioral outcomes, in developing country contexts. However, an important caveat is that it may be difficult to target specific intermediate psychological outcomes precisely, even if the intervention has the desired effects on the distal outcomes. This fact underscores

the importance of measuring psychological mechanisms to ensure external validity and translation to new contexts, even for interventions that appear to cleanly target one particular mechanism.

The fact that we found stronger effects on self-efficacy, i.e. beliefs about one's own abilities, than on time preferences and executive function, suggests that beliefs are more malleable than preferences and cognitive constraints such as planning ability. Thus, a change in beliefs, e.g. about one's ability to affect health outcomes through chlorination, may be sufficient to move behavioral outcomes, even in the absence of a chance in underlying preferences and cognitive constraints. Future work may attempt to replicate these effects, and extend the use of our interventions to other settings and behaviors of interest.

References

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

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

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

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

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.

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

Barkley, Sherry, and Nancy Fahrenwald. 2013. "Evaluation of an Intervention to Increase Self-Efficacy for Independent Exercise in Cardiac Rehabilitation." *Behavioral Medicine* 39 (4): 104–110.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova. 2012. "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India." *Science* 335 (6068): 582–586.

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

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli. 2006. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika* 93 (3): 491–507.

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

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

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

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

Carlin, Danielle, Joy Bonerba, Michael Phipps, Gene Alexander, Mark Shapiro, and Jordan Grafman. 2000. "Planning Impairments in Frontal Lobe Dementia and Frontal Lobe Lesion Patients." *Neuropsychologia* 38 (5): 655–665.

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

Charness, Gary, Uri Gneezy, and Alex Imas. 2013. "Experimental methods: eliciting risk preferences." *Journal of Economic Behavior & Organization* 87:43–51.

Cohen, Jessica, and Pascaline Dupas. 2010. "Free Distribution or Cost-Sharing?

Evidence from a Randomized Malaria Prevention Experiment." *The Quarterly Journal of Economics* 125 (1): 1–45.

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

Lezak, Muriel. 1983. *Neuropsychological Assessment.* 2nd Edition. Oxford University Press, New York, NY.

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.

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.

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

Miller, Earl, and Jonathan Cohen. 2001. "An Integrative Theory of Prefrontal Cortex Function." *Annual Review of Neuroscience* 24 (1): 167–202.

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

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.

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

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

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, Karina Lovell, Simon Gilbody, Linda Gask, David Torgerson, Michael Barkham, Martin Bland, Peter Bower, Annette Lankshear, Allison Simpson, et al. 2008. "Collaborative Care for Depression in UK Primary Care: A Randomized Controlled Trial." *Psychological Medicine* 38 (2): 279–287.

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

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

Schilbach, Frank. 2015. "Alcohol and Self-Control: A Field Experiment in India." *American Economic Review (Forthcoming)*.

Schwarzer, Ralf, and Matthias Jerusalem. 2010. "The General Self-Efficacy Scale (GSE)." *Anxiety, Stress, and Coping* 12:329–345.

Seghers, Jan, Ann-Sophie Van Hoecke, Astrid Schotte, Joke Opdenacker, and Filip Boen. 2014. "The Added Value of a Brief Self-Efficacy Coaching on the Effectiveness of a 12-Week Physical Activity Program." *Journal of Physical Activity and Health* 11 (1): 18–29.

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

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

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

Williams, Stefanie, and David French. 2011. "What Are the Most Effective Intervention Techniques for Changing Physical Activity Self-Efficacy and Physical Activity Behaviour - and Are They the Same?" *Health Education Research* 26 (2): 308–322.

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

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

Table 1: Experimental integrity

	Comparison with active control (PLA+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Time Preferences Treatment Effect	(3) Executive Function Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) TP+INF Treatment Effect	(8) EF+INF Treatment Effect	(9) PLA+INF Treatment Effect	(10) N
Baseline balance										
Age	26.37	−0.10	0.01	0.60	2975	26.62	−0.42	−0.36	−0.31	3750
	(4.56)	(0.20)	(0.20)			(4.69)	(0.22)*	(0.22)	(0.22)	
Married or cohabiting	0.89	−0.00	0.01	0.28	2975	0.90	−0.02	−0.01	−0.02	3750
	(0.32)	(0.02)	(0.01)			(0.30)	(0.02)	(0.01)	(0.02)	
Education level	5.87	−0.02	0.06	0.11	2975	5.93	−0.08	0.00	−0.05	3750
	(1.23)	(0.05)	(0.05)			(1.08)	(0.05)	(0.05)	(0.05)	
High wealth index	0.54	−0.02	−0.03	0.61	2975	0.52	0.00	−0.01	0.02	3750
	(0.50)	(0.02)	(0.02)			(0.50)	(0.02)	(0.02)	(0.02)	
Village of residence	83.26	0.84	−0.20	0.84	2975	83.31	0.80	−0.24	−0.07	3750
	(54.89)	(4.96)	(4.90)			(56.43)	(4.18)	(4.14)	(3.97)	
Delay variables										
Days between endline and baseline	68.92	0.59	0.88	0.76	2396	68.73	1.82	2.12	1.23	2984
	(21.52)	(0.98)	(0.93)			(24.07)	(1.06)*	(1.01)**	(1.00)	
Days between chlorine test and baseline	79.33	0.45	1.46	0.46	2203	81.20	0.60	1.55	0.09	2758
	(26.71)	(1.32)	(1.26)			(27.41)	(1.26)	(1.20)	(1.17)	
Attrition										
Attrited from endline	0.08	0.02	0.03	0.39	2337	0.24	−0.06	−0.04	−0.06	3750
	(0.27)	(0.02)	(0.02)*			(0.43)	(0.02)***	(0.02)*	(0.02)***	
Attrited from chlorine test	0.12	0.01	0.03	0.33	2337	0.26	−0.04	−0.02	−0.04	3750
	(0.33)	(0.02)	(0.02)			(0.44)	(0.02)*	(0.02)	(0.02)**	
Compliance										
Completed both first and second intervention	0.74	0.01	−0.02	0.35	2975	–	–	–	–	-
	(0.44)	(0.02)	(0.02)			–	–	–	–	-
Completed first intervention	0.78	0.01	0.01	0.85	2975	–	–	–	–	-
	(0.41)	(0.02)	(0.02)			–	–	–	–	-
Completed no intervention	0.22	−0.01	−0.01	0.85	2975	–	–	–	–	-
	(0.41)	(0.02)	(0.02)			–	–	–	–	-

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

Table 2: Psychological outcomes

	Comparison with active control (PLA+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Time Preferences Treatment Effect	(3) Executive Function Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) TP+INF Treatment Effect	(8) EF+INF Treatment Effect	(9) PLA+INF Treatment Effect	(10) N
Executive Function										
Main outcome										
Behavioral Activation Score (BADS) (z-score)	0.00	−0.01	0.04	0.31	2103	0.00	0.02	0.10	0.04	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)	(0.06)*	(0.06)	
Additional outcomes										
Tower of London: Total Moves (z-score)	0.00	0.01	−0.04	0.28	2103	0.00	−0.38	−0.43	−0.36	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)***	(0.05)***	(0.05)***	
Time Preferences										
Main outcome										
β^{Effort}	0.982	0.007	0.005	0.33	2068	0.953	0.007	0.012	0.009	2906
	(0.005)	(0.006)	(0.007)			(0.020)	(0.018)	(0.018)	(0.018)	
Additional outcomes										
β^{MPL}	1.05	−0.02	0.02	0.21	2103	1.02	0.00	0.03	0.02	2955
	(0.46)	(0.02)	(0.03)			(0.43)	(0.03)	(0.03)	(0.02)	
		[1.00]	[1.00]				[1.00]	[0.78]	[0.65]	
δ^{MPL}	0.98	−0.00	−0.00	0.86	2103	0.98	0.00	0.00	0.00	2955
	(0.02)	(0.00)	(0.00)			(0.02)	(0.00)	(0.00)	(0.00)*	
		[1.00]	[1.00]				[1.00]	[0.78]	[0.37]	
δ^{Effort}	0.999	−0.001	−0.002	0.16	2068	0.995	−0.001	−0.002	0.003	2906
	(0.001)	(0.001)	(0.001)**			(0.002)	(0.002)	(0.002)	(0.002)	
		[1.00]	[1.00]				[1.00]	[1.00]	[1.00]	
Risk Aversion Measure (z-score)	0.00	−0.03	−0.07	0.49	1926	0.00	0.02	0.00	0.04	2735
	(1.00)	(0.06)	(0.06)			(1.00)	(0.06)	(0.06)	(0.06)	
Self-Efficacy										
General Self-Efficacy Score (GSE) (z-score)	0.00	0.17	0.25	0.16	2103	0.00	0.17	0.17	0.01	2955
	(1.00)	(0.05)***	(0.06)***			(1.00)	(0.06)**	(0.06)***	(0.06)	
Mechanisms										
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.08	0.06	0.61	2103	0.00	0.12	0.15	0.11	2955
	(1.00)	(0.05)	(0.06)			(1.00)	(0.05)**	(0.05)***	(0.06)**	
Chlorine knowledge score (z-score)	0.00	0.06	−0.01	0.20	2103	0.00	0.11	0.07	0.08	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)**	(0.06)	(0.06)	
ANC/PNC knowledge score (z-score)	0.00	0.04	−0.05	0.08*	2103	0.00	0.35	0.28	0.34	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.06)***	(0.05)***	(0.05)***	

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

Table 3: Behavioral outcomes

	Comparison with active control (PLA+INF)					Comparison with pure control (PC)				
	(1) Active Control Group Mean (SD)	(2) Time Preferences Treatment Effect	(3) Executive Function Treatment Effect	(4) Column 2 vs. Column 3 p-value	(5) N	(6) Pure Control Mean (SD)	(7) TP+INF Treatment Effect	(8) EF+INF Treatment Effect	(9) PLA+INF Treatment Effect	(10) N
Health outcomes										
Main outcome										
Objective measure: water has been treated with chlorine (TCR)	0.23	0.05	0.02	0.15	2012	0.22	0.06	0.04	0.03	2839
	(0.42)	(0.02)**	(0.02)			(0.42)	(0.02)***	(0.02)*	(0.02)	
Additional outcomes										
Objective measure: water has sufficient chlorine to be safe (FCR)	0.21	0.04	0.01	0.16	2012	0.18	0.06	0.03	0.03	2839
	(0.40)	(0.02)**	(0.02)			(0.39)	(0.02)***	(0.02)	(0.02)	
		[0.07]*	[0.75]				[0.03]**	[0.41]	[0.61]	
Number of diarrhea incidences per child u15 in last 3 months	0.24	−0.09	−0.06	0.30	2012	0.23	−0.08	−0.06	0.01	2823
	(0.56)	(0.03)***	(0.03)**			(0.58)	(0.03)**	(0.03)**	(0.03)	
		[0.01]***	[0.15]				[0.03]**	[0.27]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.00	−0.02	0.23	1999	0.22	0.01	−0.02	0.00	2800
	(0.35)	(0.02)	(0.02)			(0.36)	(0.02)	(0.02)	(0.02)	
		[0.49]	[0.69]				[1.00]	[0.41]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.26	−0.22	0.04	0.45	200	1.19	−0.14	0.01	0.29	272
	(1.19)	(0.41)	(0.35)			(1.17)	(0.41)	(0.41)	(0.44)	
		[0.41]	[1.00]				[1.00]	[0.64]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.21	−0.04	−0.03	0.43	2004	0.17	0.00	0.01	0.03	2806
	(0.34)	(0.02)**	(0.02)			(0.31)	(0.02)	(0.02)	(0.02)*	
		[0.07]*	[0.48]				[1.00]	[0.57]	[0.49]	
Savings outcomes										
Main outcome										
Amount saved regularly (per week, KES)	93.96	24.89	3.28	0.10	2108	88.76	14.74	5.43	0.50	2972
	(230.26)	(12.37)**	(12.51)			(228.12)	(12.66)	(12.92)	(12.07)	
Additional outcomes										
Indicator: Amount saved regularly is positive	0.36	0.13	−0.02	0.00***	2108	0.32	0.12	0.01	0.03	2972
	(0.48)	(0.03)***	(0.03)			(0.47)	(0.03)***	(0.03)	(0.03)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Number of new ROSCAs joined in last 3 months	0.17	0.04	0.01	0.17	2108	0.21	0.02	−0.01	−0.02	2972
	(0.44)	(0.03)*	(0.02)			(0.46)	(0.03)	(0.03)	(0.03)	
		[0.03]**	[1.00]				[0.20]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.11	−0.01	0.00***	2108	0.17	0.09	−0.00	0.01	2972
	(0.38)	(0.02)***	(0.02)			(0.38)	(0.02)***	(0.02)	(0.02)	
		[0.00]***	[1.00]				[0.00]***	[1.00]	[1.00]	
Labor outcomes										
Main outcome										
Total hours of work in last 3 months	106.11	−6.79	−23.83	0.06*	2108	108.78	0.89	−21.49	−3.56	2972
	(174.61)	(9.54)	(8.99)***			(182.99)	(10.21)	(9.55)**	(9.73)	
Additional outcomes										
Total days of work in last 3 months	21.22	−0.59	−3.91	0.04**	2108	21.73	0.21	−3.45	−0.60	2972
	(30.09)	(1.64)	(1.60)**			(30.45)	(1.71)	(1.63)**	(1.66)	
		[1.00]	[0.03]**				[1.00]	[0.08]*	[1.00]	
Average monthly earnings in last 3 months	1094.50	3.11	−1.23	0.97	2108	1167.22	−50.02	−120.05	−149.96	2972
	(2865.35)	(147.23)	(163.19)			(3155.76)	(167.08)	(171.39)	(175.53)	
		[1.00]	[0.99]				[1.00]	[0.32]	[1.00]	
Other behavioral outcomes										
Index of investment in children's education (z-score)	0.00	−0.02	0.01	0.60	1420	0.00	0.08	0.12	0.12	1967
	(1.00)	(0.06)	(0.07)			(1.00)	(0.07)	(0.07)*	(0.07)*	

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

Table 4: Chlorine-related outcomes in dispenser vs. non-dispenser villages

	Village has no chlorine dispenser					Village has chlorine dispenser					Comparison		
	(1) Pure Control Mean (SD)	(2) TP+INF Treatment Effect	(3) EF+INF Treatment Effect	(4) PLA+INF Treatment Effect	(5) N	(6) Pure Control Mean (SD)	(7) TP+INF Treatment Effect	(8) EF+INF Treatment Effect	(9) PLA+INF Treatment Effect	(10) N	(11) TP+INF Interaction p-value	(12) EF+INF Interaction p-value	(13) PLA+INF Interaction p-value
Objective measure: water has been treated with chlorine (TCR)	0.23	0.04	0.02	−0.01	1534	0.21	0.08	0.05	0.07	1305	[0.31]	[0.42]	[0.04]**
	(0.42)	(0.03)	(0.03)	(0.03)		(0.41)	(0.03)**	(0.03)*	(0.03)**				
Objective measure: water has sufficient chlorine to be safe (FCR)	0.17	0.05	0.02	0.00	1534	0.20	0.06	0.03	0.05	1305	[0.81]	[0.80]	[0.24]
	(0.37)	(0.03)*	(0.03)	(0.03)		(0.40)	(0.03)*	(0.03)	(0.03)				
Number of diarrhea incidences per child u15 in last 3 months	0.23	−0.07	−0.06	−0.01	1493	0.17	−0.05	−0.04	0.04	1313	[0.65]	[0.68]	[0.39]
	(0.56)	(0.04)*	(0.04)	(0.04)		(0.45)	(0.03)	(0.04)	(0.04)				
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.17	0.17	0.09	1573	0.00	0.07	0.11	0.14	1382	[0.37]	[0.60]	[0.56]
	(1.00)	(0.07)**	(0.07)**	(0.08)		(1.00)	(0.08)	(0.08)	(0.08)*				
Chlorine knowledge score (z-score)	0.00	0.15	0.06	0.05	1573	0.00	0.05	0.06	0.12	1382	[0.29]	[1.00]	[0.57]
	(1.00)	(0.07)**	(0.08)	(0.08)		(1.00)	(0.08)	(0.08)	(0.08)				

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

Table 5: Salience & memory test outcomes

	(1) Chlorine word remembered	(2) Saving word remembered	(3) Future word remembered	(4) Total words remembered
Time Preferences	0.08	0.00	-0.01	-0.37
	(0.01)***	(0.01)	(0.01)	(0.23)
Executive Function	0.05	0.02	-0.01	-0.03
	(0.01)***	(0.01)	(0.01)	(0.24)
Placebo	0.02	0.01	-0.03	-0.12
	(0.01)	(0.01)	(0.01)*	(0.22)
TP x Chlorine Word Interaction			0.09	
			(0.02)***	
EF x Chlorine Word Interaction			0.06	
			(0.02)**	
PLA x Chlorine Word Interaction			0.05	
			(0.02)*	
TP x Saving Word Interaction			0.02	
			(0.02)	
EF x Saving Word Interaction			0.03	
			(0.02)	
PLA x Saving Word Interaction			0.04	
			(0.02)*	
Constant	5.15	2.27	3.36	-31.30
	(2.24)*	(2.09)	(1.09)**	(34.04)

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

Appendix

A. Additional tables and figures

Table A.1: Attrition analysis: treatments vs. active control

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

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

Table A2: Attrition analysis: active treatments vs. pure control

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

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

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

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

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

Table A.4: Baseline balance: main outcomes

	Comparison with active control (PLA+INF)				
	(1) Placebo Group Mean (SD)	(2) TP Difference	(3) EF Difference	(4) TP = EF p-value	(5) N
Baseline Score					
Tower of London: Total Moves	23.32	−0.48	−0.30	0.18	2197
	(6.83)	(0.46)	(0.42)	(0.46)	
		[1.00]	[1.00]		
General Self-Efficacy Score (GSE)	42.75	0.02	0.97	0.95	2196
	(11.62)	(0.68)	(0.73)	(0.74)	
		[1.00]	[1.00]		
β^{MPL}	0.99	0.01	0.00	−0.01	2142
	(0.37)	(0.02)	(0.02)	(0.02)	
		[1.00]	[1.00]		
δ^{MPL}	0.98	0.00	0.00	0.00	2142
	(0.02)	(0.00)	(0.00)	(0.00)	
		[1.00]	[1.00]		

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

Table A.5: Raw means of z-scored outcome variables

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

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

Table A.6: Psychological Outcomes (TOT)

	Comparison with pure control (PC)				
	(1) Pure Control Mean (SD)	(2) TP+INF Treatment Effect	(3) EF+INF Treatment Effect	(4) PLA+INF Treatment Effect	(5) N
Executive Function					
Main outcome					
Behavioral Activation Score (BADS) (z-score)	0.00	0.02	0.12	0.04	2955
	(1.00)	(0.06)	(0.06)*	(0.06)	
Additional outcomes					
Tower of London: Total Moves (z-score)	0.00	−0.43	−0.48	−0.40	2955
	(1.00)	(0.06)***	(0.06)***	(0.06)***	
Time Preferences					
Main outcome					
β^{Effort}	0.964	0.003	0.005	−0.004	2906
	(0.017)	(0.015)	(0.014)	(0.013)	
Additional outcomes					
β^{MPL}	1.02	0.00	0.03	0.02	2955
	(0.43)	(0.03)	(0.03)	(0.02)	
		[1.00]	[1.00]	[1.00]	
δ^{MPL}	0.98	0.00	0.00	0.00	2955
	(0.02)	(0.00)	(0.00)	(0.00)*	
		[1.00]	[1.00]	[1.00]	
δ^{Effort}	0.995	−0.000	0.000	0.004	2906
	(0.002)	(0.002)	(0.002)	(0.002)*	
		[1.00]	[1.00]	[1.00]	
Risk Aversion Measure (z-score)	0.00	0.03	0.00	0.05	2735
	(1.00)	(0.06)	(0.06)	(0.06)	
Self-Efficacy					
General Self-Efficacy Score (GSE) (z-score)	0.00	0.19	0.19	0.01	2955
	(1.00)	(0.07)***	(0.07)***	(0.07)	
Mechanisms					
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.14	0.17	0.12	2955
	(1.00)	(0.06)**	(0.06)***	(0.06)**	
Chlorine knowledge score (z-score)	0.00	0.13	0.08	0.09	2955
	(1.00)	(0.06)**	(0.06)	(0.06)	
ANC/PNC knowledge score (z-score)	0.00	0.39	0.31	0.38	2955
	(1.00)	(0.06)***	(0.06)***	(0.06)***	

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

Table A.7: Behavioral Outcomes (TOT)

	Comparison with pure control (PC)				
	(1) Pure Control Mean (SD)	(2) TP+INF Treatment Effect	(3) EF+INF Treatment Effect	(4) PLA+INF Treatment Effect	(5) N
Health outcomes					
Main outcome					
Objective measure: water has been treated with chlorine (TCR)	0.22	0.07	0.04	0.03	2839
	(0.42)	(0.02)***	(0.02)*	(0.02)	
Additional outcomes					
Objective measure: water has sufficient chlorine to be safe (FCR)	0.18	0.07	0.03	0.03	2839
	(0.39)	(0.02)***	(0.02)	(0.02)	
		[1.00]	[1.00]	[1.00]	
Number of diarrhea incidences per child u15 in last 3 months	0.23	−0.08	−0.07	0.01	2823
	(0.58)	(0.03)***	(0.03)**	(0.03)	
		[1.00]	[1.00]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.22	0.01	−0.03	0.00	2800
	(0.36)	(0.02)	(0.02)	(0.02)	
		[1.00]	[1.00]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.19	−0.15	0.02	0.35	272
	(1.17)	(0.31)	(0.28)	(0.34)	
		[1.00]	[1.00]	[1.00]	
Proportion of children taken for healthcare check-up in last 3 months	0.17	0.00	0.02	0.04	2806
	(0.31)	(0.02)	(0.02)	(0.02)*	
		[1.00]	[1.00]	[1.00]	
Savings outcomes					
Main outcome					
Amount saved regularly (per week, KES)	88.76	16.66	6.08	0.56	2972
	(228.12)	(13.75)	(14.08)	(13.12)	
Additional outcomes					
Indicator: Amount saved regularly is positive	0.32	0.13	0.01	0.04	2972
	(0.47)	(0.03)***	(0.03)	(0.03)	
		[1.00]	[1.00]	[1.00]	
Number of new ROSCAs joined in last 3 months	0.21	0.02	−0.01	−0.03	2972
	(0.46)	(0.03)	(0.03)	(0.03)	
		[1.00]	[1.00]	[1.00]	
Indicator: Respondent saves for productive investments	0.17	0.10	−0.00	0.01	2972
	(0.38)	(0.02)***	(0.02)	(0.02)	
		[1.00]	[1.00]	[1.00]	
Labor outcomes					
Main outcome					
Total hours of work in last 3 months	108.78	0.91	−24.43	−4.04	2972
	(182.99)	(11.10)	(10.41)**	(10.58)	
Additional outcomes					
Total days of work in last 3 months	21.73	0.22	−3.92	−0.68	2972
	(30.45)	(1.86)	(1.78)**	(1.80)	
		[1.00]	[1.00]	[1.00]	
Average monthly earnings in last 3 months	1167.22	−56.61	−135.88	−169.50	2972
	(3155.76)	(181.67)	(186.66)	(190.98)	
		[1.00]	[1.00]	[1.00]	
Other behavioral outcomes					
Index of investment in children's education (z-score)	0.00	0.09	0.13	0.14	1967
	(1.00)	(0.07)	(0.08)*	(0.08)*	

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

Table A.8: Chlorine-related outcomes in dispenser vs. non-dispenser villages: restricting to WASH villages in the Water Quality and Passive Comparison arms

	Village has no chlorine dispenser					Village has chlorine dispenser					Comparison		
	(1) Pure Control Mean (SD)	(2) TP+INF Treatment Effect	(3) EF+INF Treatment Effect	(4) PLA+INF Treatment Effect	(5) N	(6) Pure Control Mean (SD)	(7) TP+INF Treatment Effect	(8) EF+INF Treatment Effect	(9) PLA+INF Treatment Effect	(10) N	(11) TP+INF Interaction p-value	(12) EF+INF Interaction p-value	(13) PLA+INF Interaction p-value
Objective measure: water has been treated with chlorine (TCR)	0.26	−0.01	0.01	−0.04	1066	0.18	0.11	0.07	0.10	1091	[0.01]**	[0.18]	[0.00]***
	(0.44)	(0.03)	(0.03)	(0.03)		(0.38)	(0.03)***	(0.03)**	(0.03)***				
Objective measure: water has sufficient chlorine to be safe (FCR)	0.16	0.01	0.02	−0.01	1066	0.17	0.09	0.05	0.08	1091	[0.09]*	[0.64]	[0.04]**
	(0.37)	(0.03)	(0.03)	(0.03)		(0.38)	(0.04)**	(0.03)	(0.03)**				
Number of diarrhea incidences per child u15 in last 3 months	0.22	−0.05	−0.05	−0.00	1039	0.17	−0.05	−0.03	0.01	1099	[0.90]	[0.72]	[0.85]
	(0.56)	(0.05)	(0.05)	(0.05)		(0.46)	(0.04)	(0.04)	(0.04)				
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	−0.04	0.20	0.19	0.14	1092	0.01	0.07	0.09	0.11	1160	[0.37]	[0.52]	[0.89]
	(1.02)	(0.09)**	(0.09)**	(0.09)		(1.00)	(0.09)	(0.08)	(0.09)				
Chlorine Knowledge Score (z-score)	0.09	0.16	0.03	0.01	1092	0.01	0.01	0.04	0.10	1160	[0.22]	[0.92]	[0.51]
	(1.06)	(0.10)*	(0.10)	(0.10)		(0.96)	(0.08)	(0.08)	(0.08)				

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

Table A.9: Phone access & task comprehension questions

	Comparison with active control (INF)					Comparison with pure control (PC)				
	(1) Control Mean (SD)	(2) TP Treatment Effect	(3) EF Treatment Effect	(4) TP vs. EF	(5) N	(6) Control Mean (SD)	(7) TP+INF Treatment Effect	(8) EF+INF Treatment Effect	(9) INF Treatment Effect	(10) N
SMS Task Checks										
Participant uses a phone she owns	0.70	0.01	−0.03	−0.04	2387	0.70	0.00	−0.03	−0.00	2972
	(0.46)	(0.02)	(0.02)	(0.02)		(0.46)	(0.02)	(0.03)	(0.03)	
		[0.68]	[1.00]				[0.60]	[0.42]	[0.67]	
Participant uses a phone belonging to her household	0.96	0.01	−0.00	−0.01	2387	0.96	0.02	0.00	0.01	2972
	(0.19)	(0.01)	(0.01)	(0.01)		(0.21)	(0.01)	(0.01)	(0.01)	
		[0.44]	[1.00]				[0.48]	[0.70]	[0.61]	
Proportion for whom accessing a phone for 30mins is very difficult or impossible	0.12	0.03	0.01	−0.03	2386	0.13	0.02	−0.00	−0.01	2970
	(0.32)	(0.02)*	(0.02)	(0.02)		(0.33)	(0.02)	(0.02)	(0.02)	
		[0.27]	[1.00]				[0.48]	[0.70]	[0.61]	
Proportion for whom accessing a phone for 1hr is very difficult or impossible	0.16	0.04	0.03	−0.02	2386	0.19	0.02	0.00	−0.03	2970
	(0.37)	(0.02)**	(0.02)	(0.02)		(0.39)	(0.02)	(0.02)	(0.02)	
		[0.25]	[1.00]				[0.48]	[0.70]	[0.37]	
Proportion for whom accessing a phone for 4hrs is very difficult or impossible	0.31	0.02	0.01	−0.02	2384	0.35	−0.02	−0.03	−0.04	2967
	(0.46)	(0.02)	(0.02)	(0.03)		(0.48)	(0.03)	(0.03)	(0.03)	
		[0.44]	[1.00]				[0.55]	[0.42]	[0.31]	
SMS Comprehension questions correct first time	0.80	−0.03	−0.00	0.03	2372	0.72	0.04	0.07	0.07	2955
	(0.40)	(0.02)	(0.02)	(0.02)		(0.45)	(0.03)	(0.03)***	(0.03)***	
		[0.41]	[1.00]				[0.48]	[0.06]*	[0.02]**	
Number of attempts at SMS comprehension questions	0.85	0.17	0.10	−0.06	2372	1.22	−0.21	−0.27	−0.38	2955
	(1.95)	(0.12)	(0.12)	(0.12)		(2.36)	(0.13)	(0.13)**	(0.13)***	
		[0.34]	[1.00]				[0.48]	[0.12]	[0.02]**	

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

Table A.10: Summary of participation

	Dispenser villages				Non-dispenser villages				
	Control	PLA+INF	TP+INF	EF+INF	Control	PLA+INF	TP+INF	EF+INF	Total
Recruited	365	450	480	468	410	542	512	523	3750
Baselined + Intervention 1	–	360	369	363	–	417	414	414	2337
Intervention 2	–	343	349	332	–	393	385	379	2181
Endline	265	364	383	381	323	447	417	404	2984
Chlorine Test	253	342	358	352	318	429	398	389	2839
Total	883	1859	1939	1896	1051	2228	2126	2109	

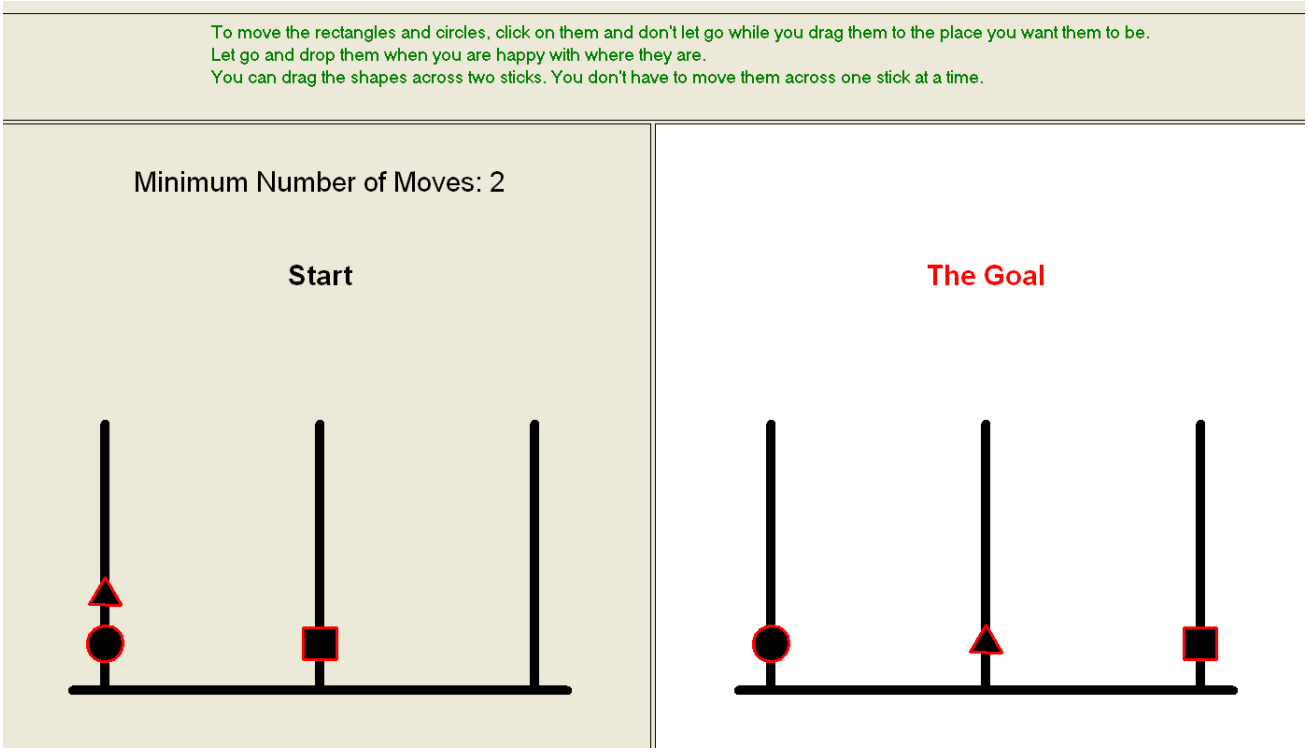
Table A.11: Temporal Discounting Decisions

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.12: 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

Figure A.1: Tower of London Example Screen



B. Schedule of tasks and treatments

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

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

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

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

Figure A.2: MPL task example screen

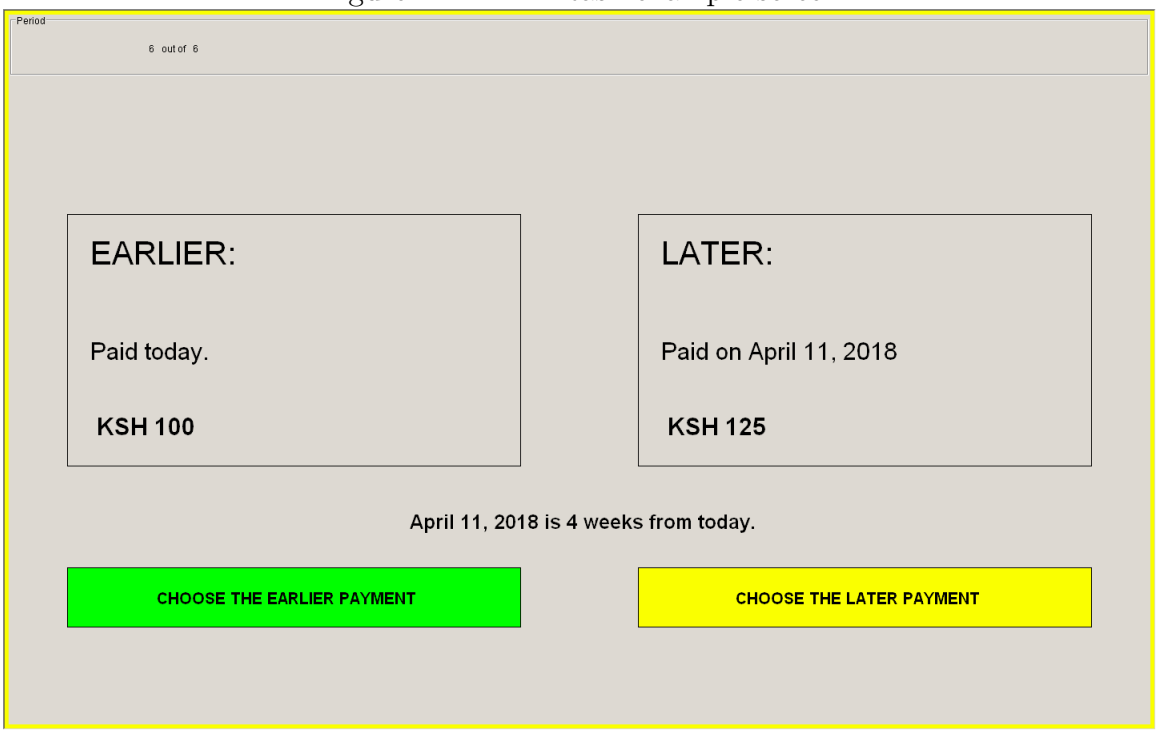
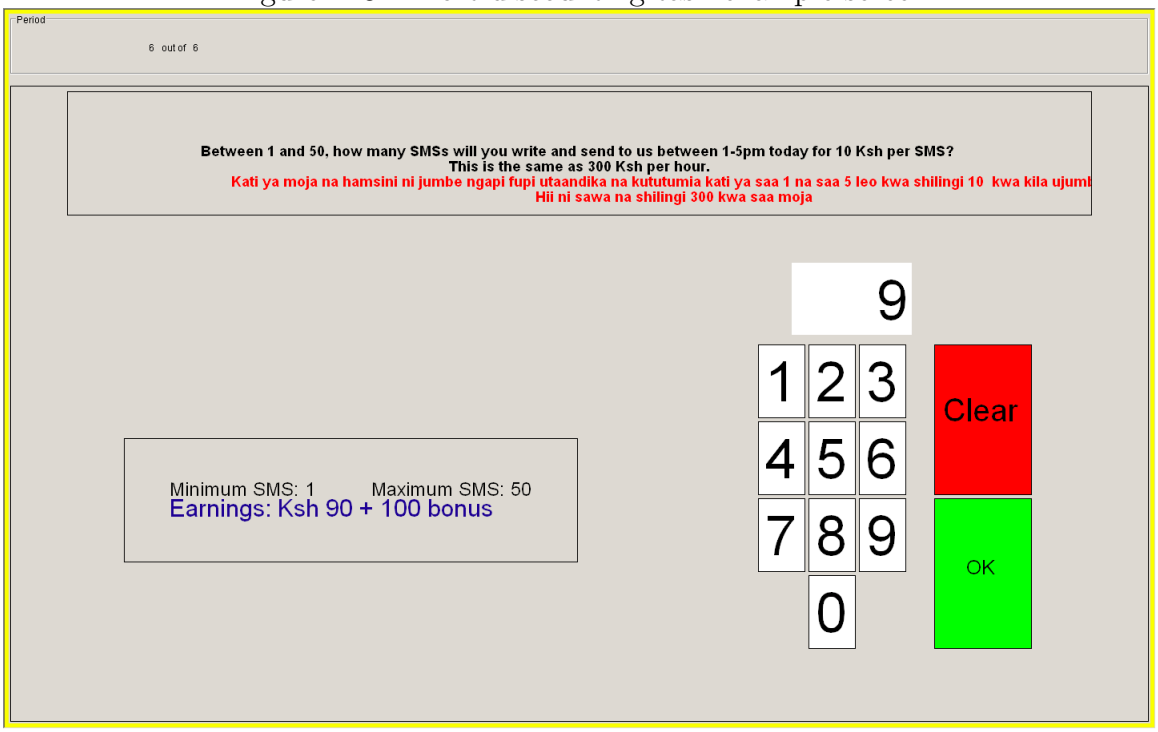


Figure A.3: Effort discounting task example screen



Baseline Session 1:

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

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

Baseline Session 2

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

Endline

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

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

C. Descriptions of other measures

C.1 Alternative mechanisms

C.1.1 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 ("TP", "EF", and "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.

C.1.2 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 treatement 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.

C.1.3 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.

C.1.4 Demand Effects (Salience of Chlorination)

We test for the possibility that our treatments differentially increased the salience of water chlorination. During the endline survey, enumerators read out three lists of nine words each to every participant, and asked her to recall as many words as possible directly after reading each list. Each list contained three categories of future-related words (chlorine, savings, and farm investment), as well as non-future related filler words. The word lists are available in original Swahili and English translation in table A.12. We estimate salience effects using equation 3:

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

where w_{im} is an indicator for participant i correctly recalling the word related to chlorine in list m; X_{im} refers to the number of words that the individual correctly recounted from that list; δ_m is a fixed effect for list m; and T_j are treatment indicators. We test $H_0 : \alpha_1 = \alpha_2 = \alpha_3$, with the null hypothesis corresponding to no differential salience of chlorine across (active) treatment groups.

In case our treatments differentially affected the salience of chlorine, we further test whether this is due to an increased salience of future-oriented behaviors in gen-

eral - which may result from our main psychological mechanisms of interest. To this end, we estimate whether the differential treatment effect on chlorine words also holds for two other future-oriented behaviors (saving and farm investment), which were not emphasized in the sessions. We estimate

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

where w_{imn} is an indicator for participant i correctly recalling the words in list m from future oriented behavior n (chlorination, savings or farm investment); and $chlorine_n$ is a dummy for the word being related to chlorine. The a_j coefficients capture increased future orientation due to treatment, while the b_j coefficients indicate that salience increased differentially for chlorination. We test $H_o : b_1 = b_2 = b_3$, with the null hypothesis corresponding to no differential salience of chlorine across (active) treatments.

D. 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). Between 2012 and 2014, 8246 pregnant women were enrolled from three counties in western Kenya: Bungoma, Kakamega and Vihiga. The six interventions were (W) improved water quality ("Water Quality"); (S) improved sanitation; (H) handwashing with soap; (WSH) combined water, sanitation and handwashing; (N) improved nutrition; (WSH+N) combined water, sanitation, handwashing and nutrition. The study also included two control arms: (A) an active control arm, who received monthly visits to measure children's arm circumferences; and (P) a passive comparison arm, who received no visits. The villages in our sample primarily came from either (W) improved water quality or (P) passive comparison arm.

In all Villages in Arm W, "Water Quality", chlorine dispensers were installed at public water sources used by study participants. All community members were able to use the dispensers free of charge. After filling water carrying containers, usually a 20 l jerrycan, users turn the knob on the dispenser to add 3 ml of 1.25 percent sodium hypochlorite (chlorine), which yields 2.5 ml/l of free chlorine residual after 30 mins for 20 l of water (Kremer et al. 2011a). This is sufficient chlorine to sanitize the

water for drinking. Community promoters encouraged use of the chlorine dispensers, which to this date are monitored and maintained by *Evidence Action's* "Dispensers for Safe Water" program. Sample households additionally received a six month supply of bottled chlorine every six months, to be used for sanitizing water at home, in case the household drank harvested rainwater or chose not to use the water source with the installed dispenser.

A coding error during randomization meant that about 20 percent of our sample was recruited uniformly from all eight WASH Benefits treatment arms (this happened in one district, Mumias district in Kakamega county). Arms WSH and WSH+N also involved the installation and maintenance of chlorine dispensers. With the exception of the installation of chlorine dispensers, which were available to the whole community, all WASH Benefits interventions were delivered at the household level: Arm S involved the improvement of compound-level sanitation through the building or upgrading of latrines, and the distribution of equipment for removing feces from the compound. Arm H installed handwashing stations for study households, at the latrine and near the cooking area. Arm WSH combined all of the interventions of W, S, and H. In Arm N, study households were delivered Lipid-based Nutrient Supplements. Arm WSH+N, the WASH + Nutrition arm, combined all of Arms W, S, H, and N. As for the entire sample, we exclude direct beneficiaries of household-level interventions. We include the concerned villages in our main estimation of treatment effects, grouping treatment arms by whether or not they received the water quality interventions. (Null et al. 2018) find no significant effects of the S, H or N interventions, which did not include any water focused interventions, on whether stored drinking water has detectable free chlorine. They find no significant differences in the effects of the W, WSH and WSHN interventions, which all included water focused interventions, on chlorine presence: all increased chlorine presence to around 40 percent in year 1 and around 20 percent in year 2. None of the interventions affect prevalence of child diarrhea. We conduct additional robustness checks, excluding participants from these villages from the heterogeneity analysis by "Water Quality" assignment. We find little difference in the sign or magnitude of coefficients, and small differences in statistical significance, likely driven by reduced power from the smaller sample size. Full results are available in Table A.8.