What Motivates Health Behavior: Preferences, Constraints, or Beliefs? Evidence from Psychological Interventions in Kenya*

Johannes Haushofer†, Anett John‡, Kate Orkin§

September 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. One group received a two-session executive function intervention that aimed to improve planning and execution of plans; a second received a two-session time preference intervention aimed at reducing present bias and impatience. A third group receives only information about the benefits of chlorination, and a pure control group received no intervention. Ten weeks after the interventions, the executive function and time preferences interventions led to significant 18 percent and 27 percent increases, respectively, in the share of households who have chlorinated their drinking water, compared to the pure control group. This increase was accompanied by significant 20–27

*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. 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 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.

§Department of Economics, University of Oxford, Oxford, UK.

percent (executive function) and 30–32 percent (time preferences) reductions in the number of diarrhea episodes in children relative to both placebo and control groups. The time preferences intervention also significantly increased the share of individuals who save regularly by 38 percent. We further study the psychological channels through which effects occur. The executive function intervention improved performance on a planning lab task relative to the placebo, and both the executive function and the time preferences intervention increased self-efficacy, i.e. beliefs about one's ability to achieve desirable outcomes. Effects are not driven by changes in information: the information treatment increased beliefs about the efficacy of chlorine, but had no effect on chlorination rates or diarrhea. Together, these results suggest that there may be important psychological barriers to health behavior, possibly including low self-efficacy.

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

1. Introduction

Individuals often do not make choices which improve their health outcomes, even when such choices cost little and individuals are aware of their benefits. A prominent example is chlorination of drinking water, which is highly effective in preventing diarrhea, particularly among young children. Among children aged 1-5, diarrhea is the second leading cause of death worldwide, contributing to nearly half a million deaths in 2015 (Wang et al. 2016), and a leading cause of morbidity, with an estimated 1.7 billion episodes occurring each year (Walker et al. 2013). Chlorine for water is readily and cheaply available, but infrequently used by individuals without access to clean water: in our study areas, only 3% of households used chlorine before any intervention (Null et al. 2018).

Standard economic explanations do not fully explain households' failure to take up chlorination. In particular, while reductions in the financial and effort cost of chlorination through home deliveries or provision of chlorine in dispensers at water points do increase chlorination (Kremer et al. 2011b; Kremer et al. 2011a), these effects dissipate over time; several years after these interventions, we find that only about a quarter of households still chlorinate their water. In contrast, promotion campaigns by members of the community significantly enhance the effects of increased access to chlorine through dispensers or home deliveries (Kremer et al. 2011a; Kremer et al. 2011b; Null et al. 2018). Together, these findings suggest that non-standard factors may play a role in determining chlorination takeup.

In this paper, we consider three potential behavioral channels which may explain why households fail to chlorinate. 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*. Second, they may have incorrect beliefs about the effectiveness of chlorination, or pessimistic beliefs about themselves and their ability: e.g., they might believe that they do not have the ability to improve their family's health outcomes through chlorination. Both of these mechanisms are accounts in terms of *beliefs*; the latter type of belief is referred to as self-efficacy in the psychology literature (Bandura 1977). Finally, people may have deficits in executive function, i.e. the ability to plan or execute the actions required to implement their preferences.[1] This possibility represents an account in terms of

[1]Executive functions are the cognitive processes required for forming goals, planning, and carrying

(cognitive) *constraints*.

In a randomized controlled trial in rural Kenya, we allocate 3,750 young women to four treatment arms. The first 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 these time *preferences* affect chlorination behavior ("TP" intervention). The second received a two-session intervention that aimed to improve executive function, i.e. the ability to plan and execute a course of action ("EF" intervention). This intervention tests the role of this specific cognitive *constraint* in decision-making about chlorination. Both interventions might plausibly have effects on an aspect of the third mechanism, self-efficacy, i.e. people's *beliefs* about their ability to succeed in specific situations or accomplish a task.

To isolate the effects of the psychologically active elements of our treatments, a third, placebo group received all elements of the intervention except the psychologically active components ("PLA" intervention). These participants also gathered as a group, but to discuss everyday topics. In addition, all three of these groups received a short information module about the benefits of chlorination ("INF" intervention). Thus, all three groups experienced the effects of interactions with facilitators and groups, and received information about the benefits of chlorination. 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 active treatment arms 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.

We had some success in designing psychological interventions that induced persistent change in targeted psychological outcomes, without inducing change in other, non-targeted outcomes. The Executive Function treatment works as we predicted theoretically: Ten weeks after the interventions, we find 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

out plans directed by goals (Lezak 1983; Miller and Cohen 2001). In most categorizations, executive function contains three processes: inhibition, which includes selective attention and self-regulation; memory; and higher-order cognitive functions, including cognitive flexibility, intelligence, and planning (Miyake et al. 2000b; Diamond 2013a; Lyon and Krasnegor 1996; Suchy 2009). We target one higher-order cognitive function, planning, i.e. the ability to generate a strategy, including the sequencing of steps, to achieve intended goals (Carlin et al. 2000).

a set of shapes on a screen into a new configuration. Effects are significant relative to both the Placebo treatment and the Time Preferences treatment. The intervention also affected everyday behavior and choices: On 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), participants in the Executive Function group scored significantly higher than those in the placebo and pure control groups. They also score somewhat higher than those in the Time Preferences treatment, although the difference is not statistically significant. In contrast, the Time Preferences treatment did not work exactly as theoretically predicted, in that it did not significantly affect its target, lab measures of time preference, relative to the Placebo group.

Intriguingly, we find large and highly significant effects of both the Time Preferences and Executive Function treatments on a self-efficacy scale, suggesting that these treatments influenced participants' beliefs about their ability to shape their life and achieve desirable outcomes. Thus, the Executive Function intervention has strong effects on its intended target as well as self-efficacy, while the Time Preferences intervention has strong effects on self-efficacy but not its intended target of time preferences. Together, these results suggest that it is possible to shift constraints in people's ability to plan and follow through on intentions with light-touch interventions, while shifting preferences may be more difficult.[2]

We next examine the effects of these interventions on water chlorination and related health outcomes, as well as economic outcomes. In the Executive Function and Time Preferences 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 water contains chlorine.[3] The effect in the Placebo group is smaller and not significant compared to the pure control group. It is also smaller than the effect of Executive Function and Time Preferences, although only the difference between Time Preferences

[2]There are two potential explanations for this difference: the treatments are not equally compelling, or it is harder to shift time preferences than the ability to plan and follow through on plans. We present some suggestive evidence that the two treatments are equally compelling: participants come back at equal rates to attend the second session of both the TP+INF and EF+INF sessions; both treatments have similar effects on beliefs and knowledge; and both treatments have similar effects on self-efficacy. Thus, the evidence is consistent with the view that it is particularly hard to shift time preferences.

[3]We collect an objective measure of whether households have increased use of chlorine in water by testing household drinking water for the presence of chlorine in unannounced household visits, an average of 11 weeks after the endline survey.

and the placebo is statistically significant. In line with these findings, the Executive Function and Time Preferences treatments significantly reduce the number of diarrhea episodes in children. This effect is statistically significant both in comparison to the pure control group (27 and 32 percent, respectively), and in comparison to the Placebo group (20 and 30 percent, respectively). In contrast, the effect in the Placebo group relative to PC is small and not statistically significant. Together, these findings suggest that our psychological treatments significantly affected health behaviors and outcomes, and that this effect goes beyond what is achieved by simply providing information.

In further support of the view that simply providing information is ineffective, all three treatment groups which received the information treatment show statistically significant increases in their belief that chlorination can prevent diarrhea, and in a two-item knowledge test about the benefits of chlorination, relative to the pure control group which did not receive the information treatment. These effects are of similar magnitude and not statistically different. In contrast, as described above, the effects of these treatments on health outcomes are significantly different, suggesting that differences in information are not the source of these differences.

The effect of our interventions is not limited to the health domain, but also resulted in significant changes in economic behavior: While the Time Preferences intervention did not affect our laboratory measure of time preferences, it caused a statistically significant 38 percent increase in the share of individuals who save regularly.

Together, these results suggest that psychological interventions targeting Executive Function and Time Preferences can affect health-related behaviors and outcomes. These effects cannot be achieved by simply changing beliefs about the effectiveness of engaging in health behaviors. The Time Preferences and Executive Function interventions were equally effective in influencing health-related outcomes, although only the Executive Function intervention affected its intended psychological target. Both interventions affected self-efficacy, i.e. beliefs about one's ability to achieve desirable outcomes. This finding creates the surprising possibility that these interventions, even though they targeted preferences and psychological constraints rather than beliefs, did take their effects on outcomes by affecting people's internal beliefs about their ability to realize desired outcomes. 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.

A possible confound of our interventions is increased salience of chlorination: the TP and EF session modules partially relied on examples and case stories. While our interventions were designed to be domain-general, and mostly prompted participants to contribute examples from their own life, the scripts also mentioned chlorination. It is possible that the mention of chlorination itself represents a nudge to chlorinate, by focusing participants' attention on this issue. 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 TP, EF, and INF all increased the salience of chlorination (but not saving or farm investment), with a stronger effect for TP and EF. This constitutes a possible alternative explanation for our treatment effects on chlorination. However, our treatment effect on savings, as well as on various other non-chlorine measures, cannot be explained by the salience of chlorination, and 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 behaviour is both non-linear and differential across domains. Alternatively, the observed increase in salience of chlorination may be a consequence, rather than a cause, of increased chlorination.

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 in which villages were randomly assigned to receiving chlorine dispensers placed at the water source (Null et al. 2018) to test if our psychological interventions have larger effects in villages where access to chlorination is easier or more difficult. Our treatment effects on chlorination are somewhat larger in villages with dispensers, although differences are not always statistically significant. Thus, when both psychological and cost/effort constraints are alleviated simultaneously, effects on behavior may be larger than when cost and access barriers remain.

Our study builds on a small literature which uses light-touch interventions to affect constraints, beliefs, or preferences and real-world behaviors. Bernard et al. (2014) 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 aspirations and beliefs about one's own ability, rather than through changes in preferences. However, Alan and Ertac (2018) use an eight-session educational intervention in Turkish primary schools to increase patience, suggesting that changing preferences may be possible. Ghosal et al. (2016) show that a short course on personal growth for sex workers improved self-esteem and "locus of control" (the belief that one is in control of one's outcomes), as well as increasing savings and attendance at health checkups. Similarly, 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 targeting each of preferences, beliefs, and constraints against independently, rather than focusing on the effect of one psychological mechanism. 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, especially education, when information about returns is provided (Jensen 2010; Jensen 2012; Dinkelman and Martínez A 2014). 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, or beliefs about oneself go beyond them.[4]

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

[4]Indeed, our results suggest some reinterpretation of past findings: some "information" interventions in this literature may combine pure information with elements targeting constraints, preferences, or beliefs about oneself: information on financial aid for university delivered through videos of role models might both enhance self-efficacy and provide information (Dinkelman and Martínez A 2014).

2. Experimental Design

2.1 Study site

Our trial areas are Bungoma and Kakamega county in rural Western Kenya. These counties were recently included in the 'WASH Benefits' study of Null et al. (2018),which provided village chlorine dispensers next to water sources, as well as community health promoters (see Section 2.5). The study sites are close to those used in Kremer et al. (2011a).

People live in fairly dispersed villages: several related households live together in fenced compounds, and compounds are interspersed with fields.

We focus on women because they are primarily responsible for household chores, including collecting water or delegating children to collect 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 26 years old, 89 percent are married or co-habiting, and they have on average 6 years of education.

2.2 Individual Level Sampling

We recruited a pool of 3750 women aged 18–35 between October 2016 and January 2017, of whom 2330 participated in the interventions. With the help of local guides, enumerators visited all households in each included village (see Section 2.5) 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) their household was not a sample household in the WASH Benefits study. The WASH Benefits study recruited women in their second or third trimester of pregnancy in 2012. In addition to village-level interventions (chlorine dispensers), roughly six households per treated village received free chlorination bottles and monthly health promoter visits. We exclude women who report having participated in the study. As a second check, we exclude households with children aged either 3-4 or 4-5, depending upon the village's WASH Benefits timing. 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.

2.3 Randomization

The pool of 3750 individuals recruited for the study were randomized into four treatment arms as follows:

1. 992 assigned to Treatment Arm 1: "Time Preferences"
2. 991 assigned to Treatment Arm 2: "Executive Function"
3. 992 assigned to the active control group: "Placebo"
4. 775 assigned to the "Pure Control" group.

We stratified the randomization on two variables collected during the census described in 2.2 :

1. Wealth Index: the total value of a limited set of assets (bicycles, cellphones, gas stoves, all livestock, radios, sofas and televisions). Participants were split at the 50th percentile into a 'high' or 'low' wealth group.
2. Village of residence

Participants were also assigned alphabetically 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 in order to minimize attrition. Randomization was conducted using the "randtreat" command in Stata, with 'misfits' equally distributed across treatment arms to ensure the target group sizes were achieved. Balance checks were conducted to ensure that randomization was successful (see Table 1).

2.4 Attrition

Attrition was a potential concern due to the need to convene participants in a central location to conduct behavioral laboratory and group intervention sessions, on three separate occasions over the course of three months. Steps taken to mitigate attrition included gathering contact details not only for participants themselves but also for their family members, neighbors and village elders. Field officers returned to villages to track down sample participants who could not be contacted by phone.

2.5 Cross-cut with WASH Benefits to examine interactions of psychological and cost constraints

Our key behavioral outcome of interest is whether households chlorinate water.

The villages included in the study are a subsample of the villages which took part in the WASH Benefits study (henceforth WASH) in Bungoma and Kakamega counties. The study was a cluster-randomized controlled trial, conducted from 2012 to 2014. Each cluster consisted of one to three neighbouring villages. The study contained eight treatment arms, which tested a variety of household-levelwater, sanitation, handwashing, and nutrition interventions – both each in isolation, and different interventions combined (Null et al. 2018). We restricted sample recruitment to villages which were assigned to either the "Water Quality" treatment arm or the "Passive Comparison" arm. [5] In the "Water Quality" villages, chlorine dispensers were installed at an average of five community water points per village cluster, and refilled as needed. Evidence Action's Dispensers for Safe Water program has since maintained these dispensers, ensured they are filled with chlorine, and retained a local promoter in each community. WASH Benefits sample households (women in the second and third trimester of pregnancy during recruitment in 2012) additionally received a free 1l bottle of chlorine every six months, and were visited by local promoters each month.[6] As outlined in Section 2.2, we excluded these sample households from our study. WASH Benefits "Passive Comparison" villages received neither dispensers nor health promoters. For more information on the WASH Benefits study, please see Appendix C.

We cross-randomize our four treatment arms across the "Water Quality" and "Pas-

[5]A coding error during randomization meant that about 20 percent of the sample was recruited uniformly from all eight WASH Benefits treatment arms (in Mumias constituency, Kakamega county). In 23 out of the 205 sampled villages, sanitation, handwashing and nutrition interventions were offered in addition to the water quality intervention. However, all interventions except water quality took place at the household level. As outlined in Section 2.2, we exclude direct sample households from our study. Consequently, the key difference between these villages is whether or not they received the water quality intervention, and thus the chlorine dispensers (three out of eight treatment arms did). We thus group these villages by their "Water Quality" treatment status, and include them in our main estimation of treatment effects. We conduct additional robustness checks, including (i) excluding them from the heterogeneity analysis by "Water Quality" assignment, and (ii) excluding them from all analyses described in section 5. See Appendix C for details.

[6]The effect of household interventions may not have persisted. Households received a free 1l bottle of chlorine every six months from 2012 to 2014, but did not receive chlorine after 2014. Local promoters visited households each month during the trial to encourage treating and safely storing water. They also tested household stored water for the presence of chlorine, and used test results to counsel households. But promoters visited all households in the community after the end of the trial.

sive Comparison" villages. This allows us to study whether psychological interventions and standard cost-reducing interventions are complements or substitutes in increasing the demand for preventive healthcare. If the cost of accessing chlorine absent free dispensers remains an important barrier to chlorine use, then treatment effects will likely be higher in WASH Benefit villages than control villages. On the other hand, our interventions may have larger effects in non-WASH villages if (i) people in WASH villages chlorinate already, and (ii) improvements in psychological targets (such as patience, executive function, and self-efficacy) can compensate for facing a higher cost of chlorination adoption.

2.6 Background on Chlorination Use

Unsafe drinking water is presumed to be a major cause of high levels of child diarrhea in the area. In the WASH study control group, diarrhea prevalence in the past 7 days was 27 percent among children aged 1 and 2 (Null et al. 2018).

Most of the population relies on communal water sources, usually wells with pumps or springs, some of which are fenced to protect them from cattle (Null et al. 2018). Women and children collect water in plastic jerry cans. Drinking water is then decanted into clay storage pots in the home, keeping water cool.

Point-of-use methods of chlorinating drinking water have been shown to improve water quality and reduce child diarrhea, thus potentially reducing child mortality (see Arnold and Colford Jr (2007) for a review of the evidence). Absent point-of-collection chlorine dispensers (as installed by WASH), the main source of dilute chlorine is the brand *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 25 KES ($0.25) per 150ml bottle. Each bottle treats 1000 L of water (approximately one month of household drinking water), and comes with instructions in Swahili and in pictures.

Uptake of chlorination is low and chlorine is irregularly used. At baseline of the WASH study in 2012, only 3 percent of households had detectable free chlorine in their water (Null et al. 2018). Similarly, before the Kremer et al. (2011a) study, 2 percent of households had detectable free chlorine in their water and only 7 percent of households reported treating the drinking water currently in their home with chlorine (6 percent used *WaterGuard*). This was despite awareness of the product: 89 percent had heard of *WaterGuard* and 29 percent had used it at least once.

Six years later, rates have increased somewhat: in our sample, 27 percent of pure control households report having chlorinated their current drinking water supply, 17 percent had detectable free chlorine in their water, and 22 percent had detectable total chlorine.[7]

Interestingly, the rate is the same in villages which received chlorine dispensers through WASH Benefits (21 percent total chlorine, 20 percent free chlorine), and in villages which did not (23 percent total chlorine, 17 percent free chlorine). For comparison, the 2015 two-year follow-up statistics from the WASH Benefits study reported 23 percent free chlorine in dispenser villages and 3 percent in non-dispenser villages, suggesting a strong convergence between WASH treatment and control villages since 2015.

Finally, high levels of diarrhea in children are affected by many factors outside of drinking water chlorination. Using the baseline data from Null et al. (2018), 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. However, 77 percent of children aged 0-3 (14 percent of those aged 3 to 8) still defecate in the open. These statistics suggest that, while other risks remain, getting households to chlorinate water may be an important barrier in reducing high levels of diarrhea.

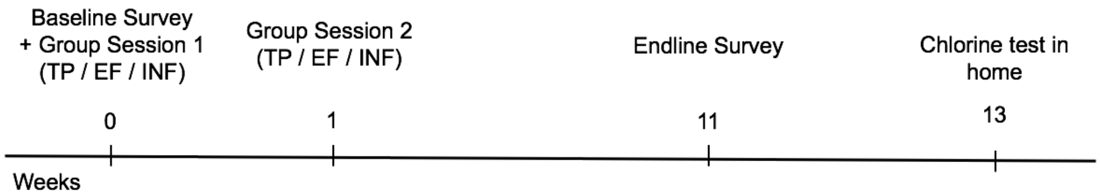
3. Interventions

The study included three active and one passive treatment arm: one arm aimed at improving planning and performance of basic tasks ("Executive Function", EF), one arm that encouraged respondents to visualize their future ("Time Preferences", TP), a placebo treatment arm on plants and birds in Kenya ("Placebo", PLA), and a passive "pure control" arm (PC). The three active treatments each contain two interactive group sessions of two hours duration each, with one week in between the sessions. The structure of each group session was held constant across treatment arms: each included a short lecture, group discussion, reflection of how the themes relate to participants' own lives, and some drawing and list-writing. Participants were split into groups of five for the sessions, which were run by a locally-trained female facilitator. Participants

[7]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.

were reconvened in the same groups for the second session. No participant was invited for the second session without having already participated in the first session.

Figure 1: Study timeline



The baseline measures were collected in Busara "mobile labs" in Bungoma and Kakamega county, which each hold up to 25 participants (five groups) at a time. The behavioral tasks and some questionnaires were administered using touch screen computers and the zTree experimental interface (Fischbacher 2007) to enable computer-illiterate respondents to participate. Enumerators read instructions to the respondents in Kiswahili to maximize comprehension.[8] At endline, individual questionnaires were administered using SurveyCTO. Respondents received KES 200 ($2) for participating in the baseline survey and first intervention session, KES 200 for the second session, and KES 300 for participating in the endline session.[9] They were additionally given a KES 50 bonus for arriving on time for each appointment. Participants were reimbursed for their transport costs, using known public transport rates from their village of residence to the mobile laboratory.

All participants recruited to the sample were invited to attend endline sessions, regardless of whether they attended the baseline and/or intervention sessions.

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

The time preference intervention is based on the idea that present utility is salient, tangible, and easy to imagine, while future utility feels vague and distant. A substantial body of evidence in psychology shows that people imagine future events in much less detail than immediately upcoming events, focusing on abstract qualities rather than

[8]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.

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

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 decisionmaker who simulates future utility by combining priors with noisy, unbiased signals. Simply assuming that the simulation noise increases in the time horizon is sufficient to generate choices *as if* she was discounting future utility. Dynamic preference reversals emerge in all but a special case (though these are caused by imperfect forecasting rather than by self-control problems). The model implies that interventions which improve forecasting ability (or forecasting efforts) will lead to more patient behavior. 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* later. The intervention used here 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, participants were encouraged 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, in order to imagine future selves in as much graphic detail as possible. Example exercises include:

- What examples you can think of, where our current behavior has an effect on the future?
- Close your eyes for one minute. Imagine the person you will be in one year. Imagine your family in one year. Use details.
- 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?
- 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?

While the script was written to make the future more tangible, the intervention carefully avoided changing participants' beliefs about which present behavior would entail which future outcome - it merely encouraged them to make the connection themselves. To distinguish the intervention from the Executive Function intervention, it also did not include any planning features.

Information module ("INF")

The intervention concluded with an information module about the benefits of chlorination. Participants were read information on chlorination and antenatal and postnatal care (ANC/PNC). These behaviors were used as real-world examples of important health behaviors in both active treatment arms.

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

This intervention targets whether people choose to make plans or set goals and make the necessary choices to execute them, rather than avoiding them. We use the term "executive function" in a loose sense.T

In the psychological literature, the term focuses largely on people's cognitive *ability* to plan.[10]

However, economists are as, if not more interested, in the more practical, behavioral dimension of planning and goal-oriented behavior:in whether people choose to make plans and execute them (Dean, Schilbach, and Schofield). Achieving a simple goal, such as chlorinating water, is unlikely to require complex cognitive planning ability. It is more likely that people avoid making plans to do it, or struggle to stick to their plans.

[10]Planning and goal-oriented behavior are higher-order cognitive functions, part of the brain's "executive functions", which cover three processes located in the pre-frontal cortex: inhibition, which includes selective attention and self-control; working memory; and higher-order cognitive functions, including cognitive flexibility, intelligence and planning (Miyake et al. 2000a; Diamond 2013b). Psychologists can improve executive function to some extent using intensive batteries of computerized exercises that progressively increase the demands on the targeted skills with weekly or daily training over months (see for example Bangirana et al. (2009), Bangirana et al. (2011) in Uganda with children). Most training is with children or adolescents, as executive function is thought to develop until roughly age 20, although some training has been conducted in adults (Dahlin et al. 2008a, Dahlin et al. 2008b, Heckhausen and Singer 2001, White and Shah 2006).

Our executive function intervention aimed simply to train respondents how to set simple goals, make clear plans, establish routines, and reduce avoidance. We drew on an approach used with patients with mild depression known as behavioral activation.[11] 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 two-session 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).[12] We also used elements of other goal-setting exercises, which often involve mental contrasting (contrasting one's present with the situation one would like), 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).

The first goal of the exercise was for participants to understand 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. They may feel a lack of energy and motivation to do things, find it difficult to get going on tasks or achieve goals. Participants listen to a story of a woman very similar to them in this position who was very tired and struggling to do her chores, including fetching water and chlorinating it. Then, if they wanted to, they sharedstories from times they had been in a similar situation.

The second goal was for participants to set some 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 list of activities, one set they enjoyed doing and one set that were necessary and important. They ranked them from most to least difficult.

The third goal was for participants to make achievable plans towards some of their goals and plan to overcome obstacles. Again in pairs, they made a weekly diary. In the first session, they picked the easiest one or two activities from each of the "necessary activities" and "enjoyable activities" list and scheduled them in the diary. They then broke the task down into steps, visualising what they would need to do to do the

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

[12] We also included some elements from A Brief Behavioral Activation Treatment for Depression (Lejuez et al. 2011).

activity, detailing small manageable steps to achieve it, anticipating potential obstacles, and making plans to overcome them. In the second 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. The first intervention session concluded with the same information module described above.

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

A third group attended baseline laboratory sessions and received an intervention titled "Nature in Kenya". The goal of this 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, a topic chosen intentionally to be psychologically inactive.

In addition, participants in this group also received the same information module as the two active 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 2017). We test both Total Chlorine Residual (TCR) and Free Chlorine Residual (FCR), using TCR as our main chlorination outcome measure of interest. TCR indicates the presence of any chlorine in the water; ie. that the household has 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; ie. that the household added *sufficient* chlorine to the drinking water. The presence of free chlorine is necessary for the water to be dependably potable (CDC 2010). Although correct usage of chlorine is an interesting outcome in its own right (and will be explored in later versions), we are primarily interested in whether households attempt to chlorinate their water at all, and thus focus on TCR as our main pre-specified outcome of interest. To conduct the tests, enumerators filled uncontaminated 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 0mg/L and 3.4mg/L. Our primary outcome measure of chlorination switches on for positive values of TCR.[13]

In the baseline and endline survey (which took place in the laboratory), as well as during home visits, enumerators additionally asked for self-reported chlorination use at present and in the last 30 days. During home visits, enumerators also noted the type of container used for storing drinking water, and whether or not it was covered.

4.2 Other Behavioral Measures

While chlorination is our primary outcome of interest, and was pre-specified at such, our interventions are in no way 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 completed severalmodules on economic and health behaviors. Outcome variables from these modules are listed in Appendix B. We pre-specified secondary and exploratory outcomes in the domains of health (diarrhea, vaccinations, and antenatal care vists), savings (internal and external margin), labor supply, and educational investment.

[13] The safety of drinking water is a function of FCR rather than TCR. We will explore this relationship in later versions of the paper.

4.3 Psychological Measures

4.3.1 Time Preferences

Following recent innovations in the elicitation of time preferences ((Andreoni and Sprenger 2012); (Augenblick, Niederle, and Sprenger 2015)), we estimate time preferences over the effort domain, using a newly developed real effort task. We use the question design 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. Only one time and one piece rate are randomly implemented at the end (described below).[14] Contrast to Augenblick (2017), we hold the time of decision constant and vary the time of effort provision (this requires us to control for weekday effects). All questions required a minimum effort allocation of one task 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 don't have access to a computer. We thus developed an innovative new effort task that is adapted to our setting: Participants complete data entry tasks by SMS, using toll-free numbers administered by the Busara Center.[15] Each SMS is a 30-digit random number string, which takes approximately two minutes to type. The participants are given a sheet which lists 50 such strings, including a counter to keep track. To ensure comprehension, participants complete one practice SMS during the baseline survey.At the end of the survey, one decision (out of 12) is randomly selected to be the "decision

[14]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 no piece rate (just the KES 100 completion bonus explained below). Unfortunately, this introduction interacted with differential attrition in the pure control group (see Table 1), and thus requires the use of more complex structural estimation methods. We therefore restrict the estimation to those who were not offered these rates at present, and will include the remaining subsample in later versions of the paper.

[15]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.

that counts": At the selected piece rate and time horizon, participants have to send the exact number of SMS they chose. If they do, they receive the full piece rate payment plus a KES 100 completion bonus. If they fail to implement the decision they made, they lose both the payment for this task and the completion bonus (see Augenblick 2017 for a full description of this method).[16] Earnings from this task were paid 14 days from the survey date, regardless of the selected effort time horizon.

We estimate time preferences over effort following the approach of Augenblick (2017) by assuming quasi-linear utility (linear in money, convex in effort) and a power cost of effort function. We additionally assume quasi-hyperbolic discounting.. Following DellaVigna and Pope (2016), 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 is motivated by the observation that participants supply non-zero amounts of effort even for low piece rates (DellaVigna and Pope (2016)). The optimal level of effort is 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.

Two concerns about the validity of the task arise from the possibilities that participants do not have access to phones, or do not understand the payment scheme. To test for the former, we include a small module in the endline survey in which participants are asked about difficulties accessing a mobile phone, particularly at the times necessary to complete the SMS task. To alleviate the concern that respondents do not understand the incentives, we include three multiple-choice comprehension questions immediately before the task that ask participants to calculate the payout in different circumstances.

[16]The field setting with SMS required some tolerance: While a laboratory computer can confirm correct and incorrect entries, and display the number of tasks still to complete, we relied on participants to do this themselves. We thus allowed for 75 percent accuracy in entering the number strings, and a tolerance of 10 in the number of completed SMS (subject to positive completion). Participants were told that there would be some tolerance for miscounting, but not how much.

Respondents could not participate in the task until they had answered the comprehension questions correctly. Table D.8 shows phone access and task comprehension by treatment arm.

In addition to the effort discounting task, we include a conventional Multiple Price List (MPL) task to measure money discounting. Participants were asked to make 10 choices between payments at earlier or later dates. The payment at the early date was always equal to KES 100, while the payment at the later date increased gradually from KES 110 to KES 300. 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 D.4. Figure **D.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 1997a), assuming linearity of utility in money.

4.3.2 Executive Function: Planning

We measure two aspects of the planning component of executive function. First, to measure the whether people choose to make a plan and follow through on it, we employ the Behavioral Activation for Depression Scale (Kanter et al. 2007). The full questionnaire contains 29 items divided into four subscales/factors: Activation, Avoidance/Rumination, Work/School Impairment, and Social Impairment. We use the short form (BADS-SF) of the scale developed by Manos, Kanter, and Luo (2011), who carry out item reduction procedures from all subscales until only 9 items remain. In these, 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" (0) to "completely" (6). Items from subscales other than Activation 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 Tower of

London 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. For each trial, we record the number of moves, the time until the participant's first move, the overall time to completion, and whether the problem is solved correctly. In all rounds, participants are limited to a maximum of 20 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, for the purpose of establishing construct validity and reliability, is computed as the total number of moves used across the four rounds, the number of rounds completed correctly, and standardized average time to complete rounds. An example of the participant's screen is shown in Figure D.1.

4.3.3 Self-Efficacy

We measure self-efficacy using the General Self-Efficacy (GSE) scale (Schwarzer and Jerusalem 2010). This self-reported scale measures individuals' optimistic self-belief: a 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" (0) to "Always true" (5). Our version contains 12 items: 10 from the generic version, and two which are repeated and reversed, to check for consistency. Higher scores indicate a higher self-perception of self-efficacy.

4.4 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 B.1 for a full description and empirical specification).

4.4.1 Salience Effects

Our interventions were designed to target time preferences and executive function on a general domain, rather than specifically to increase chlorination use. However, chlorination was mentioned in the context of case studies used in the intervention scripts. This raises the possibility of salience and attention effects: It is possible that the mention of chlorination itself represents a nudge to chlorinate, by focusing participants' attention on this issue. 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 5 KSh 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 D.5 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 at the 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. In case our treatments differentially affected the salience of chlorine, we further test whether this is due to an increased salience of future-oriented behaviors in general - which may result from our main psychological mechanisms of interest. To this end, we estimate whether the differential treatment effect also holds for two other future-oriented behaviors (saving and farm investment), which were not emphasized in the sessions (see Appendix B.1 for the empirical specification).

4.4.2 Risk Preferences

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). Assuming a constant relative risk aversion utility function, we estimate the curvature parameter for each participant as the midpoint of the implied interval.

4.4.3 Beliefs and Knowledge about Chlorination

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) 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 diarrhea cases the participant believes chlorine can avert as a measure of belief about chlorine effectiveness.

Secondly, treatment may affect chlorination by providing information about how to properly use it (as we did 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, (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 which ranges from 0 to 2.

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 as well as 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 status, 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 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 arms 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 specification for the pure control comparison is

$$Delay_i = \beta_0 + \sum_{j=1}^{3} \beta_j T_{ji} + \varepsilon_i \tag{2}$$

where $Delay_i$ is the number of days between the baseline and the endline survey, respectively between the endline survey and the chlorine test. T_j refers to treatment assignment. Standard errors are clustered by intervention group (five participants).

We test for selective attrition in attending the endline survey and the chlorination test at home, using equation 2 with the respective outcome measures, for both the placebo group comparison and the pure control comparison. 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 arms - i.e., the decision to participate in the first and second intervention session. The specification is identical to equation 2, except that the outcome variable is an indicator for session attendance, and the reference group is the placebo control group.

5.2 Main Specification: Active Treatments versus Placebo

We employ the following main specification:

$$y_{i1} = \alpha_0 + \alpha_1 T_{1i} + \alpha_2 T_{2i} + \delta y_{i0} + \mathbf{\Phi X}_i + \gamma_v + \theta_w + \eta_i \tag{3}$$

Here, y_{i1} is the outcome of interest for respondent i at time of endline, and y_{i0} is the same outcome variable at time of baseline, if applicable. The sample excludes the PC group, and is further restricted to those who participated at least at baseline, the first intervention session, and endline (either survey or home visit, for the relevant outcome measure). Thus, the 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. Standard errors are clustered by intervention group (five participants) to account for within-group dynamics.

Several outcome variables collected at endline were not included at baseline, most prominently the objective measure of chlorination behavior. For these variables we omit y_{i0} from the regressors but restrict the sample similarly. 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). In both of our main specifications, we remove outliers by censoring outcome variables at the 99th percentile.

5.3 Comparison with Pure Control Group

We also report results from comparing the active treatments (EF+INF, TP+INF) and the placebo group (PLA+INF) to a pure control group (PC). The specification is identical to that in equation 3, except that there is a third treatment indicator T_{3i} for the placebo arm, 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}. The sample includes all recruited participants who completed the endline survey, including 'non-compliers' who were assigned to the active treatment arms, but chose not to participate in the baseline survey or the interventions.

5.4 WASH Benefits Cross-Randomization

Since our treatment arms cross-cut the randomization of the WASH Benefits study described above, we are able estimate both the long-run impacts of those treatments and the differential effects of our treatments in conjunction with the "Water Quality" (chlorine dispenser) intervention. To do so, we run the pure control specification with an indicator variable for treatment status in the "Water Quality" arm of the WASH Benefits study, and interact this indicator with the treatment assignments in the present experiment. The primary outcome of interest is an indicator for objective chlorination (TCR).

Due to a coding error in sampling (see Footnote 5), some participants were drawn from treatment arms of the WASH Benefits study other than "Water Quality" or "Passive Comparison." We include these participants in the WASH regressions, grouped by whether their treatment arm received chlorine dispensers or not (this is the case for treatment arms "Water Quality, Sanitation and Handwashing" and "Water Quality, Sanitation and Handwashing and Nutrition"). We exclude these participants in a robustness check.

5.5 Multiple Hypothesis Testing (MHT) Correction

We clearly specified primary, secondary and exploratory outcomes in our pre-analysis plan, as shown in Appendix B. We use a stepdown procedure to adjust p-values for the false discovery rate (FDR) among a group of outcomes, and report the resulting "q-values." Indices are constructed following Anderson (2008). We adjust for multiple hypothesis testing within outcome groups (psychological mechanisms and behaviors) and hierarchical categories (primary and secondary), but not across them. Similarly, we consider the effects of our two active interventions to be theoretically distinct and therefore do not correct across them.

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 3 with baseline demographics as the outcome variables. Each row shows baseline balance for one demographic variable. Columns (1)–(5) show the comparison of the active treatments, targeting time preferences (TP+INF) and executive function (EF+INF), 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 three out of 30 coefficients on pairwise comparisons reaching statistical significance, all at the 10 percent level.

The second panel in Table 1 shows balance 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, both relative to the active control group, and relative to the pure control group, so that any differences between groups are driven by treatment rather than by differences in the length of time elapsing between baseline 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 18 and 24 percent, respectively. We find no differential attrition from the endline survey or from the chlorine measurement between the three active treatment groups (EP+INF, TP+INF, and PLA+INF), with very small and statistically insignificant coefficients on the pairwise comparisons. This suggests any differences between treatment groups are not driven by differential sample composition at endline.

We do find small but statistically significant imbalances in attrition between the pure control group and two of the three active arms. The pure control group were 5 percentage points more likely than the TP+INF group and 6 percentage points more likely than the PLA+INF group not to be reachable for the endline survey. We place less weight on the pure control comparisons than on the comparisons between treatment groups, as some results may be driven by differences in the samples contacted at endline in each group.

Average attrition from the chlorine measurement, conducted at people's houses, was slightly higher: 22 percent in the PLA+INF group and 26 percent in the pure control group. We find no differential attrition between the three active treatment groups (EP+INF, TP+INF, and PLA+INF) or between these groups and the pure control group with very small and statistically insignificant coefficients on the pairwise comparisons. The exception is that the PLA+INF group are 4 percentage points less likely to attrite from the chlorine measure than the pure control. This suggests any differences on the chlorine measure between treatment groups are not driven by differential sample composition at endline.

Columns (2)-(5) of Appendix Tables D.1 and D.2 further examine if the composition of our sample is similar in different treatment arms, particularly in the pure control arm, where attrition is higher. Attrition poses a more serious threat to inference if there is differential composition of each treatment group at endline, as it is not clear if differences are due to the treatment or differences in the sample composition. We run four regressions: regressions of attrition status from 1) endline and 2) the chlorine measure on demographic variables, and then regressions of attrition status from 1) endline and 2) the chlorine measure on demographic variables and the demographic variables all interacted with treatment group dummies.

The first two columns show that demographic variables do predict attrition from either endline measurement, once treatment status is controlled for. Older people, single people, less educated people and poorer people are somewhat more likely to attrite from endline, although differences are fairly small in magnitude. However, the interaction terms between demographic variables and treatment status in columns 3 and 4 show that people 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 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.

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 session 2. 74 percent of respondents completed both sessions, while 4 percent did not complete the second session.[17] Compliance is balanced across treatment groups, so any differences between treatment groups are not driven by differential rates of compliance with treatment.

[17] For TOT regressions, 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. All substantive content was covered in the first session: the second session was merely a reinforcement and repetition of the first session content.

6.2 Results for psychological targets of our intervention

We now turn to the results on the psychological and behaivoral 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 arms.

Table 2 shows results on the psychological outcome variables, estimated using equation 3. 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 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 "active" treatment arms 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. Attrition is constant across treatment arms. Finally, this comparison is more precisely estimated than the comparison to the pure control group due to the inclusion of fixed effects and control variables (we do not collect control variables for the pure control as they do not complete the baseline).

We find that the EF+INF treatment significantly improved executive function, with a 0.47 point increase in the Behavioral Activation Score relative to a PLA+INF mean of 38.14 points, significant at the 10 percent level both with and without adjustment for multiple inference. Similarly, we find a significant reduction in the number of moves required to complete the Tower of London task, indicating an improvement in the ability to make plans and execute them. This reduction in 0.62 moves relative to a control group mean of 21.29 moves is significant at the 5 percent level using conventional p-values, and at the 10 percent level after FDR adjustment. Thus, the EF+INF treatment affected the psychological outcomes it was designed to move.

These results are robust in relation to the pure control group: there is a 0.51 point increase in the Behavioral Activation Score, but this is not significant, and a 0.83 move decrease in the Tower of London task (significant at the 10 percent level). As noted above, this specification is slightly less powered, but coefficients are similar in sign and magnitude.

In contrast, 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 theTP+INF, and PLA+INF treatments on executive function outcomes relative to the pure control group.

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 (1997b); additional outcomes are the δ parameter from the same task, and corresponding parameters from the monetary discounting task. We find no statistically significant effects on any of these outcomes, with the exception of a small negative effect on effort discounting in the EF treatment, significant at the 10 percent level. One possible explanation for this unexpected finding is that the improvement in planning ability generated by the intervention makes it less cognitively costly to postpone effortful tasks into the future.

The third panel of Table 2 shows the effect of our interventions on the Generalized Self-Efficacy (GSE) scale, our measure of self-efficacy. Both the EF+INF and the TP+INF interventions generate statistically significant 0.14 SD increases 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 desirable outcomes. We find slightly smaller, and hence not statistically significant differences between the active treatment arms and the pure control group on this outcome, but coefficients are similar in magnitude to the specification which compares active treatments.

6.3 Results for beliefs and alternative psychological measures

We then show results on a range of outcomes which we measure to make the argument 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 work simply by changing people's beliefs or level of information. Our experimental design includes information in all three treatment arms. 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 the benefits of chlorination. They are much likelier to be driven by the psychologically active components in the EF+INF or TP+INF treatments.[18]

Of course, this relies on the EF+INF, TP+INF, and PLA+INF arms actually affecting beliefs and knowledge in the same way. If, for example, the placebo arm was very boring, such that people switched off and did not take on board the information component, then we would have an ineffective placebo arm. However, the last panel of Table 2 shows that the information treatment worked as anticipated and was effective in affecting beliefs about the effectiveness of chlorination and, to a lesser extent, knowledge about the benefits of chlorination, in all three treatment arms. We find that all three interventions, TP+INF, EF+INF, and PLA+INF, are effective in increasing beliefs in the efficacy of chlorine in averting diarrhea relative to the pure control group, with similar effect sizes across the three interventions.

Effects on knowledge about the benefits of chlorination are also very similar across active treatment groups and the placebo group. Compared to the pure control group mean of 1.07, the TP+INF group shows a significant increase in knowledge about chlorination of 0.08 points out of a score of 2. Effects are very similar in magnitude for the other groups (0.05 for EF+INF and 0.06 for INF), although no other pairwise comparison is significant.

Second, we tested alternative psychological mechanisms. 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.

[18]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*INF, or TP+INF compared to pure control = TP+INF+TP*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 = ENF+EF*INF). The interaction term (e.g. EF*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 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.

In sum, the EF+INF intervention improved executive function, as intended, and did not affect time preferences, which it was not intended to target. In contrast, the TP+INF intervention did not affect time preferences, as it was intended to, nor did it affect executive function. Both interventions affected self-efficacy.

We cannot rule out that the TP+INF intervention was simply not as persuasive as the EF+INF treatment. However, we present four pieces of evidence against this view. First, we examine the likelihood that participants come back to the second intervention session, having attended the first session. Table 1 shows that TP+INF intervention participants are as likely to come back for the second session as EF+INF participants, suggesting that the interventions are equally compelling. Second, both interventions have exactly the same effects on self-efficacy. Third, both interventions have very similar effects on beliefs and knowledge about chlorination: in fact, the TP+INF intervention is marginally more effective in shifting knowledge about chlorination than the EF+INF or placebo intervention. Fourth, we show below that the TP intervention does affect behaviors.

This suggests that the TP intervention was compelling and had effects on a range of psychological outcomes, but that it is difficult to shift time preferences, the main target of the intervention.Our study thus shows that it seems possible to shift constraints in one's own ability to plan and follow through on intentions, beliefs about oneself, and beliefs and knowledge about the state of the world, with light touch psychological interventions. In contrast, time preferences are more difficult to shift, even with an intervention which attempts to target them directly.

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

Both active psychological interventions have some effects on objective measures of use of chlorination in household drinking water, our primary outcome, relative to the pure control group. as measured with chlorine test kits. In comparison to the pure control group, the TP+INF group shows a 6 percentage 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.[19] 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 of households chlorinating water. The results for free chlorine (FCR, see Section 4.1) are similar in magnitude and slightly weaker in significance, suggesting that the treatments were more effective on the 'external' margin of getting people to chlorinate, than on the 'internal' margin of getting them to chlorinate the correct amount.

Importantly for our purpose of affecting end 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 30 percent reduction for TP+INF and a 25 percent reduction for EF+INF.[20]

We find no strong effects on other health outcomes, with the exception of a small reduction 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. We find no effect of any treatment on the main outcome variable, the amount of money saved regularly. However, we find a large and highly significant effect on savings on the extensive margin, with the share of respondents who save regularly increasing by 9 percentage points (25 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, 8 and 9 percentage points (47 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. We find no effects on ROSCA

[19] A self-report question on whether households treated their water in any way to make it safe to drink generated 99% 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 2017), leading to similar experimenter demand effects across all arms.

[20] 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.

membership. The EF+INF and PLA+INF intervention did not show significant effects on savings-related outcomes, and 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 argues against an experimenter demand-effect account of the impacts of our interventions: while the interventions mentioned chlorination and health-related topics, they did not discuss savings 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 17 hours (17 percent) relative to the PLA+INF group, and 21 hours (20 percent) relative to the pure control group. These effects are significant at the 5 percent level using naïve p-values, and the effect relative to the pure control group survives FDR correction at the 10 percent significance level. A similar effect is observed in the total number of days worked, which is reduced by almost 3 days or 14 percent in the EF+INF relative to the pure control group, significant at the 10 percent level after FDR correction. 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. The TP+INF and PLA+INF treatments do not show significant effects on labor outcomes.

Finally, we find a small positive effect in an index of investment in children's education in the EF+INF and PLA+INF groups relative to the pure control group.

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

We then 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 or knowledge or some other aspect of being involved in the treatment (such as gathering in groups). We conclude that, while some of the effects of our intervention undoubtedly work by altering respondents' level of information about chlorination, the psychologically active treatments have additional effects on behavior that are likely to be due solely to changes in the psychological targets of our intervention: self-efficacy (which is affected in both treatment groups) and executive function (which is affected only in the EF+INF group). We also suggest that changes are more likely to be due to changes in self-efficacy than changes in executive function.We present three pieces of evidence in line with this argument:

First, the changes in behavior and end outcomes are *larger* in the active psychological interventions than in the group that only receives information. On the primary behavioral outcome variable, whether households have chlorine present in water (TCR), the PLA+INF group mean is 23 percent. The TP intervention leads to a 4 percentage point increase in the presence of total chlorine in household drinking water. This is significant at the 10 percent level, both before and after multiple test corrections. A 1 percentage point increase in chlorination in the EF treatment group is not statistically significant. On a related variable, episodes of diarrhea per child under 15 in the last three months, the PLA+INF mean is 0.22. The TP treatment results in a 32 percent reduction. The EF treatment results in a 27 percent reduction in the number of episodes, both significant at the 1 percent level using conventional p-values, and at the 5 percent level after FDR correction.

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 argued above, there are very few differences in knowledge and beliefs between the active 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 belief or knowledge change (for example, if the psychological treatments cause some differences in how information is received and taken up) we should also see more accurate beliefs and greater knowledge in these groups. We largely do not: we saw earlier that all groups see very similar changes in beliefs about the efficacy of chlorine relative to the pure control group. Some small part of these differences may be due to differences in knowledge. In particular, there is a larger treatment effect of the TP+INF treatment on increase in knowledge of chlorine practices relative to the pure control group, mirrored by a larger effect on the presence of chlorine. Nevertheless, knowledge about chlorination practices does not differ between the EF+INF and PLA+INF groups, despitethe observed difference between these groups in effects on chlorination.

Third, we see effects of the active treatments on behaviors which are not mentioned at all in the intervention. Each psychological intervention has some effects on aspects of behavior unrelated to chlorination. 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. In contrast, the information treatment was not at all targeted at other future-oriented behaviors: it contained only information about chlorination. Accordingly, it does not have strong effects on behaviors. This suggests these interventions may affect deeper psychological characteristics of individuals, which then cause behavioral changes in multiple domains.

Fourth, a similar argument can be made that increased attention and salience of chlorination induced by the interventions cannot fully explain the results. Section 4.4.1 explains the design of a test for increased salience of chlorination, and Appendix B.1 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. On its own, this result does not prove salience effects, but may capture reverse causality: Participants become more future-oriented as a result of treatment, which in turn increases the salience of future-oriented behaviors like chlorination. If this is true, then the salience of other future-oriented behaviors should similarly increase. 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). Increased salience thus constitutes a possible alternative explanation for our treatment effects on chlorination. However, a pure salience explanation is inconsistent with the observed effect on savings behavior: As shown in Columns (2) and (3) of Table 5, the salience of savings (which was also briefly mentioned in the intervention scripts) does not increase. Finally, column (4) suggests that neither treatment arm had an effect on total words remembered, and thus on participants' memory.

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 treated with a chlorine dispenser in the WASH Benefits study that preceded ours. 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 arms 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 significant heterogeneity in the effects of the treatments on outcomes related to chlorination by village type. Specifically, the active treatment arms EF+INF, TP+INF, and PLA+INF show significant treatment effects on whether water was treated with chlorine in the dispenser villages, but not in the villages without dispensers. The interaction term is significant 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. Conversely, the TP+INF treatment has a significantly larger effect on the chlorine knowledge score in villages without compared to villages with a dispenser, lending weight to the claim that our treatment effects on chlorination are not caused by increases in knowledge about chlorination. Finally, the reduction in diarrhea episodes as a result of the TP+INF treatment is somewhat larger in non-dispenser villages compared to dispenser villages, but this effect is weak.

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 control. We find evidence for an effect of EF on the psychological constructs it targets, but little evidence of an effect of the TP intervention on time preferences. 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 a number of other outcomes.

These results suggest that light-touch psychological interventions have the potential to move intermediate psychological outcomes, as well as distal behavioral outcomes, in developing country contexts. The fact that we found stronger effects on psychological outcomes in the EF intervention than in the TP treatment may indicate that preferences are more difficult to move than psychological constraints such as executive function, and than beliefs about one's own abilities. In support of this view, both interventions affected self-efficacy, which is a set of beliefs about one's own abilities, even though the interventions were not primarily designed for this purpose. Both interventions affected chlorination and the incidence of diarrhea in comparison to the pure control group. This finding further suggests that 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. Future work may attempt to replicate these effects to shore up their statistical power, 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 a randomized educational intervention." *Journal of Political Economy.*

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Re-evaluation 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." *The American Economic Review* 102 (7): 3333–3356.

Arnold, Benjamin F, and John M 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."

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working over time: Dynamic inconsistency in real effort tasks." *The Quarterly Journal of Economics*, p. qjv020.

Bandura, Albert. 1977. "Self-efficacy: toward a unifying theory of behavioral change." *Psychological review* 84 (2): 191.

Bangirana, Paul, Peter Allebeck, Michael J Boivin, Chandy C John, Connie Page, Anna Ehnvall, and Seggane Musisi. 2011. "Cognition, behaviour and academic skills after cognitive rehabilitation in Ugandan children surviving severe malaria: a randomised trial." *BMC neurology* 11 (1): 96.

Bangirana, Paul, Bruno Giordani, Chandy C John, Connie Page, Robert O Opoka, and Michael J Boivin. 2009. "Immediate neuropsychological and behavioral benefits of computerized cognitive rehabilitation in Ugandan pediatric cerebral malaria survivors." *Journal of developmental and behavioral pediatrics: JDBP* 30 (4): 310.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2017. "Mental health and women's choices. Experimental evidence from a Randomized Control Trial."

Baron, Emily Claire, Thandi Davies, and Crick Lund. 2017. "Validation of the 10-item centre for epidemiological studies depression scale (CES-D-10) in Zulu, Xhosa and Afrikaans populations in South Africa." *BMC psychiatry* 17 (1): 6.

Beaman, Lori, Jeremy Magruder, and Jonathan Robinson. 2014. "Minding small change among small firms in Kenya." *Journal of Development Economics* 108:69–86.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, Alemayehu Seyoum Taffesse, et al. 2014. The future in mind: Aspirations and forwardlooking behaviour in rural ethiopia. Centre for the Study of African Economies conference on economic development in Africa.

Blattman, Christopher, Julian C Jamison, and Margaret Sheridan. 2017. "Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in Liberia." *American Economic Review* 107 (4): 1165–1206.

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." Technical Report, 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.

Dahlin, Erika, Anna Stigsdotter Neely, Anne Larsson, Lars Bäckman, and Lars Nyberg. 2008a. "Transfer of learning after updating training mediated by the striatum." *Science* 320 (5882): 1510–1512.

Dahlin, Erika, Lars Nyberg, Lars Bäckman, and Anna Stigsdotter Neely. 2008b. "Plasticity of executive functioning in young and older adults: immediate training gains, transfer, and long-term maintenance." *Psychology and aging* 23 (4): 720.

Dean, Emma Boswell, Frank Schilbach, and Heather Schofield. 2017. "Poverty and Cognitive Function." In *The Economics of Poverty Traps.* National Bureau of Economic Research, Inc.

DellaVigna, Stefano, and Devin Pope. 2016. "What motivates effort? Evidence and expert forecasts." *The Review of Economic Studies.*

De Quidt, Jonathan, Johannes Haushofer, and Christopher Roth. 2017. "Measuring and Bounding Experimenter Demand." Technical Report, National Bureau of Economic Research.

Diamond, Adele. 2013a. "Executive functions." *Annual review of psychology* 64:135–168.

———. 2013b. "Executive Functions." *Annual review of psychology* 64:135–168.

Dinkelman, Taryn, and Claudia Martínez A. 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 A 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.

Eaton, William W, Corey Smith, Michele Ybarra, Carles Muntaner, and Allen Tien. 2004. "Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R)."

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics* 10 (2): 171–178.

Gabaix, Xavier, and David Laibson. 2017, March. "Myopia and Discounting." Technical Report, National Bureau of Economic Research.

Ghosal, Sayantan, Smarajit Jana, Anandi Mani, Sandip Mitra, and Sanchari Roy. 2016. "Sex Workers, Stigma and Self-Belief: Evidence from a Psychological Training Programme." *Competitive Advantage in the Global Economy (CAGE) Working Paper*, no. 302:1–48.

Gilbert, Daniel T, and Timothy D 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.

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.

Heckhausen, Jutta, and Tania Singer. 2001. "Plasticity in human behavior across the lifespan." In *International encyclopedia of social and behavioral sciences*, 11497–11501. Elsevier.

Heller, Sara B, Anuj K Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A 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.

John, Anett. 2017. "When Commitment Fails-Evidence From a Field Experiment."

Jones, Michael P. 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 B Krueger, David A Schkade, Norbert Schwarz, and Arthur A Stone. 2004. "A survey method for characterizing daily life experience: The day reconstruction method." *Science* 306 (5702): 1776–1780.

Kanter, Jonathan W, Patrick S Mulick, Andrew M Busch, Kristoffer S Berlin, and Christopher R 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, Edward Miguel, Sendhil Mullainathan, Clair Null, and Alix Peterson Zwane. 2011a. "Social engineering: Evidence from a suite of take-up experiments in Kenya." *Unpublished Manuscript, Harvard University, Cambridge, MA*.

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. 1997a. "Golden eggs and hyperbolic discounting." *The Quarterly Journal of Economics* 112 (2): 443–478.

———. 1997b. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112 (2): 443–478.

Lejuez, CW, Derek R Hopko, Ron Acierno, Stacey B Daughters, and Sherry L Pagoto. 2011. "Ten year revision of the brief behavioral activation treatment for depression: revised treatment manual." *Behavior modification* 35 (2): 111–161.

Lezak, M.D. 1983. *Neuropsychological Assessment.* 2nd. Oxford University Press, New York.

Lyon, G, and Norman A Krasnegor. 1996. *Attention, memory, and executive function.* Paul H Brookes Publishing.

Manos, Rachel C, Jonathan W Kanter, and Wen Luo. 2011. "The behavioral activation for depression scale–short form: development and validation." *Behavior Therapy* 42 (4): 726–739.

Miller, Earl K, and Jonathan D Cohen. 2001. "An integrative theory of prefrontal cortex function." *Annual review of neuroscience* 24 (1): 167–202.

Miyake, A., N. P. Friedman, M. J. Emerson, A. H. Witzki, A. Howerter, and T. D. Wager. 2000a. "The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis." *Cognitive Psychology* 41 (1): 49–100 (August).

Miyake, Akira, Naomi P Friedman, Michael J Emerson, Alexander H Witzki, Amy Howerter, and Tor D Wager. 2000b. "The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis." *Cognitive psychology* 41 (1): 49–100.

Morisano, Dominique, Jacob B Hirsh, Jordan B Peterson, Robert O Pihl, and Bruce M Shore. 2010. "Setting, elaborating, and reflecting on personal goals improves academic performance." *Journal of Applied Psychology* 95 (2): 255.

Null, C, CP Stewart, AJ Pickering, HN Dentz, BF Arnold, CD Arnold, J Benjamin-Chung, T Clasen, and KG Dewey. 2018. "Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhea and child growth in rural Kenya."

Phillips, L. H., V. E. Wynn, S. McPherson, and K. J. Gilhooly. 2001. "Mental planning and the Tower of London task." *The Quarterly Journal of Experimental Psychology Section A* 54 (2): 579–597.

Radloff, Lenore Sawyer. 1977. "The CES-D scale: A self-report depression scale for research in the general population." *Applied psychological measurement* 1 (3): 385–401.

Richards, DA, Karina Lovell, Simon Gilbody, L Gask, D Torgerson, M Barkham, M Bland, Peter Bower, AJ Lankshear, A 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.

Schwarzer, Ralf, and Matthias Jerusalem. 2010. "The general self-efficacy scale (GSE)." *Anxiety, Stress, and Coping* 12:329–345.

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

Stewart, Christine P, Patricia Kariger, Lia Fernald, Amy J Pickering, Charles D Arnold, Benjamin F Arnold, Alan E Hubbard, Holly N Dentz, Audrie Lin, Theodora J Meerkerk, et al. 2018. "Effects of water quality, sanitation, handwashing, and nutritional interventions on child development in rural Kenya (WASH Benefits Kenya): a cluster-randomised controlled trial." *The lancet child & adolescent health* 2 (4): 269–280.

Suchy, Yana. 2009. "Executive functioning: Overview, assessment, and research issues for non-neuropsychologists." *Annals of behavioral medicine* 37 (2): 106–116.

Walker, Christa L Fischer, Igor Rudan, Li Liu, Harish Nair, Evropi Theodoratou, Zulfiqar A Bhutta, Katherine L O'Brien, Harry Campbell, and Robert E Black. 2013. "Global burden of childhood pneumonia and diarrhoea." *The Lancet* 381 (9875): 1405–1416.

Wang, Haidong, Mohsen Naghavi, Christine Allen, Ryan M Barber, Zulfiqar A Bhutta, Austin Carter, Daniel C Casey, Fiona J Charlson, Alan Zian Chen, Matthew M 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.

White, Holly A, and Priti Shah. 2006. "Training attention-switching ability in adults with ADHD." *Journal of Attention Disorders* 10 (1): 44–53.

Wilson, Timothy D, Thalia Wheatley, Jonathan M Meyers, Daniel T Gilbert, and Danny Axsom. 2000. "Focalism: A source of durability bias in affective forecasting." *Journal of personality and social psychology* 78 (5): 821.

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	−0.12	−0.15	0.91	2975	26.62	−0.38	−0.40	−0.25	3750
	(4.56)	(0.21)	(0.21)			(4.69)	(0.23)*	(0.22)*	(0.22)	
Married or cohabiting	0.89	−0.01	0.01	0.29	2975	0.90	−0.02	−0.00	−0.01	3750
	(0.32)	(0.01)	(0.01)			(0.30)	(0.02)	(0.01)	(0.01)	
Education level	5.87	−0.04	0.05	0.10*	2975	5.93	−0.10	−0.01	−0.06	3750
	(1.23)	(0.06)	(0.05)			(1.08)	(0.06)*	(0.05)	(0.05)	
High wealth index	0.54	−0.03	−0.04	0.61	2975	0.52	−0.01	−0.02	0.02	3750
	(0.50)	(0.02)	(0.02)*			(0.50)	(0.02)	(0.02)	(0.02)	
Village of residence	83.26	0.83	−0.21	0.68	2975	83.31	0.78	−0.27	−0.05	3750
	(54.89)	(2.52)	(2.52)			(56.43)	(2.72)	(2.72)	(2.67)	
Delay Variables										
Days between endline and baseline	68.92	0.97	0.50	0.83	2396	68.73	1.15	0.68	0.19	2984
	(21.52)	(2.13)	(2.06)			(24.07)	(1.86)	(1.77)	(1.78)	
Days between chlorine test and baseline	79.33	1.66	1.62	0.99	2203	81.20	−0.22	−0.25	−1.88	2758
	(26.71)	(2.80)	(2.72)			(27.41)	(2.38)	(2.28)	(2.26)	
Attrition										
Attrited from endline	0.18	0.01	0.03	0.50	2975	0.24	−0.05	−0.03	−0.06	3750
	(0.39)	(0.02)	(0.02)			(0.43)	(0.02)**	(0.02)	(0.02)***	
Attrited from chlorine measure	0.22	0.02	0.03	0.53	2975	0.26	−0.03	−0.01	−0.04	3750
	(0.42)	(0.02)	(0.02)			(0.44)	(0.02)	(0.02)	(0.02)*	
Compliance										
Completed both first and second intervention	0.74	−0.00	−0.02	0.38	2975	–	–	–	–	-
	(0.44)	(0.03)	(0.03)			–	–	–	–	-
Completed first intervention	0.78	0.01	0.00	0.83	2975	–	–	–	–	-
	(0.41)	(0.02)	(0.02)			–	–	–	–	-
Completed no intervention	0.22	−0.01	−0.00	0.83	2975	–	–	–	–	-
	(0.41)	(0.02)	(0.02)			–	–	–	–	-

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

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	0.02	0.09	0.21	2329	0.00	0.02	0.10	0.01	2902
	(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	0.00	−0.09	0.04**	2372	0.00	−0.02	−0.12	−0.02	2955
	(1.00)	(0.05)	(0.05)**			(1.00)	(0.07)	(0.07)*	(0.07)	
Time Preferences										
Main outcome										
β^{Effort}	0.958	−0.013	−0.025	0.35	1196	0.979	0.005	0.002	0.013	1767
	(0.022)	(0.029)	(0.030)			(0.006)	(0.008)	(0.008)	(0.008)	
Additional outcomes										
β^{MPL}	0.98	0.02	0.02	0.96	2295	0.99	−0.01	−0.01	−0.02	2869
	(0.31)	(0.02)	(0.02)			(0.37)	(0.03)	(0.02)	(0.03)	
		[1.00]	[1.00]				[1.00]	[1.00]	[1.00]	
δ^{MPL}	0.98	−0.00	−0.00	0.82	2372	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]	[1.00]	[1.00]	
δ^{Effort}	0.990	−0.002	−0.002	0.50	1196	0.999	−0.001	−0.002	−0.001	1767
	(0.003)	(0.004)	(0.004)			(0.001)	(0.001)	(0.001)	(0.001)	
		[1.00]	[1.00]				[1.00]	[1.00]	[1.00]	
Risk Aversion Measure (z-score)	0.00	−0.02	−0.05	0.65	2188	0.00	0.02	−0.00	0.04	2735
	(1.00)	(0.06)	(0.05)			(1.00)	(0.06)	(0.06)	(0.06)	
Self-Efficacy										
General Self-Efficacy Score (GSE) (z-score)	0.00	0.12	0.11	0.81	2321	0.00	0.10	0.09	−0.03	2899
	(1.00)	(0.05)**	(0.05)**			(1.00)	(0.08)	(0.08)	(0.08)	
Mechanisms										
Belief: Proportion of diarrhea incidences avoided through chlorination (z-score)	0.00	0.02	0.04	0.63	2372	0.00	0.13	0.15	0.11	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)**	(0.05)***	(0.06)**	
Chlorine Knowledge Score (z-score)	0.00	0.03	−0.00	0.51	2372	0.00	0.11	0.07	0.08	2955
	(1.00)	(0.05)	(0.05)			(1.00)	(0.05)**	(0.06)	(0.06)	

Notes: * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. Coefficients for β^{Effort} and δ^{Effort} are reported for the subsample of observations before a change in the piece rates was introduced (from KES 2, 6, 10 to KES 0, 2, 6, 10, see Footnote 14). This change was unbalanced between active and passive treatment arms, and requires the use of more complex structural estimation methods. This subsample will be added in later versions.

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	0.04	0.01	0.13	2268	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)	0.20	0.04	0.00	0.09*	2268	0.18	0.06	0.03	0.03	2839
	(0.40)	(0.02)*	(0.02)			(0.39)	(0.02)***	(0.02)	(0.02)	
		[0.10]	[0.61]				[0.03]**	[0.54]	[0.61]	
Number of diarrhea incidences per child u15 in last 3 months	0.22	−0.07	−0.06	0.78	2254	0.20	−0.06	−0.05	0.01	2806
	(0.51)	(0.02)***	(0.02)***			(0.51)	(0.03)**	(0.03)*	(0.03)	
		[0.02]**	[0.05]**				[0.06]*	[0.54]	[1.00]	
Proportion of children u15 vaccinated in last 3 months	0.23	0.01	−0.03	0.05**	2249	0.22	0.01	−0.02	0.00	2800
	(0.36)	(0.02)	(0.02)			(0.36)	(0.02)	(0.02)	(0.02)	
		[0.34]	[0.40]				[0.71]	[0.54]	[1.00]	
Number of ANC visits made in last 3 months (among pregnant women)	1.24	−0.29	−0.16	0.68	230	1.19	−0.27	−0.06	0.23	272
	(1.18)	(0.35)	(0.34)			(1.17)	(0.40)	(0.42)	(0.43)	
		[0.27]	[0.55]				[0.71]	[0.60]	[1.00]	
Number of children u15 taken for healthcare check-up in last 3 months	0.21	−0.04	−0.02	0.45	2253	0.17	0.00	0.01	0.03	2806
	(0.35)	(0.02)**	(0.02)			(0.31)	(0.02)	(0.02)	(0.02)*	
		[0.10]	[0.40]				[1.00]	[0.60]	[0.49]	
Savings outcomes										
Main outcome										
Amount saved regularly (per week)	79.78	13.40	4.13	0.40	2365	79.74	8.23	0.46	−3.85	2946
	(188.81)	(10.10)	(10.33)			(201.14)	(11.42)	(11.46)	(10.57)	
Additional outcomes										
Indicator: Amount saved regularly is positive	0.36	0.09	−0.02	0.00***	2387	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]	[0.43]	
Number of new ROSCAs joined in last 3 months	0.17	0.03	0.00	0.22	2371	0.21	0.00	−0.02	−0.03	2956
	(0.42)	(0.02)	(0.02)			(0.46)	(0.03)	(0.03)	(0.03)	
		[0.07]*	[1.00]				[0.43]	[1.00]	[0.43]	
Indicator: Respondent saves for productive investments	0.17	0.08	−0.01	0.00***	2387	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]	[0.43]	
Labor outcomes										
Main outcome										
Total hours of work in last 3 months	95.90	−1.37	−16.57	0.06*	2364	103.75	−6.18	−21.15	−5.20	2945
	(156.22)	(8.08)	(7.73)**			(173.22)	(9.50)	(9.11)**	(9.03)	
Additional outcomes										
Total days of work in last 3 months	19.20	1.10	−2.05	0.03**	2364	20.91	0.00	−2.96	−1.38	2944
	(27.27)	(1.39)	(1.37)			(29.26)	(1.61)	(1.56)*	(1.53)	
		[1.00]	[0.37]				[1.00]	[0.13]	[1.00]	
Average monthly earnings in last 3 months	918.00	12.37	−72.68	0.43	2368	906.11	26.47	−63.44	−3.66	2945
	(2313.19)	(116.85)	(122.71)			(2258.94)	(118.88)	(125.00)	(128.91)	
		[1.00]	[0.38]				[1.00]	[0.44]	[1.00]	
Other behavioral outcomes										
Main outcome										
Index of investment in children's education	0.00	−0.03	−0.01	0.63	1532	−0.05	0.07	0.09	0.09	1927
	(0.81)	(0.05)	(0.05)			(0.69)	(0.05)	(0.05)*	(0.05)*	

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

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]	[0.33]	[0.00]***
	(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	[0.11]	[0.06]*	[0.00]***
	(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	[0.08]*	[0.11]	[0.49]
	(0.56)	(0.04)*	(0.04)	(0.04)		(0.45)	(0.03)	(0.04)	(0.04)				
Belief: Proportion of diarrhea incidences avoided through chlorination	0.71	0.07	0.07	0.03	1573	0.71	0.03	0.05	0.06	1382	[0.42]	[0.57]	[0.19]
	(0.41)	(0.03)**	(0.03)**	(0.03)		(0.41)	(0.03)	(0.03)	(0.03)*				
Chlorine Knowledge Score	1.08	0.11	0.04	0.03	1573	1.06	0.03	0.04	0.07	1382	[0.03]**	[0.62]	[0.94]
	(0.71)	(0.05)**	(0.05)	(0.05)		(0.65)	(0.05)	(0.05)	(0.05)				

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

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

Baseline Session 1:

At baseline, both the demographic questionnaire and behavioral tasks 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. Tasks
    (a) Tower of London
    (b) Generalized Self-Efficacy

(c) Effort Discounting Task
(d) Monetary Discounting

5. Administration of Intervention Part 1
6. Debrief
7. Payment

Baseline Session 2

1. Welcome, Identification and Screening
2. Adminstration of Intervention Part 2
3. Debrief

Endline

1. Welcome, Identification and Screening
2. Consent
3. Salience Task
4. Group Tasks and Measures
   (a) Behavioral Activation for Depression Scale (BADS)
   (b) Generalized Self-Efficacy (GSE) Scale
   (c) Center for Epidemiological Studies Depression Scale, Reduced Version (CES-D-R)
   (d) Water Use Module
   (e) Chlorination Behavior Module
   (f) ANC/PNC Beliefs
   (g) Tower of London
   (h) Risk Measure
   (i) Effort Discounting Measure
   (j) Monetary Discounting Measure (Multiple Price Lists)
5. Individual Survey
   (a) Savings

(b) Business Development
(c) Asset Investment
(d) Labour Supply and Search
(e) Agricultural Inputs & Livestock
(f) Fertility & Antenatal/Postnatal Care
(g) Child Education & Health
(h) Participant Education
(i) Phone Access

B. List of Outcome Variables

In accordance with the hypotheses above, we divide outcomes variables into psychological mechanisms, behaviors, and tests for alternative mechanisms. The former two are enumerated below while the latter is described in Section 4.4. Within the psychological mechanism and behavior groups, we list primary, secondary and tertiary variables of interest. We apply the multiple hypothesis testing described in Section 5.5. We adjust for multiple hypothesis testing within outcome groups (psychological mechanisms and behaviors) and hierarchical categories (primary and secondary), but not across hierarchical categories or across outcome groups. Variables marked with * are available at both baseline and endline; the rest are measured only at endline.

1. Psychological Mechanisms

   (a) Primary:

      i. β^{Effort} (estimated from the effort discounting task)*
      ii. Behavioral Activation for Depression Scale - Short Form (BADS-SF)

   (b) Secondary:

      i. Generalized Self-Efficacy (GSE) scale*
      ii. $\delta^{Effort*}$
      iii. $\beta^{MPL\ *}$
      iv. δ^{MPL*}
      v. Tower of London task (outcome measure: total moves across all four rounds)*

(c) Exploratory:

   i. Center for Epidemiological Studies Depression scale, revised version (CESD-R)

   ii. Sophisticated Time-Inconsistency Self-Reports ($(tempted - ideal) \cdot \beta^{Effort}$ and $(expected - ideal) \cdot \beta^{Effort}$)

2. Economic and Health Behaviors

   (a) Primary: water chlorination (outcome measure: presence of any chlorine in household drinking water)

      i. Self-report confirmation: Indicator for any treatment of water*

   (b) Secondary:

      i. Amount saved regularly (frequency converted to weekly)

      ii. Total hours of work in last three months (includes all types of work, such as farming, casual labour, business ownership, or salaried jobs)*

      iii. Index of the following measures of investment in education:*

         - Indicator for a positive number of school days missed in last 5 days (across all school-age children in the household)
         - Total expenditure per child on children's schooling in last three months

   (c) Exploratory

      i. Savings:

         - Binary indicator: Amount saved regularly is positive
         - Number of ROSCAs joined in last 3 months
         - Indicator: Respondent saves for productive future investment (business, farming, or education)

      ii. Labor supply

         - Total days of work, paid or unpaid in last three months
         - Monthly earnings from any paid work (paid in cash and in-kind)

      iii. Health

         - Number of diarrhea incidences per child under 15 in the household in the last three months (controlled for number of children under 15 in the household)*

- Secondary
  - Number of children under 15 vaccinated in the last three months (controlling for number of children under 15 in the household)
  - Number of ANC/PNC visits made in last three months (among pregnant women)
  - Number of children under 15 taken for healthcare check-up in last three months (controlling for number of children under 15 in the household)

iv. Investment in productive assets
- Total asset expenditure in last three months, including business investment
- If significant effects are found on the total, we will examine the components
  - Total expenditure on livestock
  - Total expenditure on household durables

v. Investment in agricultural inputs
- Total expenditure on fertilizer, seeds, pesticide and renting plots
- Indicator for purchasing or leasing new agricultural plots in last three months
- If significant effects are found on the total, expenditure (separate) on each component: fertilizer, seeds, pesticide, and renting plots

vi. Business and Enterprise
- Average daily hours spent on all businesses
- Total expenditure on all businesses in last 30 days

vii. Family planning. Index of the following measures:
- Opinion on ideal number of children a woman should have
- Opinion on ideal age gap between children
- Contraceptive use wish indicator: wishing to use contraception or using it currently

B.1 Descriptions of other measures

Alternative mechanisms Evidence for alternative mechanisms is tested by regressing the relevant outcome variable on indicators for the different treatment arms, as specified in equation 2. An exception is the test for differential effects on salience, which is specified below.

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 arms ("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.

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 arms 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.

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 assume a CRRA utility function for choices in this task, which allows each choice to be rationalized by an interval of a risk parameter. We choose the mid-point of this interval as an estimate of individual risk preferences, and test for differential treatment effects as an alternative mechanism (Table 2).

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 D.5. We estimate salience effects using equation 4:

$$w_{im} = a_0 + \sum_{j=1}^{3} a_j T_{ji} + \psi_0 X_{im} + \delta_m + \theta_{im} \quad (4)$$

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

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

$$w_{imn} = a_0 + \sum_{j=1}^{3} a_j T_{ji} + \lambda chlorine_n + \psi X_{im} + \sum_{k=1}^{3} b_k T_{ki} \cdot chlorine_n + \delta_m + \theta_{imn} \quad (5)$$

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.

Other Outcome measures

Depression We include the Center for Epidemiological Studies Depression Scale - Revised, a 10-item scale intended for epidemiological research but not clinical diagnosis

(Eaton et al. 2004; Radloff 1977). The CESD-R is well validated, including for sub-Saharan African populations (Baron, Davies, and Lund 2017). Participants are asked to identify how often they felt certain emotions in the past week, from "rarely or never" (scored as 0) to "all the time or most of the time" (scored as 3). Eight items indicate greater probability of depression (e.g "I was bothered by things that usually don't bother me") while two which are negative associated with depression (e.g. "I felt hopeful about the future") are reversed for scoring purposes.

Self-Control We adapt a module on sophistication from John (2017). Participants are asked to imagine that they are given ten vouchers for a one-time consumable luxury found in their community, in this case an all-you-can-eat dinner at a *nyama choma* (Kenyan-style barbeque) restaurant, with the condition that the vouchers expire in two years. Then, they state (i) what the *ideal* distribution of this would be across the two years; (ii) how many of the 10 they will be *tempted* to use in the first year; and (iii) how many they believe they would *actually* use in the first year.

In addition, participants are asked to select the extent to which they agree with three sentences relating to self-control and time consistency. Two of these are negative (e.g "Many of my choices in the past I now regret making") and one positive ("I am willing to give up something that is beneficial for me today in order to benefit more from that in the future").

C. Description of WASH Benefits Kenya

WASH Benefits 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 (Stewart 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. After filling water carrying containers, usually a 20l jerrycan, users turn the knob on the dispenser to add 3ml of 1.25% sodium hypochlorite (chlorine), which yields 2.5mL/L of free chlorine residual after 30mins for 20l 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 receive a six month supply of bottled chlorine every six months, to be used for sanitizing water at home, in case the household drinks harvested rainwater or chooses 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 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. We conduct additional robustness checks, including (i) excluding them from the heterogeneity analysis by "Water Quality" assignment, and (ii) excluding them from all analyses described in section 5. 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.

D. Additional Tables and Figures

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

	(1) Attrited from endline	(2) Attrited from endline	(3) Attrited from chlorine measure	(4) Attrited from endline	(5) Attrited from chlorine measure
Time Preferences	0.00 (0.02)	0.00 (0.02)	−0.00 (0.02)	0.12 (0.14)	0.08 (0.15)
Executive Function	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.12 (0.15)	0.10 (0.16)
Age		−0.01 (0.00)***	−0.01 (0.00)***	−0.01 (0.00)*	−0.01 (0.00)*
Married or cohabiting		−0.07 (0.03)**	−0.07 (0.03)**	−0.06 (0.04)	−0.05 (0.04)
Education level		−0.02 (0.01)*	−0.02 (0.01)*	−0.01 (0.01)	−0.01 (0.01)
High wealth index		−0.02 (0.02)	−0.04 (0.02)*	−0.00 (0.03)	−0.03 (0.03)
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				−0.04 (0.06)	−0.06 (0.06)
EF x Married Interaction				−0.00 (0.06)	−0.00 (0.07)
TP x Education Interaction				−0.00 (0.01)	−0.00 (0.02)
EF x Education Interaction				−0.02 (0.02)	−0.02 (0.02)
TP x Wealth Interaction				−0.01 (0.04)	0.00 (0.04)
EF x Wealth Interaction				−0.04 (0.04)	−0.03 (0.04)
Constant	0.19 (0.01)***	0.55 (0.06)***	0.60 (0.07)***	0.48 (0.10)***	0.55 (0.11)***

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

Table D.2: Attrition analysis: active treatments vs. pure control

	(1) Attrited from endline	(2) Attrited from endline	(3) Attrited from chlorine measure	(4) Attrited from endline	(5) Attrited from chlorine measure
TP+INF	−0.05	−0.06	−0.04	−0.02	0.01
	(0.02)**	(0.02)**	(0.02)	(0.16)	(0.17)
EF+INF	−0.04	−0.04	−0.02	−0.04	0.01
	(0.02)	(0.02)*	(0.02)	(0.18)	(0.18)
PLA+INF	−0.06	−0.06	−0.04	−0.16	−0.09
	(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)
High wealth index		−0.01	−0.03	−0.01	−0.03
		(0.01)	(0.02)*	(0.03)	(0.03)
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	0.24	0.59	0.62	0.61	0.61
	(0.02)***	(0.06)***	(0.06)***	(0.13)***	(0.14)***

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

Table D.3: 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 D.4: 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 D.5: Word lists for salience test

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

Table D.6: 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: * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table D.7: Baseline balance: main outcomes

	Comparison with active control (PLA+INF)				
	(1) Placebo Group Mean (SD)	(2) Time Preferences Difference	(3) Execuive Function Difference	(4) TP = EF *p*-value	(5) *N*
Baseline Score					
Tower of London: Total Moves	21.29	−0.07	−0.42	−0.53	2372
	(6.71)	(0.33)	(0.33)	(0.34)	
		[1.00]	[0.41]		
General Self-Efficacy Score (GSE)	43.54	1.46	1.41	0.17	2321
	(11.05)	(0.60)**	(0.56)**	(0.64)	
		[0.06]*	[0.05]*		
β^{MPL}	1.04	−0.01	0.01	0.02	2372
	(0.44)	(0.02)	(0.02)	(0.03)	
		[1.00]	[0.64]		
δ^{MPL}	0.98	−0.00	−0.00	−0.00	2372
	(0.02)	(0.00)	(0.00)	(0.00)	
		[1.00]	[0.64]		

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

Figure D.1: Tower of London Example Screen

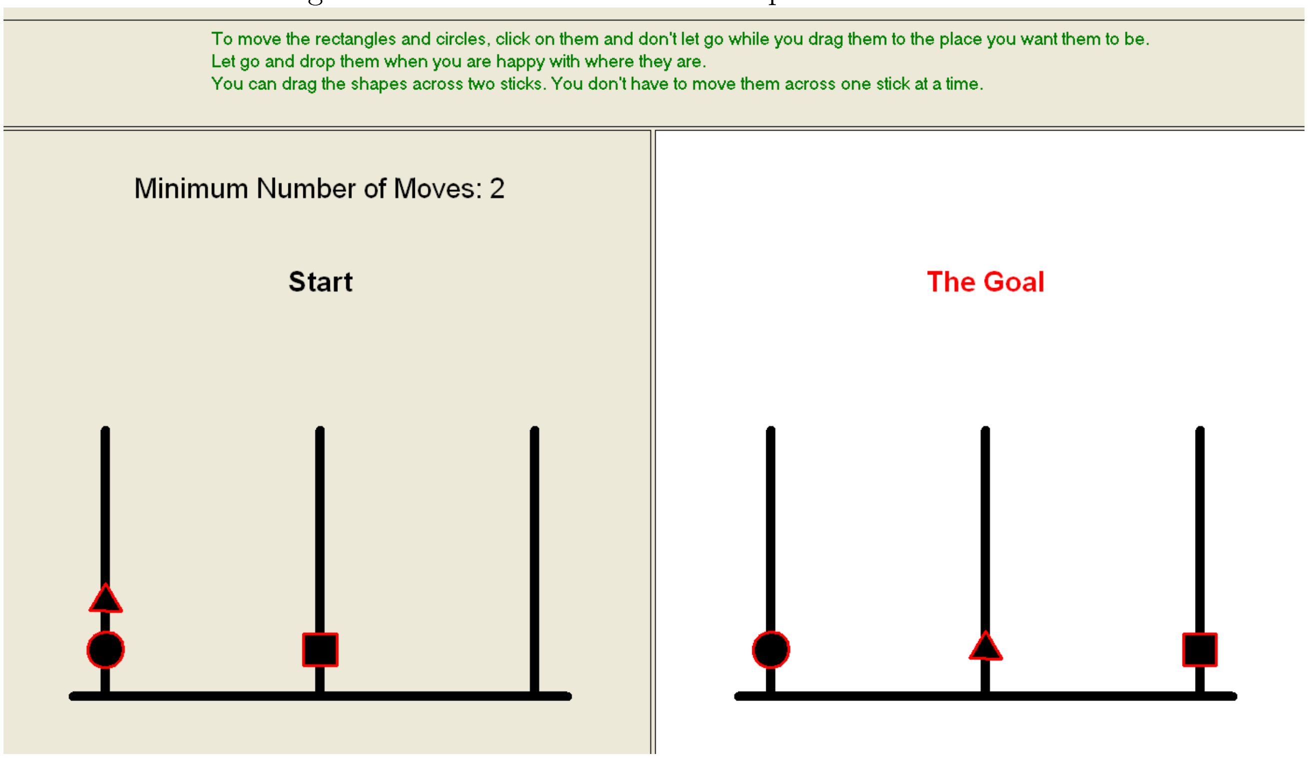


Table D.8: Phone access & task comprehension questions

	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 (0.46)	0.01 (0.02) [0.68]	−0.03 (0.02) [1.00]	−0.04 (0.02)	2387	0.70 (0.46)	0.00 (0.02) [0.60]	−0.03 (0.03) [0.42]	−0.00 (0.03) [0.67]	2972
Participant uses a phone belonging to her household	0.96 (0.19)	0.01 (0.01) [0.44]	−0.00 (0.01) [1.00]	−0.01 (0.01)	2387	0.96 (0.21)	0.02 (0.01) [0.48]	0.00 (0.01) [0.70]	0.01 (0.01) [0.61]	2972
Proportion for whom accessing a phone for 30mins is very difficult or impossible	0.12 (0.32)	0.03 (0.02)* [0.27]	0.01 (0.02) [1.00]	−0.03 (0.02)	2386	0.13 (0.33)	0.02 (0.02) [0.48]	−0.00 (0.02) [0.70]	−0.01 (0.02) [0.61]	2970
Proportion for whom accessing a phone for 1hr is very difficult or impossible	0.16 (0.37)	0.04 (0.02)** [0.25]	0.03 (0.02) [1.00]	−0.02 (0.02)	2386	0.19 (0.39)	0.02 (0.02) [0.48]	0.00 (0.02) [0.70]	−0.03 (0.02) [0.37]	2970
Proportion for whom accessing a phone for 4hrs is very difficult or impossible	0.31 (0.46)	0.02 (0.02) [0.44]	0.01 (0.02) [1.00]	−0.02 (0.03)	2384	0.35 (0.48)	−0.02 (0.03) [0.55]	−0.03 (0.03) [0.42]	−0.04 (0.03) [0.31]	2967
SMS Comprehension questions correct first time	0.80 (0.40)	−0.03 (0.02) [0.41]	−0.00 (0.02) [1.00]	0.03 (0.02)	2372	0.72 (0.45)	0.04 (0.03) [0.48]	0.07 (0.03)*** [0.06]*	0.07 (0.03)*** [0.02]**	2955
Number of attempts at SMS comprehension questions	0.85 (1.95)	0.17 (0.12) [0.34]	0.10 (0.12) [1.00]	−0.06 (0.12)	2372	1.22 (2.36)	−0.21 (0.13) [0.48]	−0.27 (0.13)** [0.12]	−0.38 (0.13)*** [0.02]**	2955

Notes: These questions were asked as validation checks for the SMS effort task. Column 5 reports the p-value of the t-test of equality between the treatment effects of TP and EF. Column 6 reports the p-value of the F-test of equality between all three treatment effects. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Figure D.2: MPL task example screen

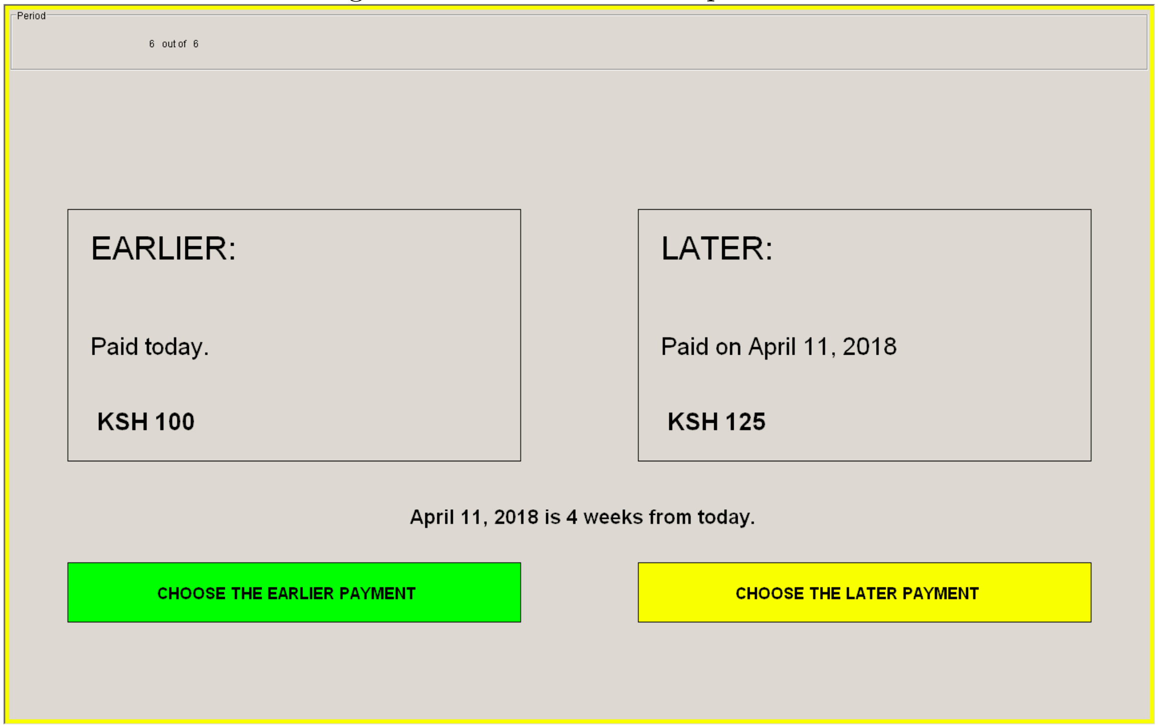


Figure D.3: Effort discounting task example screen