Effect of Light-Touch Psychological Interventions on Economic and Psychological Outcomes: Pre-Analysis Plan[*]

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

March 30, 2018

Abstract

This document describes the analysis plan for a randomized field experiment examining the effects of light touch cognitive-behavioral interventions on preferences and behavior. Our study includes women of childbearing age in Western Kenya. Respondents are randomly allocated to one of two interventions, an active control group, or a pure control group. This plan outlines the design of the study, the outcomes of interest, and the econometric approach.

[†]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, ENSAE, Paris, France.

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

1 Introduction and Design

In this study, we test the effectiveness of two light-touch psychological interventions on a number of economic and psychological outcomes. This document describes the design, outcome variables, and the estimation strategy of the study.

1.1 Sampling Strategy

We study a sample of women aged 18-35 in rural Western Kenya. We recruited a pool of roughly 3750 individuals from Bungoma and Kakamega counties between October 2017 and January 2018, of which 2330 participated in the interventions.

One behaviour we examine 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 in Bungoma and Kakamega counties, a study of a variety of interventions targeted at increasing whether people chlorinate water. We restricted sample recruitment to villages which were assigned to either the "Water Quality" treatment arm or the "Passive Comparison" arm of the WASH Benefits study (see exception below). In the "Water Quality" villages, chlorine dispensers were installed at community water points during the WASH Benefits study and sample households received a free 1l bottle of chlorine every six months. Local promoters visited households each month to encourage treating and safely storing water. Evidence Action's Dispensers for Safe Water program has since maintained these dispensers and retain a local promoter in each community. Households in WASH Benefits "Passive Comparison" villages did not receive any intervention or household visit.

By splitting our study sample across the "Water Quality" and "Passive Comparison" villages, we will be able to identify any heterogeneous effects of Evidence Action's chlorine dispensers and promotion on the efficacy of our interventions. We exclude from our sample all women from households which took part in the WASH study by not sampling women who had children the correct age to be eligible for the WASH study.

However, a coding error during randomization resulted in the inclusion of participants from villages assigned to other WASH-Benefits treatment arms (approximately

20% of sample, in Mumias constituency, Kakamega county). These villages received sanitation and nutrition interventions in addition to the "Water Quality" interventions. However, these interventions were at the household level, and as before our recruitment strategy excludes direct WASH-Benefits participants, so participants in our study did not receive them. Thus we include these villages in our main estimation of treatment effects. For the heterogeneity analysis by WASH-Benefits "Water Quality" intervention status, we group treatment arms by whether or not they received the water quality interventions (in addition to our focus arm "Water Quality"), this was also the case for arms which received "Water Quality, Sanitation and Handwashing" and "Water Quality, Sanitation, Handwashing and Nutrition"). 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 2.

In three constituencies in these counties, enumerators visited all villages assigned to the 'Water Quality' and 'Passive Comparison' arms of the WASH-Benefits study. In a fourth constituency, enumerators visited villages assigned to all arms of the WASH-Benefits study.

With the help of local guides, enumerators visited all households in each village 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) the household did not participate in the WASH-Benefits study. As a second check on (ii), enumerators also excluded households with children aged either 3-4 or 4-5, depending upon the village's WASH-Benefits block assignment, since such households were considered very likely to have participated in the WASH-Benefits study.

From this census list, all eligible women in three of the constituencies were randomized into one of the intervention arms. In the fourth constituency, a random sample of eligible women was taken..

1.1.1 Sample Size and Statistical Power

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

1. 992 assigned to Treatment Arm 1: "Imagining the Future"

2. 991assigned to Treatment Arm 2: "Get Going, Keep Doing"

3. 994 assigned to the Active Control group: "Nature in Kenya"

4. Roughly 775 assigned to the Pure Control group. Randomization Strategy

We allocated participants to one of the four arms using stratified random assignment. We stratified on two variables collected during the census described in 1.1 :

1. Wealth Index: the total value of a limited set of assets which participants reported owning in the recruitment survey. The assets included in the index are 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 randomly assigned to attend baseline and intervention sessions either in the morning or in the afternoon. While participants were first invited to and 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.

1.1.2 Attrition

Attrition is 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. Although the initial session time that participants were invited to was randomized (and is thus available as an instrument), we also offered participants flexibility in which day and time they attended sessions to better accommodate their schedules.

1.2   Treatment and Data Collection

The study included three active and one passive treatment arm: One arm aimed at improving planning and performance of basic tasks ("Get Going, Keep Doing"), one arm that encouraged respondents to visualize their future ("Imagining the Future", ITF), a placebo intervention ("Nature in Kenya"), and a passive 'pure control'. 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.

The baseline measures were collected in Busara "mobile labs" in Bungoma and Kakamega county, which each hold up to 25 participants 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.[1] At endline, individual questionnaires were administered using SurveyCTO. Respondents received KES 200 for participating in the baseline and each intervention session, and KES 300 for participating in the endline session. They were additionally given a KES 50 bonus for arriving on time

---

[1]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.

for the session. Participants were reimbursed for their transport costs from their home to the mobile laboratory.[2]

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

1.2.1 Treatment 1: Imagining the Future

This treatment arm was inspired by previous work in Turkish primary schools (Alan and Ertac, 2018), as well as Gabaix and Laibson (2017)'s theoretical work on imperfect forecasting and as-if discounting. It used interactive lectures, case stories, exercises and drawings to encourage participants to a) connect their present behavior to outcomes in the future, b) visualize alternative realizations of the future, depending on their current behaviour, and c) put themselves in the shoes of their future selves, imagine how they feel, and 'talk' to them.

1.2.2 Treatment 2: Get Going, Keep Doing

This treatment arm was inspired by a simple one-session manual for low-intensity behavioral activation called "Reach Out" (Richards and Whyte, 2011). It used the same teaching tools as the other intervention to teach participants that people sometimes become stuck in inactivity and avoid important tasks, which can have negative effects on functioning. They listen to a story of a similar woman in this position and share similar stories and possible solutions from their lives. They then make lists of pleasurable, mandatory, and important activities, rank them from most to least difficult, and schedule them in a weekly diary. In the second session, they cross off completed plans and circle uncompleted ones, and discuss barriers they faced and ways to overcome them.

1.2.3 Placebo Control: Nature in Kenya

An "Active Control" group attended baseline laboratory sessions and received a placebo group-talk session. This was implemented to control for any effects of sim-

---

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

ply attending a session and interacting with women from neighboring villages. The placebo group 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. Participants in the Active Control group also received the same chlorination and ANC/PNC information as treatment groups, discussed below.

1.2.4   Pure Control

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

1.2.5   Information Module

To hold constant the information contained in the active treatment arms, all groups except for the pure control 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. The information module was added at the end of each treatment session.

1.3   Outcome Measures

1.3.1   Self-Efficacy

We measure self-efficacy using the Generalized Self-Efficacy (GSE) scale (Schwarzer and Jerusalem, 2010). Our version contains 12 items, 10 from the generic version, and two which are repeated and reversed. 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).

1.3.2 Time Preferences

Following recent criticisms of monetary discounting elicitations (Andreoni and Sprenger, 2012; Augenblick et al., 2015), we estimate time preferences over the effort domain. We use the question design from Augenblick (2017), with a new effort task that is adapted to our field setting in Kenya. Participants make decisions over how many units of effort to supply at varying times (today, tomorrow, 7 days from now and 8 days from now) and piece rates (KES 2, 6 and 10).[3] To consider the possibility that respondents feel obligated to carry out some effort regardless of the wage, we also include a question on how many units of effort participants would supply for no piece rate (just the KES 100 completion bonus explained below) for a subsample. All questions required a minimum effort allocation of one task to control for the fixed costs of starting.

Effort is supplied in the form of SMS messages to a toll free number administered by the Busara Center. Each SMS is a 30-character numeric string, which must be keyed in with 75% accuracy and takes approximately two minutes to complete. The participants are given a sheet which lists 50 such strings (including a counter), and they complete one string for practice during the session. At the end of the session, one decision (out of 12) is selected to be the "decision that counts": At the selected piece rate and the selected time horizon, participants have to send the exact number of SMS they chose (within a tolerance). If they do, they receive the 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 (analogue to Augenblick (2017)). Earnings from this task were paid 14 days from the session date, regardless of the selected effort time horizon.

We estimate β^{Effort} and δ^{Effort} following the approach outlined in Augenblick (2017) by assuming quasi-linear utility (linear in money, convex in effort), as well as quasi-hyperbolic discounting. Like Augenblick (2017), we use a power cost of

---

[3]During the baseline, participants face one of three different sets of piece rates (KES 2, 6, 10; KES 2, 4, 6; or KES 2, 6, 10, 20). We accout for these differences by including set fixed effects. All participants made choices over the same piece rates during endline, except for the "no piece rate" subsample described below.

effort function. 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 our participants supply non-zero amounts of effort even for low or zero piece rates at baseline.

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. We will report summary statistics of the phone accessibility and comprehension questions.

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 from todays vs eight weeks from today). The list of decisions is presented in table 1. Figure **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 (Laibson, 1997), assuming linearity of utility in money.

1.3.3 Executive Function

To measure aspects of executive function, we adapt one scale and one behavioral task:

- To measure the construct most directly targeted by our BA intervention, we employ the Behavioral Activation for Depression Scale (BADS; 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 et al. (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 engaged in many different activities") and negative items (e.g. "I spent a long time thinking over and over about my problems"). 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.

- We also implemented 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 1.

1.3.4 Sophistication about Time Inconsistency

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").

1.3.5 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 et al., 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.

1.3.6 Risk Preferences

We include a modified Eckel-Grossman task to account for risk preferences (Charness et al., 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. This is used as a check for a possible alternative mechanism (see section 2.5).

1.3.7 Chlorination Behavior

During both baseline and endline surveys participants were asked "did you put chlorine in your water in the last month," with reference to popular local brands of chlorine for drinking water, which we use as a binary outcome measure. In addition, participants were visited in their homes to assess chlorination behavior, several days after coming to the laboratory for the endline survey. We test for the presence of both total and free chlorine; the former indicates the presence of any chlorine in the water while "free" chlorine refers to the residual after some chlorine has combined with nitrates in the water and is therefore unusable for sanitation purposes; therefore, the presence of free chlorine is necessary for the water to be dependably potable (CDC, 2010). At the time of home visit, enumerators again asked for self-reported chlorination today and in the last 30 days. They also noted the type of container used for storing drinking water and if it is covered.

1.3.8 Other Self-Reported Behaviors

Participants completed a demographic survey at endline and additional modules on economic and health behaviors at endline. Relevant elements of these surveys are enumerated in list of outcome variables below.

1.3.9 Salience and Demand Effects

To test whether our treatments increase the salience of chlorination, we used the following task: participants listen to three lists of nine words read out by an enumerator in a one-on-one setting. After each list, they repeat back any words they remember. Subjects are paid KES 5 for each word they remember to incentivize performance. Each list contains one word related to water chlorination, one related to savings, and one related to farm investment, along with six "filler" words which are designed to be similar as possible to the target words but do not relate to any behavior which is the target of this study. The word lists are available in original Swahili and English translation in table 2.

2 Econometric Approach

2.1 Main Specification: Treatment and Placebo Comparison

We employ the following main specification:

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

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. In this case, the active control group is the reference category, and T_{1i} and T_{2i} refer to the "Imagining the Future" and "Get Going, Keep Doing" 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. The sample is restricted to those who are in these three active treatment groups and who participated at least at baseline, the first intervention session, and endline (either survey or home visit, for the relevant outcome measure). The sample size for each regression will therefore vary among outcomes and be reported along with the coefficients. Standard errors are clustered at the

13

session level.

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 1 / 99% level.

2.2 All Treatment Arms

We use the following specification to compare all active treatment arms to the "pure" control group:

$$y_{iv} = \beta_0 + \beta_1 T_{1i} + \beta_2 T_{2i} + \beta_3 T_{3i} + \mathbf{\Phi X}_{iv} + \gamma_v + \theta_w + \varepsilon_{iv} \tag{2}$$

Here, y_{iv} is the outcome of interest for respondent i in village v at time of endline. T_j are indicators for assignment to the "Imagining the Future", "Get Going, Keep Doing", and active control groups, respectively. The pure control arm is therefore the reference category, and the β parameters are the treatment effects. $\mathbf{X}$ 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. All recruited participants for whom we recorded the outcome of interest, either during the endline survey or home visit, are included in this specification. Standard errors in each regression are clustered at the session level.

Since the sample of this specification will include both compliers and non-compliers, we run two regressions per outcome variable: one is ordinary least squares (the "intent-to-treat" estimate), and a model in which treatment status is instrumented with treatment assignment (the "treatment on the treated" estimate). We consider all those in any non-control group, including the placebo control, who attended at least the initial baseline session, though not necessarily the second intervention ses-

sion, to be complying with treatment assignment.

WASH Benefits Follow-up

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 both specifications 2 and 1 with an indicator variable for treatment status (treatment/control) in the Water Quality arm of the WASH Benefits study, and this indicator interacted with the treatment assignments in the present experiment. The outcome variable of primary interest in these specifications is an indicator for objective chlorination, with self-reported behavior run as a secondary variable of interest.

Due to an error in sampling, some participants in Mumias Constituency of Kakamega County were drawn from other treatment arms of the WASH study other than "Water Quality" or "Passive Comparison" (approximately 20% of the sample, see Section 1.1). We include these participants in the WASH regressions, grouped by whether their treatment arm received chlorine dispensers or not (this is the case for arms which received "Water Quality, Sanitation and Handwashing" and "Water Quality, Sanitation and Handwashing and Nutrition"). We exclude these participants in a robustness check.

Hypothesis Testing

The main hypotheses we intend to test are informed by our main specifications, and can be divided into three families:

1. Neither intervention has an effect relative to the placebo control group: $\alpha_1 = 0, \ \alpha_2 = 0$

2. Interventions do not have different effects relative to each other: $\alpha_1 = \alpha_2$

3. Neither intervention nor the placebo has an effect relative to the pure control group: $\beta_1 = 0, \ \beta_2 = 0, \ \beta_3 = 0$

15

4. Neither treatment enhances or decreases the effect of village chlorine dispensers

2.3   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 2.5. 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 2.4. Weadjust 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. β^{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 $(expected - ideal) \cdot \beta^{Effort}$)

16

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

2.4 Multiple Hypothesis Testing (MHT) Correction

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

2.5 Other analyses

1. **Randomization Check:** To determine whether randomization was successful, we estimate our main equation with the following demographics as outcome variables: age, years of education, marital status and location (village).

2. **Selective Attrition:** To test for differential attrition by treatment, we regress whether the participant was present at endline on treatment assignment, using equation 3 below.

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

Here, T_j refers to treatment assignment. In addition, equation 4 assesses whether attriting individuals are different in terms of the demographics described above and other control variables.

$$attrit_i = \beta_0 + \beta_1 \mathbf{X_i} + \varepsilon_i \tag{4}$$

X_i represents a vector of controls collected at baseline, as described in the randomization check.

3. **Selective Compliance:** Though participants had no way of knowing their treatment assignments prior to arriving for the session, we test for differential

compliance across treatment groups using equation 5:

$$comply_i = \alpha_0 + \sum_{j=1}^{2} \alpha_j T_{ji} + \omega_i \tag{5}$$

Here, $comply_i$ is an indicator for whether the participant came to at least the initial treatment session. Note that this specification only includes those in the placebo and two treatment arms, not those in the pure control group.

4. **Mechanisms:** We examine whether our treatments influenced chlorination behavior through mechanisms other than time preferences, self-efficacy, or executive function, using the following measures:

   (a) **Salience of chlorination (Demand Effects):** We test for the possibility that our treatments differentially increased salience of water chlorination using equation 6:

   $$w_{im} = a_0 + \sum_{j=1}^{3} a_j T_{ji} + \psi_0 X_{im} + \delta_m + \theta_{im} \tag{6}$$

   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 (there are three lists); 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.

      i. In case our treatments differentially affected the salience of chlorine, we further test whether this is due to an increased salience of future-oriented behaviours 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 behaviours (saving and farm investment), which

were not emphasized in the sessions. We estimate

$$w_{imn} = a_0 + \sum_{j=1}^{3} a_j T_{ji} + \lambda chlorine_n + \psi X_{im} + \sum_{k=1}^{3} b_k T_{ki} \cdot chlorine_n + \delta_m + \theta_{imn}$$
(7)

where w_{imn} is an indicator for participant i correctly recalling the words in list m from future oriented behaviour 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.

(b) **Risk Preferences:** To test for the possibility that the treatments unintentionally affect risk preferences, we include a modified Eckel-Grossman measure of risk preferences during the endline survey (Charness et al., 2013). Assuming a constant relative risk aversion utility function, we are able to estimate for the curvature parameter for each participant. We take the midpoint of this interval as our estimate for risk preferences. We use the main specifcations described above to assess treatment effects on risk preferences.

(c) **Beliefs about effectiveness of chlorination:** It is possible that the treatments 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 in the active treatment arms ("Imagining the Future", "Get Going, Keep Doing"| and Placebo) 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

21

chlorine effectiveness. We regress this measure on our treatment groups using the main specifications described in section 2.1.

(d) **Knowledge about how to use chlorine**: It is possible that our treatments can affect chlorination just by providing more information about how to properly use it. We ask two multiple-choice questions at endline, to which all three groups were told the correct answer at baseline:

    i. How much chlorine to add to water;

    ii. The amount of time that needs to pass after 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. We test differential belief formation across treatment groups using our main specifications described in section 2.1.

References

Alan, S. and Ertac, S. (2018). Fostering patience in the classroom: Results from a randomized educational intervention. *Journal of Political Economy*. 1.2.1

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495. 2.4

Andreoni, J. and Sprenger, C. (2012). Estimating time preferences from convex budgets. *The American Economic Review*, 102(7):3333–3356. 1.3.2

Augenblick, N. (2017). Short-term time discounting of unpleasant tasks. 1.3.2

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, page qjv020. 1.3.2

Baron, E. C., Davies, T., and Lund, C. (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. 1.3.5

CDC (2010). Chlorine residual testing fact sheet. Technical report, Centers for Disease Control and Prevention SWS Project. 1.3.7

Charness, G., Gneezy, U., and Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87:43–51. 1.3.6, 4b

DellaVigna, S. and Pope, D. (2016). What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*. 1.3.2

Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., and Tien, A. (2004). Center for epidemiologic studies depression scale: review and revision (cesd and cesd-r). 1.3.5

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178. 1.2

Gabaix, X. and Laibson, D. (2017). Myopia and discounting. Technical report, National Bureau of Economic Research. 1.2.1

John, A. (2017). When commitment fails-evidence from a field experiment. 1.3.4

Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American statistical association*, 91(433):222–230. 2.1

Kanter, J. W., Mulick, P. S., Busch, A. M., Berlin, K. S., and Martell, C. R. (2007). The behavioral activation for depression scale (bads): psychometric properties and factor structure. *Journal of Psychopathology and Behavioral Assessment*, 29(3):191. 1.3.3

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478. 1.3.2

Manos, R. C., Kanter, J. W., and Luo, W. (2011). The behavioral activation for depression scale–short form: development and validation. *Behavior Therapy*, 42(4):726–739. 1.3.3

Phillips, L. H., Wynn, V. E., McPherson, S., and Gilhooly, K. J. (2001). Mental planning and the tower of london task. *The Quarterly Journal of Experimental Psychology Section A*, 54(2):579–597. 1.3.3

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401. 1.3.5

Richards, D. and Whyte, M. (2011). *Reach Out: National programme student materials to support the delivery of training for psychological wellbeing practitioners delivering low intensity interventions.* Rethink Mental Illness. 1.2.2

Schwarzer, R. and Jerusalem, M. (2010). The general self-efficacy scale (gse). *Anxiety, Stress, and Coping*, 12:329–345. 1.3.1

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 298(1089):199–209. 1.3.3

Table 1: 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

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 is 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.

Table 2: Word Lists for Salience Task

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	Theros	Thermos	Filler
B	3	Savings Group	Chama	Savings
B	4	Baby Oil	Mafuta ya mtoto	Filler
B	5	Poultry Farming	Kilimo cha kuku	Farm investment
B	6	Petrol	Petroli	Filler
B	7	Chlorine	Klorini	Chlorine
B	8	Machete	Panga	Filler
B	9	Shoe Polish	Rangi ya viatu	Filler
C	1	Saucepan	Sufuria	Filler
C	2	Stool	Stool	Filler
C	3	Farm Lease	Kukodisha shamba	Farm investment
C	4	Transport	Transport	Filler
C	5	Dispenser	Dispensa	Chlorine
C	6	Photocopier	Photocopier	Filler
C	7	Piggybank	Benki ya nyumbani	Savings
C	8	Airtime	Airtime	Filler
C	9	Community Hall	Ukumbi wa jamii	Filler

Figure 1: Tower of London Example Screen

To move the rectangles and circles, click on them and don't let go while you drag them to the place you want them to be.
Let go and drop them when you are happy with where they are.
You can drag the shapes across two sticks. You don't have to move them across one stick at a time.

Minimum Number of Moves: 2
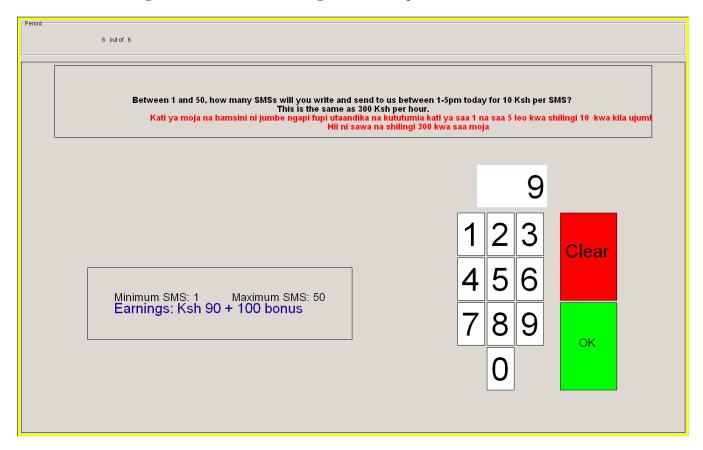
Start

The Goal

Figure 2: MPL Task Example Screen

Figure 3: Effort Discounting Task Example Screen

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