Effects of Stress on Preferences, Beliefs, and Constraints: Pre-Analysis Plan*

Johannes Haushofer[†], Prachi Jain[‡], Kristina Esopo[§], Daniel Mellow, David Ndetei[¶], Abednego Musau [‖]

February 17, 2018

Abstract

This document describes the analysis plan for a randomized laboratory experiment examining the effects of stress (implemented using four methods) on temporal discounting, effort provision, and executive control. Our study includes 1142 subjects from the informal settlements in Nairobi, Kenya. Respondents are randomly administered either a control or a treatment (stress) protocol, for seven consecutive days. This allows us to estimate the causal impacts of stress on decision-making after one and seven days of administration.

*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 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, Loyola Marymount University, Los Angeles, USA; and Busara Center for Behavioral Economics, Nairobi, Kenya.

§Department of Psychology, Princeton University, Princeton, USA; and Busara Center for Behavioral Economics, Nairobi, Kenya.

¶University of Nairobi, Nairobi, Kenya; and Africa Mental Health Foundation, Nairobi, Kenya

‖Africa Mental Health Foundation, Nairobi, Kenya

This plan outlines the design of the study, the outcomes of interest, and the econometric approach.

1 Introduction

The present project studies the effect of stress on measures of temporal discounting, self-efficacy, and executive control. We induce stress in four different ways across subjects: the Trier Social Stress Task (TSST) which induces psychosocial stress [14] , physical stress using the Cold Pressor Task [11] , financial stress using the Centipede Game [8, 16] (Rosenthal, 1981; Haushofer et al., 2017), and using hydrocortisone administration to induce the neurobiological consequences of stress [high cortisone levels, for example 4]

To test whether "acute" and "chronic" stress have different effects on these outcomes, we administer hydrocortisone or placebo for 7 consecutive days, and compare behavioral responses on the first vs. the seventh day. The stress literature has long distinguished between these two types of stress [15] , and it has recently been shown that risk aversion is affected by chronic but not acute stress [13] ; however, little is known about the relative effects on other outcomes. We hasten to add that we cannot hope to mimic the effects of truly chronic stress (over years) with this manipulation; however, it has been shown that the effects of stress differ, both behaviorally and neurobiologically, even over short periods such as several hours [12, 9, 10] or weeks [13].

2 Design

2.1 Sampling Strategy

We study a sample of Nairobi residents registered as subjects with the Busara Center for Behavioral Economics. Busara's participant pool is broadly representative of Nairobi and Kenya [7] . To be registered in the Busara subject pool, respondents must be over the ages of 18 years old, have access to a mobile phone, and have access to MPesa, a mobile money system used for payment of respondents.

The study took place between February and December 2017. Prior to the full study, we completed a number of pilot studies to perfect logistics, refine the relatively complicated protocol, and identify potential difficulties in the main experiment. Due to the resulting changes in protocol for the main study, the data from these pilot participants will not be included in the analysis described below. No treatment effects were analyzed in the pilot studies.

Participants were screened for inclusion/exclusion criteria both over the phone and in person. We restricted participation to respondents in the Busara subject pool between the ages of 18 and 40 with education greater than Standard 8. In addition, we excluded participants who did not meet the requirements of a health screening to ensure all participants could safely receive the stressors.[1]

To mitigate factors that might affect measurement of salivary cortisol, we asked participants not to drink alcohol, coffee, or smoke on the days of the study and the day before the study began. Also, we asked participants not to eat, drink any liquids other than water, or engage in strenuous physical activity, including sexual activity, during the 2 hours before the study.

2.2 Treatment and Data Collection

The treatments, tasks, and questionnaires were administered using touch screen computers and the zTree experimental interface [5] to enable computer-illiterate respondents to participate. Enumerators read instructions to the respondents in Kiswahili to maximize comprehension.[2]

2.2.1 Stressor: Hydrocortisone Administration

Hydrocortisone is a stable version of cortisol and is metabolized into cortisol upon ingestion. It is a standard approved drug used against rheumatoid and inflammatory diseases. The study was a placebo-controlled randomized controlled trial in which

[1] Detailed exclusion criteria are available upon request. For the TSST, cold pressor and centipede tasks, we excluded all participants that might be pregnant. For hydrocortisone we had a more extensive list of medical exclusion criteria.

[2] 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.

we administered either placebo or 20 mg hydrocortisone for 7 consecutive days. This dose and treatment regimen are very mild and therefore side effects were rare.

We conducted laboratory sessions with a maximum of 20 respondents per session. There are 301 respondents in the "acute condition", who were present on day 1 of the study. There are 243 respondents in the "chronic condition", who were present for day 7 of the study. The number of day 7 participants is lower than that of day 1 participants due to attrition.

Before sessions began, we randomly assigned participants to treatment and control groups. Both the laboratory administrators and subjects were unaware of their treatment status; therefore the design is double-blind. At the end of the session, we included a guessing module in which participants were asked which pill they thought they received, and indicated how confident they were that they received that pill.

Respondents received a cash compensation of KES 350 on each day. In addition, participants received KES 300 on the first day of participation, another KES 300 on the seventh day of participation, and another KES 500 if they attend all seven days of the study. For each day of the study, respondents received an additional KES 50 for arriving on time. In addition, participants could receive an additional KES 100 for correctly guessing their treatment status on the first and the last day. Finally, respondents had the opportunity to earn additional money from their choices in a series of task completed on Day 1 and Day 7, described below. The compensation and bonus were transferred to the respondents via MPesa.[3]

2.2.2 Stressor: Cold Pressor Task

We conducted laboratory sessions with a maximum of 16 respondents per session. There are 278 respondents in the "acute condition", who were present on day 1 of the study. There are 247 respondents in the "chronic condition", who were present for day 7 of the study. The number of day 7 participants is lower than that of day 1 participants due to attrition. Treatment and control groups alternated across

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

sessions.[4]

At the beginning of a session, participants were fitted with a heart rate measurement strap that was worn throughout the session. The Cold Pressor Test (Hines and Brown, 1932) consisted of immersing one's non-dominant hand in a container filled with ice water (0–4° C). Participants assigned to the control group were asked to immerse their non-dominant hand in a container filled with water heated to body temperature (37–40° C). By means of a metal partitioner, two compartments were created inside the container, one containing 6 kg of ice cubes, and another in which participants immersed their hand in water up to their wrist with outstretched fingers. A waterproof RS-2001 electrical filter pump was used to circulate the water to avoid local heat build up around the hand (Mitchell et al., 2004). Commercial-grade submersible aquarium thermometers were used to monitor and measure water temperature. Participants in both conditions submersed their hands for 60 seconds at the beginning of the session, as well as a second time for 120 seconds at the end of the session on days 1 and 7. We include a second repetition of the submersion to induce anticipatory stress and maximize the effect of the treatment (stressor) condition.

Respondents received a cash compensation of KES 350 on each day. In addition, participants received KES 300 on the first day of participation, another KES 300 on the seventh day of participation, and another KES 300 if they attended all seven days of the study. For each day of the study, respondents received an additional KES 50 for arriving on time. In addition, participants were incentivized to keep their hand in the water for the full time period, resulting in an expected payout of KES 250. [5]

[4] Specifically, the first two sessions were control, the second two sessions were treatment, the third two session were control, etc.

[5] On each day, respondents submersed their hands twice. For each instance that they kept the hand in the water for the full time period, they were entered in a lottery. At the end of the 7 days, two (if the total number of participants in the session was less than 12) or three (if the total number of participants in the session was between 12-16) of all participants were randomly chosen to receive a bonus of KES 1250 each. We varied the number of lottery winners across sessions as described above to keep the expected payout as close to KES 250 as possible.

2.2.3 Stressor: Trier Social Stress Test

We conducted laboratory sessions with a maximum of 10 respondents per session. There are 285 respondents in the "acute condition", who were present on day 1 of the study. There are 244 respondents in the "chronic condition", who were present for day 7 of the study. The number of day 7 participants is lower than that of day 1 participants due to attrition. Treatment and control groups alternated between sessions.

The TSST is designed to induce stress using two socially evaluative situations – a speech task and a mental arithmetic task. The protocol as detailed below includes slight changes to the original TSST design.[6]

Before the task began, all participants were fitted with a heart rate measurement strap that was worn throughout the session. For the treatment group, stress induction began with a panel of two judges, dressed in white, entering the room and turning on the video camera. Throughout both tasks, the judges maintained neutral expressions, remained stern, and provided little to no feedback. The speech task was a simulated job interview. The first five minutes were an anticipatory stress phase, during which participants were instructed to prepare a two-minute speech describing why they would be a good candidate for a fictitious job. Each participant then delivered their 2-minute speech without the use of their notes, immediately followed by a question-answer phase. Predesigned questions were randomly posed to the four participants in a random order, such that each participant was asked on average 4-5 questions. Participants were given 60 seconds to answer each question. The judges then exited the room for several minutes and returned for the mental arithmetic task. For this task, participants were asked to count backward from a given four-digit number (e.g. 4878, 4494, 3678) in steps of a specific number (varied

[6]Slight changes were made to the design to adapt to the Kenyan setting, and to develop a "chronic" condition in which the TSST is administered repeatedly over the course of a week. Each group consisted of five participants. The speech delivery was decreased from 5 minutes per participant to 2 minutes per participant with the addition of the question-answer phase. For the arithmetic task, instead of using the same start number and subtraction instructions, we randomized the start number for each participant and assigned different subtraction instructions per day to decrease the probability of learning effects. All protocols are available upon request.

based on day; see below) for two minutes. If a mistake was made, the participant was asked start again from the beginning. The start number and order in which the participants performed were both randomized. Throughout both tasks, participants in the treatment group had a language restriction such that they were permitted to speak in English only. If they began speaking in Swahili or another language, they were stopped and reminded of this requirement.

For the control group, the test began when a panel of two persons entered the room. Throughout both tasks, the panel maintained positive expressions, provided friendly nonverbal feedback, and created a comfortable environment. In the speech task, participants were asked to describe themselves, activities they enjoy, and their usual daily routine. The task began with a five-minute preparation phase, followed by a two-minute speech. Participants were allowed to use their notes during their speech. The speech was followed by a 20-30 minute question and answer phase, during which predesigned questions were randomly posed to the participants. They were told that if a question was difficult, they could use the 60-second answer period to talk about anything they like that would help the panel get to know him/her better. During the mental arithmetic task, the participants were asked to count forward from a particular number (e.g. 0, 5, 10) in steps of five for two minutes. The participants were neither stopped nor corrected for any mistakes. The given start number and order in which the participants perform were both randomized. Throughout both tasks, participants in the control group had the flexibility to speak in whichever language they were most comfortable, English or Swahili. They were reminded of this fact throughout the study.

Participants in both conditions completed both the speech preparation and mental arithmetic tasks at the start of the session, as well as the arithmetic task a second time at the end of the session on days 1 and 7. We included a second repetition of the arithmetic task to induce anticipatory stress and maximize the effect of the treatment (stressor) condition. At the end of each session we also include a guessing module in which we ask all participants about their rank on each task, overall, and their confidence about their rating.

Since we implemented the treatment or control protocol over seven days, we adapted the protocol across days. While participants prepared their speech on each of the seven days, only on the first or seventh day (randomized at the cohort level) were they unexpectedly asked to deliver the speech to the panel. The question-answer phase and mental arithmetic task occurred on all seven days. In addition, the starting numbers and intervals in which participants count backwards in the math task changed each day in the treatment group according to a predetermined sequence.

Respondents received a cash compensation of KES 350 on each day. In addition, participants received KES 300 on the first day of participation, another KES 300 on the seventh day of participation, and another KES 300 if they attended all seven days of the study. For each day of the study, respondents received an additional KES 50 for arriving on time. In addition, participants could receive an additional payout based on performance on the Trier Social Stress Test, resulting in an expected payout of KES 250. The recipients of this bonus were determined as follows:

In the treatment condition, each participant was given a score from 1 to 6 for their performance on each of the tasks (interview and arithmetic) for each day of participation. For any day a participant did not show up, he/she received a score of zero for both tasks for that day. Once data collection was complete, we randomly chose one day for payment, and one of the two tasks to be paid out. The two participants with the highest performance scores on the randomly chosen day and task received the bonus: the highest performer on the specific task and day received KES 1500, and the second highest performer received KES 1000.

In the control condition, each participant was given a "ticket" for each task that was completed on each day of participation. For any day a participant did not show up, he/she did not receive a ticket for either task. Once data collection was complete, we randomly chose one of the days and tasks to be paid out. Two persons whose tickets were picked for the randomly chosen task received the bonus. The first person picked for a specific day and task received KES 1500 and the second person picked received 1000 KES.

This bonus was explained to all participants at the beginning of each day, and they were reminded of it several times during the sessions.

2.2.4 Stressor: Centipede Task

We conducted laboratory sessions with a maximum of 30 respondents per session. There are 278 respondents in the "acute condition", who were present on day 1 of the study. There are 249 respondents in the "chronic condition", who were present for day 7 of the study. The number of day 7 participants is lower than that of day 1 participants due to attrition. Treatment and control groups alternated across sessions.

We used a modified real-time version of the Centipede Game first introduced by Rosenthal (1981). In our case, the game lasted for 30 rounds of 21 seconds each. In each round, a resource started at a low amount and doubled every three seconds. Player(s) were faced with a decision to "pass" or "take" the resource in each of these three-second intervals. The fact that the resource grows over time creates an incentive to pass in all three-second intervals until reaching the final interval, and to "take" in that interval. However, when the game is played with others, the player who takes first ends the round and reveives the entire resource available in that three-second interval, while the other players receive nothing. This rule creates an incentive to take early. We used two versions of the game: a 1-player (control condition) and a 4-player condition (treatment condition). Participants assigned to the 1-player condition played for themselves without partners, and their payoff thus depended only on when they themselves decided to "take" or whether they decided to wait out the round. The starting resource was KES 10, resulting in a maximum resource of KES 640. If a player did not decide to collect within the decision period of 21 seconds, they received the maximum amount of KES 640 for that round. Thus, the optimal strategy was simply to wait out each round.

Participants in the 4-player condition competed with 3 other players. The starting resource was KES 40, resulting in a maximum resource of KES 2560. Players decided simultaneously to pass or take within each three-second interval. The default decision was to pass. If all players passed, the resource remained intact. If a player took, the round ended, and that player earned the entire resource available in that three-second interval, while the other players earned nothing. If more than one player took in the same 3-second interval, those players split the resource evenly amongst themselves. If no player collected in the final round, everyone in the group split the maximum resource equally. After each round, participants were informed about their own payoff, but not that of anyone else. At the end of the study, the computer randomly chose one round as the payment for this task.

By backward induction, there exists a unique sub-game perfect equilibrium in the 4-player condition where each player takes in the first three second interval. We hypothesized that this unraveling would lead to stress. In contrast, in the one-player condition, no unraveling is predicted to take place, and as a result, stress levels should be lower. We therefore manipulated stress by randomly assigning, at the session level, some participants to the 4-player game as a stress inducer, and others to the 1-player game as the control. As described above, payoffs and endowments were carefully calibrated across the two conditions to make total payout similar in treatment and control conditions and thus mitigate income effects.

Identifying an effect of the Centipede Game on behavior, especially discounting, is potentially affected by the fact that in the "treatment" version of Centipede Game, the equilibrium strategy is to act immediately. This fact could generate a general belief that acting immediately is advantageous, and therefore spill over into the discounting task and induce participants to select the "sooner" option. We therefore generated a "reverse" version of the Centipede Game to control for this possibility. In the reverse version of the Centipede Game, each round began with a large common-pool resource that decreased over time (KES 640 for the 1-player and KES 2560 for the 4-player game). Participants assigned to the 1-player condition individually played the game and decided how much to collect. Thus, the profit-maximizing strategy for an individual in the 1-player condition was to collect the resource immediately. If the player did not collect in the final round, they received the minimum amount (1 KES). Players in the 4-player condition competed with 3 other players. The incentives were now such that it was individually optimal to take as late as possible: players who collected before the three second interval in which the last person collected received KES 0. If all players collected in the same three second interval, they split the resource among themselves. If no player collected before the time ran out, the group split the minimum amount (KES 40) evenly. In the 4-player version of this game, if each person collected immediately, the group would split the maximum resource (KES 2560) evenly. If three people took in the first interval and the fourth player took in the second three second interval, the fourth player would receive the resource of 1280, and everyone else would receive nothing; etc. The computer randomly chose one of the rounds as the payment for the game. The number of players (4 vs. 1) and direction of the game (foward vs. reverse) was randomly assigned at the session level.

Respondents in all conditions played a second series of the centipede game at the end of the session on Day 1 and 7. The second series included 15 rounds each lasting 14 seconds, and the pot of money doubled every 2 seconds. We included this repetition of the game to induce anticipatory stress and maximize the effect of the treatment (stressor) condition.

Respondents received a cash compensation of 350 KES on each day. In addition, participants received KES 300 on the first day of participation, another KES 300 on the seventh day of participation, and another KES 300 if they attend all seven days of the study. For each day of the study, respondents received an additional KES 50 for arriving on time. In addition, participants could receive an additional payout from the Centipede game, resulting in an expected payout of KES 250.[7]

2.3 Tasks

We measured behaviors relating to three constructs: self-efficacy; temporal discounting; and executive control. Each construct was measured using both behavioral tasks and self-report questionnaires. The order in which the three constructs were measured was randomized at the session level; however, within each construct, the behavioral tasks were always administered before the questionnaires.

[7] We varied endowments across conditions to ensure that the expected payout from the centipede game was KES 250. In the treatment condition, the endowment was KES 1240. In the control conditions, the endowment was KES 610. In each session, 20% respondents were randomly selected for payment, and the payout from one randomly selected round of the centipede game was paid.

2.3.1 Self-Efficacy

The concept of self-efficacy was developed by [2, 3] , who defined self-efficacy as the belief that one can perform well in specific situations [2] . Self-efficacy represents a belief about performance on a particular task. Here we used the "*slider task*" [6]. Participants were shown an on-screen "slider," a horizontal line which represents the integers from 0 to 20. They were then instructed to click the point on the line which corresponds to a randomly selected specific integer on the line (i.e. if the integer on the screen is 19, the participant must position the slider to the corresponding integer 19). The corresponding slider integer selected was then shown on the screen, and the participant could elect to move on if they correctly matched the slider or keep trying until they had made a match. An example of the participant's screen is provided in Figure 1.

Once they had selected the correct number, or elected to move on, they were presented with another randomly chosen integer to match. This "slider matching" process has the advantage of simulating effort which is purely mechanical and therefore should be not be related to age and education. After a 60-second practice round, participants proceed to a three-minute round during which they were paid KES 10 for each slider matched.

After this incentivized round, participants were asked to estimate their performance in the first round (in terms of total sliders matched), with a correct guess worth KES 50, as well as how confident they were of their estimate (unincentivized). Participants did not receive feedback on their performance to avoid changes in self-efficacy due to feedback.

Next, participants were asked to set a goal for how many sliders they wanted to match in the following two minute round, as well as their confidence level concerning that goal. Participants were informed that the task payment consisted of KES 20 times the number of sliders indicated by the goal if they achieved they goal, and zero otherwise. Thus, a participant who set a goal of x sliders and completed at least that many sliders was paid $x \times$ KES 20, even if they completed more sliders. Lastly, the final round was played and payment calculated.[8]

We consider the following outcome measures:

1. Performance in the first round of the slider task
2. Beliefs about performance in the first round of the slider task
3. Goal set for the second round of slider task
4. Compound self-efficacy measure, described below.

To compute a compound self-efficacy measure, we operationalized self-efficacy as having "high" beliefs about one's ability to complete a task, and being approximately correct about these beliefs. The rationale is as follows: first, core to the self-efficacy concept is the ability to achieve desired outcomes; hence the measure should increase in actual performance. Second, self-efficacy is distinct from overconfidence (and underconfidence): a person who has "high" beliefs about their ability, but actually has low ability, is better characterized as overconfident than has having high self-efficacy. The converse argument applies for "low" beliefs.

We define our compound measure of self-efficacy as:

$$SX = \frac{y_a}{1 + |y_a - y|}$$

Here, the goal set is denoted by y_a, and actual performance by y. Note that this measure increases in the goal set, in performance, and in accuracy about the goal.

[8] We tested for comprehension of the task by analyzing the three multiple-choice comprehension questions about the task and incentive scheme that were asked between round one and goal setting for round two. In addition, if a participant kept track of the number of sliders completed, she had no incentive to keep engaging the task after their goal is met. We will therefore use the number of participants exceeding their goal as a qualitative check on comprehension, though we avoid a formal rule as other factors (e.g. uncertainty or enjoyment in playing the game) might cause the participants to continue the task.

We also measure self-efficacy using the Pearlin Mastery Scale (PMS). The PMS is a canonical measure in self-efficacy research. It is defined by its author as measuring *mastery*, or "the extent to which one regards one's life-chances as being under one's own control in contrast to being fatalistically ruled" (Pearlin and Schooler, 1978). Participants are asked to rate seven statements about self-efficacy on a scale ranging from "strongly disagree" (0) to "strongly agree" (5). Of these, five statements are negative (e.g. "I have little control over the things that happen to me") and two positive (e.g. "what happens to me in the future mostly depends on me"), and adjusted accordingly in analysis. We use an index generated from the PMS, reverse-scoring relevant items such that they are all positive or negative and weighing each question equally, as our primary outcome measure.

2.3.2 Temporal Discounting

We used Multiple Price List (MPL) tasks in both gain and loss domains to measure temporal discounting. Participants were asked to make 48 choices between payments at earlier or later dates. The payment at the early date was always equal to KES 400, while the option at the later date changed in either ascending or descending order between eight values (KES 340, 400, 440, 600, 700, 800, 1200 and 1600). We employed three payment date combinations: today vs. two weeks from today; today vs. four weeks from today; and two weeks from today vs. four weeks from today. The loss versions of the MPL tasks were symmetric to the gains domains, with only the signs of payments reversed. The full list of decisions is presented in Table 1. Figure 2 provides examples of the participant interface for the MPL.

Participants were endowed in all four tasks with KES 1600 at both earlier and later dates, to equalize reference points across domains and time and avoid negative payment outcomes. The order of the three frames within each task was randomized at the session level. Across individuals within session, we randomly varied whether the different delayed outcomes were shown in ascending or descending order. Participants were provided with task-specific training on the interface, and encouraged to ask questions, immediately before this portion of the experiment. If the MPL money task was chosen, participants received an earlier or later payment based on one of the randomly chosen decisions made during the experiment.

A common criticism of monetary discounting tasks is that money is fungible and therefore these tasks may not capture time preferences over consumption in the presence of functioning credit markets (Augenblick et al., 2015). We therefore also implemented a task of choices over time-dated effort. In a framework similar to the MPL for money, participants had to choose between an earlier and later amount of effort, in the form of a specific number of phone calls to the Busara Center in 10 minute intervals at particular hours in the evening. Participants could choose to make two phone calls on the earlier date, or a number of between 1 and 6 at the later date, depending on the decision. Respondents were told they would be paid a fixed amount of KES 500 one month from the date of the session, conditional on completion of the task. Figure 3 provides an example of the participant interface for the effort discounting task.

Standard MPL parameter estimates assume linear utility. To allow for concave utility, we also measured risk preferences. We adopted the theoretical framework and experimental design from Tanaka et al. (2016). Participants made 27 decisions between two 50/50 lotteries, represented by two balls each in two jars. The first series of 10 lotteries included only gains and participants received an endowment of 0 KSH. The first series was followed by 10 lotteries of only losses, in which the endowment was 1000 KSH. Finally, the third series had an endowment of 350 KSH and included 7 lotteries with mixed gains and losses in each lottery. If the task was chosen for payment, one lottery was chosen at random at the session level for payment. A list of the endowments, lottery decisions, and expected payouts can be found in Table 2.

For both the MPL and the effort task, we estimate β and δ in the quasi-hyperbolic discounting model (Laibson, 1997). As a robustness check, we allow for curvature in the utility function, using CRRA utility and decisions in the risk task to obtain a curvature parameter.[9]

[9] We can use multiple switching to check consistency of answers. Will we calculate the proportion of multiple switches per lottery series, and use the simulated samples technique illustrated in Balakrishnan et al. (2015) to determine what the corresponding percentage would be if participants answered randomly. As is evident from the exposition of multiple price lists estimation,

We define an individual i as exhibiting **present bias** if $\beta_i < 1$. Intuitively, this means that they are more willing to discount consumption (or at least income) over the time period starting today than any other day in the future.

We also measure time discounting using the Consideration of Future Consequences (CFC) scale [19]. Participants were asked to indicate how much the behavior described in a statement is characteristic of them, from "not at all like me" (0) to "very much like me" (5). There are nine statements representative of forward thinking (e.g. "I am ready to sacrifice my current happiness or wellbeing in order to achieve future results") and five reverse statements (e.g. "I only act to satisfy immediate needs, thinking the future will take care of itself"), which are scored accordingly. We use an index generated from the CFC, reverse-scoring relevant items such that they are all positive or negative and weighing each question equally, as our primary outcome measure.

2.3.3 Executive Control

Executive function refers to "a set of inter-related higher-order cognitive abilities involved in self-regulatory functions" [17] such as insight, judgement, working memory, or planning [18, 1, 20]). We measured executive function using a 3 minute spatial version of the Stroop task, using congruent and incongruent directional signals (arrows) rather than words [21] . On each screen, participants saw a colored arrow that pointed either left or right, and responded by pressing a box on the left or right side of the screen. Importantly, when the arrow was red, participants were required to select the side of the screen towards which it pointed ("congruent" trials); if the arrow was blue, they were required to select the opposite side of the screen ("incongruent" trials). The sequence of arrows was randomized. Participants earned KES 25 for each correct response, but lost KES 3 for every second they took to complete

respondents' choices can be rationalized with numeric parameters (and indeed any monotonic utility function) if and only if they obey certain axioms. The clearest of these is the "law of demand" (Balakrishnan et al., 2015). We define **basic consistency** according to the following rule: as the implied interest rate rises, the chance that the later payment is chosen does not decrease, and vice versa. We will calculate the proportion of consistent responses for each possible combination of interest rates within each frame.

the task (although the total payment for this task could not go below zero). We recorded correct and incorrect responses and reaction times by trial type.

We expect the results from the spatial Stroop task to exhibit the “Stroop effect:” participants should take longer, on average, to select the correct direction for incongruent stimuli (MacLeod, 1991), and/or make more mistakes (Wühr, 2007). Since our task is time-incentivized, participants may attempt to answer the incongruent items equally quickly and therefore sacrifice accuracy. We define overall performance on the Stroop task as the ratio of number of correct responses to total time in seconds. Significantly higher response time and lower probability of correct response to incongruent stimuli will be interpreted as evidence of a Stroop Effect. [10]

We also measured executive control using a subset of the Behavior Rating Inventory of Executive Function - Adult Version (BRIEF-A). The BRIEF-A questionnaire is composed of 75 items which are grouped into two indices: behavioral regulation and metacognition. Within the indices, the items are further grouped into a total of nine subscales. We use four of the subscales within the metacognition index: (1) ability to initiate, (2) working memory, (3) ability to plan and organize, and (4) ability to monitor oneself in tasks. Together, these subscales have 32 items, but we added an additional two items with repeated content and reversed wording, for a total of 36 items.

The items consist of questions such as “I have trouble starting anything on my own” or “I don’t plan early for future activities.” The response choices range from “never” (scored as 0) to “always” (6). We reverse-score some items to make all items “positive”. Our primary outcome measure is an index combining all questions. In addition, we will also analyze indices for each subscale.

2.4 Schedule of Tasks and Treatments

Days 1 and 7:

[10] We will also access comprehension of the task using seven factual multiple-choice questions about the rules and payment calculation that participants were asked after the instructions and practice round but prior to the beginning of the task. We report the proportion of questions answered correctly.

1. Welcome
2. Consent and Pregnancy Screening (Day 1 only)
3. Nurses' Meetings to Determine Participant Eligibility (Day 1 of Hydrocortisone protcol only)
4. Heart Rate Measurement
5. Introduction to Computer Interface
6. PANAS
7. Salivette 1: baseline
8. Self-Reported Stress Question 1
9. Instructions for Time Discounting and Executive Control Tasks (occured after administration of Treatment/Control for only the Hydrocortisoneprotocol)
10. Administration of Treatment/Control
11. Salivette 2
12. Self-Reported Stress Question 2
13. *Tasks*
14. Salivette 3
15. Self-Reported Stress Question 3
16. *Tasks*
17. Salivette 4
18. Self-Reported Stress Question 4
19. *Tasks*
20. Salivette 5
21. Self-Reported Stress Question 5
22. *Tasks*

23. Salivette 6
24. Self-Reported Stress Question 6
25. *Tasks*
26. Salivette 7
27. Self-Reported Stress Question 7
28. General Questionnaire
29. Guessing Module (Hydrocortisone and TSST protocols only)
30. Final Payment Screen (payment is only shown on Day 1 if the task chosen is time discounting and the choice chosen for payment applies to the current date; otherwise shown for all participants on Day 7)
31. Administration of Treatment/Control (for Cold Pressor, TSST, and Centipede protocols only)
32. Debrief and Heart Rate Measurement

Task order was randomized by construct (self-efficacy, time discounting, executive control), and so they are listed here simply as *Tasks*. All behavioral tasks preceeded the self-report surveys. The order of the MPL, risk aversion, and effort discounting tasks were randomized within the time discounting constract. Saliva samples were collected as close as possible to 25, 50, 75... minutes after administration of the pill, but such that no tasks were interrupted

Days 2-6

1. Welcome
2. Consent
3. Heart Rate and Blood Pressure Measurement (Hydrocortisone protocol only)
4. Questionnaire: Chronic Life Stressors (Day 2 only)
5. Administration of Treatment/Control
6. PANAS

7. (Shortened) General Questionnaire
8. Guessing Module (Hydrocortisone and TSST protocols only)
9. Heart Rate and Blood Pressure Measurement (Hydrocortisone protocol only)
10. Debrief

3 Econometric Approach

3.1 Main Treatment Effects

We use the following main specification:

$$y_{it} = \beta_0 + \beta_1 T_i + \beta_2 Day7 + \beta_3 T_i \cdot Day7 + \varepsilon_{it} \tag{1}$$

Here, y_i is the outcome of interest for respondent i at time t. T_i is a treatment indicator that takes value 1 for respondents that received treatment (stressor) and 0 for respondents that received the control protocol. $Day7$ indicates whether the observation is from day 7 of the study (as opposed to Day 1). Thus, β_1 estimates the treatment effect of stress on Day 1, β_2 measures the day fixed effect, β_3 estimates whether stress has a different effect on day 7 (chronic) than on day 1 (acute), and $\beta_1 + \beta_3$ measures the treatment effect on Day 7. We will also report the joint significance of the treatment effects on Day 1 and Day 7.

3.2 Other analyses

1. **Balance, integrity, and manipulation check**

    (a) **Randomization Check:** To determine whether randomization was successful, we estimate our main equation with the following demographics as outcome variables: age, gender, years of education, marital status, average weekly spending, average monthly earnings, earnings after spending per month, BMI, number of siblings and dependents, whether the respondent fully depends on someone else financially, current unemployment, and whether currently in debt. We will also estimate this equation using baseline levels on day 1 of cortisol and self-reported stress as outcome variables.

(b) **Selection into Chronic Treatment:** To test for differential attrition by treatment, we regress whether the participant was present on day 7 on treatment assignment, using equation 2 below.

$$attrit_i = \beta_0 + \beta_1 T_i + \varepsilon_i \tag{2}$$

In addition, equation 3 assesses whether attriting individuals are different in terms of the demographics (y_i) described above.

$$y_i = \beta_0 + \beta_1 attrit_i + \varepsilon_i \tag{3}$$

We will also estimate this equation using baseline levels of cortisol and self-reported stress on day 1 as outcome variables. Finally, equation 4 measures whether the baseline characteristics of attriting individuals in the treatment group are significantly different from those in the placebo group. The sample for this regression will be restricted to individuals that were not present on Day 7, where X_i represents a vector of controls collected on Day 1, as described in the randomization check.

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

(c) **Manipulation Check:** We examine whether hydrocortisone administration affects psychological and neurobiological measures of stress, using the following measures:

i. **Cortisol:** We measure salivary cortisol at seven points during the sessions on day 1 and during those on day 7. The first salivette is taken before administration of the pills. The second through seventh salivettes samples are taken at approximately 25 minute intervals after the pills are administered; in practice the samples are not taken exactly at 25 intervals since we require that respondents provide a salivette sample *between* tasks or questionnaires. Cortisol samples are analyzed by Lancet Labs in Nairobi. We compute the area under the curve before analysis, such that the seven measurements provided by each participant in each session are collapsed to one number. Because cortisol levels are noisy, we winsorize the cortisol variables at the 95th percentile. In addition, cortisol levels are affected by time of day, smoking, consumption of food and drink, and exercise. We attempt to mitigate the influence of these factors by asking participants not to smoke, not to drink alcohol, tea, or coffee; and we ask participants not to eat or drink, nor to performance intense exercise within 2 hours of the sessions. However, participants may not always follow these rules; we therefore survey them on whether they did these activities, and include control variables for these activities on the right-hand side.

ii. **Self-Reported Stress:** On 7 occasions throughout the experiment, we asked participants to indicate (on a scale of 0-100) "In the present moment, I feel stressed." The question was asked after each salivette. As described under (a), we will compute the area under the curve for this outcome measure. We will not include control variables or top-code because this self-report variable does not have the same outlier concerns and confounds as cortisol.

iii. **Heart Rate:** Throughout the sessions for cold pressor task, centipede task, and Trier Social Stress Task, participants wore heart rate monitors that measured their heart rates in regular intervals. We winsorize and analyze area under the curve as described above for cortisol.

iv. **Beliefs about treatment:** In a guessing module for the hydrocortisone stressor, participants were asked which pill they thought they received, and indicated how confident they were that they received that pill. Similarly, in a guessing module for the TSST study, participants were asked about their performance. The outcome of interest is whether the respondent guessed correctly. To assess whether participants could correctly guess which pill they received (or their rank on the task), we construct a variable for correctly guessing, regress it on the treatment, and test whether the coefficient differs from chance performance.

3.3 Dimensions of heterogeneity

We will estimate heterogeneous treatment effects using versions of our main estimating equation in which we add a main effect and interaction terms for a binary interactant of interest. We consider the following dimensions of heterogeneity:

1. Gender

2. Above median baseline affect: To measure positive and negative general affect states, we generated a Kiswahili version of the PANAS (Positive and Negative Affect Schedule). On a scale of 0-100, respondents indicated how they felt in the present moment for each of 20 emotions. We asked this series of the questions before administration of hydrocortisone on days 1 and 7 of the experiment; we also asked this series of questions after treatment on days 2 through 6 respectively. We will use the score on day 1 of the experiment, before the administration of the treatment, to examine heterogeneity based on baseline affect.

3. Above median baseline stress level: To measure baseline stress level, we use respondent's self-reported stress (as described above) that is elicited before administration of the stressor on Day 1.

4. Above median baseline cortisol level: To measure baseline cortisol level, we use respondent's salivary cortisol (as described above) that is elicited before administration of the stressor on Day 1.

5. Above median baseline heart rate: To measure baseline heart rate level, we use respondent's heart rate, measured using heart rate monitors for the centipede, cold pressor and TSST protocols, and using blood pressure machines for the hydrocortisone protocol, that is elicited before administration of the stressor on Day 1.

6. Above median baseline chronic stress: We ask a series of questions on Day 2 of the study that will generate our measures of chronic stress.

   (a) (Modified) Holmes-Rahe stress scale (Holmes and Rahe, 1967): We modified the Holmes-Rahe survey for the Kenyan context by dropping irrelevant questions and adding context appropriate stressors. We will use two measures: (1) the Holmes and Rahe scoring for life events that are comparable across studies and (2) an index of all questions we ask, weighting each component equally.

   (b) Cohen's Perceived Stress Scale (Cohen et al., 1988): We will code variables such that they are all positive or negative and generate an index, weighting each question equally.

   (c) Self-reported Life Difficulty: The response to the question "Overall in the last 12 months, how hard would you rate your life on a scale from 1 (not at all hard) to 5 (incredibly hard)?" (not at all hard =1,...incredibly hard=5)

   (d) Self-reported Life Statisfaction: The response to the question "All things considered, how satisfied are you with your life as a whole these days?" (dissatisfied=1,...satisfied=5)

   (e) Negative Life Experiences Survey: The number of the events that that participant has faced: death of a parent, death of a close friend or immediate family member other than parents, ever experienced a serious illness/accident, ever experienced a serious illness/accident of a member of close family, ever experienced a situation of physical or emotional violence. For each, we also ask how long ago the event occured, how the event impacted the participant, and how much the participant confided in others about this experience at the time. We will generate an index based on the standardized sum of major life events experienced in the past 5-6, and 15+ years.

(f) (Subset of) Multidimensional Local of Control (Levenson, 1981): We will code variables such that they are all positive or negative and generate an index, weighting each question equally.

References

[1] Alan Baddeley, Sergio Della Sala, TW Robbins, and A Baddeley. Working memory and executive control [and discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 351(1346):1397–1404, 1996. 2.3.3

[2] Albert Bandura. Self-efficacy mechanism in human agency. *American psychologist*, 37(2):122, 1982. 2.3.1

[3] Albert Bandura. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice Hall, 1986. 2.3.1

[4] Sandra Cornelisse, Van Ast, Vanessa, Johannes Haushofer, Maayke Seinstra, and Marian Joels. Time-Dependent Effect of Hydrocortisone Administration on Intertemporal Choice. SSRN Scholarly Paper ID 2294189, Social Science Research Network, Rochester, NY, July 2013. 1

[5] Urs Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178, 2007. 2.2

[6] David Gill and Victoria Prowse. A structural analysis of disappointment aversion in a real effort competition. *The American economic review*, 102(1):469–503, 2012. 2.3.1

[7] Johannes Haushofer, Marie Collins, De Giusti, Giovanna, Joseph Muiruri Njoroge, Amos Odero, Cynthia Onyango, James Vancel, Chaning Jang, Maneesh Varghese Kuruvilla, and Conor Hughes. A Methodology for Laboratory Experiments in Developing Countries: Examples from the Busara Center. SSRN Scholarly Paper ID 2155217, Social Science Research Network, Rochester, NY, February 2014. 2.1

[8] Johannes Haushofer, Chaning Jang, and John Lynham. Stress and temporal discounting: Do domains matter? Working Paper, 2015. 1

[9] Marloes JAG Henckens, Guido A van Wingen, Marian Joëls, and Guillén Fernández. Time-dependent effects of corticosteroids on human amygdala processing. *Journal of Neuroscience*, 30(38):12725–12732, 2010. 1

[10] Marloes JAG Henckens, Guido A van Wingen, Marian Joëls, and Guillén Fernández. Time-dependent corticosteroid modulation of prefrontal working memory processing. *Proceedings of the National Academy of Sciences*, 108(14):5801–5806, 2011. 1

[11] Edgar A Hines and George E Brown. The cold pressor test for measuring the reactibility of the blood pressure: data concerning 571 normal and hypertensive subjects. *American heart journal*, 11(1):1–9, 1936. 1

[12] Marian Joëls, Guillen Fernandez, and Benno Roozendaal. Stress and emotional memory: a matter of timing. *Trends in cognitive sciences*, 15(6):280–288, 2011. 1

[13] Narayanan Kandasamy, Ben Hardy, Lionel Page, Markus Schaffner, Johann Graggaber, Andrew S. Powlson, Paul C. Fletcher, Mark Gurnell, and John Coates. Cortisol shifts financial risk preferences. *Proceedings of the National Academy of Sciences*, 111(9):3608–3613, 2014. 1

[14] Clemens Kirschbaum, Karl-Martin Pirke, and Dirk H Hellhammer. The trier social stress test–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2):76–81, 1993. 1

[15] Bruce S McEwen and Robert M Sapolsky. Stress and cognitive function. *Current Opinion in Neurobiology*, 5(2):205–216, April 1995. 1

[16] Robert W Rosenthal. Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic theory*, 25(1):92–100, 1981. 1

[17] Robert M Roth, Charles E Lance, Peter K Isquith, Adina S Fischer, and Peter R Giancola. Confirmatory factor analysis of the behavior rating inventory of executive function-adult version in healthy adults and application to attention-deficit/hyperactivity disorder. *Archives of clinical neuropsychology*, 28(5):425–434, 2013. 2.3.3

[18] Donald R Royall, Edward C Lauterbach, Jeffrey L Cummings, Allison Reeve, Teresa A Rummans, Daniel I Kaufer, W Curt LaFrance, Jr, and C Edward Coffey. Executive control function: a review of its promise and challenges for clinical research. a report from the committee on research of the american neuropsychiatric association. *The Journal of neuropsychiatry and clinical neurosciences*, 14(4):377–405, 2002. 2.3.3

[19] Alan Strathman, Faith Gleicher, David S Boninger, and C Scott Edwards. The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of personality and social psychology*, 66(4):742, 1994. 2.3.2

[20] Dimitri Van der Linden, Michael Frese, and Theo F Meijman. Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, 113(1):45–65, 2003. 2.3.3

[21] Peter Wühr. A stroop effect for spatial orientation. *The Journal of general psychology*, 134(3):285–294, 2007. 2.3.3

Figure 1: Self-Efficacy Example Screen

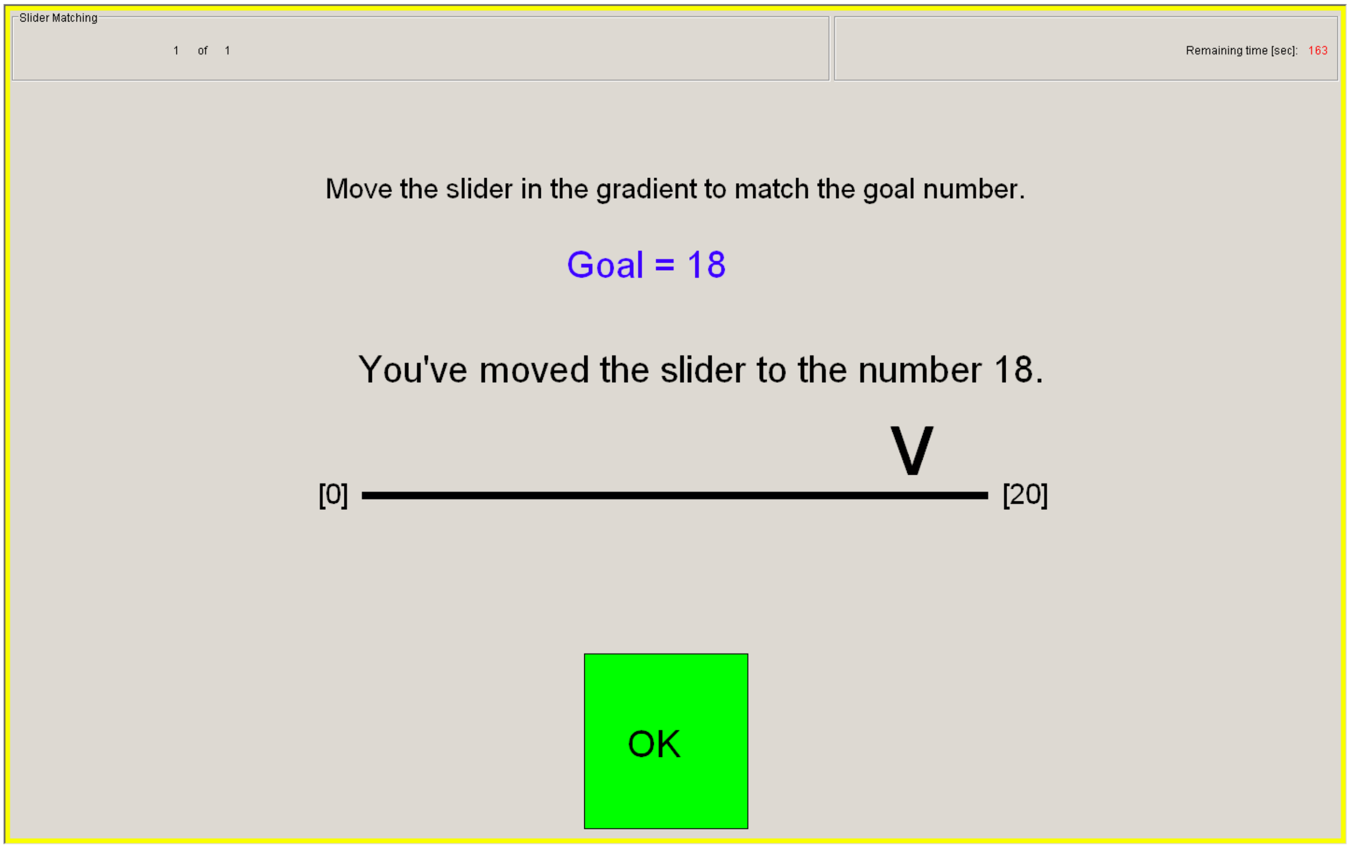

Table 1: Temporal Discounting Decisions

Front-end delay (t)	Delay between payments (k)	Early (m)	Maximum Late ($m(1+r)$)	Implied interest rate ($1+r$)
Frame 1				
0	14	400	340	.85
0	14	400	400	1
0	14	400	440	1.1
0	14	400	500	1.25
0	14	400	600	1.5
0	14	400	800	2
0	14	400	1200	3
0	14	400	1600	4
Frame 2				
0	28	400	340	.85
0	28	400	400	1
0	28	400	440	1.1
0	28	400	500	1.25
0	28	400	600	1.5
0	28	400	800	2
0	28	400	1200	3
0	28	400	1600	4
Frame 3				
14	14	400	340	.85
14	14	400	400	1
14	14	400	440	1.1
14	14	400	500	1.25
14	14	400	600	1.5
14	14	400	800	2
14	14	400	1200	3
14	14	400	1600	4

Table 2: Risk Preferences Decisions

Lottery A				Lottery B		
Red Ball	Blue Ball	Expected Value A	Green Ball	Orange Ball	Expected Value B	$E[V_a] - E[V_b]$
Series 1: Endowment 0						
300	400	350	620	50	335	15
300	400	350	650	50	350	0
300	400	350	680	50	365	-15
300	400	350	720	50	385	-35
300	400	350	760	50	405	-55
300	400	350	830	50	440	-90
300	400	350	900	50	475	-125
300	400	350	1000	50	525	-175
300	400	350	1100	50	575	-225
300	400	-50	1300	-50	625	-675
Series 2: Endowment 1000						
-300	-400	-350	-540	-50	-295	-55
-300	-400	-350	-560	-50	-305	-45
-300	-400	-350	-600	-50	-325	-25
-300	-400	-350	-620	-50	-335	-15
-300	-400	-350	-650	-50	-350	0
-300	-400	-350	-680	-50	-365	15
-300	-400	-350	-720	-50	-385	35
-300	-400	-350	-760	-50	-405	55
-300	-400	-350	-830	-50	-440	90
-300	-400	-170	300	210	255	-425
Series 3: Endowment 350						
250	-40	105	300	210	255	-150
40	-40	0	300	210	255	-255
10	-40	45	300	-160	70	-25
10	80	-35	300	-160	70	-105
10	-80	-35	300	-140	80	-115
10	-80	5	300	-100	100	-95

Figure 2: Multiple Price List Example Screen

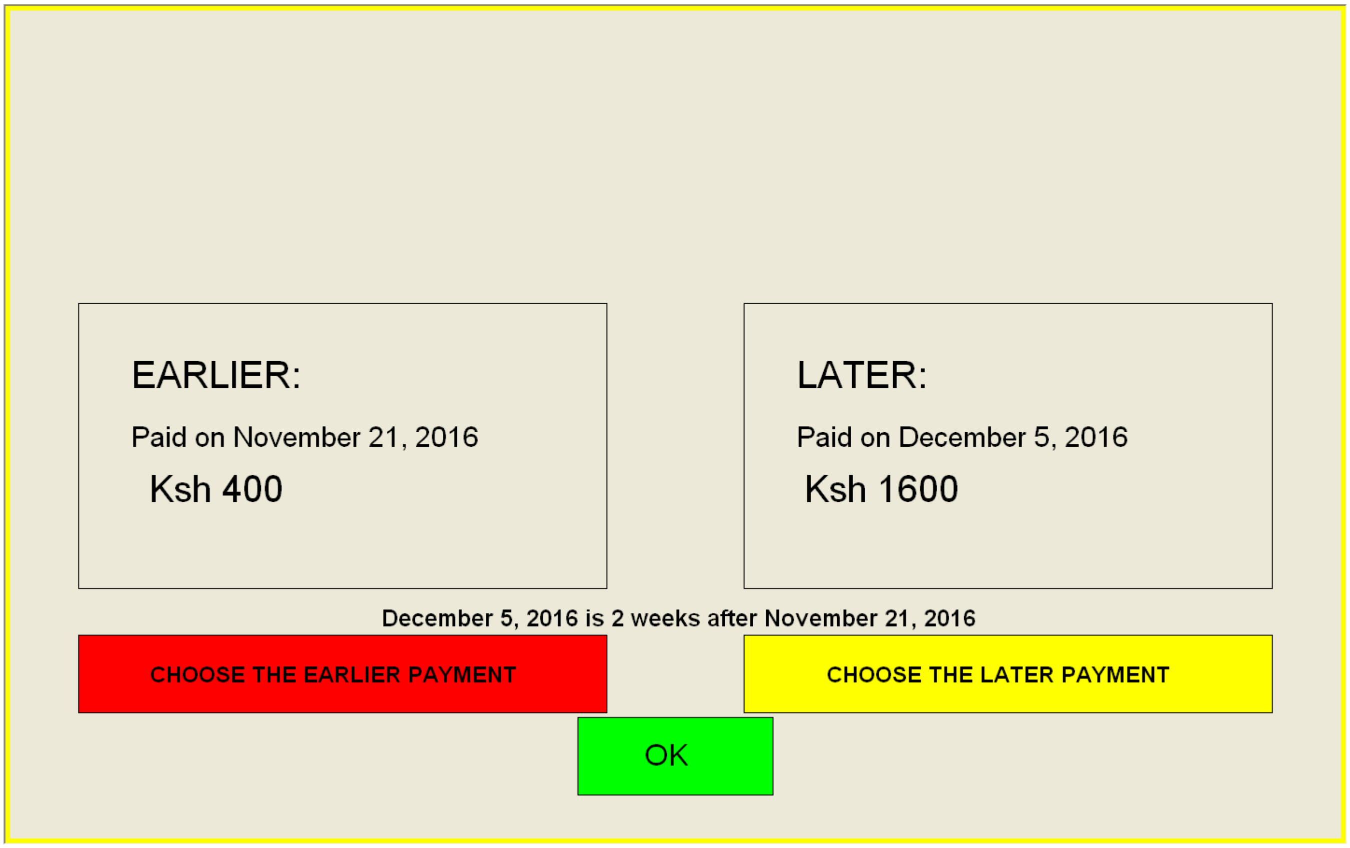

Figure 3: Effort Discounting Example Screen