Pre-Analysis Plan: Manipulation-Proof Machine Learning

Daniel Björkegren*
Joshua Blumenstock†

August 30, 2019

Abstract

An increasing number of decisions are guided by machine learning algorithms. An individuals behavior is typically used as input to an estimator that determines future decisions. But when an estimator is used to allocate resources, individuals may strategically alter their behavior to achieve a desired outcome. This paper develops a new class of estimators that are stable under manipulation, even when the decision rule is fully transparent. We explicitly model the costs of manipulating different behaviors, and identify decision rules that are stable in equilibrium. Through a large field experiment in Kenya, we test decision rules estimated with our strategy-robust method.

*Brown University danbjork@brown.edu.
†University of California Berkeley, jblumenstock@berkeley.edu.

Contents

1 Introduction . . . **3**
- 1.1 Motivation . . . 3
- 1.2 Prior Work . . . 3
- 1.3 Research Questions . . . 4

2 Study Design . . . **4**
- 2.1 Study population and recruitment . . . 4
- 2.2 Study treatments and interventions . . . 5
  - 2.2.1 Simple and Complex Challenges . . . 5
  - 2.2.2 Assignment to Treatment . . . 7
  - 2.2.3 Attrition from the Sample . . . 7

3 Data . . . **8**
- 3.1 Overview . . . 8
- 3.2 Sensing Data . . . 9
- 3.3 Survey data "response variables" . . . 12

4 Empirical Analysis . . . **14**
- 4.1 Estimating individual characteristics from Sensing data . . . 14
- 4.2 Estimating the cost of manipulation ("Simple challenges") . . . 15
- 4.3 Estimating Robust Models . . . 16
  - 4.3.1 Heterogeneous Effects . . . 16
- 4.4 Evaluating Performance . . . 17
  - 4.4.1 The Impact of Transparency . . . 17

5 Additional Research Outputs . . . **17**
- 5.1 Cost Measure Comparisons . . . 17
- 5.2 Mental Health Sensing . . . 18

6 Study Timeline . . . **19**
- 6.1 Consistency Checks . . . 23

7 Research Team . . . **23**

A1 Survey Instrument . . . **26**

1 Introduction

1.1 Motivation

Many decisions that once were made by humans are now made automatically, using algorithms trained on rich data on human behavior. Banks make lending decisions by mining the credit histories of would-be borrowers; cable and airline companies target promotions to customers who are predicted to be of high value; more recently, judicial determinations of sentencing and bail are informed by supervised learning algorithms (Kleinberg et al., 2018).

However, once people understand that their behavior informs these decisions, they may intentionally manipulate their behavior. Countless websites offer advice on how to game the airlines and hack your credit score. Search engine optimization is a $65b industry devoted to manipulating how web sites appear to search engine ranking algorithms (Borrell Associates, 2016). But manipulation is a blind spot for standard machine learning methods. With few exceptions, these methods assume that behavior is fixed, and do not consider that it may be manipulated once it is used to guide decisions.

The goal of this project is to develop a new framework for automated decision-making in settings where individuals act strategically and may manipulate their observable behavior to game the algorithm. The project has both theoretical and experimental components, as well as a concrete real-world application.

1.2 Prior Work

Manipulation is not a new topic to economists. For example, the public finance literature carefully considers how agents adjust, or distort, taxed margins of behavior. The field is well aware that relationships in observational data may cease to hold if acted upon (Lucas, 1976). The subfield of mechanism design illuminates the conditions under which individuals will strategically reveal private information, and contract theory studies the principal agent problem, where a principal sets up contracts to incentivize an agent to take particular actions Hlmstrom (cf. 1979). But modern decision rules differ from these familiar contexts. They are not just more prevalent; they also tend to be more complex, and more opaque. While economists have primarily considered simple allocation rules often based on a single variable, the availability of data and improved computation enables modern models to consume higher dimensional information, in more complex ways. (For example, mobile phone based credit scoring may select from over 5000 variables and include nonlinearities (Bjorkegren and Grissen, 2015).)

The computer science literature has also explored the potential for manipulation in supervised learning settings (Hardt et al., 2016; Hu et al., 2019; Milli et al., 2019; Kleinberg and Raghavan, 2019). The most common practical approach to dealing with manipulable data is to use opaque, backwards looking machine learning methods, and retrain the model as agents learn to manipulate the algorithm. However, this approach exposes the system to large risks: when manipulated, the system may make poor, costly decisions (for example, giving out a large batch of loans to individuals who will never repay). Some theory considers manipulation more explicitly (Bruckner et al., 2012), but does not include an empirically grounded model of behavior, and thus provides limited guidance in practical settings.

1.3 Research Questions

The main research questions we seek to address are

1. Can subject characteristics be accurately predicted from smartphone data?
2. Do consumers manipulate behavior in response to algorithmic decision rules?
3. Can we empirically test for the presence of manipulation?
4. What are the costs of manipulating different behaviors?
5. Can we estimate decision rules that are stable, even with full transparency?
6. What is the equilibrium cost of transparency?

2 Study Design

2.1 Study population and recruitment

The subject population consists of Kenyans aged 18 years or older who own a smartphone and are able to travel to the Busara center in Nairobi. We anticipate recruiting roughly 1,200 subjects for the main study, primarily through in-person solicitation in public spaces (e.g., public markets) by Busara research staff. To deal with attrition, prior to the final phase of study we will add a 'top up' group of participants to replace participants who have attrited. To ensure that this top up sample is drawn from the same population as the main subjects, we will randomize a master list of potential participants, save every third individual for the top up group, and recruit the remainder for the main group.

Informed consent is requested from each subject, following the procedures registered with the Committee for the Protection of Human Subjects at U.C. Berkeley. Subjects must actively decide to participate in the research; no pressure or undue influence will be given to induce subjects to participate. Subjects will be compensated up to 400 Ksh. each week for participating in the study.

A baseline survey will be conducted with all consenting participants in the Busara offices in Nairobi. At the time of enrollment, participants will install a 'Sensing' app on their smartphone. More details on the app and the baseline survey are provided below.

2.2 Study treatments and interventions

While enrolled in the study, each participant will be dynamically assigned to a treatment for each week. Users will be sent both a text message (SMS) and an app push-notification on a weekly basis that directs them to the app. After a user opens the app, it will ask them to opt in to a 'challenge. If they accept, they will observe a challenge which is a description of an algorithm that users may have an incentive to game — see Figure 1 for an example. These descriptions in most cases will be basic text, but in some cases will allow basic interaction, described below. Participants are paid for participation, plus any challenge payments, on a weekly basis.

Baseline weeks (**B**) will not include additional challenges, and can be thought of as control weeks. Users will receive a message such as:

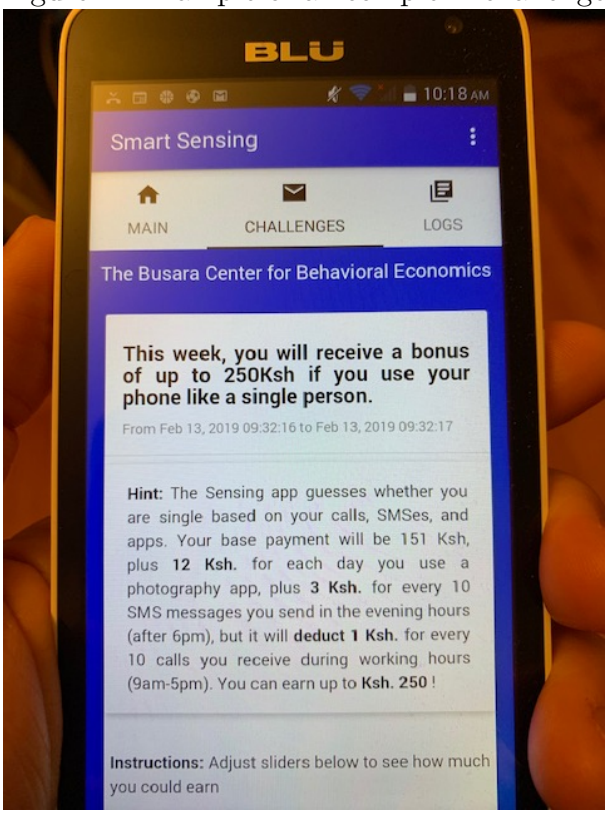
Dear user, you do not have to do anything for this week's challenge. You will receive an extra Ksh 50 for accepting this challenge. This will be paid next week Wednesday with the rest of your payout if you upload data in the upload window.

2.2.1 Simple and Complex Challenges

Simple challenges (**S**k, for $k \in \{1, ...K\}$, where K is the number of features assessed) will include challenges of the form:

- *You will receive 35 Ksh for each day you use a photography app this week, up to Ksh. 250!*
- *You will receive 3 Ksh for each text message you send in the evening hours (after 6pm) this week, up to Ksh. 250!*
- *You will receive 10 Ksh for each call you receive during working hours (9am-5pm) this week, up to Ksh. 250!*

Figure 1: Example of a "complex" challenge



- *You will receive a base payment of 100 Ksh but from that we will **deduct** 3 Ksh for each call you receive during working hours (9am-5pm) this week.*

Complex challenges will describe the outcome which is being guessed: *This week, you will receive a bonus of up to 250Ksh if you use your phone like a single person.*

These challenges may use either a naive (N) or robust (R) decision rule. These vary based on the level of transparency:.

- Opaque complex challenges (naive: **CON**j and robust: **COR**j, where $j \in \{1, ..., J\}$ is the number of outcomes assessed) will provide no more information. Since the decision rule is opaque, naive and robust decision rules are observably equivalent to participants.

- Transparent complex challenges (naive: **CTN**j and robust: **CTR**j) will also provide a hint that reveals the formula. For example: *The Sensing app guesses whether you are single based on your calls, SMSes, and apps. Your base payment will be 151 Ksh, plus 12 Ksh. for each day you use a photography app, plus 3 Ksh. for every 10 SMS messages you send in the evening hours (after 6pm), but we will deduct 1 Ksh. for every 10 calls you receive during working hours (9am-5pm). You can earn up to Ksh. 250!*

2.2.2 Assignment to Treatment

Assignment to treatment will be random, but subject to the following restrictions:

- A participant i may not be assigned to the same simple challenge more than once

- A participant i may not be assigned to an opaque complex challenge after being assigned to a transparent challenge for the same outcome.

2.2.3 Attrition from the Sample

Attrition in the context of this study has two dimensions: first, there may be participants who do not regularly upload data through their app, and second, there may be participants who do not participate in the assigned weekly challenges. (As some participants may upload data sparsely throughout the week, only those who upload within the 21-hour window at the end of the challenge-week [between 1pm Tuesday and 10am Wednesday] will be counted as having fully uploaded all of their weekly data.)

In order to minimize both such types of attrition, participants will be sent regular reminders via SMS to encourage engagement. Every participant in the study will be sent an SMS every Tuesday at 1pm to remind them to upload their data through the Smart Sensing app, and any participant who has not opted in to a challenge by Thursday at 9am (the day after challenge notifications have been sent) will be sent an additional SMS reminder about the pending challenge.

Additionally, on Wednesday and Thursday, participants who still have not uploaded data or activated their challenge respectively will be contacted by phone and surveyed by the Busara team. Specifically, the protocol is as follows:

- On Wednesday, participants who have not uploaded any data during the five day period ending on Wednesday at 12pm will be contacted and surveyed, as will those who uploaded some data in this period but not during the end-of-week upload window (between 6pm Tuesday and 10am Wednesday)

- On Thursday, participants whose phones show that they did not receive a challenge by Thursday 12pm will be contacted and surveyed, as will participants whose phones show that they did receive a challenge but who have not opted in to accept the challenge.

For all of the above categories, any participant who does not answer a survey call on the first attempt will then be re-contacted once more by the surveyor after the rest of the calls are complete.

Finally, to mitigate the effects of attrition during the analysis stage, any participant-weeks wherein the participant did not opt in and/or did not upload during the end-of-week upload window will be dropped from the sample prior to all analysis. During baseline weeks, a single passive challenge is assigned to all participants, offering a flat bonus to upload data within the upload window; in this way, we ensure that our analysis control groups will also be restricted to those who opt in to this passive challenge, and are thus a valid comparison group to the restricted panel during the challenge weeks.

3 Data

3.1 Overview

We collect data from two primary sources:

- *The Sensing App*: Participants will install an app on their smartphone, which we have developed in partnership with the Busara Center for Behavioral Economics. The *Sensing App* collects data about how each participant uses his or her phone (including call/SMS history, battery status, location, and other data accessible through the Android operating system). Additional details on this app are provided below.

- *Baseline Survey:* During enrollment in the study, participants will fill out a survey electronically which asks a variety of questions about demographics and technology usage. The survey instrument is included as Appendix A1.

3.2 Sensing Data

Individual raw data from the app will be categorized into 8 different 'probes', according to which aspect of phone usage the data describes: an app usage probe, a software information probe, a location probe, a battery probe, a call probe, an SMS probe, a Wifi probe, and a screen probe. These probes will then each be processed into a dataset of participant-week level aggregate statistics, according to the respective processing procedures described below:

- **Software Information Probe**

  The software information probe includes all information on app installations on the phone. This data is collected first in a large-sweep survey of the phone upon initial installation, and is then updated with new instances of raw data as users install or uninstall apps on their phone throughout the survey process. These instances are categorized by user ID, time, and app 'package name' (a descriptor of the app name).

  These data are first processed using a web-scraping Python script that matches these app 'package name' variables to app listings on the Google Play website, to build a dictionary mapping package names to app titles and listed app categories (e.g., 'Social', 'Photography'). For those apps that are not listed on the Google Play website, this step is skipped, and the app is added to the dictionary with the package name standing in for its title and with no category listed.

  After this processing, installation statistics are then computed for each participant-week and each app listed: a binary variable for whether an app is newly installed, a binary variable whether an app is installed (filled forward

from data on previous installations, unless superseded by a subsequent observation), a composite variable counting how many new installs have been detected, and a variable tracking how many instances of raw data are in the probe for a given participant-week.

- **App Usage Probe**

  The app usage probe includes all instances of app usage on a phone, as well as information on the duration of use for phones that support such data functionality. (Phones with Android version 5 or higher accurately capture duration of app use; phones with Android version lower than 5 do not.) These instances are categorized by user ID, time, and app package name.

  For each app in the package name-to-app dictionary created during the Software Information Probe processing (described above), the following statistics are computed: a binary indicator for any app use during the week, the weekly minutes spent on the app, the number of days the app is used, the mean daily minutes spent on the app, and the standard deviation of the daily minutes spent on the app. A value of zero is imputed for participant-weeks that do not show any instances of use for a given app. These statistics are then additionally calculated for all apps in the various app categories, along with the additional statistic of how many apps in a given category are used in a given participant-week. These categories include all categories detected from Google Play, a category for gambling apps that is constructed based on an analysis of app characteristics, and a composite category of all apps.

- **Location Probe**

  The location probe contains all of the raw data pertaining to the phone's geolocation, measured in longitude and latitude, over time. These data are set to update every 5 minutes, but may be less frequent if a phone is off, or out of service range. These data are then processed into participant-week level statistics of mean, maximum and minimum latitude and longitude, the distance from the mode coordinates to the Busara offices, and the fraction of time recorded within a certain number of kilometers from the Busara offices. Additionally, an algorithm is used to deduce the 'number of important locations' for each participant-week according to their travel patterns.

- **Battery Probe**

  The battery probe contains all of the raw data pertaining to the phone's battery levels, recorded at regular 15 minute intervals, contingent on the phone being on

and the app being active. These raw data are then processed into participant-week level statistics of the number of power sources, the fraction of charge time on AC / USB, the mean, max, min and standard deviation of battery levels, the most common battery health status, and total amount of charge time.

- **Call Probe**

  The call probe contains all raw data pertaining to incoming or outgoing phone calls, including information on duration, time of day, and caller ID, recorded at the level of individual calls. These raw data are then processed into participant-week level statistics of the number of calls; the mean, min, max and standard deviation of daily calls, split up by calls from contacts and calls from non-contacts; the mean, min, max, and standard deviation of call duration; the number of unique phone numbers interacted with; the mean, min, max and standard deviation of the daily number of unique phone numbers interacted with daily; and the fraction of calls that are between participants in the study. These statistics are then calculated over a set of restricted types of calls, defined by minimum call durations (missed calls, non-missed calls, calls that last at least 30 minutes, 60 minutes, and 300 minutes), call times (during the workday, during the weekend, during the evening, during the early morning hours, on a weekday), and whether the calls are incoming or outgoing.

- **SMS Probe**

  The SMS probe contains all raw data pertaining to incoming or outgoing SMS's, recorded at the level of individual texts. These raw data are then processed into participant-week level statistics of the number of SMS's; the mean, min, max and standard deviation of daily SMS's; the number of unique phone numbers interacted with; and the mean, min, max and standard deviation of the daily number of unique phone numbers interacted with daily. These statistics are then calculated over a set of restricted types of calls, defined by SMS times (during the workday, during the weekend, during the evening, during the early morning hours, on a weekday), and whether the SMS's are incoming or outgoing.

- **Wifi Probe**

  The Wifi probe contains all raw data pertaining to a phone's connection to Wifi hotspots. These raw data are then processed into participant-week level statistics of the number of raw data observations, the number of Wifi hotspots with no other users connected, the number of Wifi hotspots with at most one other user connected, and the maximum number of available hotspots recorded.

- **Screen Probe**

  The Screen probe contains all raw data pertaining to a phone's screen status. These raw data are then processed into participant-week level statistics of the number of times a screen is turned on or off, at all times and only in late at night, the total amount of time a screen is on or off, and the maximum single span of time that a screen is on or off, calculated over the full week, over only late-night stretches, and only during regular workdays, respectively.

These aggregate statistics will then be combined into a single merged dataset and restricted to participant-weeks wherein the participant opted in to the challenge and uploaded data within the window. The data collected through the Sensing App, described in detail above, will form the basis for the predictive models. The raw data from each of the eight probes will be processed, then combined into a single merged dataset, and then restricted to participant-weeks wherein the participant opted in to the challenge and uploaded data within the upload window. We will then use this dataset to construct an T x N x K feature matrix, where N is the number of participants, K is the number of features, and T is the number of time periods. Participant-weeks that do not appear in the final dataset due to attrition will be left as missing in this final feature matrix.

For missing data at the individual probe stage, we will fill non-observations with zeros as appropriate, e.g. when we see no calls made late at night for a given participant-week and therefore infer that the mean daily calls late at night for that participant-week are zero. For missing observations in the merged dataset (for example, participant-weeks that are present in some probes but not others) we will similarly fill missings with zero or by filling-forward and then filling-backward, as appropriate. Missings for variables that don't have a clear missing value (such as fraction variables) are filled to zero and additionally marked with a variable.isNA flag.

3.3 Survey data "response variables"

The survey instrument is included as Appendix A1 to this Pre-Analysis Plan. Of key interest will be understanding whether it is possible to predict certain characteristics of participants, as recorded in the baseline survey, from the data collected through the Sensing app. Our primarily analysis will focus on predicting these response variables (as measured via baseline survey):

- Age

- Gender
- Number of children, household members, close friends, acquaintances, people relied on, people in contact list, Facebook friends
- Last week: calls made, calls received, texts sent, texts received, people spoken to, airtime spent, cellular data amount
- Regular behaviors: gambles at least once per week, shares phone at least once per week, receives a regular salary
- Occupation status: is 'Employed', or 'Student', or 'Housewife'
- Number of days in the last month spent working
- Monthly income
- Has an electric socket in the home
- Mental health status via PHQ9: a binary indicator for whether depressed based on score, the raw score, and diagnosis: (minimal, at least mild, at least moderate, at least moderately severe, at least severe). See Section 5.2.
- Solves their own technical problems, or goes to repair shops for technical problems
- Considers their technology skills 'Expert' level, or 'Advanced' or above
- Intelligence (Raven's matrix score, and whether the person correctly answered three math problems)
- Education level
- Literacy level (measured both for reading and for writing)
- Marital status
- The first principal component of a matrix of 'popularity' associated variables, specifically consisting of indicators for having Facebook, the number of calls made and received and the number of texts sent and received the week before the survey (as estimated by the survey recipient), the number of people spoken to the week before, the number of acquaintances, and the number or close friends the survey recipient has

- The first principal component of a matrix of 'poverty scorecard' variables, specifically consisting of variables found in the Scorocs Poverty Scorecard for Kenya (available online at http://www.simplepovertyscorecard.com/)
- The first principal component of a matrix of the asset-based variables from the poverty scorecard, specifically consisting of indicators for whether the participant owns an iron, how many habitable rooms are in their home, the number of mosquito nets owned, the number of towels owned, the number of frying pans owned, and whether an electric socket is available in their home, all of which are scored according to the weights in the poverty scorecard.

4 Empirical Analysis

4.1 Estimating individual characteristics from Sensing data

The starting point for our analysis will be to use data collected during the baseline period (when subjects are not presented with challenges through the app) to predict subject characteristics using machine learning methods. Specifically, for a given characteristic Y_{it} of individual i in week t, we will fit models of the form:

$$Y_{it} = f^{(Y)}(X_{it}) + \epsilon_{it} \tag{1}$$

We use 10-fold cross-validation and LASSO regression, as described by Hastie et al. (2009). The X matrix will first be restricted to numeric features and features with at least a minimum amount of variance (those with at least 5 percent of observations not equal to the median) before analysis is performed. The penalization parameter for the LASSO regression will be chosen in two manners: first, an optimal level will be determined using 10-fold cross-validation; second, levels of the penalization parameter that produce sparse models of 3 or fewer variables will also be estimated, in case the optimal level of penalization includes more variables than can easily be communicated to users. During the cross-validation step, the X matrix restrictions will be performed by fold to preclude overfitting.

We will select a small number of response variables on which to evaluate decision rules experimentally. For each potential response variable Y, we will report the cross-validated R^2 (for continuous variables) or AUC (for binary variables) as the primary indicator of predictive performance. In addition to prioritizing Y variables that are likely to be intuitive to subjects (i.e., while subjects may respond to an incentive to use their phone like a wealthy person, focus groups indicated considerable

resistance to the idea of using their phone like someone of the opposite gender), we will prioritize the top-k best performing models and response variables from this stage of the analysis. These will form the simple and complex challenges that follow.

4.2 Estimating the cost of manipulation ("Simple challenges")

Following the procedures described above, the data features most relevant for the prediction of outcome variables will be chosen using LASSO regression across the data feature matrix, X. Thus, for a given Y, we have a subset $X^{(Y)}$ of X that are the best joint predictors of Y.

After determining $X^{(Y)}$, we then also determine an alternative subset $X'^{(Y)}$, which is a set of predictors that are highly similar to those in $X^{(Y)}$, but may have meaningfully different costs of manipulation. To find such features, we construct a list of possible alternatives from three groups:

- We consider all variables that are defined over a similar realm of behavior. For example, if a chosen indicator is the weekly number of outgoing texts at night, we consider all other indicators that consider outgoing texts at night the number of unique phone numbers texted at night, the maximum daily phone numbers texted at night, etc. We will also consider similar measures for incoming texts (which may be harder to manipulate), or WhatsApp usage (a substitute). Or, as another example. if the indicator has to do with a certain app, we consider all behaviors associated with use of that or related app.

- We consider all variables that are highly correlated (Pearson correlation coefficient ≥ 0.9) with a selected indicator.

- We consider all variables that are selected from a restricted LASSO, estimated excluding each variable in $X^{(Y)}$. This will capture potential substitute variables that explain similar variation. We will also try replacing variables that are hard to explain, or binary variables for which exist closely related non-binary variables. "Hard to explain" variables include measures based on standard deviations, measures based on certain variables showing up as missing in our data, or measures that are very finely precise in terms of timing/type (such as, number of outgoing calls of at least 30 minutes duration that occur between 9am and 5am on weekdays). Binary variables for which closely related non-binary variables exist are, in our case, variables having to do with app installation;

for these, app use variables provide smoother, non-binary representations of similar behavior.

After constructing this broad set of alternatives, we then test out how a model based on these alternative indicators would perform as compared to the optimal model. Going indicator-by-indicator, we replace each optimal indicator with a potential alternative and evaluate 1) sample R^2 from a univariate regression of that alternative on the outcome variable; 2) sample R^2 from a regression of this alternative and other optimal indicators on the outcome variable; and 3) cross-validated R^2 from a regression of this alternative and other optimal indicators on the outcome variable. We perform this analysis for each potential alternative, comparing each to the respective R^2 from the optimal model, and upon deliberation select a handful that we expect to have different manipulation costs from the optimal indicators, but that show little deterioration in terms of model predictiveness.

We plan to select 3 promising variables as potential alternatives for each Y. It is all these features ($X^{(Y)}$ and $X'^{(Y)}$) that subjects will be incentivized to manipulate during the the simple challenge phase of the experiment (Section 2.2.1).

4.3 Estimating Robust Models

4.3.1 Heterogeneous Effects

We will allow heterogeneity in estimating γ_i in a dimension such as:

- Whether the respondent considers their technology skills 'expert' level, or 'advanced' or above
- Intelligence (Raven's matrix scores, and whether three math questions were answered correctly)
- Solves their own technical problems, or goes to repair shops for technical problems
- Education is some college or above
- Age

We will assume that the decisionmaker knows this characteristic of the user and it is not manipulated.

4.4 Evaluating Performance

4.4.1 The Impact of Transparency

During the complex challenge phase of the experiment, we will randomly vary the extent to which the subject is informed of the formula used in the decision rule (see Section 2.2.1. We are specifically interested in testing:

1. **Transparent vs. Opaque ($\mathbf{CO}*_j$ vs $\mathbf{CT}*_j$)**

   Do subjects given transparent complex challenges manipulate behavior more than subjects given opaque complex challenges?

2. **Naive vs. Robust ($\mathbf{CON}_j$ vs $\mathbf{COR}_j$)**

   Does the naive model predict the outcome better under transparency than the robust model? (We expect the robust model to achieve lower performance in this case.)

3. **Naive vs. Robust ($\mathbf{CTN}_j$ vs $\mathbf{CTR}_j$)**

   Does the robust model predict the outcome better under transparency than the naive model?

5 Additional Research Outputs

5.1 Cost Measure Comparisons

Separate cost estimates will be collected in a small online survey, among a separate sample of people. We will conduct the survey with (a) Busara participants, and (b) expert academics (who may include economists, computer scientists, and statisticians for example). After a set of indicators are identified in the prephase, we will conduct a short online survey to measure their relative costs.

For each indicator k, the survey will ask:

- On a scale of 0-100, how difficult would it be to alter each of the following behaviors over the course of a week?
- (for Busara participants) We want to avoid including behaviors that are manipulable into a decision rule. Would you change the following behavior if you were paid to do so?

- (for experts) We want to avoid including behaviors that are manipulable into a decision rule. Would you exclude the following behaviors from a decision rule on the grounds of being manipulable?

We will then compare the manipulation costs implied by these surveys to those implied by the experimental estimates. We will also assess the resulting theoretical performance of robust decision rules which use these cost estimates. We can assess two measures of performance. First, theoretical performance if the true manipulation costs are given by the experimental estimates. Second, although we will not include additional complex challenges that use the robust models implied by these manipulation costs, we can assess the performance of these robust models in the manipulated environment that corresponds to the robust models estimated using experimental costs ($\mathbf{CTR}j$); this may serve as an upper bound of performance.

5.2 Mental Health Sensing

As a separate part of the experiment, we will assess whether this form of behavioral data can be used to measure mental health status. During the intake survey, participants will be asked to complete the Patient Health Questionnaire-9 (PHQ-9), which may be translated into Kiswahili. We will follow standard practice to score these questionnaires: each answer is scored 0 (not at all) - 3 (nearly every day), and scores are summed, for a total score between 0-27 for the 9 questions.

Our primary outcome will be a binary indicator for whether the person is classified as depressed. We will set a middle value of PHQ-9 score in our sample and label participants as relatively depressed if they fall above that value. We will set this value to be near the median; if the median is near the boundary between a common diagnosis category, we will adjust it to match that boundary.

We will also consider several secondary outcomes:

- Binary severity ratings: whether a participant's depression status is classified as minimal (≤ 4), at least mild (≥ 5), at least moderate (≥ 10), at least moderately severe (≥ 15), or severe (≥ 20)[1]
- The continuous PHQ-9 score itself.

We will formulate this problem as a standard supervised learning problem, using the full set of explanatory variables used elsewhere in this study, and these listed outcomes. We will train Random Forest and LASSO models using 5 fold cross

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495268/

validation, and report out of sample performance using R2, and for binary outcomes, area under the receiver operating characteristic curve (AUC).

6 Study Timeline

The study will take place over 18 weeks. The first 7 weeks are the "Phase 0" load-in weeks; the second 7 weeks are the "Phase 1" simple challenge weeks; the final 4 weeks are the "Phase 2" complex challenge weeks.

- **Phase 0: Weeks 1 - 7**

  Phase 0 is the period when participants are being on-boarded by the Busara team. Due to capacity constraints, only 200 people can be on-boarded in a given week. The full sample of 1200 will therefore take 6 weeks to on-board, and then an additional week will be required to ensure that at least a week's worth of baseline data is collected for each participant.

  This baseline data (**B**) will then be used in cross-validated LASSO regression, as described above in the 'Empirical Analysis' section, to determine the predictive power of the data features for each of the outcome variables recorded in the survey, and the overall predictive power of a restricted model based on a limited number of such features. The research team will then select three outcome variables (Y_j for $j \in \{1..3\}$) that are well-predicted by a restricted model as the variables to incentivize over the course of Phases 1 and 2. The features in the respective restricted models that predict each Y (X_{jk} for $j \in \{1..3\}$ and $k \in \{1..3\}$) will therefore be the targets of the incentive challenges in Phase 1, as will close substitutes of these features. ('Close substitutes' will be identified by inspecting the predictive power of models with optimal features swapped for similarly designed features, highly correlated features, and/or features that arise in LASSO regression with the X matrix restricted to simpler features.) For example, if optimal models include features based on outgoing calls, these close substitutes may include corresponding features based on incoming calls or text messages. Note that the first Phase 1 challenges (the Y1 challenges) will be based on only the first 6 weeks of data, since these challenges will need to be assigned before the seventh week of baseline data is available to be analyzed; the second and third Phase 1 challenges (for Y2, Y3) will be based on the full 7 weeks of baseline data.

  Additionally, in order to keep participants engaged throughout the long on-boarding period, participants who have been in the sample at least 1 week will

receive a dummy Phase 1 simple challenge every other week. These challenges will incentivize a feature that the research team determines to be predictive of a possible Y outcome given the baseline data available at that point. Specifically, weeks 3 and 5 of the onboarding period will feature dummy challenges of this type. Participant-weeks associated with dummy challenges will not be included in the baseline data used to determine the Phase 1 challenges described above.

Due to uncertainties with the number of individuals who may be onboarded in a given week, this phase 0 may extend beyond 7 weeks if that time is necessary in order to construct the complete phase 1 sample.

- **In between Phases 0 and 1**

  We will conduct cost surveys as described in Section 5.1.

- **Phase 1: Weeks 8 - 14**

  Phase 1 is the period when 'simple' challenges are assigned ($\mathbf{S}k$), which is to say, challenges that directly incentivize changes in data features. This phase will be divided into three periods, divided according to the features targeted. In each period, a control group of predetermined size will be assigned control challenges, which do not incentivize any behavior other than the timely uploading of data from their phone.

  In week 8, challenges that incentivize the data features included in the optimal restricted model for predicting Y1 will be randomly assigned to each participant, as will challenges that incentivize close substitutes for these same features. This model for predicting Y1 will be based on the data from weeks 1-6, since week 7 will not be completed while this challenge is being designed. We delay assigning challenges for the models associated with Y2 and Y3 in order to allow these models to be constructed using data from the full "phase 0" period of weeks 1-7.

  In weeks 9-13, challenges that incentivize the data features included in the models for Y1, Y2, and Y3 will all be randomly assigned to participants, as will challenges that incentivize close substitutes for these features.

  In week 14, challenges that incentivize the data features included in the model for Y3 will be randomly assigned to participants, as will challenges that incentivize close substitutes for these features. We cease incentivizing Y1- and Y2-model associated behaviors because this last week of data would not be usable the week immediately after, week 15, and we incentivize complex chal-

lenges based on Y1 and Y2 models in this week. The Y3 complex challenges are delayed a week, allowing us to use them in the final week of this phase.

The assignment of challenges is cross-randomized across weeks and across all potential orderings of challenges within the above guidelines, so that e.g. the assignment of an individual to X1, X1, X1, X2, X2, X3, X3 in weeks 8-14, respectively, is as likely as an assignment to X1, X2, X2, X1, X1, X3, X3. Specifically, there are 80 potential permutations of challenge assignment orderings across weeks, of which 30 include 3 weeks of X1, 30 include 3 weeks of X3, and 20 include 3 weeks of X2. We assign individuals equally across these three groups to ensure equal populations being assigned to X1, X2 and X3 challenges, respectively, and then randomly assign individuals to an ordering of challenges within these broader groups.

After each Y period, the treatment effects of the respective Phase 1 challenges will be used to determine the cost of inducing changes in each given feature. These costs will then be used to estimate our robust models.

In order to ameliorate the effects of attrition, an additional on-boarding period will begin in week 12 and continue through week 14, adding an additional 600 'top up' participants to the study. These participants will not be included in any Phase 1 activity and challenges, and will be assigned Phase 2 challenges as a group stratified separately from the previously on-boarded group.

- **Phase 2: Weeks 15-18**

  Phase 2 is the period when 'complex' challenges are assigned, which is to say, challenges that target the Y1, Y2 and Y3 outcomes. These challenges will be assigned according to two models, naive and robust, and two levels of transparency, opaque and transparent.

  Transparent challenges ($\mathbf{CTN}j$ and $\mathbf{CTR}j$) will be assigned along with a note that specifies precisely the mathematical formulas that will be used for payouts, along with a page of interactive sliders that will allow users to directly compute expected payouts from different hypothetical behaviors. (Slider use will be explained at length to participants during the on-boarding process.)

  Opaque challenges ($\mathbf{CON}j$ and $\mathbf{COR}j$) will only show the outcome incentivized, without any additional information on which specific features will be rewarded. Under opacity, naive and robust decision rules are observably equivalent to users until the point of payment. Thus, we will pool the opaque treatments that use these different models.

With each combination of transparency and model, four different challenges will be assigned per Y outcome.

Due to lag time in determining the cost of incentivizing Y3 variables, the first week of Phase 2 will only assign challenges for Y1 and Y2 outcomes. These challenge assignments will be randomized across participants over transparencies, models, and outcomes. The randomization will additionally be stratified across the samples, with the newly on-boarded participants in one group and the long-term participants in another.

In the subsequent two weeks, challenges will be randomized in the following manner: those who received Y1 challenges in the first week will have randomized challenge assignment across styles and models for outcomes Y2 and Y3; those who received Y2 challenges in the first week will have randomized challenge assignment across styles and models for challenges Y1 and Y3. Again, this will be stratified over the top up group and the main group. This randomization will be set so that each participant sees each Y exactly once, and exactly one of these challenges is opaque. In the fourth week, complex challenges are assigned deterministically: each participant will receive the transparent version of the opaque challenge that they earlier received. This ensures that no individual is assigned to observe the transparent decision rule prior to receiving an opaque challenge for the same outcome.

With these restrictions on randomization, each participant has 12 possible permutations of challenges that they may face, corresponding to the 2 transparent first-week choices x 4 second-and-third week choices conditional on transparent first-week choice + 2 opaque first-week choices x 2 second-and-third week choices conditional on opaque first-week choice. The randomization will therefore be designed as random assignment to 12 evenly-sized groups at the outset of the Phase 2 period, with challenges then assigned according to the following table:

	Week 15	Week 16	Week 17	Week 18
Group 1	Y1, opaque	Y2, trans.	Y3, trans.	Y1, trans.
Group 2	Y1, opaque	Y3, trans.	Y2, trans.	Y1, trans.
Group 3	Y1, trans.	Y2, opaque	Y3, trans.	Y2, trans.
Group 4	Y1, trans.	Y3, opaque	Y2, trans.	Y3, trans.
Group 5	Y1, trans.	Y2, trans.	Y3, opaque	Y3, trans.
Group 6	Y1, trans.	Y3, trans.	Y2, opaque	Y2, trans.
Group 7	Y2, opaque	Y1, trans.	Y3, trans.	Y2, trans.
Group 8	Y2, opaque	Y3, trans.	Y1, trans.	Y2, trans.
Group 9	Y2, trans.	Y1, opaque	Y3, trans.	Y1, trans.
Group 10	Y2, trans.	Y3, opaque	Y1, trans.	Y3, trans.
Group 11	Y2, trans.	Y1, trans.	Y3, opaque	Y3, trans.
Group 12	Y2, trans.	Y3, trans.	Y1, opaque	Y1, trans.

Finally, each group-week will be additionally randomly divided into two halves, one of which will receive a challenge based on the robust model and the other of which will receive a challenge based on the naive model.

6.1 Consistency Checks

This design affords us several consistency checks:

- There should be no difference in behavior between **CON**j and **COR**j if participants do not observe the decision rule
- If receiving an opaque challenge Yj prior to receiving a transparent challenge for the same outcome affects behavior, then we could find a difference in outcomes between participants assigned to that transparent challenge in weeks 15-17 and in the final week 18. If there is a substantial difference, we may analyze week 18 separately.

7 Research Team

The Principal Investigators on this study are:

- **Daniel Björkegren** is an Assistant Professor of Economics at Brown University. His research explores the opportunities generated by new technologies in the developing world. His work has been featured by NPR. He holds a Ph.D. in Economics and a Master's in Public Policy from Harvard University, a M.A. in

Economics from Stanford University, and a Bachelor's degree in Physics from the University of Washington.

- **Joshua Blumenstock** is an Assistant Professor at the U.C. Berkeley School of Information, and the Director of the Data-Intensive Development Lab. His research lies at the intersection of machine learning and development economics, and focuses on using novel data and methods to better understand the causes and consequences of global poverty. At Berkeley, Joshua teaches courses in machine learning and "big data for development". He has a Ph.D. in Information Science and a M.A. in Economics from U.C. Berkeley, and Bachelors degrees in Computer Science and Physics from Wesleyan University. His work has appeared in a variety of publications including Science, Nature, the American Economic Review, and the proceedings of KDD and AAAI.

- **The Busara Center for Behavioral Economics ("Busara")** is a non-profit research organization dedicated to furthering the understanding of human decision-making and enabling government and organizations to apply this knowledge in practice. Busara has offices in Kenya, and will be responsible for subject recruitment and surveying. Busara is separately obtaining human subject approval from the Kenya Medical Research Institute (KEMRI), the government-owned parastatal that oversees human subjects research in Kenya.

References

Bjorkegren, D. and D. Grissen (2015). Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment. *Available at SSRN 2611775*.

Borrell Associates (2016). Trends in Digital Marketing Services.

Bruckner, M., C. Kanzow, and T. Scheffer (2012). Static Prediction Games for Adversarial Learning Problems. *Journal of Machine Learning Research 13*(Sep), 2617–2654.

Hardt, M., N. Megiddo, C. Papadimitriou, and M. Wootters (2016). Strategic Classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, ITCS '16, New York, NY, USA, pp. 111–122. ACM.

Hastie, T., R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani (2009). *The elements of statistical learning*, Volume 2. Springer.

Hu, L., N. Immorlica, and J. W. Vaughan (2019). The Disparate Effects of Strategic Manipulation. *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19*, 259–268. arXiv: 1808.08646.

Hlmstrom, B. (1979). Moral Hazard and Observability. *The Bell Journal of Economics 10*(1), 74–91.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018, February). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics 133*(1), 237–293.

Kleinberg, J. and M. Raghavan (2019). How Do Classifiers Induce Agents to Invest Effort Strategically? In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, New York, NY, USA, pp. 825–844. ACM. event-place: Phoenix, AZ, USA.

Lucas, R. E. (1976, January). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy 1*(Supplement C), 19–46.

Milli, S., J. Miller, A. D. Dragan, and M. Hardt (2019). The Social Cost of Strategic Classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, New York, NY, USA, pp. 230–239. ACM. event-place: Atlanta, GA, USA.

A1 Survey Instrument

1. WELCOME SCRIPT
2. CONSENT
3. COMPREHENSION CHECKS
4. Demographics
    - What is your first name?
    - What is your middle name?
    - What is your last name?
    - What is your year of birth?
    - Gender
    - Ethnicity
    - What is the highest level of education you have completed?
        - None, or pre-school
        - Primary standards 1 to 6
        - Primary standard 7
        - Primary standard 8 or secondary forms 1 to 3
        - Secondary form 4
        - Some college
        - Completed college
        - Some graduate
        - Completed graduate
    - Can you read a letter or newspaper? [Easily / With Difficulty / Not at all]
    - Can you write a letter? [Easily / With Difficulty / Not at all]
    - What is your marital status?
    - How many children are there in your household?
    - How many close friends do you have?
    - How many acquaintances do you have?
    - How many people rely on you?
    - How frequently do you gamble? [Daily / Weekly / Monthly / Rarely / Never]
    - Are you the primary breadwinner in your household?
        - If not, what is your relation to the primary breadwinner?
    - What is your mother tongue?
    - How many biological children do you have? Biological children are directly related to you, not step children or adopted children.

5. Socioeconomic status
    - What sector do you work in?
    - Do you receive a regular salary each week?
    - In the last week, how many days of work did you do?
    - In the last month, how much income did you earn from economic activity?
    - PPI questions (https://www.povertyindex.org/country/kenya)
        - How many members does the household have?
        - What is the highest school grade that the female head/spouse has completed?
        - What kind of business (type of industry) is the main occupation of the male head/spouse connected with?

        - A. Does not work
        - B. No male head/spouse
        - C. Agriculture, hunting, forestry, fishing, mining, or quarrying
        - D. Any other
    - How many habitable rooms does this household occupy in its main dwelling (do not count bathrooms, toilets, storerooms, or garage)?
    - What material is the floor of the main dwelling predominantly made of?
        - Wood, earth, or other
        - Cement or tiles
    - What is the main source of lighting fuel for the household?
        - A. Collected firewood, purchased firewood, grass, or dry cell (torch)
        - B. Paraffin, candles, biogas, or other
        - C. Electricity, solar, or gas
    - Does your household own any irons (charcoal or electric)?
    - How many mosquito nets does your household own?
    - How many towels does your household own?
    - How many frying pans does your household own?
- Do you have electricity / an electric socket at home?
- Can you tell me approximately how much money was used for each of the following?
    - Own consumption (food, entertainment, etc.)
    - Provided as a loan or gift to a family member
    - Provided as a loan or gift to someone else in your community
    - Saved for the future
- During the last seven days, how many times did one or more people in your household not receive a regular daily meal?
- Please tell me the material that your FLOOR is made out of.
- Please tell me the material that your WALLS are made out of.
- Please tell me the material that your ROOF are made out of.
- What toilet facilities do you PRIMARILY use?
- Do you currently receive any government benefits?
    - If so, which program?
    - How much per month?
- What type of lighting do you use in your house?

6. Mobile Phone use

- Complete the following table:

Complete the Operator	How many SIM cards do you own that use this network?	What is your primary phone number on this network?	How frequently do you use this SIM card?	Do you consider this to be your main number?
Safaricom			A-E (see below)	Yes/No
Airtel			A-E (see below)	Yes/No
Telkom			A-E (see below)	Yes/No
Other			A-E (see below)	Yes/No

        - A: Multiple times per day
        - B: Roughly once per day
        - C: Multiple times per week, but not every day
        - D: Roughly once per week
        - E: Less than once per week

- How many years have you had a smartphone?
- Does your current phone typically have more than one SIM card in it?
- If you had a technical problem with your cell phone, who would you mainly ask for help? (for example if your phone would not turn on or allow you to make calls)
    - Self
    - Relative
    - Friend or neighbor
    - Repair shop
- Do you share your current mobile phone with others?
    - Typically one or more times each day
    - Not every day, but several days each week
    - Just once or twice each week
    - Very rarely (no more than once or twice each month)
    - Never
- On a scale of 1-5, with 1 being a total beginner, and 5 being an expert, what would you say your level of skill and familiarity is with digital technology (such as computers and phones)?
- Without looking at your phone, can you estimate how many different contacts you have stored in your contact list?
- Roughly how many different people do you estimate you spoke to in person in the last week?
- Without looking at your phone, can you tell me approximately…
    - How many phone calls you made in the last week?
    - How many phone calls you received in the last week?
    - How many text messages you sent in the last week?
    - How many text messages you received in the last week?
    - How many different people you spoke with on the phone last week?
    - How much money you spent on airtime in the last week?
    - How much money you spent on cellular data in the last week?
    - Do you have Facebook?
        - If so, how many Facebook friends do you have?
    - How many times you sent money using mobile money?
    - How many times you received money using mobile money?
- Subjective rating:
    - On a scale from 1-5, how easy is would it be for you to:
        - Send 10 more text messages next week than you did last week?
        - Send 10 fewer text messages next week than you did last week?
        - Make 10 more phone calls next week than you did last week?
        - Make 10 fewer phone calls next week than you did last week?

        - Call 10 people next week that you didn't call last week?
        - The number of people you call?
        - The amount of data you use?
        - The time of day that you use data?
        - The time of day that you call?

    Next, imagine a company were to provide you with a bonus for using your mobile phone differently.

    - If you were paid a bonus 0.5 Ksh per SMS you send next week, how many SMS would you send?
    - If you were paid a bonus of 50 Ksh but from that was deducted 0.5 Ksh per SMS you send next week, how many SMS would you send?

    - If you were paid a bonus 0.5 Ksh per call you make next week, how make calls would you make?
    - If you were paid a bonus 50 Ksh but from that was deducted 0.5 Ksh per call you make next week, how make calls would you make?

    - If you were paid a bonus 0.5 Ksh per person you called next week, how many people would you call?
    - If you were paid a bonus 50 Ksh minus 0.5 Ksh per person you called next week, how many people would you call?

    - If you were paid a bonus 0.05 Ksh per MB of data you used next week, how much data would you use?
    - If you were paid a bonus 50 Ksh minus 0.05 Ksh per MB of data you used next week, how much data would you use?

- How many mobile phones do you own?
    - How many are smartphones?
- Do you have a mobile money account (M-PESA, etc)?
    - Which M-PESA services do you use (check all that apply)?
- Where do you charge your mobile phone battery mostly? [home, shop, work/school, other]
    - How many times per week do you charge your mobile phone?

- How regularly do you use the following services?
    - E-mail
    - WhatsApp
    - Tuko
    - Facebook
    - YouTube
    - SportPesa or other betting service
    - Other internet sites
    - A laptop or computer (not a phone)

- Baselines
    - How many SMS do you plan to send next week?
    - How many calls do you plan to make next week?
        - How many in the morning from 5am-8am
        - How many between 8am-7pm?
        - How many in the evening from 7pm-11pm?
        - How many overnight (11pm-5am)?
    - How many different people do you plan to call next week?
    - How much data do you plan to use next week (in MB)?

7. Financial Inclusion

- Do you currently keep an account at any of the following institutions?
    - Commercial Bank
    - Microfinance institution (such as XX)
    - ROSCA (such as XX)
    - Other
- In an average month, roughly how much money do you personally earn?
- In an average month, roughly how much do you spend?
- Have you ever received a loan from any of the following sources?

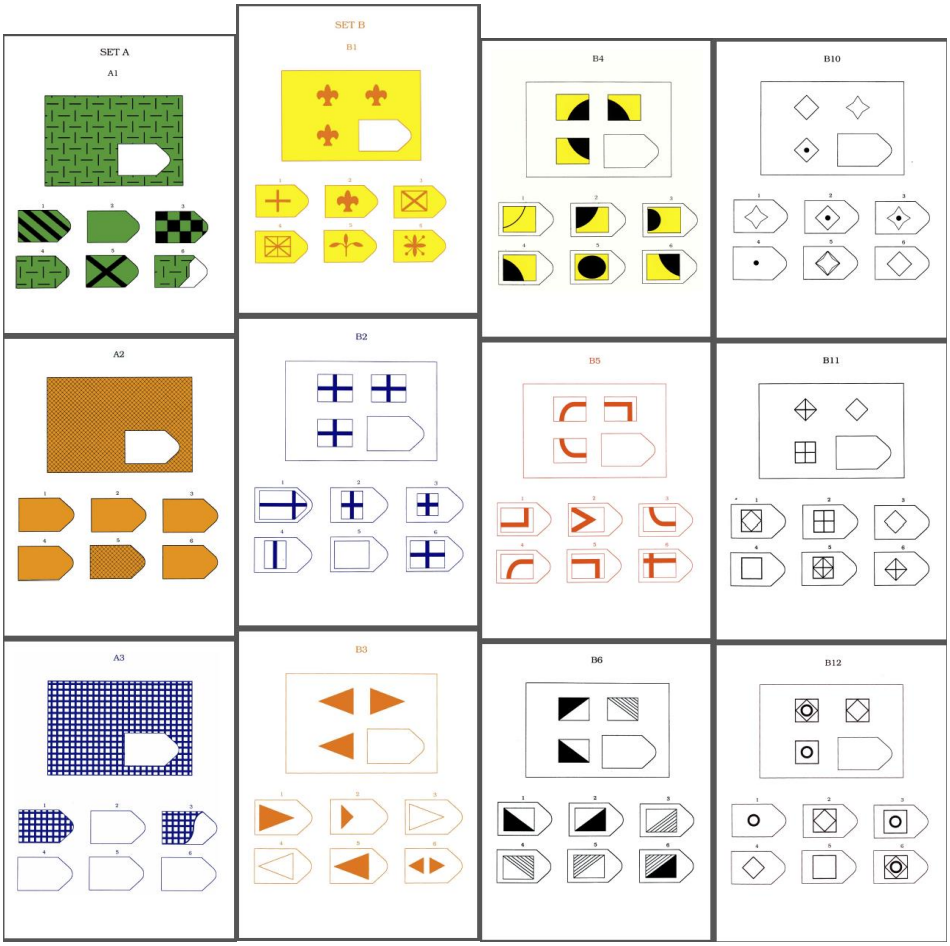
Source	If yes, roughly how many loans?	What was the largest loan you received from this lender?	Were your repayments on time or late?
Commercial bank			A-E (see below)
Local moneylender			A-E (see below)
M-Shwari			A-E (see below)
Branch			A-E (see below)
Tala			
Other (list)			

- A: Always on time
- B: Usually on time, but occasionally late
- C: Frequently late
- D: Not fully repaid by end of loan term

8. Numeracy and Raven's scores

- Suppose Joseph earns a salary of 1000 Shillings a week. He obtains a ten percent raise. How much exactly will his income be after the raise?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 10 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
- If it takes five machines five minutes to make five widgets, how long does it take 100 machines to make 100 widgets?
- Raven's matrix: Here is a pattern with a piece missing. Below are six pieces, choose the one that completes the pattern.

- A1 – example (doesn't count)
- A2 – example (doesn't count)
- A3
- B1 – B12 [see attached PDF for enlargement]



9. Present Bias and Risk Aversion

We are interested in understanding how Kenyans make decisions involving uncertain outcomes and some normal risks that people face every day. We would like to ask you some hypothetical questions that will help us understand these decisions. There is no real money involved and you will not receive any money for answering these questions . Are you willing to answer these questions? [Yes (Proceed with survey)) / No (Conclude Surevy)]

- Suppose someone was going to pay you 4000 Shillings 13 months from now. He/she offers to pay you a lower amount in 12 months time. What amount in 12 months would make you just as happy as receiving 4000 Shillings in 13 months?
    - [Comparison ladder]
- Suppose someone was going to pay you 4000 Shillings 6 months from now. He/she offers to pay you a lower amount in 5 months time. What amount in 5 months would make you just as happy as receiving 4000 Shillings in 6 months?
    - [Comparison ladder]
- Suppose someone was going to pay you 4000 Shillings 1 month from now. He/she offers to pay you a lower amount today. What amount today would make you just as happy as receiving 4000 Shillings in 1 month?
    - [Comparison ladder]

- [Show Card] First we will ask you a hypothetical question over an amount for certain, or an amount that will be awarded depending on which of ten numbers you draw from a bag. We have deposited 10 cards numbered 1 through 10 into a bag. You have an even chance of drawing any of the 10 numbers. The numbers in parentheses indicate the winning number. For each Option No., please indicate whether you would prefer Choice 1 or Choice 2. For each option No. there will be 10 numbers in the bag and you are only able to draw one. This is not for real money and we are not asking you to make a gamble, we just want to understand how you would respond to naturally occurring risk.
    - [Comparison ladder]
- Now we will ask you which of two lotteries you would prefer. We will place 10 even sized cards numbered 1 through 10 into a bag and ask you to draw one. You have an even chance of drawing any of the cards. The numbers in parentheses indicate the winning numbers.
    - [Comparison ladder]

10. Mental Health

Patient Health Questionnaire-9 (PHQ-9)

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The Phq-9. Journal of general internal medicine, 16(9), 606-613.

ENUMERATOR: Over the last two weeks, how often have you been bothered by any of the following problems?

TRANSLATION: Kwa juma viwili zilizopita mara ngapi umesumbuliwa na matatizo haya?

Code	Question	Values
	Little interest or pleasure in doing things	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day
	Mwelekeo mdogo au kukosa raha wa kufanya vitu	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Feeling down, depressed or hopeless	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day
	Kujisikia kama huwezi kuchangamka,kusikia, huzuni au kukosa tumaini	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Trouble falling or staying asleep, or sleeping too much	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day
	Tatizo kupata usingizi au tatizo kuendelea kulala baada ya usingizi,ama kulala kupita kiasi	1 Hapana kabisa 2 Siku kadhaa 3 Zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Feeling tired or having little energy	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day

	Kujisikia kuchoka au kuwa na nguvu kidogo	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Poor appetite or overeating	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day
	Hama ya kula ni mbaya, au kula kupita kiasi	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Feeling bad about yourself -- or that you are a failure or have let yourself or your family down	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day
	Kusikia vibaya kuhusu binafsi, au kuskia kama umeshindwa, Au umejishusha, ama umeshusha chini familia yako	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Trouble concentrating on things, such as reading the newspaper or watching television	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day
	Tatizo kutuliza akili kwenye vitu kama kusoma gazeti au kusilikiliza radio	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Moving or speaking so slowly that other people could have noticed? Or the opposite -- being so fidgety or restless that you have been moving around a lot more than usual.	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day
	Kutembea au kuzungumza pole polesana hata ingeweza kuonekana kwa watu wengine. Ama kinyume-kuwa namashaka/wasiwasi au kutotulia kiasi hata umekuwa ukitembea tembea sana kuliko kawaida	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	Thoughts that you would be better off dead or of hurting yourself in some way	1 Not at all 2 Several Days 3 More than half the days 4 Nearly every day

	Fikira kwamba ni heri ukifa, au fikira za kujiumiza kwa njia fulaniFulani	1 Hapana kabisa 2 Siku kadhaa 3 zaidi ya nusu ya siku hizi 4 Karibu kila siku
	If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?	1 Not difficult at all 2 Somewhat difficult 3 Very difficult 4 Extremely difficult
	Ikiwa umejibu kuhusu shida yoyote, ni kwa kiasi gani haya mashida yamefanya iwe vigumu kwako kufanya kazi yako, kutunza vitu nyumbani au kuelewana na watu wengine?	1 Sio vigumu hata kidogo 2 Vigumu kiasi 3 Vigumu sana 4 vigumu kupita zaidi