Pre-analysis plan

Aspirations and forward-looking behaviour in rural Ethiopia: long-term impacts of field experiment*

Tanguy Bernard[†], Stefan Dercon[‡], Kate Orkin[§], Alemayehu Seyoum Taffesse[¶]

15th February 2017

1 Introduction

This document outlines our pre-analysis plan for analysis on a documentary screening intervention. Individuals in a poor, remote part of rural Ethiopia were randomly invited to watch short documentaries about people from similar backgrounds who had succeeded in agriculture or business, without help from government or NGOs. A placebo group, in the same villages, watched an Ethiopian entertainment programme. A control group, in the same villages, were

*The authors thank research assistants Marc Witte, Fanaye Tadesse, and Geetika Nagpal, fieldwork managers Bezabih Tesfaye, Kibrom Hirfrot and Tewodros Abate; Mekamu Kedir, IFPRI Addis Ababa staff; Doba Woreda Administration, Next Studio and Tadesse Fayissa for producing documentaries, and staff at the Centre for Study of African Economies and Oxford Department for International Development. The baseline and midline were funded by the United Kingdom Department for International Development (DFID) as part of the Institutions for Pro-Poor Growth Consortium (iiG), while the endline was funded by USAID. Production of the documentaries used in the randomized controlled trial was funded by SEVEN (Social Equity Venture Fund). Bernard and Taffesse acknowledge the support of the International Food Policy Research Institute (IFPRI) Development Strategy and Governance and Markets, Trade and Institutions divisions.

†Department of Economics, University of Bordeaux, International Food Policy Research Institute; t.bernard@cgiar.org.

‡Blavatnik School of Government, Department of Economics, Centre for the Study of African Economies and Jesus College, University of Oxford; stefan.dercon@economics.ox.ac.uk.

§Department of Economics, Centre for the Study of African Economies and Merton College, University of Oxford; kate.orkin@merton.ox.ac.uk.

¶International Food Policy Research Institute; a.seyoumtaffesse@cgiar.org.

simply surveyed. In addition, the number of people invited to the documentary screenings was varied by village to assess the importance of peer effects.

Here we outline plans to estimate the long-term impacts of this intervention, five years after the intervention. We present our main research hypotheses regarding impacts on psychological outcomes targeted by the intervention (H1) and economic choices (H2). We also examine two secondary, separate, sets of hypotheses. We examine psychological outcomes not targeted by the intervention (H3), which we do not expect to be affected. Secondly, if we see changes in household economic behaviour (H2), we will explore if there are any long-run changes in household welfare (H4).

Data were collected via face-to-face household surveys in 2010 (baseline), immediately after screenings in 2010 (midline 1), after six months in 2011 (midline 2), and after five years in 2015 (endline) and through surveys with the *kebele* (village) leader at baseline and endline. We also use administrative data at village level.

A core part of the analysis on midline 1 and midline 2 is written up in a working paper (Bernard et al., 2014). No pre-analysis plan was formally deposited before this short-run analysis, although all plausible variables were analysed and reported in that paper, as can be ascertained from the questionnaires (available here: https://sites.google.com/site/kateorkin/).

Scientific integrity requires that we report on the long-term impacts of the intervention for the outcomes analysed in that paper. We use the same questionnaires, work with the same fieldwork partner and conduct the same training. In addition, data collection and analysis is extended in the endline:

1. We extend the questionnaires to include variables measuring household welfare and more variables measuring economic choices.

2. We report findings against two counterfactuals.

   a) In the midline, we can only compare outcomes of the intervention among treatment households to outcomes among placebo and control households in the same villages. In the endline, we conduct this same comparison. There may be some spillovers between treated, placebo and control households. For these estimates to be valid, within-village spillovers need to be small.

   b) We are also able to compare treated households and households in "pure control" villages in which no treatment or placebo interventions or surveys took place until the endline. Respondents' aspirations, expectations or beliefs are not affected by visits by outsiders.

In Section 2, we describe the experimental design and data collection and give information on the intervention. In Section 3, we describe tests of experimental integrity, including tests for balance and survey attrition. In Section 4, we report the overall empirical strategy, including analysis of spillover effects. In Section 4, we discuss the theory of change and the construction of outcome variables. The instruments are available here: https://sites.google.com/site/kateorkin/.

They include A: Baseline survey questionnaire for households and individuals, B: Midline 2 survey questionnaire for households and individuals; C: Endline survey questionnaire for; D: Community survey questionnaire at baseline; E: Community survey questionnaire at endline.

2 Experimental design and data

2.1 Sampling

The intervention was designed to test for the existence of a psychological mechanism, so the site was chosen to provide a remote setting where even a relatively light-touch and inexpensive intervention might have effects. The documentaries feature rural inhabitants in fairly mountainous grain-growing areas in Oromia region. We thus chose Doba, a poor, rural, remote, mountainous, grain-growing administrative district

Most surveyed villages were accessible only by 4x4 vehicle and some required camel transportation. There is limited exposure to television: at baseline only 10 per cent of respondents watched TV once a week or more, 29 per cent watched at least once a month and 61 per cent watched about once a year or never. Doba residents would thus be likely to find a television show memorable. The point estimates of the effect of the intervention may not be replicated in less remote areas where participants have more exposure to media.

We used the Central Statistical Agency's list of rural villages for the district to create a list of villages with 50 to 100 households in them and randomly selected 84 villages. No villages have overlapping boundaries. The sample is designed to be approximately representative of households living in villages of this size. We limited heterogeneity in village size so there is similar potential for spillovers in different villages. We exclude very small villages to obtain equal-sized village clusters. Sampling probability is slightly higher for households in smaller villages. We will conduct all analyses without sampling weights to account for differing sampling probabilities in our main estimates.

2.2 Assignment to treatment

This study is a two-level cluster-randomized controlled trial. We randomly assigned the 84 villages to three groups. Twenty villages were assigned to be pure control villages and received no surveys or interventions until the endline. We had initially intended to conduct baseline surveys in the pure control villages but did not, even though we were conscious that this would be at the expense of statistical power, because of financial constraints.[1] In the remaining 64

[1] Other studies have subsequently pursued such a strategy deliberately to avoid the effects of surveys on behaviour (Zwane et al., 2011; Bidwell, Casey, and Glennerster, 2016; Haushofer and Shapiro, 2016). Some studies have included two control groups, one receiving the baseline and endline and one only receiving endline, to identify the magnitude of survey effects (Zwane et al., 2011; Bidwell, Casey, and Glennerster, 2016). We do not have two control groups and cannot identify the effects of surveys on their own. In our study, the difference between pure control and within-village control respondents at endline is the effect of

villages, 32 villages were assigned to be "intense-treatment" villages and the other 32 to be "intense-placebo" villages. These were grouped using GIS data into 16 screening sites of four neighbouring villages, two intense-treatment villages and two intense-placebo villages. Figure 1 shows the intense-treatment, intense-placebo villages and pure control villages.

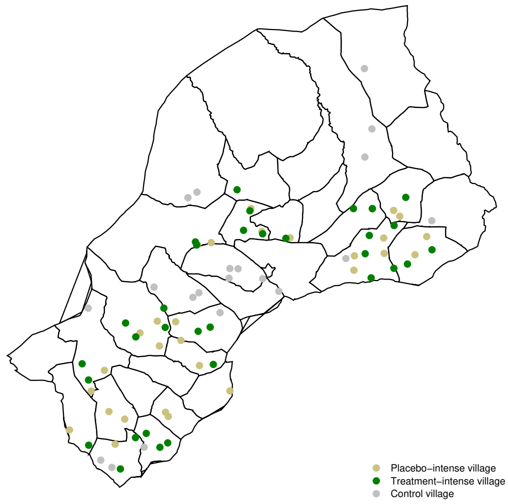

Figure 1: Distribution of villages within the Doba Woreda

In intense-treatment and intense-placebo villages, the enumerators conducted a census to compile a list of all households at baseline, administered with the assistance of the village elder. From this list, the enumerators randomly selected 6 households for each of the treatment, placebo and control groups using a public lottery. We refer to the within-village control group as the "within-village spillover" group to distinguish them from the "pure control" group, respondents in the pure control villages. Randomisation to treatment, placebo and within-village spillover groups was at household rather than individual level, so household heads and their spouses had the same treatment status. The researchers could not conduct the randomisation because the villages largely could not be reached by mobile phone and there was no internet data connection.

The treatment group of six households were invited to a screening of the four 15-minute documentaries. The placebo group of six households were invited to watch four 15-minute segments of an Ethiopian comedy TV show about rural life. The six within-village spillover households were surveyed but did not receive invitations. Enumerators conducted the baseline in all four villages and then conducted the treatment and placebo screenings on one day at different times, with treatment and placebo randomly allocated to morning or afternoon screen-

receiving the baseline survey, having a video screening occur near the village, and spillovers from treatment and/or placebo households in the village.

ings, to avoid within-village treatment effects being confounded by weather, market days or other factors. The within-village spillover households treatment villages were not invited to any screening session.

In addition, in the intense-treatment villages, we randomly selected 18 additional households (roughly 36 individuals) to be invited to the documentary but did not survey these individuals. In the intense-placebo villages, we similarly randomly selected 18 additional households to be invited to the placebo session. In both types of villages, six households are treated, six households are given the placebo and six households receive the control. The only difference between these is exogenous variation in the extent to which an individual's network was exposed to the treatment. This potentially allows us to explore the role social interactions play in changes in psychological variables and behaviour in response to the treatment.[2]

2.3 Intervention

A goal of this study was to assess whether individuals revise their aspirations after a "vicarious experience" (Bandura (1977b), Bandura (1977a)), where they are exposed to the lives of potential role models from a similar background to theirs who have improved their socio-economic position. To make the documentaries, development agents and NGO staff in rural areas were invited to submit descriptions of the life stories of ordinary individuals who had improved their socio-economic well-being despite adverse initial conditions. Ten individuals were selected to have short documentaries made about their lives by Next Studios, an Ethiopian production company. Four were selected for the experiment, two about men and two about women. Each documentary is 15 minutes long and in Oromiffa, the local language in the study site. The placebo treatment consisted of a screening of four fifteen-minute segments of an Ethiopian comedy TV show in the language Oromiffa about rural life. The theory of change is described in Bernard et al. (2014) and summaries of two documentaries and one placebo segment are provided. The documentaries, with English subtitles, and one of four placebo segments are at https://www.youtube.com/channel/UCqfoNjCzt8YPjTRWQaMQfAg.

2.4 Data collection

In the intense-treatment and intense-placebo villages, we have four data collection points with the treatment, placebo and within-village spillover households. The household head completed a **baseline** household interview and individual interview and their spouse (if they had one) completed an individual interview. Enumerators interviewed spouses separately, usually by interviewing them at the same time in different locations. At the end of their individual baseline interview, the household head and spouse in treatment and placebo households received tickets for a screening session in a few days' time.

[2] Villages are about 30 minutes walk apart and individuals' networks of close friends were mostly in the same village, as discussed in Bernard et al. (2014). We thus focus on spillovers between individuals in the same village.

Second, there was a short **follow-up survey** on aspirations and expectations directly after the screening intervention, conducted with treatment and placebo households at the screening venue and with within-village spillover households at their homes. This mainly ensured within-village spillover households did not try to come to the screening venue as they had appointments to meet enumerators at their homes. The baseline survey and screening took place between September and November 2010. Third, the **midline** survey in the intense-treatment and intense-placebo villages occurred six months after the baseline, between March and May 2011. Finally, in all 84 villages, the **endline** survey was conducted five years after the baseline survey, between December 2015 and March 2016. In the intense-treatment and intense-placebo villages, we followed up the households who had already been sampled. The order in which midline and endline villages were surveyed followed the same order as the baseline.

In the pure control villages, there is only one data collection point, at endline. The enumerators compiled a list of all households before starting endline. The census was conducted in the same fashion in intense-treatment/placebo and pure control villages. We work with the same fieldwork manager and fieldwork provider. Enumerators then randomly selected approximately 15 households to be surveyed, using the same public lottery as in the original villages. The samples in pure control and intense treatment/placebo villages are both chosen at random, but at different points at time. If attrition is at random, we would expect the sample to be representative of the village in both types of village. We address the timing difference in detail in Section 3.

2.5 Administrative data on distances

We merge GPS data on village locations with administrative data from the Central Statistical Authority and various aid agencies.[3] The data comes from the sources listed in Table 1 below. The data from this table is used to calculate the distance from the villages in our sample to the next geographic feature, e.g. to the next river, as further specified in section 4.4.

3 Tests for experimental integrity

3.1 Tests for balance

For all balance tests, we will display p-values and q-values corrected for multiple testing using the Benjamini and Hochberg (1995) procedure. We will also test the joint equality of all three group means and the maximum pairwise difference between any two group means and divide this by the standard deviation of the variable, following Imbens (2015).

[3]We took GPS readings for the centre of the village, defined as where the *kebele* authority office is located, at both baseline and endline. We cross-check between these two points and the GPS data for household locations within the village to identify any incorrect location observations, using whichever of the baseline or endline observation is closest to the majority of households.

Table 1: Administrative data sources

GIS object	Source	Year
Cities	1994 population census	1994
Health Centers	FAO Environment and Natural Resources Service (SDRN)	2007
Market Centers	IFPRI/FAO Environment and Natural Resources Service (SDRN)	2004
Rivers	FAO Environment and Natural Resources Service (SDRN)	2007
Roads	Woody Biomass Inventory and Strategic Planning Project (WBISPP), Ministry of Agriculture and Rural Development	2004

3.1.1 For intense-treatment and intense-placebo villages

At individual level at baseline We already ran and reported analysis of balance between the treatment, placebo and control groups on individual and household-level controls in Bernard et al. (2014). We will reproduce these variables in the long-run paper. We test if all three groups (treatment, placebo, within-village control) have equal means on:

- gender;
- age;
- highest completed school grade;
- the frequency with which individuals watch TV, listen to the radio, travel outside the district or have lived outside the district;
- the asset measure, answered by the household head, captured which of a list of durable goods – tools, furniture, electrical goods and carts or bicycles – the household owned and their estimated resale value. Land and houses were not included;
- variables measuring the composition of the household: number of household members, number of children below age 6, between ages 6-9, 10-12, 13-15, >15, number of adults over 60, as well as number of males and females in household;

- the baseline values of our outcome variables (either individual- or household-level) described in Section 5.

We will control as far as possible for any outcome variables where there are significant differences over all three groups (at 10% significance level after multiple testing correction) in our analysis. Our strategy for accounting for missing data at baseline in the pure control villages is detailed in Section 4.4.

At village level at baseline At baseline, we only collect village-level data for intense-treatment and intense-placebo villages. We report tests for balance to demonstrate experimental integrity in the village-level randomisation by testing whether the intense-treatment and intense-placebo villages have equal means across the following village-level characteristics: # of inhabitants in village; # of households in village; a dummy for main ethnicity = Oromo; % share of inhabitants belonging to the main ethnicity; % of inhabitants that are: Muslim, Orthodox, other Christians; hectares of: agricultural land, cultivated land, irrigated land, grazing land, forest; a dummy for most important crop = sorghum; a dummy for most important livestock = Oxen; dummies for most important income: subsistence, cash crops; % of inhabitants that are: farmers, farm laborers, non-agricultural business owners, non-agricultural business workers; dummies for inhabitants sell products: in public, in local shop, in nearby markets; % of inhabitants that do subsistence farming; dummies for only walking access to village in: the rain season, the dry season; a dummy for the most important transportation = mules/donkey/horse; costs for trip to nearest market (in ETB); walking time to nearest market (in min.); distance to nearest market (in km); agricultural wages (for men); a dummy for village has electricity; a dummy for main source of drinking water = pond/river/stream.

3.1.2 Between pure control, intense-treatment and intense-placebo villages

To obtain completely comparable samples in intense-treatment, intense-placebo and pure control villages, we would have censused pure control villages at baseline. However, we would have entered villages and had some effects on villagers' expectations and behaviour. We thus conduct tests of whether pure control villages, for which we only have endline data, are similar to the intense-treatment and intense-placebo villages that we observe over time (from baseline to endline).

Variables collected prior to baseline using GIS data for all villages These are collected before baseline in the years stipulated in Table 1.

- Distance to next closest village in the sample
- Distance to next closest intense-treatment village in the sample
- Distance to next city (in km); distance to next health center (in km)

- Distance to next market place (in km)
- Distance to next river (in km); distance to next road (in km)

Variables collected in community-level data # of inhabitants; # of households; a dummy for main ethnicity = Oromo; % share of inhabitants belonging to the main ethnicity; % of inhabitants that are: Muslim, Orthodox, other Christians; hectares of: agricultural land, cultivated land, irrigated land, grazing land, forest; a dummy for most important crop = sorghum; a dummy for most important livestock = oxen; dummies for most important income: subsistence, cash crops; % of inhabitants that are: farmers, farm laborers, non-agricultural business owners, non-agricultural business workers; costs for trip to nearest market (in ETB); walking time to nearest market (in min.); distance to nearest market (in km); dummies for whether the village has: electricity, a kindergarten, a first cycle school, a second cycle school, a secondary school, a preparatory school; dummies for whether the village receives: radio transmission, TV transmission, mobile network; % of households with: radio, TV, mobile phone; a dummy for main source of drinking water = pond/river/stream; a dummy for the most important transportation = mules/donkey/horse.

3.2 Accounting for attrition

Our main specifications will be estimated for respondents in intense-treatment and intense-placebo villages who are surveyed in all three rounds and control respondents who are surveyed in the endline.[4] We will use the following approaches to test if attrition could have influenced the estimated treatment effects:

- We will estimate our main regression specification using indicators of attrition as the outcome variable to test whether households in different treatment groups attrited differentially. We exclude households in pure control villages, who did not complete baseline. We will use the following indicators of attrition:
  - Between baseline and midline 2, after six months;
  - Between baseline and endline, after five years;
  - Between midline 2 and endline.[5]
- We will regress the control variables (listed in Section 4.4), including total assets, on attrition to assess whether attrited households differ in baseline characteristics.

[4]Some households are sampled for inclusion in the baseline but do not complete the baseline. These households are excluded from treatment and therefore are not included in any of the analysis.

[5]We will also examine attrition between the baseline and immediate follow-up. However, attrition may differ between treatment/placebo and control households as interviews were in different locations: treated and placebo households were interviewed at the screening venue and had to travel there, whereas control households were interviewed at home. The main purpose of this data collection was to keep control households at their homes during the video screening, so we treat estimates, including of attrition, with some caution.

We will then modify our main analysis in the following

- We will estimate our main specification using the baseline characteristics of attriting households as the outcome variable, excluding non-attriters and the control group.
- If we find worrying levels of differential attrition, we will report results using bounds assuming positive and negative selection into attrition (Lee, 2009). We will trim the outcome distribution (respectively from above and below) until the proportion of non-missing and non-trimmed observations is the same for all groups.
- If we find attrition above roughly 15 percent of our baseline sample, we will report weighted regression results, where the weights equal the inverse probability of attrition, estimated from a logit regression of the attrition indicator on the set of variables that significantly predict attrition. We will winsorize weights at the 5th and 95th percentiles. These adjustments will account for the fact that control households have fewer "opportunities"' to attrit than treatment households and so may have a different probability of attrition.

3.3 Adjustments for noncompliance

All models will be estimated in intention-to-treat format: the indicators for treatment will reflect assigned treatment. As shown in Bernard et al. (2014), we have high levels of compliance with treatment assignment.[6] We thus provide the Intention-to-Treat (ITT) effect of the intervention. Given the high rates of compliance, these effects are unlikely to differ substantially from the Average Treatment Effects on the Treated (ATT). We use the entire sample of respondents who were given tickets, including those non-compliers who missed the screening or attended the incorrect screening.

4 Estimation of treatment effects

We will focus our analysis on a set of three primary outcomes defined in Section 2. We might expect that the error terms across the eight regressions are correlated, in which case estimating the system of seemingly unrelated regressions (SUR) improves the precision of the coefficient estimates (Zellner, 1962). Simultaneous estimation also allows us to perform tests of joint significance on the treatment coefficient across equations. SUR is equivalent to OLS when the

[6] Respondents were told that the screening was an entertainment show, tickets were non-transferable, they could only attend the screening at the time and place on the ticket, and each respondent would receive a bag of sugar after the screening if they attended the correct screening. Screening venues were usually farmers' training centres located between villages. On average, people walked 29 minutes travel time to the screening venue, so people not invited to the screenings were unlikely to walk to the centre. Ticketing was tightly controlled at the screening to ensure compliance . Tickets had the name and survey identifier of each respondent on them.

error terms are in fact uncorrelated between regressions or when each equation contains the same set of regressors.

4.1 Midline treatment effects

The following regressions estimate the impact of the intervention after six months, when we do not have any data on households from pure control villages. We estimate effects on all individuals present in both baseline and midline in the 64 intense treatment and intense placebo villages. In all of the following regressions, we will cluster the standard errors at household level and include village fixed effects to take care of time-invariant village-specific characteristics.

Our main and preferred specification is regression 1.1, including the treatment T_i and placebo P_i dummies. The regression is run on data from the midline follow-up period $t = 1$.

$$y_i = \alpha + \beta_1 T_i + \beta_1' P_i + X' \tau_1 + u_v + \epsilon_i \tag{1.1}$$

where

$$T_{it} = \begin{cases} 1 & \text{if } i \text{ was invited to watch the documentary} \\ 0 & \text{if } i \text{ was not invited to watch the documentary} \end{cases}$$

and

$$P_{it} = \begin{cases} 1 & \text{if } i \text{ was invited to watch the placebo movie} \\ 0 & \text{if } i \text{ was not invited to watch the placebo movie} \end{cases}$$

and where y_i is the outcome variable of interest of individual i at time t, X' is a vector of control variables, and ϵ_{it} is an error term. We will include village fixed effects u_v in the short-run regressions to control for potential unobservable heterogeneity across our sample villages.

The three hypotheses we intend to test in the short run are based on the main specification 1.1:

1. The treatment has no effect relative to the within-village control: $\beta_1 = 0$
2. The placebo treatment has no effect relative to the within-village control: $\beta_1' = 0$
3. The treatment has no effect relative to the placebo treatment: $\beta_1 = \beta_1'$

4.2 Long run treatment effects

Our main specification is Equation 1.

$$y_i = \alpha + \beta_1 T_i + \delta_1 C_i + \beta_1' P_i + X' \tau_2 + \epsilon_i \tag{1}$$

where

$$T_i = \begin{cases} 1 & \text{if } i \text{ was invited to watch the documentary} \\ 0 & \text{if } i \text{ was not invited to watch the documentary} \end{cases}$$

and

$$P_i = \begin{cases} 1 & \text{if } i \text{ was invited to watch the placebo movie} \\ 0 & \text{if } i \text{ was not invited to watch the placebo movie} \end{cases}$$

and

$$C_i = \begin{cases} 1 & \text{if } i \text{ is living in a treatment village, but was not invited to watch the documentary} \\ & \text{(i.e. a spillover household)} \\ 0 & \text{otherwise} \end{cases}$$

Again, y_i is the outcome variable of interest of individual i at endline, X' is a vector of control variables (cf. section 4.4), and ϵ_i is an error term. This specification includes the individual-level treatment T_i, placebo P_i and within-village spillover C_i dummies. The regression is run on individual-level data from the endline $t = 2$.. Standard errors will be clustered at village level.

We will not include village fixed effects in our long-run specifications, since the lack of treatment variation in our pure control villages potentially leads to biased estimators when including village fixed effects. Instead, we will control for a range of village characteristics described in Section 4.4.

4.2.1 Hypothesis testing

The principal hypotheses we intend to test in the long run regressions are based on the specification 1:

1. The treatment has no effect relative to the pure control group: $\beta_1 = 0$
2. There is no spillover effect on untreated households in treatment villages: $\delta_1 = 0$
3. The placebo treatment has no effect relative to the pure control group: $\beta_1' = 0$
4. The treatment has no effect relative to the placebo treatment: $\beta_1 = \beta_1'$

4.3 Coping with outliers

We trim the sample for all continuous outcome variables used in the paper. Individuals who report values on the outcome variable which are four standard deviations or more above or below the sample mean have that outcome variable replaced as missing. We use the same procedure on the other outcome variables considered. We only lose small percentages of observations through trimming (1-3%).

4.4 Control variables

4.4.1 Midline analysis

In the individual-level and household-member-level regressions, we will control for age at baseline, gender, marital status and highest completed school grade. In the household-level regressions, we will control for age and gender of the household head at baseline and the highest completed schooling grade of the household head. These controls are selected as they are likely to be time-invariant and we can use the endline values in the pure control group at endline. We will also show robustness of our results to including total assets, marital status, and any individual- or household-level characteristics that are imbalanced at baseline, as specified in section 3. as control variables.[7] We exclude these from the main specification as they are time variant and are not available for the pure control group at baseline for the long-run regressions.

To ensure maximum comparability between short-run and long-run estimations, we do not include the baseline value of the outcome variables (y_{it-1}) in our main regressions, as we do not have these data for the pure control villages in the long-run regressions and wished to run a simple regression without any imputation. We will, however, show that our results in the short-run are robust to including the lagged outcome variable (as in an ANCOVA specification) (McKenzie, 2012).

In the long-run specification, we use village-level control variables instead of village fixed effects. We will show that our short-run results are robust to including these instead of village fixed effects.

4.4.2 Endline analysis in all villages

We control for the same basic, largely time-invariant individual-level variables as in the short run (only 6 percent of the sample is not married at baseline and there are very few changes in marital status between baseline and midline).

We do not have a baseline survey in the pure control villages. As with the short-run estimations, we wish to test robustness to including values of the outcome variable at baseline, and in addition to including control for total assets and marital status. For these observations and variables, we set the values of the variables equal to the sample means. This generates classical measurement error in the baseline measures and so attenuates their coefficient estimators. However, this does not affect the point estimates of the treatment group indicators (in expectation) because treatment assignment is random.[8] We estimate the variance-covariance matrix of the estimated coefficients using a clustered bootstrap. We resample village clusters, stratifying by treatment status, and impute the values of the missing baseline variables within

[7]The asset measure, answered by the household head, captured which of a list of durable goods – tools, furniture, electrical goods and carts or bicycles – the household owned and their estimated resale value. Land and houses were not included;

[8]See Jones (1996) for a proof of this claim and background discussion.

each resample. This approach follows Shao (1996).[9]

As we do not include village fixed effects, we examine a set of village-level controls. None are likely to be influenced by the video intervention as they are largely determined by external factors. Nonetheless, the distance variables are collected from administrative data collected before baseline, so we conduct a robustness check to including just those variables. These are:

- Number of inhabitants
- Hectares covered by forest
- Dummy for whether sorghum is the main crop
- Costs of trip to nearest market
- Dummy for whether the village has a first cycle school
- Percentage of households with radio
- Distance
  - to the next market place
  - school
  - farmers training center
  - health center
  - river.

4.5 Heterogeneity

We will estimate heterogeneous treatment effects in both short-run and long-run data. This allows us to understand which sub-groups contribute disproportionately to average treatment effects. We estimate these on the following dimensions:

1. Respondent is female
2. A dummy variable for whether the respondent's education level is above/below the sample median level of education
3. A dummy variable for whether the respondent's value of household assets at baseline is above/below the sample median level.
4. Respondent's expectations index at baseline

[9] See the simulations attached to the pre-analysis plan for Orkin et al. (2017) at https://www.socialscienceregistry.org/trials/991.

5. Respondent's internal locus of control score at baseline

Variables 1 and 2 are assumed to be time-invariant for the adults in the sample and so are observed for the control group. We use the values at endline as these are measured in the same way across all four treatment groups. Heterogeneity by variable 3, 4 and 5 can only be tested for the treatment, placebo and within-village controls as this variable is not available at baseline. These will be estimated by augmenting equation 1.1 and 1 to include a vector of interactions between the vector of treatment dummies and the measure of interest, as well as a term for the measure of interest itself.

In the short run, we report only tests for heterogeneity in effects for the treatment group compared to the within-village controls as we do not anticipate being powered to detect heterogeneity in placebo effects. In the long run, we report only tests for heterogeneity in effects between the treatment individuals, compared to people in pure control villages. We discuss multiple testing below in Section 5

We may employ a supervised learning approach for an exploratory analysis of heterogeneous effects. The causal tree (CT) method estimates the conditional average treatment effect $\tau(\mathbf{X})$ by partitioning the covariate space $\mathbf{X}$ with a modified regression tree algorithm (Athey and Imbens, 2016). The CT method confers several advantages over the analysis described above for exploratory analysis. First, it avoids overfitting by optimizing a cross-validation criterion for comparing causal effects. Second, the method allows us to identify heterogeneity without specifying baseline characteristics ex ante. Third, we can conduct valid inference over a large set of covariates without concern for multiple comparisons. However, we will be informed by the extent to which we find main effects.

For heterogeneous effects, we will correct p-values on our interaction terms for multiple testing using FDR-adjusted q-values. We report standard p-values and sharpened q-values that control the false discovery rate (FDR), adjusted based on the number of outcomes tested. Rather than pre-specifying a single q, we report the minimum q-value at which each hypothesis is rejected, following Anderson (2008a) and Benjamini, Krieger, and Yekutieli (2006). These are calculated across the number of outcomes per interaction term, most relevant for determining whether heterogeneous effects are statistically significantly different than zero. We do not include FDR adjusted q-values where we correct for the number of interactions (dimensions of heterogeneity) for a given outcome, which is most relevant for determining if the magnitude of heterogeneous effects varies across dimensions of heterogeneity, as we do not anticipate being powered for such tests.

4.6 Peer effects

As mentioned previously, the number of people invited to the documentary screenings was varied by village to assess the importance of peer effects. We have data on the four closest friends of each individual i in the treatment villages and hence know how many of her friends were also invited to the documentary or placebo screenings. For both the short-run and the long-run regressions, we will interact our treatment and placebo indicators with the number of peers invited to the respective screening in order to capture peer reinforcement mechanisms behind the main treatment effects of the interventions. We will lodge an addendum to this pre-analysis plan to specify this additional analysis. But robustness to different peer effects specifications is of less interest for variables where there are no main treatment effects, so we propose to use the initial analysis to limit the number of tests run.

5 Theory of change and primary outcomes of interest

For some groups of outcomes, we group several related variables into index variables. We construct the indices following Anderson (2008b). First, for each outcome variable y_{jk}, where j indexes the outcome group and k indexes variables within outcome groups, we re-code the variable such that high values within the index are similar in meaning. We then compute the covariance matrix $\hat{\sum}_{\mathbf{j}}$for outcomes in outcome group j which consists of elements:

$$\hat{\sum}_{jmn} =[i = 1]N_{jmn} \sum \frac{y_{ijm} - \bar{y}_{jm}}{\sigma^y_{jm}} \frac{y_{ijn} - \bar{y}_{jn}}{\sigma^y_{jn}} \tag{2}$$

Here, $N_{jmn}$is the number of non-missing observations for outcomes n and m in outcome group j, $\bar{y}_{jm}$and $\bar{y}_{jn}$ are the means for outcomes n and m in outcome group j, and σ^y_{jn} and $\sigma^y_{jm}$are the standard deviations for the outcomes in the whole sample at baseline (we do not create indices for variables which we only measure at endline).

Second, we invert the covariance matrix. We define weight w_{jk} for each outcome k in outcome group j by summing the entries in the row of the inverted covariance matrix corresponding to that outcome:

$$\hat{\sum}_{\mathbf{j}}^{\mathbf{-1}} = \begin{bmatrix} c_{j11} & c_{j12} & \cdots & c_{j1K} \\ c_{j21} & c_{j22} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ c_{jK1} & \vdots & \ddots & c_{jKK} \end{bmatrix} \tag{3}$$

$$w_{jk} = \sum_{l=1}^{K_j} c_{jkl} \tag{4}$$

Here, $K_j$is the total number of outcome variables in outcome group j. Finally, we transform each outcome variable by subtracting its mean and dividing by the standard deviation, and then weighting it with the weights obtained as described above. We denote the result $\hat{y}_{ij}$ as this transformation yields a generalized least squares estimator:

$$\hat{y}_{jk} = \left(\sum_{k \epsilon K_{ij}} w_{jk} \right)^{-1} \left(\sum_{k \epsilon K_{ij}} w_{jk} \frac{y_{ijk} - \bar{y}_{jk}}{\sigma^y_{jk}} \right) \tag{5}$$

We show robustness to the index construction method of Kling, Liebman, and Katz (2007a), as in Casey, Glennerster, and Miguel (2012, 1784–5). In this approach, estimation of the mean treatment effect similarly 1) orients each outcome so higher values indicate "better outcomes"; 2) standardises outcomes by translating each into standard deviation units (i.e. subtracting the mean and dividing by the standard deviation); 3) imputes missing values at the treatment assignment group mean; 4) compiles a summary index that gives equal weight to each individual outcome component. Where an outcome value is missing for a respondent, we will omit this outcome from the index construction, following Kling, Liebman, and Katz (2007b).

The major difference is whether or not outcomes are weighted by the inverse of the covariance matrix. In the Anderson (2008b) approach, if two measures are strongly correlated, then this gives each of them less weight than two measures which are not strongly correlated. This is appropriate if strongly correlated measures are measures of the same underlying concept, which seems plausible within families but may not empirically be the case. The Kling, Liebman, and Katz (2007a) method gives equal weight to all three measures. This is perhaps more appropriate for measures of different underlying concepts that may be grouped together but may be less closely correlated.

5.1 Theory of change and primary hypotheses

In this section we list the outcome variables of interest. Organising our reporting of these outcome variables is informed by a theory of change of how we may think about how the intervention affects beliefs and psychological characteristics, how in turn the intervention may affect actual behaviour, and how changes in behaviour may (or may not) affect economic outcomes. The RCT design ensures that any changes in both psychological characteristics and beliefs and in behaviour in the short run can be attributed relatively cleanly to exposure to the intervention, not to other changes in opportunities and constraints in treatment villages.

The interventions were chosen because they are likely to have a direct effect on specific psychological outcomes, consistent with evidence from psychological and economic research. The experiment identifies causal links between the psychological intervention and changes in psychological outcomes. The concepts used in our work are aspirations and locus of control. We expected our interventions to improve these in the short run, as they did. In the long run, these effects may or may not persist, but are unlikely to reverse. We test this in sub-hypotheses tested under hypothesis 1: that the intervention improves self-beliefs.

However, even if changes in self-beliefs do not persist in the long run, they may result in changes in behaviour in the short or long run. The experiment does not establish that changes in psychological outcomes cause changes in economic outcomes without further assumptions.

The theory of change specifies these assumptions. It suggests that the interventions, via changes in psychological outcomes such as aspirations and locus of control will affect decision-making about investment and other forward-looking behaviours in the short run. We expected that behavioral decisions that can be made without the need for substantial additional resources or inputs were more likely to be affected in the short-run, and thus focused short run measurement on such variables, examining hypothetical decisions (such as professed demand for credit) as well as decisions on how to spend or save resources (savings and investing in education of children). The experiment identifies causal links between exposure to the psychological intervention and changes in behaviour. We saw evidence of changes in labour supply and education investment in the short run. In the long run, we examine if these changes in behaviour persist, informing H2, that the intervention increases labour supply and human capital investments.

We saw some evidence of changes in future-oriented economic behaviour on the limited set of variables we measured: professed demand for credit, actual credit and savings. In addition, we examine if there are changes in investment-oriented behaviour (the flow of inputs), such as expenditure on agricultural inputs. We did not collect data on these variables in the short run: no effects may mean either that the intervention did not change these outcomes at all in the face of severe resource constraints, or that changes occurred in the short run but did not persist. These inform H3, that the intervention affects future-oriented economic behaviour.

These variables are the eight primary outcomes of our analysis and will be included in multiple comparison adjustments. We run all tests above for these outcomes/indices. Outcomes flagged by * can only be analysed in the endline data. We test variables of three types: psychological outcomes, in particular self-beliefs; labour supply and human capital investments, and economic behaviour.

Each sub-hypothesis (e.g. H5a) consists of either one index or one summary measure. The summary measure performs the same function as an index. It is either an aggregate measure (e.g. the constructed consumption aggregate) or one variable selected as being the most important of variables in the group (e.g. total savings).

We show naïve p-values in all cases. We use the Benjamini-Hochberg resampling procedure to increase power.[10] First, we show q-values which correct for the fact that we conduct eight tests across the eight primary hypotheses. Second, we calculate sharpened q-values as if we had corrected only over the indices or focal outcomes within the hypothesis (four in H1, two in H2, and two in H3). We present both q values as we have stronger evidence in favour of H1 and H2 than H3.

Within each sub-hypothesis (i.e. over the individual outcomes that make up the index, or the set of outcomes other than the focal outcome), we follow the same procedure and report standard p-values and sharpened q-values. In correcting for multiple comparisons, we adjust across outcomes within each of the sub-hypotheses, but not across them.

[10]The FDR does not depend on the size of the family and has a lower probability of under-rejecting while still controlling a reasonable object (i.e. the rate of false rejections among all rejections. The family-wise error rate (FWER) is more conservative as it controls the probability of at least one false rejection among the tested hypotheses.

We clearly mention which variables were available in the midline analysis in Bernard et al. (2014). For all long-term analysis, results will be reported that clearly refer to those that we conducted in the short-run in this midline report.

5.1.1 H1: The intervention affects self-beliefs

- H1a: The intervention increases aspirations for the future
- H1b: The intervention increases expectations for the future
- H1c: The intervention increases people's beliefs in their ability to control their own circumstances
- H1d: The intervention decreases people's belief in the extent to which their lives are controlled by chance

5.1.2 H2: The intervention increases labour supply and human capital investments

- H2a: The intervention increases labour supply to work
- H2b: The intervention increases investment in education

5.1.3 H3: The intervention affects future-oriented economic behaviour

- H3a: The intervention affects savings and credit choices
- H3b: The intervention affects investment-oriented behaviour (the flow of inputs)*

5.2 Secondary hypotheses

Within our theory of change (and consistent with psychological priors) we expect that the intervention itself is unlikely to change other psychological outcomes, such as risk or time preferences, beliefs about the returns to current opportunities, or subjective wellbeing, at least in the short run. We aimed to exclude alternative explanations for any effects we see and improve our understanding whether the psychological mechanisms we theorise are in fact present. We did not find results in the short run. We examine them in the long run for completeness, but do not have a clear hypothesis on their behaviour and are not particularly interested in them, so we exclude them from any multiple testing correction. If there are behavioural changes, such as in the form of investment or activity changes consistent with forward-looking behaviour, then indirectly, these outcomes may change in the long-run.

A related potential mechanism is that households may simply have derived information on the particular activities discussed in the videos and undertaken these activities, rather than changing aspirations, self-beliefs or locus of control and undertaking a broader set of future-oriented behaviours. We test whether treated respondents are more likely to undertake the particular activities mentioned in the videos. However: this is not a perfect test. Respondents may also have diversified more to reduce risk or invested in more information gathering, and as a result may be more likely to undertake these economic activities. In either case, such changes are likely to be small and are very much second order.

H5 is also of secondary interest and we consider it to be a more exploratory hypothesis. Cumulatively, if the intervention changed self-beliefs, if it changed behaviour, and if households changed behaviour in ways that improved their economic position, the intervention might have led to higher asset stocks or better welfare or wellbeing. It is unlikely that these outcomes are affected in the short run, and the time frame and intensity of intervention involved may not be enough to detect such effects even in the long run. Nonetheless, examination of these variables is of exploratory interest.

We do not aggregate variables from the four sub-hypotheses in H5 as our theory of change does not predict that all outcomes in H5 will move in one direction, nor in which direction they might move. Some of them (e.g. consumption and investment in assets) can even plausibly be expected to move in opposite directions in response to an intervention that promotes more future-oriented behaviour. Rather, we calculate both naïve p-values and FDR-adjusted p-values over the five focal outcomes and indices listed in H5. We suggest that the naïve p-values are most illuminating for these variables as we have prior interest in the specific, separate outcomes (Kling, Liebman, and Katz, 2007b; Casey, Glennerster, and Miguel, 2012).

- H4: The intervention does not affect other psychological channels
  - H4a: The intervention does not affect risk aversion
  - H4b: The intervention does not affect discount rates
  - H4c: The intervention encourages respondents to undertake activities mentioned in the videos
- H5: The intervention affects household welfare
  - H5a: The intervention affects household consumption
  - H5b: The intervention affects household food security*
  - H5c: The intervention affects household income*
  - H5d: The intervention affects the stock of assets*
  - H5e: The intervention affects subjective wellbeing

5.3 Omnibus tests

We will also conduct "omnibus tests" as a summary measure of the effectiveness of the interventions, following only the Kling, Liebman, and Katz (2007b) index construction method. We will conduct these tests on a single outcome index. This outcome index will be constructed by averaging the eight primary outcomes/indices listed in Section 5. Before averaging, the individual outcomes will be standardized to have mean zero and standard deviation one. In the long run, we will test if the aspiration treatment has a different effect to the placebo and if there are are significant differences between within-village spillovers and pure controls.

5.4 Outcome variables

In this section, we list the primary psychological outcome variables. We anticipate these will be most strongly affected by the intervention. All variables in this section 5.4 are measured at individual level.

5.4.1 Self-beliefs

Aspirations and expectations We use survey data to construct measures of aspirations in four specific dimensions: income, wealth, social status and children's educational attainment. For each of these dimensions, respondents were asked two questions: what level on this dimension they would like to achieve (which we refer to as "aspirations") and what level they thought they would reach in ten years (which we refer to as "expectations"). The survey instrument's validity and reliability was tested in 2009 in 16 villages in central Ethiopia (Bernard and Taffesse, 2014).[11] Income, measured in Ethiopian birr (ETB) includes cash income from all activities. Wealth focuses on durable wealth (including housing, vehicles, furniture and other valuable durables). Education was measured in the years of schooling the respondent wanted their eldest child to complete. We include codes for different types of post-school education available in Ethiopia, so completing a three-year university degree is 15 years of education, while a one-year diploma is 13 years. Social status was measured as the percentage of community members who would ask for the respondent's advice at times of important decisions.

We construct an aspirations index and an expectations index based on a standardised average of aspirations or expectations measured over four dimensions: income, wealth, education and status. We will also conduct a robustness check where we construct the indices using respondents' weights. We asked respondents to weight the four dimensions according to their own assessment of each dimension's significance for them, which accounts for heterogeneity in valued attributes of life. We used these weights to aggregate the standardised responses to each of the four dimensions into an aspirations index. In particular, let a_i^k be individual

[11] The validity and reliability tests were performed on the aspiration indicator only and rested on a slightly different wording, namely "what is the level that (they) would like to achieve in their life". The phrase "in your life" was removed so respondents would report the highest achievement they sought rather than the level at the end of their life. Results from (Bernard and Taffesse, 2014) suggest the measure had high reliability and validity, provided experiences enumerators are used. The enumerators in this study were all experienced. Two days of the two weeks of survey training were dedicated to the administration of the aspiration-related questions.

i's aspiration for dimension k. w_i^k is the weight that individual i assigned to this dimension. μ_i^k and σ_i^k measure the sample mean and standard deviation at baseline on dimension k. The aspiration index is thus $A_i = \sum_k((a_i^k - \mu_k)/\sigma_k)w_i^k$. Analogously, the expectation index is$E_i = \sum_k((e_i^k - \mu_k)/\sigma_k)w_i^k$. Overall, our measurement approach is similar to , except that each aspiration constituent is numerical (as opposed to categorical) and weights are specific to the individual.

Variable	Definition	Source
Aspirations index	Annual income score, (durable) wealth score, social status and oldest child's education level (four components)	
Annual income	Annual income is the amount of cash income you earn from all agricultural and non-agricultural activities, and money from PSNP or other programmes. What is the level of income that you would like to achieve?	2.1.2
(Durable) wealth	The value of your assets is the worth of your house, your furniture, consumer goods like a TV and fridge and any transport vehicles. What is the level of assets that you would like to achieve?	2.2.2
Social status	Someone has high social status if all people in the village ask their advice for an important decision. Someone has medium social status if half (50%) of the people in the village ask their advice for an important decision. What is the level of social status that you would like to achieve?	2.3.2
Oldest child's education level	What is the level of education that you would like your oldest child to achieve?[12]	2.4.2
Expectations index	Annual income score, (durable) wealth score, social status and oldest child's education level (four components). Same definitions as under aspirations above.	
Annual income	What is the level of income that you think you will reach within ten years?	2.1.3
(Durable) wealth	What is the level of assets that you think you will reach within ten years?	2.2.3
Social status	What is the level of social status that you would like to achieve?	2.3.3
Oldest child's education level	What is the level of education that you think your oldest child will achieve?[13]	2.4.3

[12]For households without children, we calculate our aspirations index as an average of the three remaining dimensions.

[13]For households without children, we calculate our aspirations index as an average of the three remaining

People's belief in their ability to control their own circumstances In this index and the next, we use sub-scales from two existing scales. Locus of control is defined as "a generalised expectancy pertaining to the connection between personal characteristics and/or actions and experienced outcomes" (Lefcourt, 1991, 414). We use the Internal and Chance scales of the Internal, Powerful Others and Chance scale (Levenson, 1981), omitting those which were not appropriate to the rural Ethiopian context.

From sociology and political science, we use the Attributions for Poverty scale (Feagin, 1972, 1975) to measure people's perceptions of the causes of poverty among people in general, rather than only in their own lives. We use a version adapted for China (Shek, 2003) (a shorter but less accurate version is included in the World Values Survey (Abramson and Inglehart, 1995)). The scale assesses the extent to which respondents agree with each of three types – Individualistic, Structural and Fatalistic – of explanations for poverty. These groupings echo the groups used in the IPC scale.

To measure people's belief in their ability to control their own circumstances, we use the Internal scale from the IPC scale, which captures if people see outcomes as contingent on their behaviour. We use the Individualistic Causes of Poverty scale to measure if individuals use Individualistic explanations for poverty.

People's belief in the extent to which their lives are controlled by chance The Chance scale from locus of control captures whether individuals think chance, luck or fate affects their outcomes, The Fatalistic explanations for poverty scale captures if individuals explain poverty as being caused by fate, chance or luck.

Variable	Definition	Source
Individual agency beliefs index	Internal locus of control score and Causes of poverty score - individual (2 components)	
Internal locus of control sub-scale	Five items scored from 1 to 4. Scores are summed.	3.4.3, 3.11, 3.4.12, 3.4.13, 3.4.15

dimensions.

Individual causes of poverty sub-scale	Triple factor structure - belief in individualistic, structural and fatalistic causes of poverty (Hunt, 2004).[14] Items 1, 2, 3 and 11 load onto the individualistic factor, items 8, 9 10 and 12 load onto the fatalistic factor, and items 4, 5, 6 and 7 load onto the structural factor, which we do not use. Four items scored from 1 to 4. Scores are summed.	3.6.1, 3.6.2, 3.6.3, 3.6.11
Beliefs regarding destiny index	Chance locus of control score and Casuses of poverty score - fate(2 components)	
Chance locus of control sub-scale	Five items scored from 1 to 4. Scores are summed.	3.4.1, 3.4.4, 3.4.5, 3.4.6, 3.4.9
Fate causes of poverty sub-scale	Defined as above. Four items scored from 1 to 4. Scores are summed.	3.6.8, 3.6.9, 3.6.10, 3.6.12

List of items originally in the questionnaire

Source	Definition
Locus of control	*I am going to read you some statements people often say about their lives. Please say whether you strongly disagree, disagree, agree or strongly agree.*
3.4.1	To a great extent my life is controlled by accidental/chance happenings.
3.4.3	When I make plans, I am almost certain/guaranteed/sure to make them work.
3.4.4	Often there is no chance of protecting my personal interests from bad luck happenings.
3.4.5	When I get what I want, it's usually/mostly because I'm lucky.
3.4.6	My experience in my life has been that what is going to happen will happen.
3.4.9	It's not always wise for me to plan too far ahead because many things turn out to be a matter of good or bad fortune.
3.4.11	I can mostly determine what will happen in my life.

[14]The original paper does not specify any restrictions on the covariances between error terms or factors, nor on factor loadings or error coefficients.

3.4.12	I am usually able to protect my personal interests (I can usually look after what is important to me)
3.4.13	When I get what I want, it's usually because I worked hard for it.
3.4.15	My life is determined by my own actions.
Perceptions of causes of poverty	*I am going to read you some reasons people give for why poor people are poor. Please say whether you strongly disagree, disagree, agree or strongly agree whether these reasons are reasons that people are poor.*
3.6.1	They lack the ability to manage money or other assets
3.6.2	They waste their money on inappropriate items (e.g. alcohol, cigarettes, gambling)
3.6.3	They do not actively seek to improve their lives
3.6.8	They have bad fate/destiny
3.6.9	They lack luck
3.6.10	They have encountered misfortunes
3.6.11	They are not motivated because of food aid (e.g. direct support programme, food parcels from NGOs not during famine)

Note on construction of psychological scales The questionnaire contained scales to assess the impact of treatment on psychological outcomes. The psychological scales were refined via item analysis. Items that met any of the following criteria were removed: low corrected item-total correlation (0.25); increased Cronbach's α if item removed; low item variation (80% identical responses on the item); low loading on primary unrotated factor (< 0.30), and high cross-loading (> 0.30) (Lamping et al., 2002). In addition, the proportion of missing responses across all respondents for a given item will be taken as an indication of poor comprehension and acceptability of the item, with 20% item non-response leading to the removal of that item from the scale. We score each scale according to the instructions in the original literature. If respondents did not answer all items in an index, we code the items they do not answer as missing and average over the items they respond to.

5.4.2 Labour supply and human capital investment

Household-member-level outcomes are denoted by *; household-level outcomes are denoted by $^{+}$. Otherwise, outcomes are at individual-member level, drawing on household roster data.

Education investment We repeat the analysis conducted in the midline data analysis for the sample of children of school-going age as our main analysis, with the number of children of school-going age who are enrolled as the focal outcome.[15]

Focal outcome:Children 6-15 enrolled	Number of school-aged children in the household who are enrolled in school	A.14, A.19
Expenditure on children's schooling (in ETB) during past year	Expenditure on children's schooling (in ETB) during past two months, including uniforms and shoes, pens/pencils/other stationary and exercise books, textbooks, and schooling fees/donations to school	Sum of : school fees, school books, uniforms, money sent students etc; U1 Consumption: U20*U21 (161 only) – 12 month recall (binary yes/no variable *value)

In addition, we conduct a secondary analysis of variables at household member level for all children and young adults who might have been affected by their parents being affected by the aspirations intervention i.e. those between age 0 and age 20. We correct over the tests conducted within this hypothesis, but not over household-member and household level tests.

	Is child enrolled now?	A.14, A.19
Absenteeism	Absenteeism in past two weeks	A.15[16]
Time in school	Hours per day (during past month) spent on schooling and formal education	A.21
Time studying	Hours per day (during past month) spent studying outside of school	A.21

Labour supply In terms of human capital investments, we will consider consider as well as whether more labour was put into income-generating activities.

Variable	Definition	Source

[15]Education spending is reported as part of the consumption module at endline. It is reported as expenditure on children's schooling over the past 12 months. In the baseline and midline, it is asked as being about the period between September and December in the most recent school year, so for a period of only 3 months. We will conduct analysis on the raw data in the endline survey but are cautious in comparing the magnitude of the coefficients between midline and endline.

[16]This outcome does not exist for baseline due to SurveyCTO error. Standardization of this outcome for the creation of an index will be done with respect to within village controls in the six month follow up round.

Focal outcome: Labour supply to income-generating activities	Hours per day (during past month) spent on paid activities, work on family farm/business	A.21
Leisure	Hours per day (during past month) spent in play time/ general leisure (including time taken eating, drinking, bathing, sleeping)*	A.21

5.4.3 Economic behaviour

In this section, we document a series of behavioural outcomes. In line with the theory of change, we focus on those outcomes one can relatively enter, such as hypothetical decisions (not requiring actual commitment, such as well as outcomes reflecting decisions on how to spend or save resources (savings or flows into assets, investing in education of children). We include some others, such acquiring new land (within the confines of Ethiopian legal framework for land acquisition), although they are less likely to show changes in the short run. Some of these were either not available in the short run (midline) analysis within the survey instruments. The variables below are ordered to reflect first human capital investments followed by other capital investments.

Savings and credit* A first group are variables linked to savings and credit. Savings would arguably be easier to engage in than acquiring credit, as the latter depends on lenders willing to lend to the households involved.

Variable	Definition	Source
Focal outcome: Total savings	Total amount saved (in ETB)	sum of 6.8
Savings	Does the respondent have any savings?	6.1
Saving outside home	Is respondent saving outside the home?	6.4
Borrowing	Did respondent borrow from an institution, group or individual during the past 6 months? (all loans, including borrowing money from the iqqub or from family or neighbours)	8.1
Credit	During year 2015, did respondent receive any crop inputs or agricultural equipment on credit?	8.17
Loans	Total amount taken in loans (in ETB)	sum of 8.7

Hypothetical loan repayable in 1 year (in ETB)[17]	Someone from a microfinance institution came to you and offered to lend you any amount of money you ask without charging interest or service charge. How much would you ask for if the loan is payable in 1 year?	D.6.1
Hypothetical loan repayable in 5 years (in ETB)	Someone from a microfinance institution came to you and offered to lend you any amount of money you ask without charging interest or service charge. How much would you ask for if the loan is payable in 5 years?	D.6.3
Hypothetical loan repayable in 10 years (in ETB)	Someone from a microfinance institution came to you and offered to lend you any amount of money you ask without charging interest or service charge. How much would you ask for if the loan is payable in 10 years?	D.6.5

Investment-oriented behavior - flow$^+$ As these are farmers, one way of investing will be on the land, and if possible, in land, as well as in livestock (as this is a mixed farming context). Hence we consider spending on inputs in agriculture, including seeds, fertiliser, etc, as well as labour inputs and livestock inputs. Land is not privately owned, but farmers can have user rights, and may cultivate more land by engaging in land rental or sharecropping. All crop-related expenditure in this section is for the last long rains growing season. This analysis cannot be performed in the short-run as less than one seasonal cycle had passed by then (compared to pre-intervention - midline was only 6 months after intervention).

Variable	Definition	Source
Aggregate Investment in Livestock and Agriculture	Sum of two components below: expenditure on agricultural inputs and expenditure on inputs for livestock and poultry.	
Total expenditure on agricultural inputs in the most recent long rains season	Expenditure on this season, on hired and purchased farm inputs, and the imputed value of seeds saved from the last harvest.	J6v(seed purchased) + J7v(fertiliser) + J8(pesticide) + J9(tractor+other non-labour) + [J4v(seed saved) + J5v(seed bartered)]* price of seed from seed purchased

[17] The three hypothetical loan levels are measured on household level and answered by the household head.

Total expenditure on inputs for livestock and poultry activities	Expenditure over the past 12 months, on the purchase of animals, feed, and other services required by the activity incl. hired labor.	Q9+Q10+Q11 (Q9: animal purchase, Q10: feed, Q11: other inputs)
Total labour supply in person days in agriculture in the most recent long rains season		Family labor (J10 - days * wages) across all plots and crops
Total labour supply in person days from hired labor in agriculture in the most recent short and long rains season		Hired labor (J11 - days * wages) across all plots and crops
Total land area under cultivation		I1A10
Total land area rented or sharecropped in for agriculture	Land allocations to households in Ethiopia are fairly fixed because of state land ownership, but households can rent land	Sum of H2 (in hectares) if H6 = 5 or 6 (H2 = land area; H6 = 5 $\Rightarrow$ rented land; H6 = 6 $\Rightarrow$ sharecropped in land)

5.4.4 Eliminating alternative psychological mechanisms

We go some way toward linking any changes in particular psychological outcomes – aspirations and locus of control – to changes in behaviour by employing very targeted, theoretically motivated psychological interventions that have been shown in other contexts to change these psychological outcomes and also alter behavior. In this section, we also test if the treatment causes changes in other psychological mechanisms (changes in risk/time preferences and changes in behaviour that simply mimic the characters in the videos). Our prior is that we would not find effects in these mechanisms. If we do not, it strengthens the argument that any changes in behaviour occur because of changes in aspirations and locus of control. All variables in this section 5.4.4 are measured at individual level.

Subjective discount factor We use a survey-based measurement tool used by Cole et al. (2013a) in India and Hill, Hoddinott, and Kumar (2013) in Ethiopia to construct individual subjective discount factors. The scale and logistics of the study meant that a survey-based tool was chosen over an experimental tool. We find a very similar distribution over categories to Hill, Hoddinott, and Kumar (2013) in Ethiopia. The outcome variable is the subjective discount factor $\beta = \frac{1}{1+\delta}$, where δ is the rate of time preference. We ask if respondents would prefer receiving 100 ETB now or 125 ETB in one month. To those who chose 125 ETB, one ETB in one month is worth between 0.8 and 1 ETB today: they have a monthly discount factor between 1 and 0.8. We assign them the mid-point of 0.9. If they chose 100 ETB, they then choose between 100 ETB now or 150 ETB in one month. If they choose 150 ETB, they have a monthly discount factor between 0.8 and 0.667 and are assigned the midpoint of 0.733. Those who have a discount factor lower than 0.667 are asked how much they would need to be given in one month to choose to wait.[18]

Risk preferences We use survey-based instruments to calculate risk preferences, following the line of enquiry by Binswanger (1980) and in line with Cole et al. (2013a) and Hill, Hoddinott, and Kumar (2013). Individuals were presented with two hypothetical decisions. The first asked which of five hypothetical payouts they would choose if the payout was determined by a coin toss Cole et al. (2013a). The second asked about the amount of price risk individuals would choose when selling surplus grain output Vargas Hill (2009). It had the same structure of payouts as the first question but multiplied by 100. All payouts had the same, constant probability, as in a coin toss, which is simple to explain to respondents. The measure is outlined in detail in the appendices to Bernard and Taffesse (2014).

Information about new activities, technologies or returns Provided impacts can be detected, it is of interest to explore possible hypotheses about mechanisms. We can explore whether the decisions taken by farmers in terms of investment are at all related to the decisions that the subjects in the documentaries had told to camera. A transcription of the stories told allows us to check whether certain actions identical to those in the documentaries were shown. Again, this is not treated as a outcome, but merely as an exploration of possible mechanisms.

Variable	Definition	Source

[18] This measurement assumes a linear utility function, and will estimate a discount rate which is biased upwards (and a discount factor which is downward biased) if the function is actually concave. More complex measures are available to address this (Andersen et al. (2008); Andreoni and Sprenger (2012)), but we do not use these because of the very limited numeracy of our respondents.

Information index	Seven components: as specified below	
Earns any income from trading	In Teyiba video	R2, code 102 (charcoal), 103 (agricultural), 104 (other trading)
Earns any income from grain milling	In Teyiba video	R2, code 105
Took advice on agricultural extension index	In Beshir video	Index of eight dummies: M28 (visited demonstration plot in last year), M20 (attended community meeting on agriculture), M24 and M25 (expert visited or requested a visit), M4 (expert visited), I2.9 a), b) or c) (received advice from extension agent)
Planting cash crops, vegetables or fruits	In Beshir video	I1A.9 - is crop grown?
Used pump	In Beshir video	I2.10 is any plot irrigated?
Building stone bands and terracing	In Immortal Treasure video	M.1, code a)
Water conservation/water harvesting	In Immortal Treasure video	M7 or 12.11, option 6
Water storage	In Immortal Treasure video	12.12 option 5
Grows maize, teff, mango, sugarcane or coffee	In the exemplary achievement video	I1A.9 - is crop grown?
Took advice on agricultural extension index	In the exemplary achievement video	Index of eight dummies: M28 (visited demonstration plot in last year), M20 (attended community meeting on agriculture), M24 and M25 (expert visited or requested a visit), M4 (expert visited), I2.9 a), b) or c) (received advice from extension agent)
Crop rotation and shifting	In the exemplary achievement video	K1.11
Uses water pump	In the exemplary achievement video	E.4.10
Irrigation practise	In the exemplary achievement video	I2.10 is any plot irrigated? or 12.10, 12.11

Keeping cattle	In the exemplary achievement video	Q2 for code 82
Earns any income from grain milling	In the exemplary achievement video	R2, code 105

5.4.5 Welfare

In this section, we present a set of welfare, income and asset outcomes that may be affected in the long-run. One should be conscious of two points: the time period is not necessarily long, so it may not be enough for any welfare outcomes to be observed. Furthermore, even if some improve, one would not necessarily expect that all would move in the same direction. For example, it is possible for food security to improve but not overall consumption or vice-versa, or households to have invested more but not necessarily consuming more as well, potentially even initially less. Individual-level outcomes are denoted by *; household-level outcomes are denoted by $^{+}$.

Household welfare^{+}

Variable	Definition	Source
Focal outcome: Constructed Consumption Aggregate	Sum of the consumption expenditure of the household for a 12 month period. Includes consumption from the following sources: purchases, production for self consumption, and barter, gifts, loans and wages in kind.	Constructed as the aggregate of food, non-food small scale and non-food lumpy consumption over the past 12 months.
Food Consumption	We ask respondents about food consumption over the past 7 days. This includes food purchased, received via barter, gifts, loans, wages in kind and self production.	[(U3q*U3u)+(U4q*U4u)+(U6q*U6u)]*U5 (Constructed as (qty*qty unit)*price; wherein qty unit refers to a conversion factor for standardisation)
Non-food small scale consumption	Sum of frequent non-food consumption, with a recall period of one month.	(U17*U18) (U17: binary yes/no variable for whether consumed, U18: total value)
Non-food lumpy consumption	Sum of non-frequent non-food expenditure with a recall period of 12 months. Includes expenditure on education.	(U20*U21) (U20: binary yes/no variable, U21: total value)

Consumption of sin goods	Value of cigarette and tobacco consumed in the past 30 days.	(U17*U18): for 191 only (U17: binary yes/no variable, U18: total value)
General economic position	Self reported measure of general economic position of the household.	In general, would you describe your household as: 1 = Doing well - able to meet household needs by own efforts and making some extra for stores, savings and investment; 2= Doing just ok - able to meet household needs by own efforts but with nothing extra to save or invest; 3= Struggling - managing to meet household needs but by depleting productive assets and/or sometimes receiving support. Productive assets are used to generate income e.g. plough, donkey for transport; 4= Unable to meet household needs, dependent on support from community or government

Food security^{+} We use a version of the United States Department of Agriculture's food insecurity questionnaire (Bickel et al. (2000), Andrews et al. (1999)) adapted for Ethiopia (Hadley et al., 2008).

Variable	Definition	Source

Food security index	Standardized index of food security items	Standardized index constructed using the score of all the items. The sum of the scores of households with children and without children is weighted according to the definition in Bickel et al. (2000).
	We worried whether our food would run out before we got money to buy more.	All items are scored as: Always - on nearly all of last 30 days = 4; Often true - on more than 7 days in the last 30 = 3; Sometimes true - on less than 7 days = 2; Never true = 1
	The food that we bought just didn't last, and we didn't have money to get more.	
	We relied on only a few kinds of low-cost food (e.g. no vegetables or meat) to feed the children because we were running out of money to buy food.	
	We had to eat some food we did not want to eat because we could not afford to buy other food (e.g. wild food, immature crops, discarded food)	
	Adults in the household cut portion sizes or skipped meals because there wasn't enough money for food	
	Children in the household cut portion sizes or skipped meals because there wasn't enough money for food.	
	We had to ask others outside the household for food or money to buy food.	
	Adults in the household went for a whole day without eating because there was not enough food (do not include fasting for religious reasons)	

	Children in the household went for a whole day without eating because there was not enough food (do not include fasting for religious reasons)	
Food shortage in lean season	How many months in the last 12 Ethiopian months did you have problems satisfying the food needs of the household?	B.1

Revenue from production+[19]

Variable	Definition	Source
Revenue Aggregate (Household Level)	Recall period: 12 months, constructed using the four outcomes described below.	Sum of all outcomes outlined below.
Paid labour income	Sum of agricultural and non-agricultural wages received by the household over the past 12 months	R2*R4*R5: for items 110 and 111
Revenue from livestock and poultry in the last 12 months	Total revenue from livestock and poultry in the last month	Livestock Sale: (Q6*Q7); Q6: Qty sold, Q7: Average price received; Livestock Activities: Q8 ; Q8: Includes rental, dairy, wool, egg sale etc.
Revenue from crops in the most recent short rains and long rains season	Total revenue from crops in the most recent short and long rains season	Computed as the quantity harvested*price.[20]

[19] As a robustness check, we will analyse the impact on net income from production, which would be calculated using the revenue from production and expenditure on inputs defined in the variable Investment-oriented behaviour: flow.

Non-farm revenue	Total revenue from all non-farm activities	R2*R4*R5; Multiplicative result for items: 101-109, where R2 indicates whether this form of income exists for the household; R4: number of operational months; R5: average income per month

Investment-oriented behavior/Assets - stock$^+$

Variable	Definition	Source
Focal Outcome: Asset Aggregate	Aggregate constructed using the assets owned by the household, as outlined below.	Sum of three components outlined below.
Value of productive durables	Sum of present purchase value of productive goods used in agriculture and enterprise owned by the household.	E4.1+E4.2+E.4.3+ E.4.4 +E.4.5+E.4.6 +E4.8+E4.15+E4.17 +E4.18+E4.19+E4.20 = [A]*[B] (Calc as price*quantity owned)
Value of non-productive durable assets	Sum of present purchase value of non-productive goods owned by the household.	E4.7+E4.9+E.4.10 +E.4.11+E.4.12+ E.4.13+E.4.14+E.4.16 +E.4.21 +Others= [A]*[B] (Calc as price*quantity owned)
Value of livestock and poultry	Total value of livestock currently owned by the household	Q4a (Qty) * Price[21]

Subjective wellbeing *

[20] We follow the methodology in Beegle et al(2012), i.e. using median kebele, village and sample prices in case the household level price is not available.

[21] We follow the methodology in Beegle et al (2012), i.e. using median kebele, village and sample prices in case the household level price is not available.

Variable	Definition	Source
Subjective well-being index	Best life, happiest life (2 components)	
Best life	We measure life satisfaction by showing respondents a picture of a ladder with 10 steps (Cantril, 1966). They are told the top of the ladder represents the best possible life for them and the bottom step represents the worst possible. They are then asked, "Where on the ladder do you feel you personally stand at present?"	3.11.1
Happiest life	The above question was repeated with the top and bottom of the ladder representing the happiest and most miserable possible life.	3.11.2

5.4.6 Outcomes analysed in midline working paper

We have already analysed the midline data and written up the results in a working paper (Bernard et al., 2014). No pre-analysis plan had been formally deposited before this short-run analysis. For the purpose of scientific integrity, the table below reports all outcomes analysed in this midline working paper. If we analyse a variable in the midline and endline data, we construct the variable in the same way. There is one exception: Education spending is reported as part of the consumption module at endline. It is reported as expenditure on children's schooling over the past 12 months. We will conduct analysis on the raw data in the endline survey but are cautious in comparing the magnitude of the coefficients.

Variable	Definition	Source
Aspirations index	Annual income score, (durable) wealth score, social status and oldest child's education level (four components)	
Annual income	Annual income is the amount of cash income you earn from all agricultural and non-agricultural activities, and money from PSNP or other programmes. What is the level of income that you would like to achieve?	2.1.2
(Durable) wealth	The value of your assets is the worth of your house, your furniture, consumer goods like a TV and fridge and any transport vehicles. What is the level of assets that you would like to achieve?	2.2.2

Social status	Someone has high social status if all people in the village ask their advice for an important decision. Someone has medium social status if half (50%) of the people in the village ask their advice for an important decision. What is the level of social status that you would like to achieve?	2.3.2
Oldest child's education level	What is the level of education that you would like your oldest child to achieve?[22]	2.4.2
Expectations index	Annual income score, (durable) wealth score, social status and oldest child's education level (four components). Same definitions as under aspirations above.	
Annual income	What is the level of income that you think you will reach within ten years?	2.1.3
(Durable) wealth	What is the level of assets that you think you will reach within ten years?	2.2.3
Social status	What is the level of social status that you would like to achieve?	2.3.3
Oldest child's education level	What is the level of education that you think your oldest child will achieve?[23]	2.4.3
Time in farm work	Hours per day (during past month) spent with work on family farm/business	A.21
Leisure	Hours per day (during past month) spent in play time/ general leisure (including time taken eating, drinking, bathing, sleeping)	A.21
Savings	Does the respondent have any savings?	6.1
Total savings	Total amount saved (in ETB)	sum of 6.8
Borrowing	Did respondent borrow from an institution, group or individual during the past 6 months? (all loans, including borrowing money from the iqqub or from family or neighbours)	8.1
Loans	Total amount taken in loans (in ETB)	sum of 8.7
Children 6-15 enrolled	Number of school-aged children in the household who are enrolled in school	A.14, A.19

Expenditure on children's schooling (in ETB) during past two months	Expenditure on children's schooling (in ETB) during past two months, including uniforms and shoes, pens/pencils/other stationary and exercise books, textbooks, and schooling fees/donations to school	C.5.1+C.5.2 +C.5.3 +C.5.4; Household: Section C; Sum of : school fees, school books, uniforms, money sent students etc; U1 Consumption: U20*U21 (161 only) – 12 month recall (binary yes/no variable *value)
Hypothetical loan repayable in 1 year (in ETB)[24]	Someone from a microfinance institution came to you and offered to lend you any amount of money you ask without charging interest or service charge. How much would you ask for if the loan is payable in 1 year?	D.6.1
Hypothetical loan repayable in 5 years (in ETB)	Someone from a microfinance institution came to you and offered to lend you any amount of money you ask without charging interest or service charge. How much would you ask for if the loan is payable in 5 years?	D.6.3
Hypothetical loan repayable in 10 years (in ETB)	Someone from a microfinance institution came to you and offered to lend you any amount of money you ask without charging interest or service charge. How much would you ask for if the loan is payable in 10 years?	D.6.5

Subjective discount factor	We ask if respondents would prefer receiving 100 ETB now or 125 ETB in one month. To those who chose 125 ETB, one ETB in one month is worth between 0.8 and 1 ETB today: they have a monthly discount factor between 1 and 0.8. We assign them the mid-point of 0.9. If they chose 100 ETB, they then choose between 100 ETB now or 150 ETB in one month. If they choose 150 ETB, they have a monthly discount factor between 0.8 and 0.667 and are assigned the midpoint of 0.733. Those who have a discount factor lower than 0.667 are asked how much they would need to be given in one month to choose to wait	Different variables in 4.1
Constant relative risk aversion parameter: Risk aversion (coin toss) & market game	We use survey-based instruments to calculate risk preferences, following the line of enquiry by Binswanger (1980) and in line with Cole et al. (2013b) and Hill, Hoddinott, and Kumar (2013). Individuals were presented with two hypothetical decisions. The first asked which of five hypothetical payouts they would choose if the payout was determined by a coin toss (Cole et al., 2013b). The second asked about the amount of price risk individuals would choose when selling surplus grain output (Vargas Hill, 2009). It had the same structure of payouts as the first question but multiplied by 100. All payouts had the same, constant probability, as in a coin toss, which is simple to explain to respondents.	Different variables in 4.3
Chance locus of control sub-scale	Five items scored from 1 to 4. Scores are summed.	3.4.4, 3.4.5, 3.4.9, 3.4.6, 3.4.1
Internal locus of control sub-scale	Five items scored from 1 to 4. Scores are summed.	3.4.3, 3.4.11, 3.4.12, 3.4.13, 3.4.15

Fate causes of poverty sub-scale	Triple factor structure - belief in individualistic, structural and fatalistic causes of poverty (Hunt, 2004). Items 1, 2, and 3 load onto the individualistic factor, items 4, 5, 6 and 7 load onto the structural factor and items 8, 9 and 10 load onto the fatalistic factor. Three items scored from 1 to 4. Scores are summed.	3.6.8, 3.6.9, 3.6.10
Individual causes of poverty sub-scale	Defined as above. Four items scored from 1 to 4. Scores are summed.	3.6.1, 3.6.2, 3.6.3, 3.6.11
Best life	We measure life satisfaction by showing respondents a picture of a ladder with 10 steps (Cantril, 1966). They are told the top of the ladder represents the best possible life for them and the bottom step represents the worst possible. They are then asked, "Where on the ladder do you feel you personally stand at present?"	3.11.1
Happiest life	The above question was repeated with the top and bottom of the ladder representing the happiest and most miserable possible life.	3.11.2
Assessment of documentaries and placebo	Enjoyed watching what I saw, Discussed film I saw a lot with my neighbours, Discussed film others saw a lot with my neighbours, Discussed film I saw at least once with neighbours in the past two weeks, What I saw generated a lot of discussion within village	

References

Abramson, P. R. and R. F. Inglehart. 1995. *Value Change in Global Perspective.* Ann Arbor, Michigan: University of Michigan Press. 5.4.1

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutstrom. 2008. "Eliciting Risk and Time Preferences." *Econometrica* 76 (3): 583–618. 18

[22] For households without children, we calculate our aspirations index as an average of the three remaining dimensions.

[23] For households without children, we calculate our aspirations index as an average of the three remaining dimensions.

[24] The three hypothetical loan levels are measured on household level and answered by the household head.

Anderson, M. L. 2008a. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495. 4.5

———. 2008b. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495. 5, 5

Andreoni, J. and C. Sprenger. 2012. "Estimating Time Preferences from Convex Budgets." *American Economic Review* 102 (7): 3333–3356. 18

Andrews, M., M. Nord, G. Bickel, and S. Carlson. 1999. "Household Food Security in the United States." *Food Assistance and Nutrition Research Report* 8. 5.4.5

Athey, S. and G. Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *PNAS* 113 (27): 7353–7360. 4.5

Bandura, A. 1977a. "Self-Efficacy: Toward a Unifying Theory of Behavioural Change." *Psychological Review* 84 (2): 191–215. 2.3

———. 1977b. *Social Learning Theory*. Englewood Cliffs, New Jersey: Prentice-Hall. 2.3

Benjamini, Y. and Y. Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the royal statistical society. Series B (Methodological)* : 289–300. 3.1

Benjamini, Y., A. M. Krieger, and D. Yekutieli. 2006. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika* 93 (3): 491–507. 4.5

Bernard, T., S. Dercon, K. Orkin, and A. S. Taffesse. 2014. "The Future in Mind: Aspirations and Forward-Looking Behaviour in Rural Ethiopia." *Bureau for Research and Economic Analysis of Development Working Paper* 429: 1–42. 1, 2.3, 2, 3.1.1, 3.3, 5.1, 5.4.6

Bernard, T. and A. S. Taffesse. 2014. "Aspirations: An approach to measurement with validation using Ethiopian data." *Journal of African Economies* 23 (2): 189–224. 5.4.1, 11, 5.4.4

Bickel, G., M. Nord, C. Price, W. Hamilton, and J. Cook. 2000. "Guide to measuring household food security." *US Department of Agriculture, Food and Nutrition Service, Office of Analysis, Nutrition, and Evaluation http://www fns usda gov/fsec/FILES/Guide% 20to% 20Measuring% 20Household% 20Food% 20Security (3-23-00) pdf* . 5.4.5, 12

Bidwell, K., K. Casey, and R. Glennerster. 2016. "Debates: Voting and Expenditure Responses to Political Communication." *Unpublished* (Stanford Graduate School of Business): 1–42. 1

Binswanger, H. P. 1980. "Attitudes toward risk: Experimental measurement in rural India." *American journal of agricultural economics* 62 (3): 395–407. 5.4.4, 16

Cantril, H. 1966. "The human design." *The Pattern of Human Concerns (Rutgers University Press, New Brunswick)* . 15, 16

Casey, K., R. Glennerster, and E. Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *Quarterly Journal of Economics* 127 (4): 1755–1812. 5, 5.2

Cole, S., X. Gine, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery. 2013a. "Barriers to Household Risk Management: Evidence from India." *American Economic Journal: Applied Economics* 5 (1): 104–135. 5.4.4, 5.4.4

Cole, S., X. Giné, J. Tobacman, P. Topalova, R. Townsend, and J. Vickery. 2013b. "Barriers to household risk management: Evidence from India." *American Economic Journal: Applied Economics* 5 (1): 104–135. 16

Feagin, J. R. 1972. "Poverty: We Still Believe that God Helps Those Who Help Themselves." *Psychology Today* 6: 101–129. 5.4.1

———. 1975. *Subordinating the Poor: Welfare and American Beliefs*. Englewood Cliffs, New Jersey: Prentice Hall. 5.4.1

Hadley, C., D. Lindstrom, F. Tessema, and T. Belachew. 2008. "Gender bias in the food insecurity experience of Ethiopian adolescents." *Social science & medicine* 66 (2): 427–438. 5.4.5

Haushofer, J. and J. Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *Quarterly Journal of Economics* 131 (4): 1973–2042. 1

Hill, R. V., J. Hoddinott, and N. Kumar. 2013. "Adoption of weather-index insurance: learning from willingness to pay among a panel of households in rural Ethiopia." *Agricultural Economics* 44 (4-5): 385–398. 5.4.4, 5.4.4, 16

Imbens, G. 2015. "Matching Methods in Practice: Three Examples." *Journal of Human Resources* 50 (2): 373–419. 3.1

Jones, M. 1996. "Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression." *Journal of the American Statistical Association* 91: 222–230. 8

Kling, J. R., J. B. Liebman, and L. F. Katz. 2007a. "Experimental analysis of neighborhood effects." *Econometrica* 75 (1): 83–119. 5

———. 2007b. "Experimental Analysis of Neighbourhood Effects." *Econometrica* 75 (1): 83–119. 5, 5.2, 5.3

Lamping, D. L., S. Schroter, P. Marquis, A. Marrelb, I. Duprat-Lomon, and P.-P. Sagnier. 2002. "The Community-Acquired Pneumonia Symptom Questionnaire: a New, Patient-Based Outcome Measure to Evaluate Symptoms in Patients with Community-Acquired Pneumonia." *Chest* 122 (3): 920–929. 5.4.1

Lee, D. S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3): 1071–1102. 3.2

Lefcourt, H. M. 1991. "Locus of Control." In *Measures of Personality and Social Psychological Attitudes*, edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightsman. San Diego: Academic Press, 413–499. 5.4.1

Levenson, H. 1981. "Differentiating Among Internality, Powerful Others, and Chance." In *Research with the Locus of Control Construct Vol.1*, edited by H. M. Lefcourt. New York: Academic Press, 15–63. 5.4.1

McKenzie, D. 2012. "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2): 210–221. 4.4.1

Orkin, K., R. Garlick, J. Haushofer, R. Sedlmayr, and S. Dercon. 2017. "Short-Term Economic Effects of Psychological Affirmation and Goal-Setting Interventions." *Pre-analysis plan* (https://www.socialscienceregistry.org/trials/991). 9

Shao, J. 1996. "Bootstrap Model Selection." *Journal of the American Statistical Association* 91 (434): 655–665. 4.4.2

Shek, D. 2003. "Chinese People's Explanations of Poverty: The Perceived Causes of Poverty Scale." *Research on Social Work Practice* 13 (5): 622–640. 5.4.1

Vargas Hill, R. 2009. "Using stated preferences and beliefs to identify the impact of risk on poor households." *The Journal of Development Studies* 45 (2): 151–171. 5.4.4, 16

Zellner, A. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57: 348–68. 4

Zwane, A. P., J. Zinman, E. Van Dusen, W. Pariente, C. Null, E. Miguel, M. Kremer, D. Karlan, R. Hornbeck, X. Giné, E. Duflo, F. Devoto, B. Crepon, and A. Banerjee. 2011. "Being Surveyed can Change Later Behavior and Related Parameter Estimates." *Proceedings of the National Academy of Sciences* 108 (5): 1821–1826. 1