Not compromising: using experiments to measure women's power *

Vilas J. Gobin[†] Paulo Santos[‡]

This draft: March 2017

Comments are welcome

Please do not cite without authors' permission

Abstract

We propose an experimental approach to the measurement of women's power, which we also apply in the context of the evaluation of poverty graduation program. The experimental measure is compared with traditional survey-based measures of empowerment. We find that (1) the two measures are not correlated. We also find (2) a positive impact of participation in the program on empowerment when it is measured using an experimental indicator, but not when using traditional survey measures and that (3) the experimental indicator is better correlated with indicators of well-being usually associated with more empowered women.

*The data used in this study were supplied by The BOMA Project. We thank staff of The BOMA Project especially Kathleen Colson, Ahmed Omar, Meshack Omarre, Fredrick Learapo, Sabdio Doti, Bernadette Njoroge, Alex Villec, and Sean Kelly for their assistance in data collection as well as in facilitating the first author's stay in Kenya. All analyses, interpretations or conclusions based on these data are solely that of the authors. The BOMA Project disclaims responsibility for any such analyses, interpretations or conclusions.

[†]Economics, Monash University, Melbourne, Australia. Corresponding author (e-mail: vilas.gobin@monash.edu)

[‡]Economics, Monash University, Melbourne, Australia.

1 Introduction

In 2015, the Sustainable Development Goals included the achievement of gender equality and empowerment of all women and girls as its fifth goal. This decision reflects both the intrinsic value of gender equity and female autonomy and its perceived association with a variety of positive outcomes such as reduced fertility, increased child survival rates and the allocation of resources in favor of children within the household (see, for example, Duflo (2012) and the references therein).

These associations led many poverty alleviation programs to target women as their main beneficiaries and to indicate women's empowerment as a key output (van den Bold, Quisumbing, & Gillespie, 2013). The commonly assumed causal pathway between such interventions and empowerment is that the provision of additional resources (such as transfers) will increase women's bargaining power within the household, leading to an increased ability to make choices, from which the achievement of outcomes strongly aligned with their preferences then follow (Kabeer, 1999).

Because bargaining power is not directly measurable, most of the existing analysis relies on proxy measures such as self-reported participation in household decisions or control over assets (Carter et al., 2014). Although these measures have often been criticized for lack of rigour (van den Bold et al., 2013), measurement problems due to social desirability bias (Jejeebhoy, 2002), and potential inadequacy to local context (Schatz & Williams, 2012; Upadhyay & Karasek, 2012) their use has been widespread in the evaluation of the impact of many programs on women's empowerment.

One alternative to these measures is to use incentivized decision making tasks that create an environment in which bargaining power can be directly measured. In the analysis of de Palma, Picard, and Ziegelmeyer (2011), Carlsson, Martinsson, Qin, and Sutter (2013) and Braaten and Martinsson (2015), bargaining power is inferred by examining the influence of individual preferences on a couple's joint preferences and the analysis of differences between decisions made individually and jointly by a husband and wife. Almas, Armand, Attanasio, and Carneiro (2015) take a different approach, and use women's willingness to pay to receive a transfer instead of their partners as a measure of their capacity to make decisions.

The approach we propose provides a more direct measure of a woman's power, as we focus on the capacity to change her spouse's decision. In the experiment, women (and their spouses) play a modified version of the Gneezy and Potters (1997) risk elicitation task, in which the investment decision takes place in two stages. In the first stage, the first mover decides how much to invest in the risky lottery. The second mover then has the opportunity to change the decision. Because all participants first act as the first mover and then as the second mover in response to every possible first move by their spouse, we can define an indicator of power that reflects the woman's decision to not compromise from her own preferred investment decision. This indicator reflects traditional concepts of power in the social sciences which Dahl (1957), for example, defines as "A has power over B to the extent that he can get B to do something that B would not otherwise do": the woman, in making a decision to change the husband's investment decision to one that her husband would not otherwise choose (but she would) is exercising her power over her husband. [1]

We implement this task in the context of the evaluation of a poverty graduation program targeted at ultra-poor women in northern Kenya, the Rural Entrepreneur Access Project (REAP), which is briefly presented in the next section. The program was rolled-out in three funding cycles, and survey measures of women's decision making power were collected at baseline and at two follow-up surveys. In addition, and shortly after the last follow-up survey, the beneficiaries (and their spouses) from the first and last funding cycle, who benefited from the program one year apart from each other, were invited to a set of decision making experiments designed to directly measure women's bargaining power within the household. The two instruments (the experiment and the survey questions) are presented in detail in section 3, where we also contrast them with each other. We find that the experimental measure has little correlation with survey measures, naturally raising the question of which is preferable.

We address that question in two ways. In section 4, we estimate the impact of REAP on these measures of power. Participation in REAP leads to an increase in women's capacity to decide when such capacity is measured using the experimental

[1] See also Harsanyi (1962) or Russell (1938) for similar definitions of power.

measure, but not when it is measured using the traditional survey measures. Hence, our work also contributes to the literature on the impact of poverty graduation programs, as it raises the possibility that the muted impacts that have been identified in the literature may simply reflect an inadequate measurement of women's power. [2] In section 5 both the experimental and survey measures are correlated with indicators of well-being that one would expect to improve with women's power. The experimental measure is correlated with our measures of food security (number of nights a child goes hungry at night) and food consumption, but the survey measures are not, providing further evidence in favor of the latter. We conclude, in section 6, with a discussion of the limitations of this approach, in particular the need for further work in different contexts in order to validate this instrument.

2 The Rural Entrepreneur Access Project (REAP)

The Rural Entrepreneur Access Project (REAP) was designed as a poverty graduation program, combining multiple interventions with the aim of providing ultra-poor women in northern Kenya with a localized "big push" that would allow them to increase their incomes by simultaneously addressing the overlapping set of constraints that they might face. [3] The program was implemented in 14 locations in northern

[2] In the case of poverty graduation programs, the general conclusion from the use of survey measures of power is that there is little evidence of an impact. Banerjee et al. (2015), in an analysis of six poverty graduation programs across six countries, find that at the end of the program women have more say in decisions on health expenditures and home improvements but this gain does not persist after the end of the program, even though consumption and income gains do. They also show that the impact depends on the cultural context: the effect of these programs is larger in South Asian countries (India and Pakistan) compared to African (Ethiopia and Ghana) and Latin American countries (Peru and Honduras), which may either reflect initial differences in bargaining power or the inadequacy of the proxies used to measure empowerment, given that many of the survey items that form these measures are grounded in formative research from South Asia where women's empowerment may manifest differently (Malhotra, Schuler, & Boender, 2002; Schatz & Williams, 2012; Upadhyay & Karasek, 2012). Bandiera et al. (2016) similarly find no impact of a poverty graduation program, implemented in Bangladesh, on women's empowerment.

[3] This multifaceted approach to poverty was pioneered by BRAC through the program Challenging the Frontiers of Poverty Reduction – Targeting the Ultra-Poor (CFPR/TUP) (Goldberg & Salomon, 2011; Matin, Sulaiman, & Rabbani, 2008).

Kenya, a region where more than 80% of the population are estimated to live below the national poverty line and where gender inequality is estimated to be more than 25% higher than the national average (Kenya National Bureau of Statistics and Society for International Development, 2013). [4] These numbers likely reflect the importance of social and cultural norms that restrict women's ownership rights of livestock, the main livelihood option in the region. Despite women playing a key role in the management of livestock, their contribution usually goes uncompensated and their ability to dispose of livestock is greatly limited, with men usually having the final say in this decision (Fratkin, 2004).

During a limited period (2 years), participants of REAP would benefit from a set of interventions which included the development of business plans and mentoring, grants and the promotion of savings mechanisms. The sequence of these interventions is presented in Figure 1 and each intervention is briefly described below.

In November 2012 local committees, formed specifically for the targeting of this program, were asked to identify women who were among the poorest in the community, considered to be responsible and entrepreneurially minded, and willing to run a business with two other women. The committees, under the guidance of trained business mentors, identified 1755 women as being eligible for REAP, and once the women accepted the invitation to participate in REAP, the business mentor proceeded to form business groups of three women.

Due to a lack of capacity to simultaneously enrol all eligible women, the potential beneficiaries were divided into three groups to be successively funded over three cycles (March/April 2013, September/October 2013 or March/April 2014, hereafter groups A, B and C respectively). In order to be transparent and fair, and to be seen as such, a public lottery was used in each location to assign eligible women to one of the three funding cycles.

[4] Although the 2010 Kenyan Constitution explicitly gave women the same legal rights as men, women in Kenya continue to suffer from gender inequities which are exacerbated in rural areas (Nature Conservancy, 2013). In Marsabit County, where REAP was implemented, the Gender Inequality Index (GII) in 2013 was estimated to be 0.69 compared to the national average of 0.55 (Kenya National Bureau of Statistics and Society for International Development, 2013). The United Nations Human Development Report, 2015 ranks Kenya 126 out of 155 countries based on the GII.

In the month leading up to program enrolment, business mentors met with beneficiaries to assist with the development of a business proposal. Then, on the day of program enrolment, participants were required to attend a short business skills training session, delivered by mentors. At the end of this training sessions business groups were provided with a cash grant of USD 100 (PPP USD 237.97 at 2014 prices) to be used to establish their business, an amount which is equivalent to approximately 7.5 months of expenditure per capita. [5] The distribution of this initial grant was followed by a period during which a mentor regularly met with the business group to monitor its progress and offer advice and training.

Six months after the start of the business, groups were eligible for an additional grant of USD 50 (PPP USD 118.98) conditional on meeting the following criteria: two or more original members remained involved in the business, business assets were held collectively, and the business value (defined as the sum of cash on hand, business savings and credit outstanding, and business stock and assets) was equal to or greater than the initial grant. Participants were also required to participate in a short training session on savings, designed to provide a basic understanding of the formation and operation of savings groups. After this session and the distribution of grants, participants were encouraged to form a savings group (SG) or join existing ones. The savings group model introduced to participants during the training most closely resembled Village Savings and Loans Associations (VSLA), also known as Accumulating Savings and Credit Associations (ASCA), described in Allen (2006). The groups are self-managed and allow members to save money and access loans, which are paid back with interest. These groups would meet monthly, and during the first nine meetings mentors delivered additional training sessions on savings.

All eligible women who were selected in November 2012 were interviewed at baseline (November 2012) and at two follow-up surveys, conducted at six month intervals and timed to coincide with the beginning of each new funding cycle. Using this data,

[5] From hereon, all monetary values reported in this paper are in PPP terms at 2014 prices unless otherwise stated. We use the following PPP exchange rates to convert Kenya Shillings to USD PPP: 36.83 (2012), 38.38 (2013), 40.35 (2014). These values are then converted to 2014 prices by multiplying the ratio of the 2014 US Consumer Price Index (CPI) to the US CPI for the relevant year.

Gobin, Santos, and Toth (2016) take advantage of the sequential roll-out of the program, the randomized allocation to each funding cycle, the perfect compliance of observations to treatment and control groups, and the low attrition rate to identify the program impacts on a variety of outcomes. [6] They find that participation in REAP resulted in significant improvements in income, savings and asset accumulation, with the increase in income and savings being driven by women's participation in microenterprises and savings groups. Here, we estimate the effect of participation in REAP on women's empowerment, a hypothesized consequence of improvements in income earned by the woman, if such improvements translate into a stronger bargaining position within the household.

We consider two ways of measuring power. The first comes from a lab-in-the-field experiment to which all married women in groups A and C were invited (see section 3.1). [7] These experiments took place in June/July 2014, approximately 14 months (2 months) after group A (group C) participants were enrolled in REAP, and two months after the endline surveys were conducted (see Figure 1). During these two months participants in group C started to benefit from REAP: they participated in a business skills training session and received the initial grant to implement their business plan. In light of this, the impact of REAP on the experimental measure of empowerment estimated in this study will likely understate the true impact of REAP on empowerment. The second measure of power is derived from women's reports of their participation in decisions on five domains (purchase of food, purchase of household items, purchase of livestock, and paying for schooling of for medical fees) which was collected during the different surveys designed to evaluate the impact of this program. This information is used to construct a household decision making index (HDMI), a commonly used indicator of empowerment (see section 3.2). We describe both measures in more detail in the next section.

[6] Survey attrition is very low in both follow-up rounds of survey, with less than 4% of women not re-interviewed at the first follow-up (midline) and less than 6% of women not re-interviewed at the second follow-up (endline). Gobin et al. (2016) also check the validity of the experimental design. They find that the groups that were assigned to different funding cycles are balanced on baseline characteristics. They also show that spillover and program anticipation effects, if at all present, are of limited importance.

[7] The term married is used to refer to women who are either married or in a relationship.

3 Measuring power: experiments and surveys

3.1 Experimental measure

The experimental measure of power that we define is based on a modification of the risk elicitation task proposed by Gneezy and Potters (1997), which is changed to a sequential move game between the beneficiary and her spouse. In this modified task, a first mover decides how much to invest in a risky lottery (from a 200 Ksh endowment) and a second mover has the opportunity to accept or change the decision of the first mover. [8] Investments are restricted to 0, 50, 100, 150 or 200 Ksh. The investment is doubled with a probability of two-thirds and lost with probability one-third. [9] The Gneezy and Potters (1997) method was used because it is easier to understand and implement than other approaches, which seemed of central importance in the setting studied, where beneficiaries have less than one year of formal education. [10]

To be concrete, participants in the experiment (i.e, REAP beneficiaries and their spouses), were asked the following sequence of questions: [11]

- Suppose you are the first mover, how much will you choose to invest?
- Suppose you are the second mover, would you accept or overrule the first mover if they choose to invest [0, 50, 100, 150, 200] Ksh?

[8] At the time of the experiment the exchange rate was 1 USD to 85 Ksh and the average daily wage for menial labor was approximately 200 Ksh. The average daily consumption per capita, in our sample and at endline, was approximately 58 Ksh.

[9] We also use the Gneezy and Potters (1997) method in the more usual way, to elicit attitudes towards risk. In this case, the lottery is played individually, and participants have to decide how much to invest out of 100 Ksh, with possible investments restricted to 0, 25, 50, 75 or 100 Ksh. There are no other differences between these tasks. This version is played before the sequential version we just described.

[10] Other risk elicitation methods such as those used by Eckel and Grossman (2002) and Holt and Laury (2002) were also considered in a pilot study conducted in June 2013 but it was found that many women were making inconsistent choices when the Multiple Price List method was used, which was likely due to their lack of understanding of the task. See Charness, Gneezy, and Imas (2013) for a review of commonly used methods.

[11] See Appendix A for more details of the experimental procedure and Appendix B for the detailed instructions.

and, if the respondent decided to overrule,

- How much would you invest?

Participants were not allowed to communicate with each other in this task and they each played the role of the first and second mover. The strategy method (in which a responder makes decisions conditional on each possible element of the information set) was used primarily due to feasibility considerations, as it was easier to implement than the direct response method. [12] Although the scenario presented in this experiment is neutral with respect to the nature of the investment, it may approximate several instances when women have to make decisions about household expenditure. [13]

By focusing on decision making under risk, our design is similar to several other experiments on household decision making. However, while much of the previous work typically elicits individual and joint preferences, and then examines the influence of individual preferences on joint preferences as a way to determine relative bargaining power within the household (for example, Braaten & Martinsson, 2015; Carlsson et al., 2013; de Palma et al., 2011), the task we use measures women's decision making power within the household based on a comparison between the choices as the second mover (as a response to the potential decisions of their spouse) and the choice as first mover (which indicates their own preferred investment). As such, it closely reflects classic definitions of power such as, for example, Dahl (1957) where person "A has power over B to the extent that he can get B to do something that B would not otherwise do": women reveal their power by changing the investment

[12] Comparisons between the two methods find no meaningful differences between them in the case of sequential move games (see Brandts and Charness (2011) for a survey of this literature), although it is argued that hot vs cold decisions might make a difference in experiments in which emotions are involved. For example, Brosig, Weimann, and Yang (2003) find higher punishment by second movers in response to a selfish play by the first mover when the direct response method was used.

[13] It is not uncommon in northern Kenya for husbands to rely on women to execute their wishes with regards to household expenditure when they migrate with livestock herds (or to towns). In many cases the woman is left with some livestock (or is sent money) and has the choice to either follow the husband's wishes or to go against them, a decision that is similar to the one faced in this experiment. In this case, the experimental scenario can be seen as capturing a woman's decision to go against the wishes of her husband regarding what items to purchase.

decisions after both the husband's preference and her own are clearly stated.

As a second mover, a woman must first decide to either accept or change the potential investment decision made by her spouse (I_1^H) when this decision differs from her preferred investment (revealed while being the first mover, I_1^W). After that, and if a decision is made to change the husband's decision, the woman faces a second decision, of how much to invest (I_2^W). The use of the strategy method in this task results in five second mover decisions for each woman, in response to the five possible first mover investments by her spouse, i.e. I_1^H=$\{$0, 50, 100, 150, 200 Ksh$\}$, although only four of these decisions are informative about intrahousehold differences in power (as we can learn nothing about these differences when $I_1^W = I_1^H$).

We can then ask the following question: facing different investment decisions by her husband, does a woman choose her own preferred investment or does she compromise with the husband's decision? For those cases when $I_1^W \neq I_1^H$, we can summarize these decisions through a new variable, *compromise*, as follows:

$$compromise = \begin{cases} 0 & \text{if } I_2^W = I_1^W \neq I_1^H \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

Here, *compromise* is equal to zero if the woman decides to stick with her original investment decision. Alternatively, the variable *compromise* is equal to one if the woman chooses to accept her husband's decision or changes her husband's investment to some amount that is not equal to her original preference.

In addition to excluding those observations for which $I_1^W = I_1^H$ we also exclude inconsistent observations where, similarly, nothing can be learned from these decisions. This includes those observations where $I_1^H > I_1^W > I_2^W$, $I_1^W > I_1^H > I_2^W$, $I_2^W > I_1^H > I_1^W$, or $I_2^W > I_1^W > I_1^H$, as these decisions involve the woman choosing an amount I_2^W that lies outside of the interval between her investment as the first mover, I_1^W, and her husband's investment as first mover, I_1^H. The number of excluded observations for these reasons is, however, relatively small (6.5% of the total number of observations left after we disregard the cases where $I_1^W = I_1^H$).

We use the multiple observations of *compromise* for each woman to construct a

single index of *power*, which we define as a binary indicator that takes the value of one if, more often than not, the wife does not compromise with her husband (i.e. the average is less than 0.5), and zero otherwise. This index is the experimental measure of women's power.

3.2 Survey measures

As part of the household survey designed for the evaluation of the impact of REAP, we collected information on household decision making, through questions phrased as follows: "When you have to buy food, how is the decision made? Who has the final say?" Possible responses to this question were: husband only, wife only, husband and wife, or other. In addition to food, similar questions were also asked about decisions over the purchase of livestock, the purchase of household items, and the payment of children's medical expenses and of school fees.

The answers to questions such as these are widely used as proxy indicators for women's decision making power. In quantifying women's power, frequently identified with being the sole decider over household decisions, the answers to these questions are usually coded as one if the woman has the final say in the decision or zero if otherwise and then summed to produce a single index. This approach leads to the first household decision making index, which we label as *HDMI1*.

Recent work by Peterman, Schwab, Roy, Hidrobo, and Gilligan (2015) finds that substantially different conclusions on empowerment can be reached depending on whether one considers joint decision making and not only sole decision making, with women's power identified with having at least some say in household decisions. Given this result, the same responses are used to construct a second index of power, in which the response to each question is coded as one if the woman has the sole or joint final say in the decision and zero if otherwise, which we label as *HDMI2*.

3.3 Measuring power: comparing experiments and survey measures

Summary statistics for the experimental and survey measures of women's power are presented in Table 1. We start our discussion of these two measures by focusing on the decisions made by women in the experimental task.

As mentioned in section 3.1, only married women and their spouses were eligible for the experimental task, leaving us with a potential sample of 946 women who reported having a husband or being in a relationship (out of the 1167 women that participated in REAP in the two funding cycles analysed here). A total of 700 eligible women (and 229 of their spouses) accepted the invitation to participate in the experiment. [14]

As a first mover, the woman must decide how much of her endowment (of 200 Ksh) to invest in the lottery where her spouse has the chance to change her investment decision. Women's first mover decisions are summarised in Figure 2 where we show that, on average, they invest half of the endowment (99.8 Ksh). The decision of women, as second movers, to change the spouse's decisions, conditional on the restrictions identified in section 3.1 is reported in Table 2. A greater proportion of women change the spouse's investment to their own preference when the spouse's decision is at an extreme (i.e. $I_1^H = 0$ or 200), a result which is expected as women prefer investments that are away from the extremes (see Figure 2). On average, women chose to change their spouse's investment to their own preferred investment in approximately 40% of the cases (column (1)), and more than 25% of women change more than half of the spouse's possible first mover decisions to their own preferred investment (i.e., do not compromise in more than half of the cases).

Turning to the survey measures, we can conclude that women are, on average, involved in X decisions as sole decision maker, and Y decisions either as sole or joint decision maker. These averages mask important differences across decisions. For

[14] See Table C2 for more details on sample size. Note that, although the experiment was designed for couples, they were not required to make decisions simultaneously. Women who were unsure whether their spouses would show up to take part in the experiment, were still allowed to participate. Summary statistics for the male respondents are presented in **??**.

example, almost no women are sole deciders on the purchase of livestock, and only 16% report being involved at all in that decision. Perhaps surprisingly, women's involvement in deciding on payment of school fees or of medical expenses is not too different: roughly, only 1 in 5 women report being involved in these decisions. And although women's power is greater with respect to purchasing food, it is still the case that men are reported to be the sole deciders in almost 60% of the households.

A natural starting point when comparing these different measures is the examination of the correlation between experimental and survey measures of power. If the two measures were largely positively correlated, it is likely that preference should be given to measuring empowerment using survey measures, given both past experience with these questionnaires and an easier and less expensive implementation. The correlation coefficients are reported in Table 3. We find a high correlation between *HDMI1* and *HDMI2*, which was to be expected given the way these indexes are constructed. However, *power* is not correlated with either of the survey measures. [15]

There are two explanations for this result. The first is that both experimental and survey measures capture substantially different dimensions of empowerment. The second is that one of the measures does not appropriately reflect bargaining power, either because the experiment is too alien to women's decisions to meaningfully capture it, or because that is true for survey measures, developed and validated in a context that is substantially different from the one we study. [16]

We address these alternative explanations in two ways. Firstly, by quantifying the impact of REAP on the different measures of women's power. If the causal pathway from assets to agency that underlies much of the discussion of empowerment is real, then the identification of an impact of REAP on one of the measures of power would

[15] This conclusion is similar to Almas et al. (2015) who also use an incentivized task to derive a quantitative measure of empowerment and find that the correlation between the different measures goes against the expected direction.

[16] For example, the module on women's empowerment in the Demographic and Health Surveys, which is similar to the one used in this study, has been criticized for its inadequacy to Sub-Saharan Africa (Schatz & Williams, 2012; Upadhyay & Karasek, 2012), possibly reflecting the fact that survey items are grounded in formative research from South Asia (Malhotra et al., 2002), where women's empowerment may manifest itself differently.

provide some support for its validity. Alternatively, the identification of an impact on both would suggest that the two measures capture different aspects of women's power. Secondly, and in a way that is similar in spirit to the approach just outlined, we estimate the correlation between the different measures of power and two welfare indicators that are usually expected to be associated with women's power, food expenditure and child hunger.

4 The impact of REAP on women's power

To estimate the impact of REAP on women's power, we specify the following regression:

$$Y_i = \theta + \beta T_i + \varphi X_i + \epsilon_i \tag{2}$$

where Y_i is the outcome of interest for woman i (which is either *power*, *HDMI1* or *HDMI2*), T_i is a binary variable equal to one if woman i benefited from REAP in March/April 2013 and zero if she benefited in March/April 2014 (hereafter referred to as treatment and control groups respectively), and X_i is a matrix of control variables measured at baseline (collected prior to the first group of women being enrolled in REAP) that includes sub-location fixed effects as well as controls for the woman (and spouse's) age, her literacy and numeracy, whether she is in a polygamous marriage, the number of children and adults in the household, and the livestock and other durable assets owned by the household. [17] When Y_i is *power*, control variables for the husband's attendance at the experiments as well as the woman's investment decision in the individual risk task are also included. [18]

Several comments are in order before we present our estimates. Firstly, and given the nature of the data we have, collected during the first of a two year program, the estimates should be interpreted as the short-to-medium term impact of REAP. It is possible that the full impact would be larger for women who benefit from

[17] Stratification took place at the sub-location level (77 sub-locations).

[18] As mentioned above, we use the Gneezy and Potters (1997) task to elicit women's attitudes towards risk.

the program in its entirety. Secondly, it is important to notice that we are likely to underestimate the true impact of REAP on women's empowerment when we measure it using the variable *power* given that, at the time of the experiments, all participants had received funding. [19] Although we may expect empowerment to be a gradual process (that reflects participation in REAP, as women grow their enterprises, incomes and agency), we cannot dismiss the possibility that changes in women's power may take place quite early on in the program, especially given the findings of Gobin et al. (2016) who show that after just six months of participation in REAP beneficiaries' incomes are significantly higher, with this effect being driven by income earned from the REAP enterprise.

Finally, and again in the case of the experimental measure of power, a causal interpretation of the estimates that we present requires us to address the possible bias associated with the fact that not all eligible REAP beneficiaries participated in the experiments. As mentioned above, out of the 946 women that were eligible to participate in the experiments, only 700 attended. In Appendix C we show that participants are in fact observationally different from non-participants at baseline. To address the potential bias from the self-selection into the experiments, and following Nichols (2007, 2008), observations are reweighted such that the distribution of observable pre-treatment characteristics is identical for participants and non-participants. See Appendix C for the details of this correction.

Table 4 presents estimates of the impact of REAP on the experimental and survey measures of power (columns (1)–(4) and columns (5)–(12), respectively). For completeness, we present the estimates of the impact of REAP with and without additional control variables. In the case of *power*, we present both weighted and unweighted estimates. In the case of *HDMI1* and *HDMI2*, we present estimates of REAP for the entire sample of married women (i.e. including those who did not participate in the experiment but for whom survey measures were collected during the endline survey and for which concerns about possible selection bias are not relevant) and, to facilitate the comparison with the experimental measure of power, the

[19] As noted previously, when the experimental task took place, the treatment group had been enrolled in REAP for approximately 14 months and the control group for two months.

weighted estimates of the impact of REAP when we limit ourselves to the sub-sample of women who participated in the experiments. Our preferred specifications are presented in column (4) for the variable *power* and columns (6) and (10) for *HDMI1* and *HDMI2*, respectively.

Focusing first on the impact of REAP on *power*, our preferred estimates show that women who received funding from REAP one year earlier are 7% more likely to not compromise with their husbands in at least half of their decisions as the second mover, and this effect is significant at the 10% level. This conclusion is robust to whether we address concerns about self-selection into the experiments or not, but relies on the inclusion of control variables, even if individual effect of these control variables is, in general, insignificant at the usual levels of statistical significance.

This conclusion about the effect of REAP on women's power would not be supported had we measured this outcome solely using survey measures, as shown in columns (6) and (10). Turning first to the impact of REAP on *HDMI1*, the index that reflects power of the woman as the sole decision maker and focusing on our preferred specification (column 6), we find that this measure of power increases by approximately 5% but this effect is not precisely estimated, and is statistically insignificant at the usual levels of statistical significance. This conclusion is robust to the exclusion of control variables and, when focusing on the sub-sample of participants in the experiment, to the reweighting of the observations. This conclusion of a lack of impact of REAP on women's power does not change when we focus instead on the *HDMI2*, the index that reflects the power of a woman as the sole or joint decision maker.

There are several possible explanations for the differences in the estimates of the impact of REAP on women's empowerment depending on how we measure it. It can be that the time horizon might be too short to see changes in women's role in decisions over the various items that comprise these indexes, even if women seem now more capable of imposing their will in the experimental task. Alternatively, responses in the survey may be subject to social desirability bias or they may not capture empowerment in this context - in which case, a relatively neutral task such as the one we implemented would be more adequate. As an additional piece of evidence

in this comparison between the two measures we examine in more detail, in the next section, whether the experimental and survey measures of power are correlated with household outcomes that are expected to be positively correlated with empowered women, namely food expenditure and child hunger.

5 Women's power and welfare outcomes

The emphasis on empowering women as a development strategy reflects its perceived association with several positive outcomes, including the allocation of resources to food or in favor of children. If this expectation is valid, we should be able to conclude that higher levels of women's power are associated with better household outcomes, specifically, in our data, an increase in monthly food consumption per capita and a reduction of the number of nights that a child has gone to bed hungry in the past week. We use the quantification of these relations as an additional test of the validity of the two types of measures.

The OLS estimates of these correlations, after controlling for other variables that might be associated with the indicators of well-being (including the age of both the woman and her spouse, her literacy and numeracy, whether she is in a polygamous marriage, the number of children and adults in the household, and the livestock and other durable assets owned by the household) are presented in Table 5. The estimates presented in Table 5 are robust to the exclusion of these control variables. Taken as a whole, these estimates suggest that the experimental measure (*power*) is a better indicator of empowerment than either *HDMI1* or *HDMI2* in this context, although, as with other studies, our results are somewhat mixed. [20]

As shown in columns (1) and (3), there is a positive and precisely estimated relation between *power* and *HDMI2* and food consumption. That said, that rela-

[20]For example, Peterman et al. (2015) compare two indexes that are similar to *HDMI1* and *HDMI2*, to household consumption, and find that for a sample of women from Ecuador, there are significant correlations between the decision making indexes and household consumption but no associations between the indexes and dietary diversity but find the opposite result in Yemen, with no associations between the indexes and household consumption but significant correlations between the indexes and dietary diversity.

tion is more important, in both economic and statistical terms, in the case of the experimental measure. Although this would seem to suggest that *HDMI2* could be an appropriate and simpler way to measure women's power, the results in column (6) suggest otherwise, as this measure is the only one that is positively correlated with an *increase* in the number of nights that a child goes to bed hungry in the past week (a result that goes against the expectation that more empowering women lead to reductions in the incidence of children's hunger). The experimental measure, however, has no statistically significant effect on this outcome (column (4)).

6 Conclusion

This study presents an incentivized decision making investment task specifically designed to elicit a measure of intra-household bargaining power. In this task, a woman is allowed to change her husband's investment decision to a decision that she favors. Such decision is closely aligned with classic conceptualizations of power such as Dahl (1957) and others, with more powerful women choosing their own preferred investment levels, even when the husband's preference is clearly stated. Despite this advantage, experimental measures are often more difficult and more expensive to implement than traditional survey based questions, hence it seems important to understand whether there is, practically, an advantage in using the former over the latter.

We contrast the two measures in three steps. The first is a simple analysis of the correlation between the different indicators of power, both survey-based and experimental. We conclude that they are not correlated. There are two alternative explanations for this lack of correlation: either one measure is an adequate measure of power and the other is not, or they measure different (and orthogonal) components of women's agency. We explore the evidence in favor of one of these alternative explanations in two ways.

The first way relies on the conceptualized causal pathway from assets to agency that underlies much of the expectation of positive effects of programs such as microfinance, cash transfers or poverty graduation programs on empowerment. We use

data from a poverty graduation program, the Rural Entrepreneur Access Project (REAP), to quantify the impact of the set of interventions under this program on women's power. We conclude that the program empowers women when power is measured using the experimental measure that we develop, and that we would not be able to detect that effect had we limited ourselves to quantify empowerment using traditional survey measures. Hence, our analysis contributes to the literature on the impact evaluation of poverty graduation programs, by suggesting one explanation for earlier results that found no impact on female empowerment despite increases in income.

The second way relies on the possible extrapolation of the behavior in the experiment to other decisions outside this environment. If we are correct in the interpretation of the decision making process during the experiment, more powerful women (as revealed in this task) may also be more likely to bargain with their husbands when it comes to decisions where their preferences may not be aligned. We estimate the correlation between this measure and household outcomes to check the validity of this assumption and conclude that, in this context, our experimental task is a better indicator of women's power than either *HDMI1* or *HDMI2*.

Additionally, concerns have been raised about the adequacy of survey items to capture empowerment in different cultural contexts since they were grounded in formative research from South Asia (Malhotra et al., 2002). [21] With this criticism in mind, we argue that the neutrality of the experiment may allow for the estimation of intra-household bargaining power across different contexts. This, however, remains to be tested.

References

Allen, H. (2006). Village Savings and Loans Associations: sustainable and cost-

[21] This inadequacy might explain the discrepancies found by Banerjee et al. (2015), where poverty graduation programs, similar to REAP, were found to have a larger impact on women's empowerment in Asian countries (India and Pakistan) compared to African (Ethiopia and Ghana) and Latin American (Peru and Honduras) countries.

effective rural finance. *Small Enterprise Development*, *17*(1), 61–68.

Almas, I., Armand, A., Attanasio, O., & Carneiro, P. (2015). *Measuring and changing control: Women's empowerment and targeted transfers* (Working Paper No. 21717). Cambridge, MA: National Bureau of Economic Research.

Bandiera, O., Burgess, R., Das, N. C., Gulesci, S., Rasul, I., & Sulaiman, M. (2016). *Labor markets and poverty in village economies* (Economic Organisation and Public Policy Discussion Papers Series). London, UK: LSE, STICERD.

Banerjee, A. V., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parient, W., ... Udry, C. (2015). A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science*, *348*(6236), 1260799.

Braaten, R., & Martinsson, P. (2015). *Experimental measures of household decision power* (Working Paper No. 02/2015). Oslo, Norway: CREE.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics*, *14*(3), 375–398.

Brosig, J., Weimann, J., & Yang, C. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics*, *6*(1), 75–90.

Carlsson, F., Martinsson, P., Qin, P., & Sutter, M. (2013). The influence of spouses on household decision making under risk: An experiment in rural China. *Experimental Economics*, *16*(3), 383–401.

Carter, J., Byrne, S., Schrader, K., Kabir, H., Uraguchi, Z., Pandit, B., ... Fendrich, P. (2014). Learning about women's empowerment in the context of development projects: Do the figures tell us enough? *Gender and Development*, *22*(2), 327–349.

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

Dahl, R. (1957). The concept of power. *Behavioural Science*, *2*(3), 201–215.

de Palma, A., Picard, N., & Ziegelmeyer, A. (2011). Individual and couple decision behavior under risk: Evidence on the dynamics of power balance. *Theory and Decision*, *70*(1), 45–64.

Duflo, E. (2012). Women empowerment and economic development. *Journal of*

Economic Literature, *50*(4), 1051–1079.

Eckel, C., & Grossman, P. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, *23*(4), 281–295.

Fratkin, E. M. (2004). *Ariaal pastoralists of Kenya: studying pastoralism, drought, and development in Africa's arid lands* (Second ed.). Pearson Education, Incorporated.

Gneezy, U., & Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631–645.

Gobin, V., Santos, P., & Toth, R. (2016). *Poverty graduation with cash transfers: A randomised evaluation* (Discussion Paper No. 23-16). , Australia: Department of Economics, Monash Business School, Monash University.

Goldberg, N., & Salomon, A. (2011). *Ultra Poor Graduation Pilots: Spanning the gap between charity and microfinance* (Commissioned Workshop Paper). Valladolid, Spain: 2011 Global Microcredit Summit.

Harsanyi, J. C. (1962). Measurement of social power, opportunity costs, and the theory of two-person bargaining games. *Behavioral Science*, *7*, 67–80.

Holt, C., & Laury, S. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Jejeebhoy, S. (2002). Convergence and divergence in spouses' perspectives on women's autonomy in rural India. *Studies in Family Planning*, *33*(4), 299–308.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change*, *30*(3), 435–464.

Kenya National Bureau of Statistics and Society for International Development. (2013). *Exploring kenya's inequality: pulling apart or pooling together?* Nairobi, Kenya: Kenya National Bureau of Statistics.

Malhotra, A., Schuler, S., & Boender, C. (2002). *Measuring women's empowerment as a variable in international development* (Paper prepared for the World Bank Workshop on Poverty and Gender). Washington, DC: World Bank.

Matin, I., Sulaiman, M., & Rabbani, M. (2008). *Crafting a graduation pathway for*

the ultra poor: Lessons and evidence from a BRAC programme (Working Paper No. 109). Dhaka, Bangladesh: BRAC Research and Evaluation Division.

Nature Conservancy. (2013, July). *Kenya's national gender context and its implications for conservation: A gender analysis.* Nairobi, Kenya: Nature Conservancy, Central Science. Retrieved from `http://www.nature.org/science-in-action/leading-with-science/kenya-gender-analysis.pdf`

Nichols, A. (2007). Causal inference with observational data. *The Stata Journal*, *7*(4), 507–541.

Nichols, A. (2008). Erratum and discussion of propensity-score reweighting. *The Stata Journal*, *8*(4), 532–539.

Peterman, A., Schwab, B., Roy, S., Hidrobo, M., & Gilligan, D. (2015). *Measuring women's decision making: Indicator choice and survey design experiments from cash and food transfer evaluations in Ecuador, Uganda, and Yemen* (Discussion Paper No. 01453). Washington, DC: IFPRI.

Russell, B. (1938). *Power: a new social analysis.* London: Allen & Unwin.

Schatz, E., & Williams, J. (2012). Measuring gender and reproductive health in Africa using Demographic and Health Surveys: The need for mixed-methods research. *Culture, Health and Sexuality*, *14*(7), 811–826.

Upadhyay, U., & Karasek, D. (2012). Women's empowerment and ideal family size: An examination of DHS empowerment measures in sub-Saharan Africa. *International perspectives on sexual and reproductive health*, *38*(2), 78–89.

van den Bold, M., Quisumbing, A., & Gillespie, S. (2013). *Women's empowerment and nutrition: an evidence review* (Discussion Paper No. 01294). Washington, DC: IFPRI.

Figures and tables

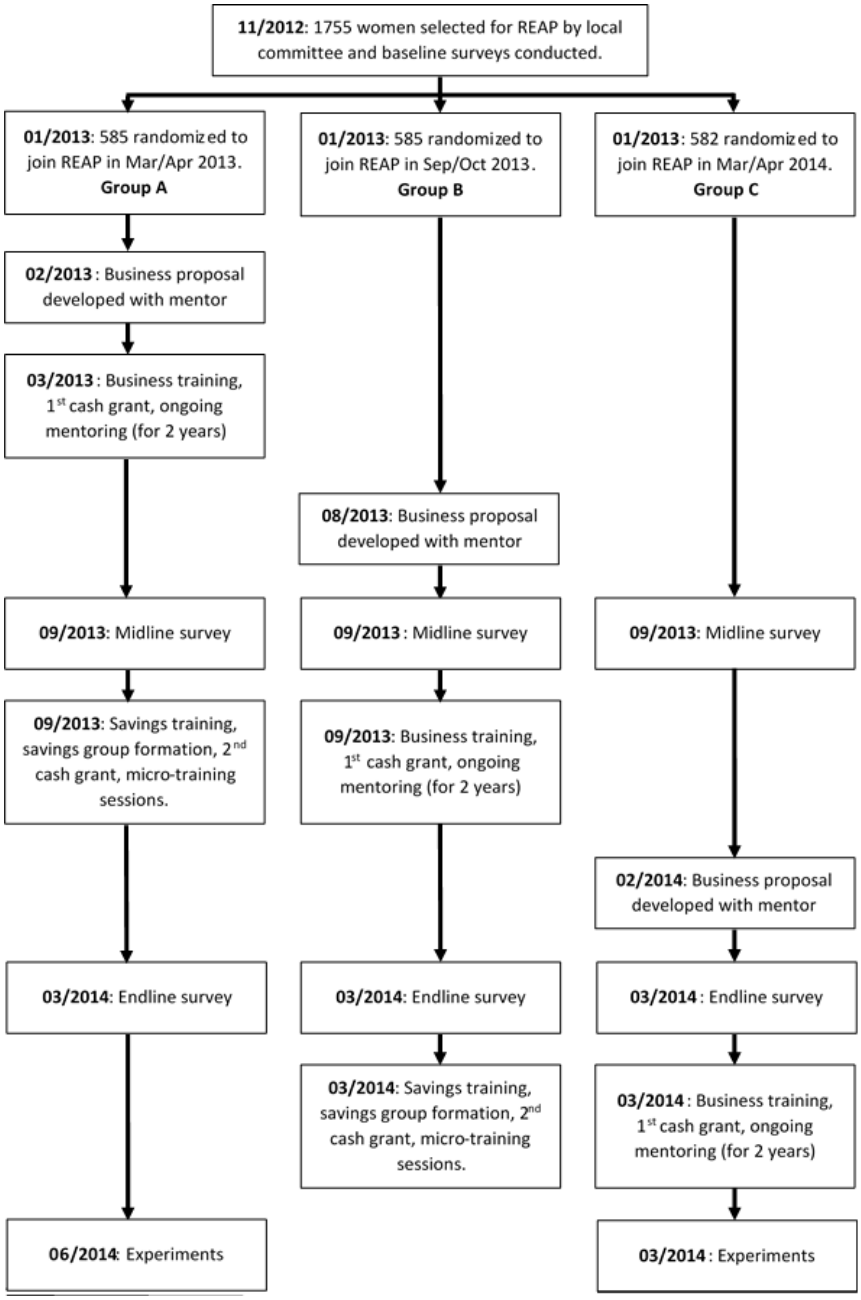
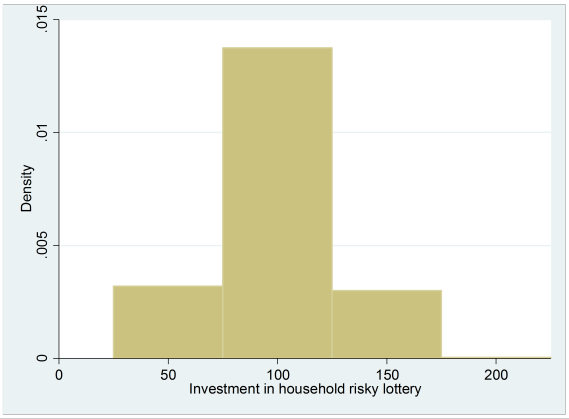


Figure 1: Description of the intervention

	Women
Mean	99.786
(Std Dev)	(28.367)
N	700
Distribution of investment decisions	

Note: The initial endowment given to participants to invest is 200 Ksh. Two women chose to invest the maximum of 200 Ksh.

Figure 2: Summary of investments by women in the household risky lottery

Table 1: Experimental and survey measures of women's power: summary statistics

	(1)	(2)	(3)	(4)
	compromise	*power*	*HDMI1*	*HDMI2*
Mean	0.583	0.266	0.272	0.411
(Std Dev)	(0.493)	(0.442)	(0.239)	(0.334)
N	2619	700	687	687

Note: The variable *compromise* is a binary indicator that is equal to one if the woman chooses to accept her spouses decision or if she changes her spouses decision to some amount not equal to her original investment (and equal to zero if the woman decides to stick with her original investment decision which differs from her spouses decision). The variable *power* is coded as one if, as the second mover, the woman does not compromise at all with her husband in more than 50% of decisions. The components of the dependent variable *HDMI1* are coded as one if the woman has the final say in the decision (and zero, if otherwise) whereas those of *HDMI2* are coded as one if the woman has the sole or joint final say in the decision (and zero, if otherwise).

Table 2: Women's decisions as second movers

I_1^H	(1) $I_2^W = I_1^H$	(2) $I_2^W \neq I_1^H$ and $I_2^W \neq I_1^W$	(3) $I_2^W \neq I_1^H$ and $I_2^W = I_1^W$
	Compromise		No compromise
0	0.233	0.165	0.602
50	0.691	0.037	0.272
100	0.873	0.000	0.127
150	0.586	0.079	0.336
200	0.230	0.244	0.527

Note: This table reports the proportion of women who, as second movers, choose to either accept the spouse's decision ($I_2^W = I_1^H$), change it to some amount not equal to her original preference ($I_2^W \neq I_1^H$ and $I_2^W \neq I_1^W$) or stick with her original preference ($I_2^W \neq I_1^H$ and $I_2^W = I_1^W$), for each possible first move decision by the spouse (I_1^H). We exclude those observations where $I_1^W = I_1^H$, $I_1^H > I_1^W > I_2^W$, $I_1^W > I_1^H > I_2^W$, $I_2^W > I_1^H > I_1^W$, or $I_2^W > I_1^W > I_1^H$.

Table 3: Correlation between different measures of women's power

	power	*HDMI1*	*HDMI2*
power	1.000		
HDMI1	0.039 (0.313)	1.000	
HDMI2	0.012 (0.748)	0.557 (0.000)	1.000

Note: Values within parentheses are p-values of the test of the null hypothesis that the two variables are independent. See note to Table 1 for the definition of *power*, *HDMI1* and *HDMI2*.

Table 4: The impact of REAP on experimental and survey measures of power

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent Variable:	*power*				*HDMI1*				*HDMI2*			
Treatment	0.058*	0.071**	0.053	0.071*	0.026	0.048	0.007	-0.014	-0.012	0.033	-0.059	-0.052
	(0.034)	(0.035)	(0.036)	(0.037)	(0.067)	(0.067)	(0.076)	(0.076)	(0.065)	(0.065)	(0.077)	(0.077)
Age		-0.001		0.001		0.005		-0.005		0.012**		0.000
		(0.003)		(0.003)		(0.005)		(0.006)		(0.005)		(0.006)
Age of spouse		0.001		-0.001		0.010*		0.012**		-0.006		-0.007
		(0.002)		(0.003)		(0.005)		(0.005)		(0.005)		(0.006)
Literacy		-0.032		0.079		0.094		-0.031		0.022		-0.144
		(0.088)		(0.130)		(0.178)		(0.268)		(0.173)		(0.272)
Numeracy		-0.024		-0.058		0.324**		0.384***		-0.025		0.060
		(0.066)		(0.069)		(0.125)		(0.143)		(0.122)		(0.144)
Wife in polygamous marriage		0.006		-0.041		-0.052		-0.000		0.205***		0.258***
		(0.039)		(0.040)		(0.075)		(0.082)		(0.073)		(0.083)
# children in household		0.002		-0.002		-0.048**		-0.030		-0.042**		-0.045*
		(0.011)		(0.012)		(0.022)		(0.025)		(0.021)		(0.026)
# adults in household		-0.029		-0.059*		-0.280***		-0.247***		-0.175***		-0.106
		(0.031)		(0.034)		(0.060)		(0.071)		(0.059)		(0.072)
TLU per capita		0.033		0.056*		-0.040		-0.027		-0.116**		-0.074
		(0.031)		(0.031)		(0.060)		(0.064)		(0.058)		(0.065)
Durable asset index		0.002		0.003		0.014		-0.001		0.021*		0.006
		(0.005)		(0.006)		(0.012)		(0.011)		(0.012)		(0.012)
Husband attends experiment		0.020		0.003								
		(0.043)		(0.045)								
Investment in individual risk task		-0.001		-0.000								
		(0.001)		(0.001)								
N	700	687	677	665	888	875	665	665	888	875	665	665
R-squared	0.137	0.154	0.120	0.143	0.146	0.189	0.146	0.248	0.223	0.250	0.153	0.288

Note: Results in columns (1) to (4) are from a linear probability model where the dependent variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not compromise at all with her husband. The dependent variable in columns (5) to (8) and columns (9) to (12) are *HDMI1* and *HDMI2*, respectively. The components of the dependent variable *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise whereas those of *HDMI2* are coded as one if the woman has the sole or joint final say in the decision or zero if otherwise. *HDMI1* and *HDMI2* are standardised using the control group mean and standard deviation, allowing the estimate to be interpreted as the effect size relative to the control group. Regressions in columns (1), (2), (5), (6), (9) and (10) are unweighted and those in columns (5), (6), (9) and (10) use the sample of all married women in groups *A* and *C* regardless of participation in the experiment. Regressions in columns (3), (4), (7), (8), (11) and (12) include the weights generated from the propensity score estimates as reported in Appendix C, Table C1. All regressions include sub-location fixed effects. Demographic and household controls are at baseline levels. Standard errors are shown in parentheses. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Table 5: Women's power and welfare indicators

Dependent variable:	(1)	(2)	(3)	(4)	(5)	(6)
	# nights that child has gone to bed hungry in past week			Monthly food consumption per capita		
	HDMI1	*HDMI2*	*power*	*HDMI1*	*HDMI2*	*power*
Measure of power	0.038	0.282***	0.088	1.040	1.725**	7.429***
	(0.052)	(0.050)	(0.108)	(0.869)	(0.857)	(1.776)
Age	0.024***	0.023***	0.023***	0.129	0.123	0.114
	(0.008)	(0.008)	(0.008)	(0.134)	(0.134)	(0.133)
Age of spouse	-0.017**	-0.014**	-0.016**	-0.421***	-0.396***	-0.404***
	(0.007)	(0.007)	(0.007)	(0.116)	(0.115)	(0.114)
Literacy	-0.252	-0.196	-0.263	13.776**	13.994**	13.141**
	(0.348)	(0.340)	(0.348)	(5.638)	(5.626)	(5.564)
Numeracy	0.092	0.090	0.110	-25.651***	-25.345***	-24.884***
	(0.179)	(0.174)	(0.178)	(3.008)	(2.983)	(2.950)
Wife in polygamous marriage	0.267**	0.190*	0.271***	2.149	1.703	2.470
	(0.103)	(0.102)	(0.103)	(1.728)	(1.738)	(1.706)
# children in household	-0.008	0.004	-0.009	-7.331***	-7.285***	-7.353***
	(0.032)	(0.031)	(0.032)	(0.536)	(0.535)	(0.528)
# adults in household	0.142	0.160*	0.140	-3.899**	-3.971***	-3.721**
	(0.092)	(0.089)	(0.092)	(1.509)	(1.493)	(1.477)
TLU per capita	-0.012	0.009	-0.018	1.517	1.611	1.118
	(0.080)	(0.078)	(0.080)	(1.334)	(1.333)	(1.319)
Durable asset index	0.000	-0.001	-0.000	0.481**	0.469*	0.459*
	(0.014)	(0.014)	(0.014)	(0.239)	(0.238)	(0.236)
N	654	654	654	665	665	665
R-squared	0.153	0.196	0.153	0.613	0.615	0.623

Note: The dependent variable is either the number of nights that a child is reported as going to bed hungry in the past week, or monthly food consumption per capita which is measure in USD PPP at 2014 prices. All regressions include sub-location fixed effects and are reweighted using the weights generated from propensity scores. Demographic and household controls are at baseline levels. The components of the *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise whereas those of *HDMI2* are coded as one if the woman has the sole or joint final say in the decision or zero if otherwise. *HDMI1* and *HDMI2* are standardised using the control group mean and standard deviation. The variable *power* is coded as one if in more than 50% of decisions as the second mover, the woman does not compromise at all with her husband. Standard errors are shown in parentheses. *, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

Appendix A Experimental procedure

In June 2014 married REAP beneficiaries in groups A and C, along with their husbands, were invited to participate in the household decision making experiment described in section 3.1. Several steps were taken to maximize the attendance. Invitations were made by the business mentors, who were provided with a checklist of eligible beneficiaries to ensure that participants were from either group A or C. On the day before the experiment, mentors reminded all eligible persons of the experiment. The team of enumerators arrived in each village the day before the experiment, which also served as a reminder.

The experiments were run separately in each location by a team of four enumerators and a research assistant who was in charge of overseeing all experimental procedures. [1] Three enumerators focused on conducting experiments with women and one focused on experiments with men. Two sets of tasks were conducted with the first set focused on household decision making and the second set on decision making within business groups, and as such not relevant for this analysis.

The first set of experimental tasks was run with both women and men and included the risk taking task designed by Gneezy and Potters (1997) and our modification of this task, designed to measure women's power, as described in section 3.1. The second set of experiments included a trust game and a coordination game, in addition to tasks designed to elicit risk and time preferences, and was run with women only. Women were only made aware of the second set of tasks after they had finished making their decisions in the first set, hence latter decisions are not expected to confound the decisions made beforehand. [2] An outline of the sequence of tasks that took place during the experimental sessions is presented in Table A1.

Before beginning the experiments women separated from men as they waited to enter the venue, naturally eliminating communication between spouses. The mentors also ensured that participants who were waiting to take part in the experiments were kept separate from those who had completed the experiments. Once the participants consented to take part in the experiment, the rules of the first task (the Gneezy and Potters (1997) task) were explained, first orally and then visually. [3] Two bowls were used to represent the money kept by the participant and the money invested in the

[1] Eight local language enumerators were trained for three days on the experimental procedures before being divided into two teams. Experiments took place in churches, schools or meeting halls which were divided into five separate spaces for each enumerator and the research assistant.

[2] If the woman was married and her spouse was present or she was uncertain if he would attend then the enumerator conducted both sets of experiments with her. If she knew that her spouse was not attending or she was single then the enumerator started with the second set of experiments.

[3] No eligible participants who attended the sessions declined to participate in the experiments.

lottery, with real money being used in the demonstration. The probabilities for the high and low payoffs in each part were demonstrated using four white and two blue balls. Several examples were used to further illustrate the task, before presenting the participant with a scenario to check their understanding. If the task was still not clear to the participant, the explanation was repeated until it was clear that the participant fully understood the task. The second task (our modification of the Gneezy and Potters (1997) task, explained in section 3.1) was similarly explained first orally and then visually. The participants were informed that together, these two tasks would take approximately 20 minutes to complete and that at the end of the experiment they would (blindly) pick a numbered ball from a bag, to determine which task they would be rewarded for. The enumerators stressed that the payments would take place in private at the end of the sessions and that their decisions would not be revealed to other participants, including their spouses.

Participants were informed that a coin toss would determine if their first mover decisions or second mover decisions would be used in calculating their payoffs. Once a couple had completed both tasks of the first set of experiments they were invited separately to pick a numbered ball from a bag to determine which task from the first set of experiments would be played for real. If the first task was chosen for payment then the participant picked a colored ball from a bag to determine if their investment was doubled or lost. If the second task was chosen, then a coin toss first identified whose first mover and second mover decisions would be used to determine payoffs. A colored ball was then picked to determine if the final amount invested was doubled or lost. If a woman was uncertain as to whether her husband would show up, she was asked to wait until he arrived. If he did not arrive then she would only receive payment for the first task (in addition to any other payment she received from the second set of tasks). The full set of instructions used by the enumerators is presented in Appendix B.

Table A1: Sequence of tasks in the experimental sessions

Task	Women				Men
	Single	Married			
		Husband will not attend	Not sure if husband will attend	Husband attend	
Risk 1			✓	✓	✓
Household task			✓	✓	✓
Household decision making survey					✓
Risk 2	✓	✓	✓	✓	
Trust game	✓	✓	✓	✓	
Coordination game	✓	✓	✓	✓	
Payment	2 of 3	2 of 3	1 of first 2 and 2 of last 3 if husband attends **OR** Risk 1 and then 2 of last 3 if husband does not attend	1 of first 2 and 2 of last 3	1 of 2
Time preference: hypothetical	✓	✓	✓	✓	

Note: In this analysis, we focus on the item in bold print, while also using data from the Risk1 and household decision making survey fielded to men who attended the experiment. The other tasks are part of the second set of experiments which were designed to understand decisions and preferences among business group members and are not relevant to the analysis of household decision making.

Appendix B Instructions for household experiment

Instructions

This section of the study will take approximately 20 minutes. There are 2 parts to this section and each will be explained at the appropriate time. Your earnings for Parts 1 and 2, and your total earnings for the study will be determined by the decisions you and the other players make in each part.

You are free to make as much money as you can. How your rewards for Parts 1 and 2 will be determined is explained below. At the end of this tasks you will be presented with a bag with balls numbered 1 and 2. You will be asked to select ONE ball and the number on this ball will correspond to the Part of the section that you will receive money for. For example, if you pick balls numbered 1 then you will be rewarded for Part 1.

You will be paid in cash in private at the end of the session. We will not tell anyone about the decisions you make or the amount of money you receive.

Task 1

In this part of the study you will work individually. You will be given 100 Ksh and will be asked to make an investment decision. You may choose to invest 0 Ksh, 25 Ksh, 50 Ksh, 75 Ksh or 100 Ksh.

If you choose to invest 0 Ksh then you will just keep the 100 Ksh you are given. If you choose to invest any amount greater than zero then you will be presented with an opportunity to double your investment. How?

This bag is filled with 6 balls: 4 WHITE and 2 BLUE. You will be given a SINGLE chance to pick ONE ball (without looking) and the returns to the investment are based on the colour of the ball that is blindly selected from a bag. There are two possible outcomes of this investment: 1) If you pick a WHITE ball then the money you invested will be doubled and returned to you. 2) If you pick a BLUE ball then you will lose your investment.

Note, you will always retain the amount that is NOT invested regardless of the colour of the ball that is picked from the bag.

Example 1:

Suppose you invest 50 Ksh and keep 50 Ksh. You pick a WHITE ball. Your investment will automatically double to 100 KSH and you will receive a total of 150 Ksh i.e. 100 Ksh plus the 50 Ksh that was not invested.

Example 2:

Suppose you invest 50 Ksh and keep 50 Ksh. You pick a BLUE ball. Your investment

will be lost and you will receive a total of 50 Ksh i.e. the amount that was not invested.

Question:
How much money would you receive in total if you invested 75 Ksh and you pick a WHITE ball?
Remember that in this task the bag will be filled with 4 WHITE balls and 2 BLUE balls

How much money do you choose to invest? _____ KSh

You will pick a ball from the bag at the end of the session if this task is selected for payment.

Task 2
In this task you will be paired with your spouse/partner and any rewards will be divided in half and you will receive half of the overall reward.
You will be asked to make an investment decision like you did in the previous task, i.e. whatever you invest is doubled if a WHITE ball is picked but the investment is lost if a BLUE ball is picked. The bag will contain 4 WHITE balls and 2 BLUE balls.
This time the amount you and your spouse/partner are given to invest is 200 Ksh.
You and your spouse/partner will be assigned as either "first mover" or "second mover" and this will be determined by flipping a coin with "heads" resulting in you being assigned as "first mover" and your spouse/partner being nominated as "second mover". If the coin lands on "tails" then the assignment will be reversed.

The game will be played as follows:

- The first mover will choose an amount to invest. They can choose to invest 0 Ksh, 50 Ksh, 100 Ksh, 150 Ksh or 200 Ksh.
- The second mover will then have a chance to accept or overrule the investment decision made by the first mover.
- If the second mover chooses to accept the investment decision of the first mover then a ball is chosen to determine if the investment is doubled or lost.
- If the second mover chooses to overrule the decision then they will be asked to select an alternative amount to invest. A ball will then be picked to determine if this new investment is doubled or lost.
- The reward for you will be half of the amount that is not invested plus half of the returns from the investment.

There will be no communication between the first and second mover.

Example:
The first mover chooses to invest 50 Ksh. The second mover then decides to overrule this decision and invest 150 Ksh instead. A WHITE ball is picked. The 150 Ksh investment is doubled and the total reward, including the amount that is not invested, is 350 Ksh which is divided in half.
The first mover chooses to invest 50 Ksh. The second mover then decides to accept this decision. A WHITE ball is picked. The 50 Ksh investment is doubled and the total reward, including the amount that is not invested, is 250 Ksh which is divided in half.

Question:
The first mover decides to invest 100 Ksh. The second mover decides to accept this decision. What is the total earned by the pair if a BLUE ball is chosen?
The first mover decides to invest 100 Ksh. The second mover decides to overrule this decision and invest 50 Ksh instead. What is the total earned by the pair if a BLUE ball is chosen?

Remember, there are 4 WHITE balls and 2 BLUE balls in the bag.

Suppose you are the first mover, how much will you choose to invest? _____ KSh

Suppose you are the second mover, would you accept or overrule the first mover if they choose to invest 0 Ksh? 1 = Yes 2 = No
[If overrule, ask:] Then how much would you invest? _____ KSh

Suppose you are the second mover, would you accept or overrule the first mover if they choose to invest 50 Ksh? 1 = Yes 2 = No
[If overrule, ask:] Then how much would you invest? _____ KSh

Suppose you are the second mover, would you accept or overrule the first mover if they choose to invest 100 Ksh? 1 = Yes 2 = No
[If overrule, ask:] Then how much would you invest? _____ KSh

Suppose you are the second mover, would you accept or overrule the first mover if they choose to invest 150 Ksh? 1 = Yes 2 = No
[If overrule, ask:] Then how much would you invest? _____ KSh

Suppose you are the second mover, would you accept or overrule the first mover if they choose to invest 200 Ksh? 1 = Yes 2 = No
[If overrule, ask:] Then how much would you invest? _____ KSh

The first mover and second mover will be determined at the end of the session if this task is selected for payment. You will also be asked to pick a ball at the end of the session to determine if the investment is doubled or lost.

Appendix C Propensity score estimates

As mentioned in section 3.3, not all eligible women and their spouses accepted the invitation to participate in the experimental tasks that we conducted, as shown in Table C2.

Table C2: Sample size

	(1)	(2)
	Treatment	Control
Funding Cycle:	Group A	Group C
Funding date:		
Eligible REAP beneficiaries	585	585
Final REAP beneficiaries	585	582
Beneficiaries eligible for experiments	483	463
Beneficiaries that attend experiments	336	364
Husbands that attend experiments	118	111

As a first step in addressing concerns about potential selection bias, we check whether, conditional on being eligible, participants in the experiments are different from non-participants in terms of baseline characteristics. The results of that comparison are presented in Table C3. Panel A presents summary statistics (means and standard deviations) of both household and individual characteristics of participants and non-participants, while panel B presents the $t-$tests of the null hypothesis of equality at baseline.

Table C3: Participants vs non-participants in the experiment: summary statistics at baseline

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Variable:	Monthly income per capita	Monthly expenditure per capita	Monthly food expenditure per capita	Monthly non-food expenditure per capita	Total savings per capita	TLU per capita	Durable asset index	Meals per day	Nights that child has gone to bed hungry in past week	Proportion of children in school	Household Size	# children	Years of education	Business Experience	Benefitting from HSNP	Participating in CARE VSLA	HDMI1	Funding cycle
Panel A: Means and standard errors of variables at baseline																		
Participated	32.919	23.781	9.137	21.468	3.705	0.655	-0.153	1.948	0.580	0.419	5.899	3.887	0.379	0.556	0.130	0.090	-0.052	0.480
(Std Err)	(1.056)	(0.845)	(0.501)	(0.781)	(0.316)	(0.027)	(0.156)	(0.014)	(0.026)	(0.010)	(0.070)	(0.065)	(0.060)	(0.019)	(0.013)	(0.011)	(0.038)	(0.019)
Observations	700	700	700	700	700	700	700	694	693	698	700	700	700	700	700	700	684	700
Did not participate	31.073	20.052	11.021	19.430	4.405	0.667	0.300	1.894	0.534	0.391	5.797	3.813	0.252	0.500	0.081	0.098	0.148	0.598
(Std Err)	(1.803)	(1.027)	(1.308)	(1.201)	(0.987)	(0.045)	(0.278)	(0.027)	(0.040)	(0.019)	(0.111)	(0.106)	(0.084)	(0.032)	(0.017)	(0.019)	(0.063)	(0.031)
Observations	246	246	246	246	246	246	246	246	242	245	246	246	246	246	246	246	242	246
Panel B: t test comparison of means of baseline characteristics based on participation in experiments																		
p-value	0.377	0.005***	0.180	0.156	0.500	0.807	0.156	0.075*	0.339	0.177	0.437	0.551	0.222	0.134	0.025**	0.729	0.007***	0.001***

Panel C: F-test from regression of participation in experiments on variables above.[a]

F-Stat	p-value
2.65	0.000***

Note: All monetary values are reported in 2014 USD, PPP terms. TLU refers to tropical livestock unit which is a standardised way of measuring the size of a mixed herd: 1 head of cattle is equivalent to 0.7 camels, 10 sheep/goats, or 2 donkeys. A description of the durable asset index can be found in Gobin et al. (2016). The components of *HDMI1* are coded as one if the woman has the final say in the decision or zero if otherwise. Funding cycle equals one (zero) if assigned to Mar/Apr 2013 (2014). [a]Monthly food and non-food expenditure per capita are excluded from this regression.*, ** and *** stand for significant at the 10%, 5% and 1% level of significance, respectively.

It is clear from the analysis of this table that participants and non-participants are similar along many dimensions but that some notable differences also exist. Non-participant households spent less on food per capita, are less likely to have benefited from cash transfer programs targeted at the poor (the Hunger Safety Net Program (HSNP)), reported having more say in household decisions, and are more likely to belong to the treatment group. Overall, it seems that participant households are observationally different from non-participant households, a finding that is reinforced by the results of a F-test of the joint effect of these variables on participation in the experiments, reported in panel C.

Following Nichols (2007, 2008), we estimate the relation between participation in experiments and observable pre-treatment characteristics, namely baseline levels of the household decision making index, income, expenditure, savings and assets, food security indicators, household composition, characteristics of the woman and her spouse (age, literacy, business experience, whether in a polygamous marriage), participation in other NGO programs, and cycle of assignment to REAP. The estimates of the logit model used in explaining the participation decision are reported in Table C1. [4] The estimates of the conditional probability of being a participant (i.e. the propensity score), $\hat{p}$, and the non-participation odds, $\hat{p}(1-\hat{p})$, are then used to reweight the participant observations to estimate the average treatment effect on the treated.

[4]The –pscore– command in Stata is used to estimate the propensity scores ensuring that the balancing property is met. The analysis of the balancing property is restricted to participants and non-participants in the region of common support. Twenty-three participant observations are dropped due to either missing values or because their propensity scores lye outside the region of common support. The final sample comprises of 349 treatment and 328 control individuals.

Table C1: Estimation of the propensity score

Variable	Coefficient estimate
1^{st} cycle of REAP	0.372**
	(0.168)
Literacy	-1.065**
	(0.537)
Business experience	0.493
	(0.302)
Age	0.003
	(0.016)
Age of spouse	-0.022
	(0.014)
First wife in polygamous marriage	0.341
	(0.215)
Second wife in polygamous marriage	0.239
	(0.243)
HDMI1	-0.258**
	(0.108)
Household size	-0.297**
	(0.148)
# children in household (logged)	1.191*
	(0.691)
Proportion of school aged children in school	-3.358***
	(0.971)
Proportion of school aged children in school (squared)	3.355***
	(0.965)
Food expenditure per capita	-0.013**
	(0.006)
Non-food expenditure per capita	0.009
	(0.006)
Total income per capita (logged)	-0.299***
	(0.104)
Total savings per capita	0.008
	(0.013)
Durable asset index	0.017
	(0.035)
Durable asset index (squared)	-0.010
	(0.006)
Durable asset index (cubed)	0.001
	(0.000)
TLU per capita	-0.402
	(0.661)
TLU per capita (squared)	0.517
	(0.445)
TLU per capita (cubed)	-0.111
	(0.080)
Benefited from HSNP cash transfer	-1.102**
	(0.448)
Participated in CARE savings groups	0.114
	(0.341)
Location dummies	Yes
Pseudo R-squared	0.103
N	908
Likelihood Ratio chi-squared (36)	104.57

Note: Propensity score is estimated in Stata using –pscore– command with a logit model. The dependent variable equals one if the woman did not participate in the experiments, and zero otherwise. All monetary values are reported in 2014 USD, PPP terms. The *HDMI1* is standardised using the mean and the standard deviation of the overall sample of married women at baseline. Standard errors are shown in parentheses. *, ** and *** denote significance at the 10%, 5% and 1% levels, respectively.