Does Poverty Change Labor Supply? Evidence from Multiple Income Effects and 115,579 Bags*

Abhijit Banerjee†
Dean Karlan‡
Hannah Trachtman§
Christopher Udry¶

December 17, 2023

Abstract

The income elasticity of labor supply is a central parameter of many economic models. We test the response of labor supply and effort to exogenous changes in income using data from a randomized evaluation of a multi-faceted grant program in northern Ghana combined with a researcher-implemented bag-making operation. We find evidence of a *positive* "income effect" on labor supply. We argue that simple models with either labor or capital market frictions cannot explain the results, whereas a model that allows for positive physiological or psychological productivity effects from higher income fits with our findings.

Keywords: poverty, labor supply, income elasticity
JEL Classifications: H31, J22, O12

*Approval from the Yale University Human Subjects Committee, IRB 0705002656, 1002006308, 1006007026, and 1011007628; and from the Innovations for Poverty Action Human Subjects Committee, IRB Protocol 19.08January-002, 09December-003, 59.10June-002, and 10November-003.494. Thanks to the Ford Foundation, and 3ie for funding. Thanks to Nathan Barker, Caton Brewster, Abubakari Bukari, David Bullon Patton, Sébastien Fontenay, Angela Garcia, Yann Guy, Samantha Horn, Sana Khan, Hideto Koizumi, Matthew Lowes, Elizabeth Naah, Michael Polansky, Elana Safran, Sneha Stephen, Rachel Strohm, and Stefan Vedder for outstanding research assistance and project management, and in particular Bram Thuysbaert for collaboration. The authors would like to thank the leadership and staff at Presbyterian Agricultural Services (PAS) for their partnership. Thanks to Frank DeGiovanni of the Ford Foundation, Syed Hashemi of BRAC University, and Aude de Montesquiou and Alexia Latortue of CGAP for their support and encouragement of the research. No authors have any real or apparent conflicts of interest, except Karlan is on the Board of Directors of Innovations for Poverty Action, which participated in oversight of the implementation. All data and code will be available upon publication at the IPA Dataverse (doi pending).

†MIT, CEPR, NBER and Jameel Poverty Action Lab (J-PAL): banerjee@mit.edu
‡Northwestern University, CEPR, NBER, IPA, and J-PAL: karlan@northwestern.edu
§Hebrew University of Jerusalem: hannah.trachtman@mail.huji.ac.il
¶Northwestern University, CEPR, NBER, and J-PAL: christopher.udry@northwestern.edu

1 Introduction

The income elasticity of labor supply is one of the central parameters of economic models. Under the standard assumption that consumption and work are not strong complements, it is easy to derive the prediction that any increase in income will reduce labor supply. This has important implications for the design of social support policies, because, for example, such a reduction in labor supply would limit the net income gains.

The basic argument for why we should expect this negative labor supply response is well-known. If utility from consumption is $u(c)$, the disutility of labor supply is $v(l)$ and the relation between consumption and labor supply is $c = f(l) + t$, where f is income and is some increasing concave function of labor supply and t is a transfer, we immediately get a first order condition

$$u'(f(l) + t)f'(l) = v'(l)$$

from which it follows that any increase in t will reduce the marginal utility of income and therefore labor supply. Of course, several important assumptions are being made here. First, as pointed out by Benjamin (1992), we need that t does not directly raise the marginal product of labor. In other words, we cannot have $f(l,t)$ with $f_{lt}(l,t) > 0$. As Benjamin (1992) also points out, this is typically ruled out by either the assumption of perfect capital markets (in which case t should not enter $f(l,t)$) or by the assumption that household labor and market labor are perfect substitutes at the margin (in which case $f_l(l,t)$ equals the market wage). However neither of these assumptions seem particularly plausible for low income families in developing countries (LaFave et al., 2020). A transfer may directly raise the marginal product of labor, making this kind of *investment productivity effect* quite relevant.

A second reason why the expected income effect may be absent for the very poor is that consumption (or income) and labor supply may be complements, so that the disutility of effort takes the form $v(l,c)$ with $v_{lc}(l,c) < 0$. The idea that a mechanical nutrition-productivity relationship generates complementarity between consumption and work lies at the heart of the earliest models of a poverty trap (Leibenstein, 1957; Dasgupta and Ray, 1986). In these models, a better-fed worker provides more effort. More generally, higher consumption may result in better health, which in turn may reduce the disutility of work (Strauss and Thomas, 1998). We call this a *physiological productivity effect*.

More recently, psychological models of poverty traps have made a similar case for why, at low levels of psychological well-being, positive income shocks may boost labor supply—what we will call a *psychological productivity effect*. People living under scarcity may exhibit "tunnel vision," focusing intently to allocate scarce resources at the expense of other margins (Mullainathan and Shafir, 2013; Fehr et al., 2022), including, in some cases, productivity (Fink et al., 2020; Pinto, 2021; Kaur et al., 2022). Poverty may also affect risk-taking and time-discounting (Haushofer and Fehr, 2014).[1] Finally, both individuals (Dalton et al., 2016) and economies (Genicot and Ray, 2017) can get trapped in poverty when aspirations and outcomes are jointly determined (Bernard et al., 2020).

Consistent with this set of theories, the evidence from a number of recent field experiments suggests that the income effect on labor supply is often non-negative (Baird et al., 2018). Using data from cash transfer

[1]Positive income shocks can reduce risk aversion (Tanaka et al., 2010), and negative income shocks can increase impatience (Haushofer and Fehr, 2019). Relatedly, poverty has been shown to cause negative affect and stress (Haushofer and Shapiro, 2016), which can in turn influence risk-taking (Kandasamy et al., 2014).

programs around the world, Banerjee et al. (2017) shows that cash transfers to low income households have no effect on labor supply, either at the intensive margin or at the extensive margin. Banerjee et al. (2015) and Bandiera et al. (2017) report on a six-country study and a one-country study, respectively, of the Graduation program, a multi-faceted program built around an asset transfer to very poor households, and both find that the intervention led to higher incomes and labor supply. The positive impact persisted three years after the intervention began (and one year after it ended), and subsequent work has found impacts persisting at positive levels up to ten years later (Banerjee et al., 2021; Balboni et al., 2022; Barker et al., 2023). While the graduation program potentially changes the entire life circumstances of the beneficiaries, there is also evidence that more temporary positive shocks also have a *positive* effect on labor supply of low-income households: in a field experiment with piece-rate workers in India, Kaur et al. (2022) finds higher levels of productivity on days when the workers are cash-rich, and argues this is indicative of improved cognition and focus due to lower levels of financial stress.

This body of evidence has three important potential limitations. First, labor supply measurement is difficult. For example, if much of the labor supply response is in the form of reduced (unmeasured) effort on a job, it could be that the person is doing less and eventually will be fired, but we do not observe this long-term outcome.

Second, the Graduation experiments were not designed to distinguish experimentally between different possible mechanisms. The intervention involved both the transfer of a productive asset to households who are very plausibly credit constrained (so an increase in t, which may shift $f(l,t)$) as well as encouragement and training intended to shift their $v(.)$ functions. The physiological or psychological effect of extra income and the non-monetary components of the Graduation program may be an important part of what is happening, but there is no way to tell from the data.

Finally, the productivity effects of temporary income shocks studied in Kaur et al. (2022) are indeed likely to be psychological: the immediacy of the impact makes it unlikely to be due to investment in a productive asset (unlike in Gertler et al. (2012)). Our study complements the Kaur et al. (2022) lab-in-the-field experiment, investigating whether productivity effects sustained over a longer period of time may result from an intervention that substantially and durably increases incomes.[2] It is precisely these larger and more prolonged transfers that policy-makers worry about.

With this context in mind, we make two contributions by building on our study of the Ghana Graduation program, (also called "Graduating the Ultra Poor", and here onward referred to as "GUP"), one of the sites reported on in Banerjee et al. (2015). First, we provide better measurement of labor supply in the context of a durable increase in the earnings of a household and still find a non-negative income effect on labor supply over a non-trivial period of time (six months). Second, we provide evidence that what we call the psychological or physiological productivity effect is driving the observed departure from the traditional income effect.

A key to both contributions is a novel measurement exercise involving a bag-making operation. Treatment and control villages were randomly chosen to have bag production units. Those who were invited to work in these units were offered piece rate contracts to produce bags, and all inputs were provided. The number of bags they produced as well as their quality was carefully graded and the piece rate de-

[2]Indeed, the position of the literature, as summarized by Baird et al. (2018) is that "the income effect underlying the labor-leisure trade-off appears most apparent when transfers are large and/or prolonged (as with lottery winnings and pensions), and appears to be much weaker for one-time transfers."

pended on quality, so we have a reliable measure of how much effort individuals put into bag-making. Each bag-making unit was also randomly assigned to produce either simple or more complex bags, to test for differential effects.

For those in the bags production sub-groups, the comparison of GUP and control households tells us that GUP increases participation in bags, bags production, and earnings from bags by 20-30%. These effects are individually statistically significant, and the q-values after adjusting for multiple hypothesis testing are each 0.15. If anything, the effects are stronger for those assigned complex bags: GUP households make 48% more complex bags than Control households.

Of course, it could be that this increased effort reflects the fact that the household has cut back on other activities. In terms of reported hours worked, we estimate that GUP households supply only about two percent fewer hours to all forms of productive labor than do Control households, and this difference is nowhere near statistically significant at conventional levels. Survey data on agricultural outcomes, business outcomes, and wage income suggest that GUP households do not make large labor-saving investments or neglect their agricultural or non-agricultural businesses; in other words, our measures of time use are not missing important dimensions of effort.

It is striking that GUP households supply more overall effort because these households earn substantially (and statistically significantly) more than the Control households. Summing up across all the sources of earnings plus any cash transfers, during bag-making GUP households earned $19 more per month than Control households, excluding bags earnings ($p<0.01$). This implies that they earned more than triple the Control monthly earnings of $8 while spending roughly the same amount of time on productive labor and delivering more effort to bags production and no less effort elsewhere. In other words there is prima facie evidence of a *positive* GUP effect on labor supply.

Turning to our second question, we argue that this is evidence for what we have called a physiological or psychological productivity effect, rather than an investment productivity effect. This is because bag-making offered no scope for additional investment by the households: all capital was provided by us, the researchers.

This still leaves the question of whether the GUP effect is merely an income effect, given the multifaceted nature of the program. The experimental design included several arms that allow us to investigate the role of access to savings as well as the role of pure income. First, we show that access to savings does not boost labor supply in this context. Second, we document that GUP participants with randomly larger cash transfers show more evidence of a non-negative labor supply response to GUP. Ultimately, the evidence for a pure income effect is not dispositive, and the encouragement component of GUP may contribute to the higher labor supply we observe among its participants.

This paper contributes to a large literature on labor markets in developing countries (e.g. Lewis (1954); Rosenzweig (1988); Foster and Rosenzweig (1996); Goldberg (2016); Guiteras and Jack (2018)). It relates to work on the relationship between credit constraints and labor supply (e.g. Kochar (1999); Rose (2001); Jayachandran (2006); Fink et al. (2020)), and most directly builds on work understanding the effects of positive income shocks, through transfers or other mechanisms, on labor supply (e.g. Baird et al. (2018); Kaur et al. (2022)). Finally, it contributes to the large body of work that attempts to unpack the determinants of effort (e.g. Breza et al. (2018); Brune (2016); Brune et al. (2019); Kaur et al. (2015)), including the potential importance of psychological well-being and its link to income (Mani et al., 2013; Shah et al., 2012).

2 Experimental Design

We partnered with Presbyterian Agricultural Services (PAS), a local NGO in northern Ghana, whose field agents engaged in the direct field implementation of the Graduation program (GUP), the Savings-Only program (SOUP), and the employment program (Bags).

2.1 GUP and SOUP

Table 1 Panel A shows the assignment of households and villages to GUP, SOUP and control, and the cross-cutting bags measurement village assignments. Each village was assigned GUP, SOUP, or Control, and then within each treatment village, half of sample households actually received the treatment intervention, and half served as control households within treatment villages.[3]

In GUP villages, half of sample households were assigned to the GUP treatment, and the remainder were assigned to the GUP Control group. The GUP program included six components, all of which were directed to the female household head: (1) the transfer of a productive asset; (2) skills training for the management of the asset, (3) life skills training and mentorship, via weekly household visits over two years, (4) a weekly cash stipend for consumption support, worth between $6 and $9 PPP depending on family size, during each lean season, (5) some basic health services and health education, and (6) access to a savings account at a local bank and deposit collection. Additional details are described in Appendix A.1.

The sixth component of GUP, access to a savings account and weekly deposit collection by PAS field agents, was the entirety of the SOUP intervention. In SOUP villages, slightly more than half of sample households were assigned to the SOUP treatment. These households received a visit from the field officer to collect savings, but did not receive any other components of the program. The remaining households in SOUP villages were assigned to the SOUP Control group.

For half of the households assigned to GUP, the weekly collection of savings deposits was randomly withheld. For the GUP households who did receive deposit collection services, the treatment is equivalent to the combination of GUP and SOUP.[4]

2.2 Bag-Making

We designed an employment program ("Bags") offering wages for the production of cloth bags, and implemented it such that it cross-cut GUP, SOUP, and control assignments. Half of the villages were randomly assigned to Bags, as shown in Table 1 Panel A. In GUP and SOUP villages assigned to Bags, all sample households assigned to GUP or SOUP were invited to participate. In Control villages assigned to Bags, half of sample households were invited to participate. For logistical reasons, we assigned all pure Control villages with fewer than 30 compounds to No-Bags; we thus restrict all of our analysis to villages with more than 30 compounds. This leaves 93 Bags villages and 72 No-Bags villages. We have 896 total Bags households: 397 Control, 221 GUP, and 278 SOUP.

Table 1 Panel B presents the details of two sub-treatments within the bags measurement exercise. First, among Bags villages, half were assigned to produce a simple bag, and half were assigned to produce a

[3]Note that this is the same sample that was studied in Banerjee et al. (2015) and Banerjee et al. (2022).

[4]We find no evidence that the presence or absence of savings collection makes a difference to the impact of GUP on consumption or income; see Banerjee et al. (2022).

complex bag. While the simple bag has basic "running stitches" on the hem and the strap, the complex bag alternates one "running stitch" with four more difficult "chain stitches," in a pattern that requires counting. (See Appendix Figure 1.) Second, we varied the amount of unconditional consumption support, in the form of a cash transfer, received by GUP-Bags households. This was varied at the village level, and was either \$1.31 or \$3.92.[5]

The bag-making employment program ran for six months, from June through November 2012, and was designed to coincide with the lean season. The GUP and SOUP programs, which launched in July 2011, had already been running for almost a year by the time the bag-making began, and continued to run until July 2013. (See Appendix Figure 2 for a full timeline.)

Immediately prior to the start of the employment program, bag-making field agents invited female household heads to participate in a community-level training, which lasted for four days. During production, GUP, SOUP, and Control field agents visited each community on a weekly basis. At each visit, they collected new bags, distributed replacement fabric (according to the number of bags collected), and paid wages for bags submitted two weeks prior. Households could submit a maximum of ten bags per week. In the two weeks between when bags were collected and when wages were paid, quality checks were carried out by bag-making field agents, and bags were classified as high, mid, or low quality (see Appendix A.2 for details).

Wages were paid with a two-week lag. Upon each visit, GUP, SOUP, and Control field agents informed households of the composition of high, mid, and low quality bags submitted two weeks prior, and distributed payment accordingly. Baseline piece-rates for mid-quality bags were randomly assigned to be either \$0.40 or \$0.91. Bags judged to be high/low quality earned the baseline wage plus/minus \$0.13. The wage was not affected by whether the bag was simple or complex, except through quality scores, which were generally lower for complex bags. Every four weeks, bag-making field agents returned to communities to give feedback and remedial training.

To the extent possible, activities connected to the bag-making program were conducted by separate bag-making field agents (including training, quality-checking, feedback, and announcements about wage changes). However, the regular GUP/SOUP field agents were responsible for collecting the bags and delivering payment each week. As a result, the two programs were certainly not perceived as separate, and experimenter demand effects may contribute to the effects we observe.

3 A Model of Labor Supply

We provide a simple model to clarify the interplay between the GUP, SOUP and Bags programs and to explain how we use them together to conclude that there was a non-negative (and perhaps positive) income effect on labor supply, driven by a psychological or physiological productivity effect.

The utility from a certain income c is given by $\lambda u(\frac{c}{\lambda}))$, where λ is a shifter for the utility function. A higher λ is meant to capture the impact of the savings component of the GUP intervention, which makes it possible to spread the extra consumption over a longer future, hence raising the marginal utility of income. For results 1 and 2 we keep λ fixed. The household production function is $f(l,t)$, where the inclusion of t

[5]This amount was chosen to make expected weekly transfers, inclusive of bags earnings (\$2.6 on average), roughly equivalent to the \$6-\$9 received by GUP-No-Bags households.

represents the possibility that the transfers raise the marginal product of labor. In other words we assume that $f_l(l,t) > 0$, $f_{ll}(l,t) < 0$, $f_t(l,t) \geq 0$ and $f_{lt}(l,t) \geq 0$. As noted, a necessary condition for this is that there are imperfections in both the capital market and the labor market. This framework captures the three components of the GUP program (the productive asset transfer, the skills training, and the consumption support) that have the potential to raise the marginal product of labor.

The disutility of labor supply l is given by $v(l,T)$, where the inclusion of T is aimed to capture the relation between the various interventions and labor supply, via physiological or psychological channels. In other words it is possible that $T = t$, but we want to allow for possibility of interventions that shift labor supply without providing an income transfer (such as through encouragement). We assume that $v_l(l,T) > 0$, $v_{ll}(l,T) > 0$, $v_T(l,T) \leq 0$ and $v_{lT}(l,T) \leq 0$. One case where we might expect $v_T(l,T) < 0$ and $v_{lT}(l,T) < 0$, is when $T = t$, income transfers boost consumption and greater consumption raises labor supply. Another would be a coaching/encouragement treatment, where $t = 0$ but $T > 0$. Within this framework, two components of the GUP program (life skills training, and basic health services and health education) may have direct effects on the disutility of work via T. Finally we assume that $c = f(l,t) + t$. The asset transfer and consumption support components of GUP enter the budget constraint through t.

The first order condition for utility maximization is

$$u'\left(\frac{f(l,t)+t}{\lambda}\right) f_l(l,t) = v_l(l,T).$$

Suppose that $t = t(T)$ with $t'(T) > 0$. It is evident that $\frac{dl}{dT} < 0$ as long as $f_{lt}(l,t) = 0$ and $v_{lT}(l,T) = 0$. However $\frac{dl}{dT}$ can be positive if either $f_{lt}(l,t) > 0$ or $v_{lt}(l,T) < 0$. As before we call these two sources of a non-traditional income effects the investment productivity effect and the psychological/ physiological productivity effect.

Result 1: As long as λ is fixed, a necessary condition for the income effect on labor supply not to be negative is that there has to be either the investment productivity effect or the psychological/physiological productivity effect.

For our second result, we permit the household to have access to two production technologies, so that

$$c = f^a(l^a,t) + f^b(l^b,t) + t,$$

where $f^a(.)$ represents the bag making opportunity.

The household now maximizes

$$\lambda u\left(\frac{f^a\left(l^a,t\right) + f^b\left(l^b,t\right) + t}{\lambda}\right) - v(l^a + \gamma l^b, T)$$

by choosing l^a and l^b. γ represents the relative cost of effort in the two tasks. Now suppose $f^a_{l^a t}(l^a,t) = 0$. The first order condition with respect to l^a yields

$$u'\left(\frac{c}{\lambda}\right) f^a_{l^a}(l^a,t) = v_l(l,T)$$

We wish to compare $l^a(T)$ with $l^a(T')$ where $t(T) > t(T')$. Suppose $c(T) \geq c(T')$ and therefore $u'(c(T)) < u'(c(T'))$. Moreover let $l^b(T) \geq l^b(T')$. Then if it also true that $l^a(T) \geq l^a(T')$ then $l(T) \geq l(T')$.

Now if $v_{lT}(l,T) = 0$, then $v_l(l(T),T) \geq v_l(l(T'),T')$. In this case the only way to satisfy the first order condition is for $f^a_{l^a t}(l^a,t) > 0$. Conversely, if $f^a_{l^a t}(l^a,t) = 0$ then it must be the case that $v_{lT}(l,T) < 0$. We summarize this as:

Result 2: As long as λ is fixed, if there is one activity where there is no investment productivity effect, and the labor supply to that activity is greater despite the fact the household is richer and is working no less, then there must be a psychological/physiological productivity effect on the disutility of labor.

The last observation is about λ. If λ goes up, say because of savings collection, the household's marginal utility of income goes up and therefore both its labor supply and its income must both go up.

Result 3: If λ goes up, the household's labor supply and its income must both go up.

4 Data and Empirical Methods

4.1 Data

We have three sources of data. First, we have weekly administrative data on labor supply (the number of bags submitted), the quality of each bag, and the resulting earnings. Of the 896 clients in the Bags sample, 88.7% chose to make bags at some point over the six months. Over the course of the study, we collected 115,579 bags.[6] On average, Bags participants produced 3.9 bags per week. Among clients who participated in a given week, the average number of bags submitted was 6.7. Most people submitted either zero or 10 bags, as demonstrated in Appendix Figure 3. Appendix Figure 4 shows the distribution of earnings, broken down by complex and simple bags. Both have a mode at zero (consistent with Figure 3), and the simple bags do earn more (undoubtedly because the task was easier, which led to both more bag production and higher quality scores).

Second, we have time use surveys in which the female household head reported how she spent her time the previous day. We administered these surveys five times monthly during the Bags program, to 1,238 households, including almost all 896 Bags households and roughly 20% of No-Bags households. On average, 80% of Bags households were found and surveyed each month. In our time use survey, rather than asking about time spent on bags directly, we asked only about "wage labor (including bags)" in order to maintain separation between the evaluation team and the team that was implementing the Bags program. We thus impute time on bags by taking reported time on wage labor, and subtracting average time on wage labor from the Control-No-Bags, GUP-No-Bags, and SOUP-No-Bags households for each Bags group, respectively. See Appendix B for details.

Third, we have a series of standard and comprehensive household surveys that were part of the larger program evaluation of the Graduation program (Banerjee et al., 2015), including a baseline survey, three shorter midline surveys (conducted with one third of the sample), a two-year follow-up survey and a three-year follow-up survey. These surveys included questions about income, consumption, agricultural outcomes, business outcomes, and welfare. The second midline survey is used heavily, as it took place during the bags program. It includes 865 households: 288 Bags households and 577 No-Bags households.[7]

[6] At the end of the program, the bags were sold to other research teams, who used them as participant gifts.

[7] We discuss survey attrition in Appendix C. We do not find differential attrition by treatment group for the key comparisons used in this paper.

4.2 Orthogonality

Appendix Tables 1 and 2 show baseline indicators, including a wealth index that aggregates the seven variables related to overall wealth, across treatment groups. We fail to reject differences for 9 out of 14 variables examined, but do have imbalance on 5: average age, land area, monthly per capita consumption, monthly household income, and the food security index. We had intended to re-randomize, but due to a coding error, it did not happen. To be conservative, we control for these five baseline variables in all regressions.[8] Moreover, in each table, we report "baseline p-values" that correspond to treatment-control differentials, estimated using the same specification, for the baseline value of the outcome.

4.3 Method of Analysis

We use two main specifications for our three types of data: one for the analysis of individual-level outcomes measured in our surveys (Equation 1); and one for the analysis of individual-month level time use outcomes, or individual-week level bag-making outcomes, measured during the Bags program (Equation 2). Any deviations from these specifications or additional details will be reported in table notes.

$$Y_i = \alpha + \beta T_i + \gamma Y_i^0 + W_i^{strata} + \theta^{interviewer} + \epsilon_i \tag{1}$$

$$Y_{it} = \alpha + \beta T_i + W_i^{strata} + \rho^{station \times t} + \epsilon_{it} \tag{2}$$

$Y_{i(t)}$ is outcome Y for individual i at either month or week t, T_i is a treatment dummy, Y_i^0 is the baseline value of outcome Y for individual i (only used in Equation 1 since we do not have baseline data for time use or bag-making), W_i^{strata} is a vector of baseline controls that consists of the variables we intended to use for re-randomization plus the five variables that were imbalanced at baseline, $\theta^{interviewer}$ are interviewer fixed effects, and $\rho^{station*t}$ are either $station \times week$ or $station \times month$ fixed effects. We cluster standard errors at the village level, since both GUP/SOUP and Bags were assigned at the village level.

We use the Benjamini-Hochberg procedures(Benjamini and Hochberg, 1995) put forward in Anderson 2008 to compute q-values that correct for the multiple hypotheses within each table (and sometimes within panels). We do not extend these corrections beyond the boundary of an individual table (or panel) because the substantive aspects of the hypotheses we test change dramatically across tables.

5 Results

5.1 Effects of GUP on income and labor supply

Table 2 reports the main results, the GUP and SOUP treatment effects for Bags households on both Bags and non-Bags labor supply. We focus on Bags households, so from now on we will refer to these households simply as "GUP" "SOUP" or "Control." In this section we focus on Panels A and B, which compare GUP, SOUP, and Control households.

[8] In Appendix Tables 9, 10, and 11, we show that our main results are robust to the exclusion of these controls.

In column 1 we look at monthly household income (excluding bags earnings) during the bag-making program. This is the sum of self-reported income at the midline that took place during Bags, plus monthly unconditional consumption support, which was received in varying amounts. GUP households earn more than triple their Control counterparts.

Columns 2 through 5 show impacts on bag-making. GUP participants produce 30% more bags, are 21% more likely to submit any bags, and earn 27% more from bags production than Control. On the other hand SOUP participants are actually less likely to produce bags, produce fewer bags and earn less from bags production than Control. The difference with Control is not statistically significant, but SOUP participants under-perform GUP participants on almost every measure (for example, there is a 23 percentage point gap in participation).

Column 6 shows impacts on time spent on non-bags activities, including farming, businesses, animals, and home labor. GUP and SOUP households may spend a bit less time (9 minutes) on daily non-bags productive labor, but this difference constitutes only 1.5% of the control group mean and is not statistically significant.[9] In Table 3 columns 5-7 we see that this effect is composed of a reduction in time spent on agriculture combined with smaller increases in time spent on non-agricultural businesses and home labor. In column 9 we do see that GUP households are also spending slightly more time on leisure (13 minutes), which, given how many more bags they produce, may indicate improved productivity.

The differences in bag-making become more stark when we focus on complex bag production in Panel B of Table 2. GUP households produce 48% more complex bags than control households on a weekly basis.[10] SOUP households, on the other hand, produce 57% fewer complex bags than Control and a fortiori than GUP.

5.2 Which components of GUP drive the result?

What is the source of the labor supply effect? We first consider the role of savings collection, since Result 3 from our model shows that access to savings can boost labor supply by raising the marginal utility of income. The dramatic gap in bag-making between GUP and SOUP households displayed in Table 2 suggests that savings collection is unlikely to be the source of GUP's positive effect on bag-making. This conclusion is further supported by the fact that the impacts of GUP, including those on bag-making, are not significantly different for households with and without savings collection (see Appendix Tables 7 and 8).[11]

Second, we explore the role of income versus encouragement. Since the Graduation program aims to simultaneously shift income and encourage more labor through psychosocial improvements, we take advantage of experimental variation in the amount of the unconditional cash transfer to isolate the cash income effect.[12] Panel C of Table 2 compares the outcomes of GUP participants receiving transfers of $3.9 weekly with those receiving transfers of $1.3 weekly. While this is a fairly small treatment difference, it does matter: Column 1 shows that high-UCT participants earn about 50% more than their low-UCT counterparts (though this difference is not statistically significant because of high standard errors on self-

[9] We additionally replicate the basic finding that, considering self-reported labor supply among the full sample at the end of the program, GUP does not produce a classic negative income effect (see Appendix Table 3)

[10] In Appendix E we explore these effects further.

[11] The one exception is that GUP-no-savings households spend more time on leisure than GUP-savings households (but not more than Control households).

[12] The external validity of this test may be limited if there is complementarity between GUP and the cash transfer.

reported monthly income). Column 2 shows that the bags production index is higher for GUP households receiving high UCT than for low UCT, but the difference between the two is not statistically different from zero. The same pattern is seen across the components of the index, suggesting that high UCT households are important drivers of the positive effect on bag production. Moreover, column 6 shows that any reduction in time on non-bags labor for GUP participants is likely to be driven by low UCT households. All together, it seems unlikely to be the case that this pure income shock generates a classical negative income effect.

5.3 Are we missing the effect on effort?

Of course, labor supply consists of more than labor hours, and there could also be a reduction in effort in other activities. To get at a measure of effort we start from the fact that there is minimal wage labor in our context. Individuals either work on their own farms or run their own businesses. In both cases the household is the residual claimant and the effective labor supply, including any differences in effort, should be reflected in the income from the activity. We also look at labor-saving investments, which would allow earnings to go up even when effort has gone down.

Tables 3 and 4 report GUP-induced changes in agriculture (the dominant household enterprise) and non-farm enterprises for Bags households. There is no difference in the amount of hired labor used by GUP households compared to Control households (Table 3 column 1), despite the fact that GUP recipients do spend somewhat less time on their farms compared to Control households (Table 3 column 5). At the same time we see only minimal evidence of labor-saving expenditures, the most important of which would be herbicide. Column 2 shows a statistically significant increase in expenditure on herbicide among GUP household, which is large relative to the control mean, but the absolute magnitude is very small. As a point of comparison, the increase in herbicide equals about two percent of the average use by farmers in this region.[13] Moreover, there is a more sizable increase in fertilizer expenditure (10 percent of the average use in the region), which is labor-using rather than labor-saving because of its effects on weed growth and output. Table 4 shows that agricultural revenue is no lower for GUP households—the point estimate is positive (column 1). Moreover GUP has no impact on residual productivity, which is the residual from regressing harvest value on input expenditure, acreage and labor time, and is an attempt to measure the effort the household is putting into agriculture (column 2). In other words there is no evidence that GUP households are neglecting their agricultural business, despite the fact that they are spending less time on them.

The same holds for their other businesses—the effect on business revenue and earnings (Table 4 columns 3 and 4) is positive, albeit not statistically significant—and the effect on time spent on the business is positive and statistically significant (Table 3 column 6). We do not have ways to measure labor substitution for these businesses, but given the (tiny) scale of the businesses, this seems unlikely.

Wage labor is uncommon in our sample.[14] Table 4 column 5 shows that for Bags households during the bags program, GUP households earned $1.11 less in monthly wage income relative to Control households (who only earn $1.52 in wage income monthly). Thus there may be substitution away from wage labor, but this is small relative to the increases in earnings across the other sources.

[13] This is calculated from data from the same agroclimatic zone from a representative set of farmers in villages with fewer than 50 compounds (Udry, 2019). Households in our sample, who were classified as "ultra-poor," on average use much smaller amounts.

[14] See Appendix F for details.

Finally, it is possible that GUP households do less household work or shift toward more relaxing or flexible labor in ways that would be difficult to measure.[15] There is no difference in the time spent on household work among GUP, SOUP, and Control households assigned to Bags (Table 3). Although we cannot rule out the possibility that GUP households reduced effort on unmeasured margins, it seems equally likely that it could be the opposite: for example, GUP households might be exerting more effort in housework due to the presence of additional livestock. In net, these effects are likely to be second order.

Importantly, the high UCT GUP households do not drive any differences in time spent on non-bags labor (Table 2 Panel C column 6) or expenditure on labor-saving substitutes (Table 3 Panel B columns 1-3). Their harvest value and residual productivity are statistically significantly higher than the low UCT households (Table 4 Panel B columns 1-2), suggesting that if there is any crowd out of farming effort due to the GUP intervention, it is happening only for the low UCT households.

6 Conclusion

The idea that there may be positive rather than negative income effects on labor supply has a long pedigree. We provide support for this view based on a field experiment designed for this purpose.

Specifically, we find that GUP has a positive effect on income, but does not reduce labor supply, and in fact raises production of bags and especially production of complex bags. From Results 1 and 2 in our theoretical model, these results are consistent with a psychological/physiological effect on labor supply brought about by the GUP intervention. The weak impacts of GUP on consumption and health shown in Appendix Table 3 Panel B suggest that nutritional or other physiological mechanisms are less likely to explain the observed increases in labor supply, especially given that SOUP generated similar impacts on consumption without corresponding labor supply effects. That said, we do not rule out a physiological channel, since even small amounts of measurement error in self-reported consumption may make it difficult to detect such effects.

The observed labor supply effects are not driven by the savings component, and seems to be driven at least in part by pure income. That said, we urge caution in interpreting the entirety of the GUP impact as a pure income effect, since several components (specifically, the informational and encouragement aspects of household visits from a case worker to the household) were not experimentally tested.[16] Overall, our results should strengthen the case for well-designed transfer programs to those living in poverty.

[15] For example, GUP households might buy milled grain instead of grinding it themselves.

[16] In other contexts, such visits have not produced an additive effect on income, although this remains under-researched (Blattman et al., 2016).

References

Baird, S., McKenzie, D., and Özler, B. (2018). The effects of cash transfers on adult labor market outcomes. *IZA Journal of Development and Migration*, 8(1):22.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M., and Heil, A. (2022). Why do people stay poor? *The Quarterly Journal of Economics*, 137(2):785–844.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2017). Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, 132(2):811–870.

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

Banerjee, A., Duflo, E., and Sharma, G. (2021). Long-term effects of the targeting the ultra poor program. *American Economic Review: Insights*, 3(4):471–86.

Banerjee, A., Karlan, D., Osei, R., Trachtman, H., and Udry, C. (2022). Unpacking a multi-faceted program to build sustainable income for the very poor. *Journal of Development Economics*, 155:102781.

Banerjee, A., Rema, H., Gabriel, K., and Benjamin, O. (2017). Debunking the myth of the lazy welfare recipient: Evidence from cash transfer programs worldwide. *World Bank Research Observer*, 32(2):155–184.

Barker, N., Karlan, D., Udry, C., and Wright, K. (2023). The fading treatment effects of a multi-faceted asset-transfer program in ethiopia. *Buffett Institute Global Poverty Research Lab Working Paper*.

Benjamin, D. (1992). Household composition, labor markets, and labor demand: Testing for separation in agricultural household models. *Econometrica*, 60(2):287–322.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Bernard, T., Dercon, S., Orkin, K., and Seyoum Taffesse, A. (2020). The future in mind: Long-run impact of an aspirations intervention in rural Ethiopia. Working paper.

Blattman, C., Green, E. P., Jamison, J., Lehmann, M. C., and Annan, J. (2016). The Returns to Microenterprise Support among the Ultrapoor: A Field Experiment in Postwar Uganda. *American Economic Journal: Applied Economics*, 8(2):35–64.

Breza, E., Kaur, S., and Shamdasani, Y. (2018). The morale effects of pay inequality. *The Quarterly Journal of Economics*, 133(2):611–663.

Browning, M. and Chiappori, P.-A. (1998). Efficient intra-household allocations: A general characterization and empirical tests. *Econometrica*, 66(6):1241–1278.

Brune, L. (2016). The effect of lottery-incentives on labor supply: Evidence from a firm experiment in malawi. Working paper.

Brune, L., Chyn, E., and Kerwin, J. T. (2019). Pay me later: A simple employerbased saving scheme. Working paper.

Dalton, P. S., Ghosal, S., and Mani, A. (2016). Poverty and aspirations failure. *The Economic Journal*, 126(590):165–188.

Dasgupta, P. and Ray, D. (1986). Inequality as a determinant of malnutrition and unemployment: Theory. *The Economic Journal*, 96(384):1011–1034.

Deaton, A. (2005). Measuring poverty in a growing world (or measuring growth in a poor world). *Review of Economics and Statistics*, 87(1):1–19.

Fehr, D., Fink, G., and Jack, B. K. (2022). Poor and rational: Decision-making under scarcity. *Journal of Political Economy*, 130(11):2862–2897.

Fink, G., Jack, B. K., and Masiye, F. (2020). Seasonal liquidity, rural labor markets and agricultural production. *American Economic Review*, 110(10):3351â3392.

Foster, A. D. and Rosenzweig, M. R. (1996). Comparative advantage, information and the allocation of workers to tasks: Evidence from an agricultural labour market. *The Review of Economic Studies*, 63(3):347–374.

Genicot, G. and Ray, D. (2017). Aspirations and inequality. *Econometrica*, 85(2):489–519.

Gertler, P., Martinez, S., and Rubio-Codina, M. (2012). Investing cash transfers to raise long term living standards. *American Economic Journal: Applied Economics*, 4(1):164–192.

Goldberg, J. (2016). Kwacha gonna do? Experimental evidence about labor supply in rural Malawi. *American Economic Journal: Applied Economics*, 8(1):129–49.

Guiteras, R. P. and Jack, B. K. (2018). Productivity in piece-rate labor markets: Evidence from rural Malawi. *Journal of Development Economics*, 131:42–61.

Haushofer, J. and Fehr, E. (2014). On the psychology of poverty. *Science*, 344(6186):862–867.

Haushofer, J. and Fehr, E. (2019). Negative income shocks increase discount rates. Working paper.

Haushofer, J. and Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *The Quarterly Journal of Economics*, 131(4):1973–2042.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy*, 114(3):538–575.

Kandasamy, N., Hardy, B., Page, L., Schaffner, M., Graggaber, J., Powlson, A. S., Fletcher, P. C., Gurnell, M., and Coates, J. (2014). Cortisol shifts financial risk preferences. *Proceedings of the National Academy of Sciences*, 111(9):3608–3613.

Karlan, D., Savonitto, B., Thuysbaert, B., and Udry, C. (2017). Impact of savings groups on the lives of the poor. *Proceedings of the National Academy of Sciences*, 114(12):3079–3084.

Kaur, S., Kremer, M., and Mullainathan, S. (2015). Self-control at work. *Journal of Political Economy*, 123(6):1227–1277.

Kaur, S., Mullainathan, S., Schilbach, F., and Oh, S. (2022). Do financial concerns make workers less productive? Working paper.

Kochar, A. (1999). Smoothing consumption by smoothing income: Hours-of-work responses to idiosyncratic agricultural shocks in rural India. *Review of Economics and Statistics*, 81(1):50–61.

LaFave, D. R., Peet, E. D., and Thomas, D. (2020). Farm profits, prices and household behavior. Working Paper 26636.

Leibenstein, H. (1957). *Economic backwardness and economic growth*. Wiley, New York.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *Manchester School*, 22(2):139–191.

Lybbert, T. J. and Wydick, B. (2017). Hope as aspirations, agency, and pathways: Poverty dynamics and microfinance in Oaxaca, Mexico. In *The Economics of Poverty Traps*, NBER Chapters, pages 153–177. National Bureau of Economic Research, Inc.

Lybbert, T. J. and Wydick, B. (2018). Poverty, aspirations, and the economics of hope. *Economic Development and Cultural Change*, 66(4):709–753.

Mani, A., Mullainathan, S., Shafir, E., and Zhao, J. (2013). Poverty impedes cognitive function. *Science*, 341(6149):976–980.

Mullainathan, S. and Shafir, E. (2013). *Scarcity: Why having too little means so much*. Times Books/Henry Holt and Co.

Pinto, R. (2021). Financially stressed employees. SSRN Working Paper 3823267.

Rose, E. (2001). Ex ante and ex post labor supply response to risk in a low-income area. *Journal of Development Economics*, 64(2):371–388.

Rosenzweig, M. R. (1988). Labor markets in low-income countries. *Handbook of Development Economics*, 1:713–762.

Shah, A. K., Mullainathan, S., and Shafir, E. (2012). Some consequences of having too little. *Science*, 338(6107):682–685.

Strauss, J. and Thomas, D. (1998). Health, nutrition, and economic development. *Journal of Economic Literature*, 36(2):766–817.

Tanaka, T., Camerer, C. F., and Nguyen, Q. (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review*, 100(1):557–71.

Udry, C. (2019). Information, market access and risk: Addressing constraints to agricultural transformation in Northern Ghana. *Draft Report*.

Table 1: Experimental Design

Panel A: Intervention and Bags Assignments

Intervention Village Assignment	Bags Village Assignment	# Villages	Household Assignment	# Households
control	No Bags	15	Control-No-Bags	282
	Bags	42	Control-No-Bags	376
			Control-Bags	397
GUP	No Bags	29	Control-No-Bags	276
			GUP-No-Bags	294
	Bags	25	Control-No-Bags	215
			GUP-Bags	221
SOUP	No Bags	28	Control-No-Bags	208
			SOUP-No-Bags	283
	Bags	26	Control-No-Bags	203
			SOUP-Bags	278
TOTAL		165		3033

Panel B: Bags Sub-Treatment Assignment

Intervention Village Assignment - Bags	Bags Simple/Complex Sub-treatment	Bags UCT Sub-treatment	# Villages	# Households
Control-Bags	Simple	n/a	21	189
	Complex	n/a	21	208
GUP-Bags	Simple	High UCT	8	59
		Low UCT	7	71
	Complex	High UCT	7	61
		Low UCT	3	30
SOUP-Bags	Simple	n/a	13	141
	Complex	n/a	13	137
TOTAL			93	896

Panel A shows intervention treatment assignments (GUP, SOUP, and control) and assignment to the Bags program. Both were assigned at the village level. Within each village assigned to GUP or SOUP, about half of sample households were treated with GUP or SOUP, respectively. All treated households in Bags villages received the Bags program. In Control villages assigned to Bags, about half of sample households were selected to receive the Bags program. Panel B shows sub-treatments within the Bags program. All sub-treatments were randomized at the village level such that all individuals within a village who received the Bags program received identical sub-treatment assignments. Control-Bags = intervention control villages assigned to Bags. GUP-Bags = GUP intervention villages assigned to bags. SOUP-Bags = SOUP intervention villages assigned to bags. Simple = assigned to sew the simple bag. Complex = assigned to sew the complex bag. High UCT = GUP intervention households with Bags who received an unconditional cash transfer of $3.92 each week. Low UCT = GUP intervention households with Bags who received an unconditional cash transfer of $1.31 each week. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Table 2: Effects of GUP, SOUP, and UCT on Income and Labor Supply Among Bags Households During Bags Program

Panel A: Effects of GUP and SOUP

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
GUP	ITT	18.95	0.27	1.14	0.12	0.66	-8.74
	SE	(5.94)	(0.12)	(0.50)	(0.05)	(0.35)	(12.75)
	p-val	0.00	0.03	0.03	0.02	0.06	0.49
	q-val	0.12	0.15	0.15	0.15	0.15	0.55
SOUP	ITT	-4.19	-0.18	-0.65	-0.11	-0.33	-8.64
	SE	(2.94)	(0.13)	(0.49)	(0.06)	(0.38)	(12.76)
	p-val	0.16	0.16	0.18	0.06	0.39	0.50
	q-val	0.27	0.27	0.28	0.15	0.45	0.55
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Obs		288	18816	18816	18816	18816	3442
GUP - SOUP	ITT	23.13	0.45	1.79	0.24	0.99	-0.10
GUP - SOUP	p-val	0.00	0.00	0.00	0.00	0.03	0.99

Panel B: Effects of GUP and SOUP for Households Assigned Complex Bags

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
GUP-Complex	ITT	7.70	0.29	1.43	0.12	0.59	-4.40
	SE	(3.78)	(0.19)	(0.77)	(0.10)	(0.43)	(18.39)
	p-val	0.06	0.13	0.07	0.23	0.17	0.81
	q-val	0.15	0.24	0.16	0.31	0.27	0.86
SOUP-Complex	ITT	-8.80	-0.28	-1.08	-0.15	-0.60	-0.48
	SE	(4.12)	(0.14)	(0.52)	(0.08)	(0.35)	(15.01)
	p-val	0.05	0.06	0.04	0.07	0.09	0.97
	q-val	0.15	0.15	0.15	0.16	0.18	0.98
Ctrl Mean		14.44	-0.16	3.01	0.54	1.88	541.23
Obs		160	9156	9156	9156	9156	1608
GUP - SOUP	ITT	16.49	0.56	2.50	0.27	1.20	-3.92
GUP - SOUP	p-val	0.00	0.00	0.00	0.01	0.01	0.81

Panel C: Effects of UCT

		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
High UCT	ITT	22.26	0.35	1.46	0.16	0.86	2.69
	SE	(8.56)	(0.16)	(0.67)	(0.07)	(0.49)	(12.94)
	p-val	0.01	0.04	0.03	0.03	0.09	0.84
	q-val	0.15	0.15	0.15	0.15	0.18	0.86
Low UCT	ITT	14.49	0.18	0.77	0.08	0.44	-21.89
	SE	(8.28)	(0.15)	(0.66)	(0.07)	(0.41)	(19.90)
	p-val	0.09	0.24	0.24	0.22	0.28	0.27
	q-val	0.18	0.31	0.31	0.31	0.34	0.34
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Obs		288	18816	18816	18816	18816	3442
High - Low	ITT	7.76	0.17	0.70	0.07	0.42	24.58
High - Low	p-val	0.53	0.43	0.44	0.41	0.49	0.26

Treatment effects among Bags households during the Bags program. In all panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP-Bags and SOUP-Bags. Panel B shows effects of GUP-Bags and SOUP-Bags for households assigned complex bags. Panel C shows effects of GUP-Bags separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). For column 1, we use household-level data from the midline survey during bag-making, and we include surveyor fixed effects as well as the baseline value of the outcome when possible. For columns 2-5, we use household-week-level bag production data, and include station-week fixed effects. The bags production index is a standardized index of the variables in columns 3-5, centered around the Control-Bags mean. For column 6, we use household-month-level data from time use surveys during bag-making, and include station-month fixed effects. Daily time non-bags labor is minutes spent yesterday on farming, business operations, animal production and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Table 3: Effects of GUP, SOUP, and UCT on Inputs Among Bags Households During Bags Program

Panel A: Effects of GUP and SOUP

		Exp. Labor (1)	**Exp. Herbicide** (2)	**Exp. Fertilizer** (3)	**Time Bags** (4)	**Time Field** (5)	**Time Business** (6)	**Time Home** (7)	**Time All Labor** (8)	**Time Leisure** (9)
GUP	ITT	2.59	3.83	20.06	-2.12	-20.96	9.22	1.18	-12.63	12.89
	SE	(3.54)	(2.11)	(13.35)	(11.64)	(9.80)	(4.80)	(11.79)	(12.35)	(7.22)
	p-val	0.47	0.08	0.14	0.86	0.04	0.06	0.92	0.31	0.08
	q-val	0.67	0.22	0.35	0.88	0.19	0.22	0.92	0.49	0.22
	Bsl p-val	0.536	0.199	0.045	.	.	.	.	.	.
SOUP	ITT	-0.59	1.27	17.49	-21.85	-24.02	10.06	4.41	-23.30	16.24
	SE	(2.18)	(2.32)	(20.01)	(12.37)	(10.86)	(4.74)	(11.42)	(11.58)	(9.25)
	p-val	0.79	0.59	0.39	0.08	0.03	0.04	0.70	0.05	0.08
	q-val	0.86	0.71	0.58	0.22	0.19	0.19	0.79	0.21	0.22
	Bsl p-val	0.437	0.879	0.039	.	.	.	.	.	.
Ctrl Mean		4.02	3.83	57.58	112.94	206.61	28.48	299.27	603.49	127.17
Ctrl SD		15.67	10.82	88.11	106.61	185.65	78.23	155.09	168.19	140.45
Obs		272	272	272	1978	3442	3442	3442	3442	3442
GUP - SOUP	ITT	3.17	2.56	2.58	19.74	3.06	-0.84	-3.23	10.68	-3.35
GUP - SOUP	p-val	0.35	0.36	0.90	0.19	0.79	0.88	0.81	0.33	0.72

Panel B: Effects of UCT

		Exp. Labor (1)	**Exp. Herbicide** (2)	**Exp. Fertilizer** (3)	**Time Bags** (4)	**Time Field** (5)	**Time Business** (6)	**Time Home** (7)	**Time All Labor** (8)	**Time Leisure** (9)
High UCT	ITT	-2.62	-0.32	26.20	-28.90	-25.13	14.85	11.73	-13.91	12.51
	SE	(3.68)	(1.64)	(14.74)	(12.55)	(11.82)	(5.95)	(11.38)	(11.87)	(9.38)
	p-val	0.48	0.85	0.09	0.02	0.04	0.01	0.31	0.24	0.19
	q-val	0.67	0.88	0.22	0.19	0.19	0.19	0.49	0.42	0.38
	Bsl p-val	0.417	0.212	0.077	.	.	.	.	.	.
Low UCT	ITT	9.45	9.36	11.91	18.19	-16.15	2.75	-10.95	-11.14	13.33
	SE	(7.69)	(3.38)	(22.04)	(13.66)	(12.74)	(6.07)	(17.66)	(18.73)	(9.29)
	p-val	0.23	0.01	0.59	0.19	0.21	0.65	0.54	0.55	0.15
	q-val	0.41	0.19	0.71	0.38	0.40	0.76	0.71	0.71	0.35
	Bsl p-val	0.938	0.513	0.124	.	.	.	.	.	.
Ctrl Mean		4.02	3.83	57.58	112.94	206.61	28.48	299.27	603.49	127.17
Ctrl SD		15.67	10.82	88.11	106.61	185.65	78.23	155.09	168.19	140.45
Obs		272	272	272	1978	3442	3442	3442	3442	3442
High - Low	ITT	-12.07	-9.68	14.30	-47.08	-8.98	12.10	22.68	-2.77	-0.83
High - Low	p-val	0.21	0.03	0.56	0.01	0.55	0.11	0.22	0.88	0.94

Treatment effects on inputs among Bags households during the Bags program. In both panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP-Bags and SOUP-Bags. Panel B shows effects of GUP-Bags separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). Columns 1-3 use household-level data from the midline survey during bag-making, and we include surveyor fixed effects as well as the baseline value of the outcome when possible. Expenditure on labor, herbicide, fertilizer are expenditures in the last 12 months. Columns 4-9 use household-month-level data from time use surveys (minutes spent on each activity yesterday) during bag-making, and include station-month fixed effects. Home labor includes childcare, cleaning, cooking, collecting firewood/water, and shopping. All labor includes time spent on bags and other wage labor, agriculture, businesses, animals (which is minimal, so we do not show separately), and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Table 4: Effects of GUP, SOUP, and UCT on Non-Bags Outputs Among Bags Households During Bags Program

Panel A: Effects of GUP and SOUP

		Monthly Harvest Value (1)	**Monthly Residual Prod.** (2)	**Monthly Business Revenue** (3)	**Monthly Business Income** (4)	**Monthly Wage Income** (5)	**Monthly Non-Bags Income** (6)
GUP	ITT	59.31	-4.52	14.97	8.76	-1.11	6.74
	SE	(55.62)	(64.00)	(11.66)	(6.05)	(0.56)	(5.88)
	p-val	0.30	0.94	0.21	0.16	0.06	0.26
	q-val	0.42	0.94	0.42	0.38	0.34	0.42
	Bsl p-val	0.061	.	$0.093^{(+)}$	$0.026^{(+)}$	0.304	0.927
SOUP	ITT	-22.20	-60.59	-9.47	-2.94	-0.73	-3.59
	SE	(70.42)	(54.77)	(6.16)	(2.56)	(0.91)	(2.78)
	p-val	0.75	0.28	0.13	0.26	0.43	0.21
	q-val	0.79	0.42	0.38	0.42	0.57	0.42
	Bsl p-val	0.136	.	0.181	$0.006^{(+)}$	0.007	0.832
Ctrl Mean		408.47	-13.50	13.94	6.54	1.52	8.01
Ctrl SD		461.91	349.39	55.61	24.83	4.66	25.04
Obs		272	266	287	287	288	288
GUP - SOUP	ITT	81.51	56.07	24.44	11.71	-0.38	10.34
GUP - SOUP	p-val	0.18	0.39	0.05	0.06	0.63	0.08

Panel B: Effects of UCT

		Monthly Harvest Value (1)	**Monthly Residual Prod.** (2)	**Monthly Business Revenue** (3)	**Monthly Business Income** (4)	**Monthly Wage Income** (5)	**Monthly Non-Bags Income** (6)
High UCT	ITT	138.60	94.92	10.58	6.42	-0.79	5.21
	SE	(54.79)	(59.36)	(18.88)	(8.52)	(0.53)	(8.56)
	p-val	0.02	0.12	0.58	0.46	0.15	0.55
	q-val	0.34	0.38	0.63	0.57	0.38	0.63
	Bsl p-val	0.024	.	0.638	$0.010^{(+)}$	0.807	0.773
Low UCT	ITT	-46.51	-136.67	21.01	12.00	-1.52	8.81
	SE	(64.83)	(68.40)	(11.59)	(8.23)	(0.69)	(8.28)
	p-val	0.48	0.06	0.08	0.16	0.03	0.30
	q-val	0.57	0.34	0.38	0.38	0.34	0.42
	Bsl p-val	0.524	.	$0.058^{(+)}$	0.352	0.008	0.135
Ctrl Mean		408.47	-13.50	13.94	6.54	1.52	8.01
Ctrl SD		461.91	349.39	55.61	24.83	4.66	25.04
Obs		272	266	287	287	288	288
High - Low	ITT	185.11	231.60	-10.43	-5.58	0.73	-3.60
High - Low	p-val	0.02	0.01	0.65	0.64	0.21	0.77

Treatment effects on non-bags outputs among Bags households during the Bags program. In both panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP and SOUP. Panel B shows effects of GUP separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). We use household-level data from the midline survey during bag-making, and we include surveyor fixed effects as well as the baseline value of the outcome when possible. Harvest value is the total revenue from harvest, six months after the bag-making period. Residual productivity contains the residuals from a regression of harvest value on input expenditure, acreage, and average time on field. Business revenue and income pertain to all non-agricultural businesses. Wage income does not include earnings from bags. Non-Bags income includes income from agriculture, business, wages, and animals. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Online Appendix

A Program Implementation Details

A.1 GUP and SOUP

In this section we elaborate on the implementation of the GUP and SOUP programs.

Presbyterian Agricultural Services (PAS), our implementation partner, has experience doing extension work and promotion of savings groups, including a prior randomized controlled trial with Innovations for Poverty Action (Karlan et al., 2017). PAS field agents engaged in the direct field implementation, while Innovations for Poverty Action coordinated the implementation with senior management of PAS.

Participants in the program come from three areas of Northern Ghana corresponding to three agricultural "stations" run by PAS: Tamale, Langbensi, and Sandema. PAS first identified poor communities, and in each identified community, staff members then facilitated a Participatory Wealth Ranking (PWR) in which members of the community ranked households by economic status. Finally, PAS staff members returned for a verification of the households judged to be the poorest. In order to be eligible, households were required to have a female between the ages of 18-24, who we will call the female household head, to be the direct recipient of all treatment activities and the primary respondent for surveys.

As mentioned in the text, the GUP program included six components, all of which were directed to the female household head: (1) the transfer of a productive asset; (2) skills training for the management of the asset, (3) life skills training and mentorship, via weekly household visits over two years, (4) a weekly cash stipend for consumption support, worth between \$6 and \$9 PPP depending on family size, during each lean season, (5) some basic health services and health education, and (6) access to a savings account at a local bank and deposit collection. The first component, the productive asset transfer, was provided at the beginning of the program, and households were permitted to choose a package of productive assets from a set list. Most households chose a package that included four goats.[17] The skills training, in which participants learned how to take care of the asset (e.g., when to vaccinate goats), took place at the start of the program, and then also as part of weekly household visits by the PAS field officer. The household visits also provided the backbone for delivering components three through six. The third component, a "hand-holding" or life-skills component, provided nudges to help the household focus on building productive assets to generate positive change in long-term outcomes, and more generally, to set aspirations and plans for coping with current problems and improving the future. The consumption support was explicitly intended to help this process in the short-run, by helping to absorb short-run shocks that could lead to households consuming the transferred assets. The fifth component included basic education on health and hygiene as well as enrollment in the national health insurance scheme (about \$2 per month).

As mentioned in the text, the sixth component of GUP, access to a savings account and weekly deposit collection by PAS field agents, was the entirety of the SOUP intervention. Among households assigned to SOUP, there was an additional sub-treatment: half received savings accounts and deposit collection without a match ("SOUP without match") and half received savings accounts and deposit collection with a 50% match ("SOUP-match"). Specifically, for every GHC 1 deposited, households in this group received

[17] Other assets included hens, pigs, and inputs for the production of shea butter, maize, and sorghum.

a matching contribution of GHC 0.50. At the onset of the program, there was a maximum match of GHC 1.50 GHC per week (for a GHC 3 deposit) but this cap was eventually removed.

A.2 Bag-Making

In this section, we elaborate on the quality checks on bags conducted by field agents. There were 18 quality standards for simple bags, and 25 quality standards for complex bags, which were reviewed thoroughly during the training. Bags were assigned one point for meeting the quality standards at the "excellent" level, half a point for "satisfactory," and zero points for "unsatisfactory." At the end of the quality check, the final quality score was calculated and the bag was classified as high, mid, or low quality. Over the course of the study, 35% of bags collected were low quality, 34% were mid quality, and 31% were high quality.

B Imputing Time Spent on Bags

In our time use survey, rather than asking about time spent on bags directly, we asked only about "wage labor (including bags)" in order to maintain a strong separation between the evaluation team and the team that was implementing the bags program. We thus impute time on bags by taking the answer to a question about time on wage labor, and subtracting average time on wage labor from the Control-No-Bags, GUP-No-Bags, and SOUP-No-Bags households for each bags group, respectively.

In Appendix Table 12 we report levels of monthly wage income and time spent on bags and/or wage labor across treatment groups. In Column 1, we can see that within each treatment group—control, GUP, and SOUP—there is very little difference in wage income between bags and No-Bags within each treatment group. Since wage income does not include bags earnings, we infer from this that there were not significant differences in non-bags wage labor across Bags and No-Bags groups. Therefore, any differences in time spent on "time bags and/or wage labor" within each treatment group, between bags and No-Bags, can be attributed to time spent on bags. We thus impute time spent on bags by taking the time spent on "time bags and/or wage labor" for each bags participant, and subtracting the mean time spent on "time bags and/or wage labor" from the corresponding No-Bags treatment group. For example, for a GUP-Bags participant, we take "time spent on bags and/or wage labor" and subtract the mean time spent on "bags and/or wage labor" in GUP-No-Bags, to impute time spent on bags.

C Attrition

In Appendix Table 13 we report attrition by treatment status for the second midline, the two-year follow-up survey, and the time use surveys. For the two-year survey, GUP-No-Bags had slightly higher participation than Control-No-Bags ($p = 0.10$), but participation was between 94% and 100% across all groups, and these are not our principal outcomes. In our time use midlines, we do find that Bags households were substantially more likely to participate than No-Bags households across all treatment groups ($p < 0.01$). (Surveyors reported that they were easier to find, since they were also anticipating visits from field agents who collected their bags on a weekly basis.) Fortunately, we do not focus on this comparison in our analysis,

and we do not find significant differences between Any GUP-Bags and Control-Bags, between Any GUP-Bags and SOUP-Bags, or between high and low UCT. (We do not show attrition results by high versus low UCT in this table, but have checked that there are no significant differences in participation for any of the three surveys.)

D Overall Impacts of GUP and SOUP

In this section, we discuss overall impacts of GUP and SOUP for the full sample of participants. In Appendix Table 3 Panel A we report on the basic treatments, GUP and SOUP, including both Bags and No-Bags households. Columns 1-5 report data collected at the end of the two-year program (7-9 months after the end of bag-making); columns 6 and 7 report time use data collected in the same survey. Consistent with our results from the bag-making sample in Table 2, GUP and SOUP households spend the same amount of time providing productive labor as do control households, and report the same amount of leisure time (each of the estimated treatment effects is smaller than four percent of the control mean, and statistically indistinguishable from zero at any conventional level of significance). The GUP treatment raised the value of livestock owned by the household by more than 30 percent relative to control (itt = $73, s.e. = 18). SOUP households also acquire more livestock (itt = $32, s.e. = 15), but the net increase is significantly less than that for the GUP households. On the other hand, as column 2 reports, SOUP has as large an effect on total asset value as GUP (and both are statistically different from control). The pattern for self-reported monthly income at the two-year mark (in column 3) is similar: both SOUP and GUP have positive point estimates, but the GUP effect is almost twice as large as the SOUP and is the only one that is significant. SOUP households, like GUP households, do not report more time spent on leisure. There are no statistically significant effects on consumption or health (columns 4 and 5).[18]

We next describe the results for the GUP-No-Bags and SOUP-No-Bags interventions, reported in Appendix Table 3 Panel B. Here, GUP-No-Bags is the classic "graduation" intervention. GUP-No-Bags households report statistically significantly lower amounts of leisure than control No-Bags households, and also that they spend more time on productive labor (although this later effect is not statistically significant at conventional levels).[19] SOUP-No-Bags households also report less leisure time and more productive labor supply than control No-Bags households, but neither coefficient is statistically significant (nor can either be distinguished from its corresponding GUP effect). The effects of GUP-No-Bags and SOUP-No-Bags on livestock, total assets and income parallel those of GUP and SOUP overall: GUP-No-Bags has a stronger effect on livestock than SOUP-No-Bags, they have similar impacts on total assets, and GUP-No-Bags has the largest and only statistically significant impact on income. Neither GUP-No-Bags nor SOUP-No-Bags has a noticeable impact on health, but SOUP-No-Bags does increase consumption.

These program impacts indicate, first, that self-reported income was higher among GUP households, both with and without bags, at the end of the two-year program. Second, they show no evidence of a reduction in labor supply.

[18] The discrepancy between household income and consumption that we find is common, see Deaton (2005) for in depth analysis.
[19] The effect of Control-Bags on time spent on productive labor is driven by time spent on bags.

E Exploring Effects on Complex Bags

In Appendix Tables 4-6, we take a closer look at the effects on complex bag-making displayed in Table 2 Panel B. In Appendix Table 4, we show that the additional complex bag production by GUP households does not come at the expense of quality. In Appendix Table 5 we show that this effect is consistent over the course of the bags program, suggesting that the difference in production between GUP and Control households is not due to differential learning rates in the early weeks of the program. Finally, in Appendix Table 6, we show that the effect holds for both high-UCT and low-UCT GUP households.

F Wage Labor

In this section, we present data on wage labor in our sample. In Control-No-Bags, average monthly wage labor earnings (excluding bags) are $1.13, and only 16% of households have positive wage earnings in a month. In terms of time, in Control-No-Bags, average time spent on wage labor is 9.7 minutes daily, and only 7.1% of households spent any time on wage labor yesterday. Demand for wage labor is also low: in Control-No-Bags, yearly expenditure on wage labor is $4.06 and only 9.7% of households demand any labor from the market in a year.

G Intra-Household Allocation of Labor

Here, we address the possibility that our treatments shifted intra-household allocation of labor, which our simple model did not account for. However, a close analogue to Result 2 holds for any collective household (Browning and Chiappori, 1998). The primary beneficiary of all of the GUP interventions was the female head of household, and she was the direct recipient of the bag-sewing training. If the receipt of the GUP intervention does not decrease the bargaining power (Pareto weight) of the recipient, and the recipient supplies more labor to the bags activity despite the fact that the household is richer and she is working no less, then there must be a psychological/physiological productivity effect on the bags activity. We have shown in Tables 2 that GUP increased household income and did not reduce the recipient's overall labor supply, nor is there evidence that it increased her leisure (Table **??**). Overall, 95% of bags were made by the female head of household and 96% were made by an adult female (over age 15). In Appendix Table 14, we find no evidence that GUP, SOUP, or high consumption support affected the fraction of bags made by adult females within the household. Shifts in intrahousehold labor allocation, therefore, do not provide an alternative to a physiological or psychological effect of GUP on the supply of labor.[20]

H Inter-Temporal Substitution of Labor Supply

Importantly, any intertemporal substitution of labor supply induced by the Bags treatment is likely similar for GUP and Control, and thus cannot explain the main effects that we document. Of course, it is possible that GUP interacted with the Bags treatment in ways that produced intertemporal changes in incentives and

[20]The bags program itself may have shifted the intra-household allocation of labor but our evidence focuses on differences between GUP-Bags and Control-Bags households.

opportunities more subtle than the investment productivity effect we describe. Indeed, though such effects are part of what we intend to capture, even if they do not fit perfectly under the labels of physiological or psychological productivity effects.

I Wage Elasticities

As part of the experiment design, we also introduced variation in the wage at the village level over time. Every four weeks, villages were assigned a different baseline wage: \$0.40 or \$0.91. At the start of that week, participants were informed of the payment per bag they would be receiving for bags made in the subsequent four weeks. As discussed in Section 2.2, bags produced in week 1 of a given wage rotation would be collected at the end of week 1 and inspected for quality over the course of weeks 2 and 3. Payment for the bags produced from the beginning of week 1 and collected at the end of week 1 would be made at the end of week 3. For this reason, there may be a lag between when the wage rate changes and when individuals start receiving higher wages.

We explore wage elasticities in Appendix Table 15. Panel A provides evidence that participants were responsive to wages they were receiving for bags submitted previously, as opposed to the correct relevant wage for the bags they were making. We look at three sub-groups. First, we look at participants who were randomly assigned two consecutive high wage months and two consecutive low wage months (39/120 villages, and 363/1098 participants). Second, we look at participant-weeks that were the fourth week in the wage month. If wages changes were only fully internalized upon receiving new wages, then they should take the new wage into account only for bags produced in the fourth week of the month. (The new wage is active in the first week of production; wages for these bags are paid in the third week, and thus only bags collected in the fourth week are produced with experience of new wage.) Finally, we define "experience" to mean either the fourth week of the month, or for "consecutive" participants, any week in the second consecutive month with the same wage. Given this evidence, Panel B shows positive but small elasticity estimates with respect to the 3-week lagged wage. Given the fuzzy understanding of wage changes and when they would take effect, we do not focus on these results in the main part of the paper.

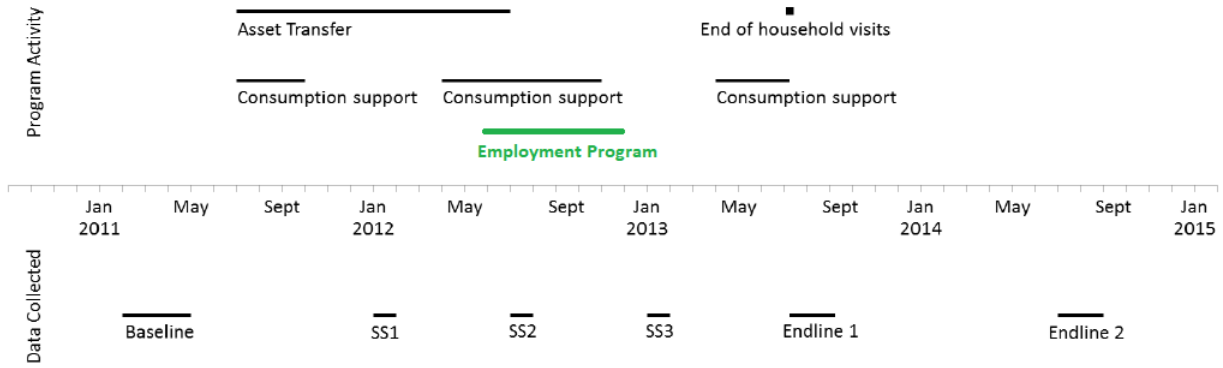
J Additional Figures and Tables

Appendix Figure 1: Simple Bag (left) and Complex Bag (right)

(a) Simple bag *(b) Complex bag*

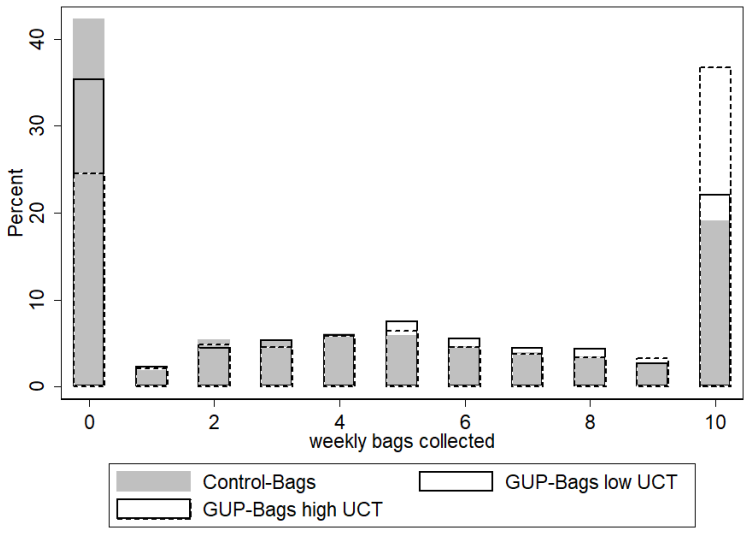
The simple bag has "running" stitches on the hem and strap. The complex bag has a more complicated pattern on the hem and strap: a sequence of four "chain" stitches alternating with one "running" stitch.

Appendix Figure 2: Timeline



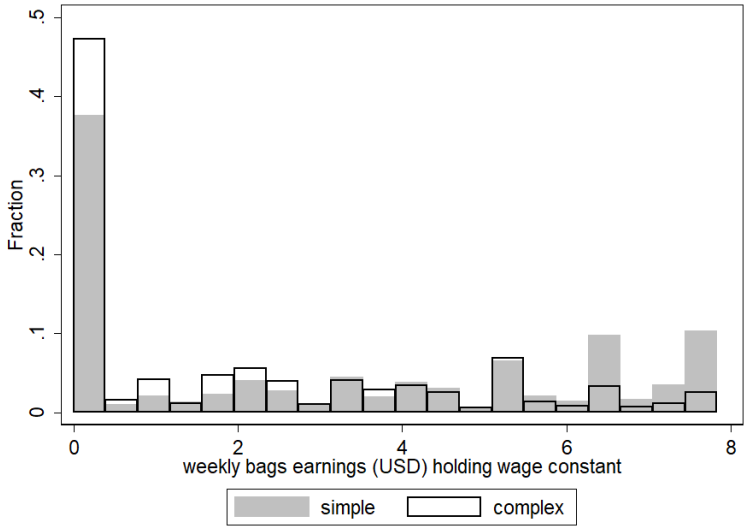
In the top part of the timeline we show program activities, and in the bottom part we show data collection. During the employment program we conducted additional time use surveys each month, over five months.

Appendix Figure 3: Bags by GUP & UCT



The distribution of weekly bags produced by the Bags sample, shown separately for control households, GUP households with high UCT, and GUP households with low UCT.

Appendix Figure 4: Bags Earnings by Complexity



The distribution of weekly earnings from bags produced by the Bags sample, shown separately for households assigned simple bags and households assigned complex bags. Here, earnings are computed holding the baseline piece-rate constant (ignoring the randomized assignment), but naturally including variations that occurred in the piece-rate due to variations in quality.

Appendix Table 1: Household Level Baseline Indicators, Intervention Treatments (Means and Standard Deviations)

	Ctrl No-Bags (1)	**GUP No-Bags** (2)	**SOUP No-Bags** (3)	**Ctrl Bags** (4)	**GUP Bags** (5)	**SOUP Bags** (6)	**p-value F-test Joint Sig** (7)
Household Size	7.16	7.61	7.28	7.18	7.29	7.48	0.82
	(3.88)	(4.16)	(3.72)	(3.50)	(3.62)	(3.52)	
Average Age, Household	25.57	25.02	25.13	24.77	24.37	23.04	0.02
	(10.90)	(10.55)	(10.69)	(9.69)	(9.13)	(8.12)	
Land Area (Acres)	4.54	4.50	4.84	4.65	5.03	4.82	0.81
	(3.92)	(4.02)	(4.12)	(3.77)	(3.88)	(3.94)	
Monthly Per Capita Cons (USD)	58.96	54.00	62.61	58.58	59.24	58.55	0.53
	(38.68)	(34.49)	(40.88)	(39.18)	(41.96)	(37.03)	
Monthly HH Income (USD)	42.34	42.60	48.74	41.12	45.07	45.49	0.87
	(56.76)	(58.84)	(56.77)	(53.13)	(50.23)	(56.72)	
Savings Balances (USD)	2.40	1.63	2.78	1.65	2.46	3.11	0.50
	(11.98)	(8.54)	(13.91)	(9.87)	(11.09)	(14.00)	
Food Security Index	0.00	0.03	0.16	-0.14	0.14	0.19	0.04
	(1.01)	(1.02)	(1.11)	(0.91)	(1.05)	(1.07)	
Asset Value Index	0.03	0.03	0.09	0.03	0.00	0.02	0.99
	(1.07)	(1.01)	(1.12)	(1.02)	(0.73)	(0.77)	
Financial Inclusion Index	0.00	0.00	0.05	-0.08	-0.08	0.19	0.24
	(0.96)	(0.94)	(1.07)	(0.84)	(0.76)	(1.36)	
Wealth Index	0.01	0.00	0.16	-0.06	0.07	0.15	0.50
	(1.01)	(0.94)	(1.13)	(0.93)	(0.89)	(1.04)	
Physical Health Index	-0.06	0.00	0.00	0.03	0.13	0.00	0.12
	(1.02)	(1.07)	(1.04)	(0.98)	(0.93)	(0.99)	
Mental Health Index	0.00	-0.04	0.14	-0.08	0.03	0.08	0.38
	(1.02)	(0.97)	(1.00)	(1.02)	(1.02)	(1.02)	
Political Involvement Index	0.00	0.09	-0.16	0.06	0.00	0.01	0.27
	(1.00)	(0.99)	(1.01)	(0.99)	(1.00)	(1.00)	
Female Empowerment Index	0.00	0.11	-0.09	0.02	-0.12	0.00	0.57
	(0.99)	(1.03)	(0.99)	(1.01)	(1.03)	(1.00)	

Means and standard deviations for key indicators at baseline. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms. Indices are centered around mean baseline value. The wealth index aggregates the seven previous indicators (land area, consumption, income, savings, food security, asset value, and financial inclusion). The last column contains the p-value from an F-test of joint significance of all treatments. See Appendix Table 16 for components of all indices.

Appendix Table 2: Household Level Baseline Indicators, Bags Treatments (Means and Standard Deviations)

	Ctrl-Bags Simple (1)	**Ctrl-Bags Complex** (2)	**GUP-Bags Simple Low UCT** (3)	**GUP-Bags Simple High UCT** (4)	**GUP-Bags Complex Low UCT** (5)	**GUP-Bags Complex High UCT** (6)	**SOUP-Bags Simple** (7)	**SOUP-Bags Complex** (8)	**p-value F-test Joint Sig** (9)
Household Size	6.96 (3.38)	7.38 (3.61)	6.59 (3.21)	7.71 (4.33)	8.13 (4.31)	7.28 (2.83)	7.58 (3.38)	7.38 (3.66)	0.83
Average Age, Household	24.20 (8.53)	25.28 (10.63)	26.87 (11.44)	23.02 (7.43)	24.45 (9.27)	22.74 (6.78)	22.72 (7.45)	23.36 (8.76)	0.34
Land Area, Acres	4.76 (4.22)	4.54 (3.31)	3.45 (3.11)	5.01 (4.14)	6.50 (3.89)	6.17 (3.85)	5.01 (4.05)	4.63 (3.84)	0.04
Monthly Per Capita Cons. (USD)	51.98 (33.12)	64.57 (43.18)	46.03 (35.43)	51.17 (34.99)	73.18 (49.59)	75.58 (44.50)	63.31 (37.35)	53.64 (36.18)	0.00
Monthly HH Income (USD)	41.37 (58.35)	40.89 (48.04)	25.69 (30.27)	43.46 (44.11)	71.51 (79.01)	56.17 (48.62)	53.74 (67.07)	36.99 (42.18)	0.02
Savings Balances (USD)	1.46 (8.19)	1.82 (11.21)	4.27 (15.26)	1.43 (6.25)	1.76 (9.61)	1.73 (9.55)	2.54 (12.41)	3.70 (15.48)	0.69
Food Security Index	-0.28 (0.79)	0.00 (0.99)	-0.13 (0.86)	0.28 (1.16)	0.08 (1.03)	0.36 (1.11)	0.31 (1.05)	0.06 (1.08)	0.00
Asset Value Index	-0.10 (0.88)	0.15 (1.13)	-0.12 (0.67)	0.00 (0.71)	0.12 (0.98)	0.10 (0.68)	0.06 (0.78)	0.00 (0.75)	0.59
Financial Inclusion Index	0.00 (0.89)	-0.14 (0.79)	0.00 (0.91)	-0.05 (0.76)	-0.21 (0.58)	-0.10 (0.67)	0.08 (1.18)	0.30 (1.52)	0.14
Wealth Index	-0.15 (0.93)	0.02 (0.91)	-0.28 (0.89)	0.05 (0.70)	0.35 (1.06)	0.37 (0.84)	0.24 (1.01)	0.05 (1.07)	0.15
Physical Health Index	0.00 (0.98)	0.06 (0.99)	0.12 (1.02)	0.27 (0.62)	-0.09 (1.14)	0.11 (0.95)	0.00 (1.02)	0.03 (0.97)	0.30
Mental Health Index	-0.20 (0.96)	0.02 (1.06)	-0.10 (1.03)	0.18 (1.11)	0.00 (0.98)	0.07 (0.94)	0.14 (1.07)	0.01 (0.96)	0.35
Political Inv. Index	0.06 (0.99)	0.06 (0.99)	0.03 (1.00)	0.12 (0.99)	-0.14 (1.02)	-0.08 (1.01)	0.06 (1.00)	0.00 (1.01)	0.91
Female Emp. Index	0.05 (0.98)	0.01 (1.04)	0.10 (1.07)	-0.10 (0.96)	-0.39 (1.09)	-0.27 (0.97)	-0.06 (1.03)	0.00 (0.97)	0.50

Means and standard deviations for key indicators at baseline. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms. Indices are centered around mean baseline value. The wealth index aggregates the seven previous indicators (land area, consumption, income, savings, food security, asset value, and financial inclusion). The last column contains the p-value from an F-test of joint significance of all treatments. See Appendix Table 16 for components of all indices.

Appendix Table 3: Intervention Treatment Effects at Two Years

Panel A: Impacts of GUP and SOUP

		Livestock Value (1)	Asset Value (2)	Monthly HH Income (3)	Monthly Cons./Cap. (4)	Health Index (5)	Time Prod. Labor (6)	Time Leisure (7)
GUP	ITT	72.92	74.85	7.49	1.13	0.04	-1.52	0.25
	SE	(17.61)	(38.64)	(2.54)	(1.62)	(0.03)	(16.23)	(9.01)
	p-val	0.00	0.05	0.00	0.48	0.24	0.93	0.98
	q-val	0.00	0.22	0.05	0.74	0.46	1.00	1.00
	Bsl p-val	0.416	0.312	0.967	0.765	0.128	.	.
SOUP	ITT	32.09	83.33	4.00	2.91	-0.01	6.12	2.50
	SE	(14.65)	(39.53)	(2.49)	(1.64)	(0.05)	(22.62)	(10.48)
	p-val	0.03	0.04	0.11	0.08	0.85	0.79	0.81
	q-val	0.17	0.18	0.33	0.27	0.95	0.94	0.95
	Bsl p-val	0.797	0.269	0.913	0.930	0.164	.	.
Ctrl Mean		240.17	589.48	36.59	44.15	-0.16	330.83	84.63
Ctrl SD		348.59	764.13	43.08	30.15	0.84	269.48	127.04
Obs		2909	2900	2907	2880	2767	1130	1130
GUP - SOUP = 0	p-val	0.05	0.86	0.25	0.35	0.36	0.76	0.86

Panel B: Impacts by Bags Treatment

		Livestock Value (1)	Asset Value (2)	Monthly HH Income (3)	Monthly Cons./Cap. (4)	Health Index (5)	Time Prod. Labor (6)	Time Leisure (7)
Control-Bags	ITT	-3.16	56.57	2.48	2.98	-0.02	7.45	11.39
	SE	(16.21)	(46.06)	(2.49)	(2.01)	(0.05)	(22.42)	(10.30)
	p-val	0.85	0.22	0.32	0.14	0.65	0.74	0.27
	q-val	0.95	0.45	0.56	0.38	0.91	0.93	0.49
	Bsl p-val	0.161	0.464	0.615	0.646	0.641	.	.
GUP-No-Bags	ITT	77.20	104.70	9.40	2.73	0.01	12.04	9.93
	SE	(23.75)	(47.92)	(3.33)	(2.07)	(0.04)	(38.38)	(22.98)
	p-val	0.00	0.03	0.01	0.19	0.76	0.75	0.67
	q-val	0.04	0.17	0.05	0.44	0.93	0.93	0.91
	Bsl p-val	0.427	0.830	0.949	0.945	0.807	.	.
GUP-Bags	ITT	64.73	72.67	6.44	0.63	0.05	0.34	6.02
	SE	(21.75)	(54.46)	(3.32)	(1.63)	(0.05)	(21.59)	(10.31)
	p-val	0.00	0.18	0.05	0.70	0.35	0.99	0.56
	q-val	0.05	0.44	0.22	0.93	0.57	1.00	0.83
	Bsl p-val	0.148	0.627	0.388	0.876	$0.055^{(+)}$	.	.
SOUP-No-Bags	ITT	26.77	107.61	4.97	5.09	0.03	48.41	0.10
	SE	(18.55)	(49.79)	(3.14)	(1.99)	(0.05)	(41.40)	(14.46)
	p-val	0.15	0.03	0.12	0.01	0.63	0.24	0.99
	q-val	0.39	0.17	0.33	0.09	0.90	0.46	1.00
	Bsl p-val	0.526	0.234	0.763	0.271	$0.016^{(+)}$	.	.
SOUP-Bags	ITT	34.92	91.45	4.22	1.99	-0.07	2.24	10.99
	SE	(20.01)	(50.65)	(3.36)	(1.61)	(0.08)	(27.95)	(11.74)
	p-val	0.08	0.07	0.21	0.22	0.37	0.94	0.35
	q-val	0.27	0.27	0.45	0.45	0.59	1.00	0.57
	Bsl p-val	0.726	0.236	0.724	0.799	0.281	.	.
Ctrl Mean		242.93	578.33	35.94	43.64	-0.17	325.83	83.81
Ctrl SD		356.19	760.35	42.14	29.43	0.85	267.81	128.06
Obs		2909	2900	2907	2880	2767	1130	1130

Treatment effects among all sample households (Bags and No-Bags) two years after the start of the GUP and SOUP interventions. Panel A shows average effects of GUP and SOUP; the omitted group is Control households (Bags and No-Bags) in any village. Panel B shows effects by Bags sub-treatment; the omitted group is Control No-Bags households in any village. The sample is restricted to villages with more than 30 compounds. We control for stratification variables, imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income), and whether or not household was treated with bags (Panel A only). We use household-level data from the two-year survey and include surveyor fixed effects and the baseline value of the outcome when possible. Livestock value is the total number of livestock owned times the median reported price for each animal. Asset value is the total number of assets (including livestock, household and productive assets, and stocks), valued using asset prices relative to the price of goats from other countries. Monthly household income is the sum of income from the household's business, farm, wage labor, and (revenue from) animals. (Income from wage labor does not include bags earnings.) Monthly consumption per capita includes both food and non-food expenditure. Health index includes two variables: average daily living score (mean of capacity bathing, lifting, walking, and working) and sick day (1 if the member did not miss a day of work due to illness in the last year, 0 otherwise). Time productive labor is minutes spent yesterday spent on bags, agriculture, business, home labor, and animals. Time leisure is minutes spent yesterday on religious activities, social activities, ceremonies, traveling, personal care, and resting. (These time use measures are taken at the two-year mark, not during the bags program.) Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 4: Effect of Complex vs. Simple on Bag Quality

		Weekly High Bags (1)	**Mean Quality Score (out of 1)** (2)	**Quality Bag Index** (3)
Control Complex	ITT	-0.68	-0.03	-0.33
	SE	(0.29)	(0.03)	(0.09)
	p-val	0.02	0.30	0.00
	q-val	0.05	0.44	0.00
GUP Simple	ITT	0.37	-0.01	-0.02
	SE	(0.52)	(0.03)	(0.12)
	p-val	0.48	0.68	0.84
	q-val	0.62	0.76	0.84
GUP Complex	ITT	-0.79	-0.06	-0.44
	SE	(0.30)	(0.04)	(0.09)
	p-val	0.01	0.12	0.00
	q-val	0.03	0.21	0.00
Ctrl Mean		1.84	0.73	1.20
Ctrl SD		3.20	0.19	0.70
Obs		18816	10854	10854
GUP Complex - Ctrl Complex	ITT	-0.11	-0.03	-0.11
GUP Complex - Ctrl Complex	p-val	0.65	0.43	0.19

Treatment effects by complexity on bag quality among Bags households during the Bags program. The omitted group is Control-Bags households with simple bags. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). We use household-week-level bag production data, and include station-week fixed effects. Column 1 shows the number of "high quality" bags submitted, column 2 shows the mean quality score (the percent of criteria fulfilled, which was used to determine whether the bag was "low," "mid," or "high" quality) over all of the bags submitted, and column 3 shows a composite measure of bag quality that awards 2 points for every "high" quality bag submitted and 1 point for every "mid" quality bag submitted. Standard errors are clustered at village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel.

Appendix Table 5: Effect of Complex vs. Simple on Bags Labor Supply: Dynamics

		Bag Production Index, First Four Weeks (1)	**Bag Production Index, Final 17 Weeks** (2)
Control Complex	ITT	-0.16	-0.24
	SE	(0.15)	(0.13)
	p-val	0.26	0.08
	q-val	0.46	0.23
GUP Simple	ITT	0.44	0.14
	SE	(0.19)	(0.17)
	p-val	0.02	0.40
	q-val	0.13	0.48
GUP Complex	ITT	0.18	0.08
	SE	(0.18)	(0.21)
	p-val	0.31	0.70
	q-val	0.46	0.70
Ctrl Mean		0.15	0.19
Ctrl SD		1.03	1.06
Obs		3584	15232
GUP Complex - Ctrl Complex	ITT	0.35	0.32
GUP Complex - Ctrl Complex	p-val	0.06	0.11

Treatment effects by complexity on bag-making labor supply for Bags households, looking separately at the first four weeks (after 5 weeks of initial piloting) versus the final 17 weeks. The omitted group is Control-Bags households with simple bags. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). We use household-week-level bag production data, and include station-week fixed effects. The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Standard errors are clustered at village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel.

Appendix Table 6: Effect of UCT for Simple vs. Complex Bags

		Bag Production Index (1)
Control Complex	ITT	-0.23
	SE	(0.13)
	p-val	0.08
	q-val	0.20
GUP Simple Low UCT	ITT	-0.01
	SE	(0.20)
	p-val	0.94
	q-val	0.94
GUP Simple High UCT	ITT	0.46
	SE	(0.20)
	p-val	0.02
	q-val	0.11
GUP Complex Low UCT	ITT	0.27
	SE	(0.22)
	p-val	0.22
	q-val	0.37
GUP Complex High UCT	ITT	0.02
	SE	(0.26)
	p-val	0.93
	q-val	0.94
Ctrl Mean		0.12
Ctrl SD		1.08
Obs		18816
GUP Simple High UCT - GUP Simple Low UCT	ITT	0.48
GUP Simple High UCT - GUP Simple Low UCT	p-val	0.07
GUP Complex High UCT - GUP Complex Low UCT	ITT	-0.25
GUP Complex High UCT - GUP Complex Low UCT	p-val	0.43

Treatment effects by complexity and UCT on bag-making labor supply for Bags households. The omitted group is Control-Bags households with simple bags. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). We use household-week-level bag production data, and include station-week fixed effects. The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Standard errors are clustered at village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel.

Appendix Table 7: Bags Labor Supply among Bags Households During Bags Program, GUP-No-Savings vs. GUP-Savings

		Bags Prod. Index (1)	Number of Bags (2)	Participates (0/1) (3)	Bags Earnings (4)
GUP-No-Savings	ITT	0.22	0.97	0.09	0.53
	SE	(0.13)	(0.53)	(0.06)	(0.35)
	p-val	0.09	0.07	0.11	0.14
	q-val	0.12	0.12	0.13	0.14
GUP-Savings	ITT	0.33	1.32	0.15	0.80
	SE	(0.13)	(0.51)	(0.05)	(0.38)
	p-val	0.01	0.01	0.01	0.04
	q-val	0.03	0.03	0.03	0.07
Ctrl Mean		0.00	3.76	0.58	2.45
Ctrl SD		1.00	3.97	0.49	3.01
Obs		18816	18816	18816	18816
GUP-Sav - GUP-No-Sav = 0	p-val	0.18	0.25	0.12	0.22

Treatment effects of GUP on bag-making labor supply for Bags households during the Bags program, separately for GUP households with savings collection and GUP households without. The omitted group is Control-Bags households. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). We use household-week-level bag production data, and include station-week fixed effects. The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Standard errors are clustered at village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel.

Appendix Table 8: Intervention Treatment Effects at Two Years, GUP-No-Savings vs. GUP-Savings

		Livestock Value (1)	Asset Value (2)	Monthly Household Income (3)	Monthly Cons. per Capita (4)	Physical Health Index (5)	Time Prod. Labor (6)	Time Leisure (7)
GUP-No-Savings	ITT	80.32	103.30	4.76	1.02	0.05	-2.09	6.12
	SE	(22.84)	(45.34)	(3.13)	(2.08)	(0.05)	(18.34)	(11.84)
	p-val	0.00	0.02	0.13	0.62	0.25	0.91	0.61
	q-val	0.01	0.09	0.37	0.73	0.58	0.97	0.73
GUP-Savings	ITT	65.18	45.20	10.35	1.25	0.02	-0.92	-6.01
	SE	(20.86)	(42.83)	(3.53)	(1.84)	(0.05)	(26.38)	(11.74)
	p-val	0.00	0.29	0.00	0.50	0.60	0.97	0.61
	q-val	0.01	0.59	0.02	0.73	0.73	0.97	0.73
Ctrl Mean		240.17	589.48	36.59	44.15	-0.16	330.83	84.63
Ctrl SD		348.59	764.13	43.08	30.15	0.84	269.48	127.04
Obs		2909	2900	2907	2880	2767	1130	1130
GUP-Sav - GUP-No-Sav = 0	p-val	0.56	0.18	0.19	0.92	0.64	0.97	0.43

Treatment effects among all sample households (Bags and No-Bags) two years after the start of the interventions, focusing on GUP-Savings and GUP-No-Savings. The omitted group is Control households (Bags and No-Bags) in any village. The sample is restricted to villages with more than 30 compounds. We control for stratification variables, imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income), and whether or not household was treated with bags. We use household-level data from the two-year survey and include surveyor fixed effects and the baseline value of the outcome when possible. Livestock value is the total number of livestock owned times the median reported price for each animal. Asset value is the total number of assets (including livestock, household and productive assets, and stocks), valued using asset prices relative to the price of goats from other countries. Monthly household income is the sum of income from the household's business, farm, wage labor, and (revenue from) animals. (Income from wage labor does not include bags earnings.) Monthly consumption per capita includes both food and non-food expenditure. Health index includes two variables: average daily living score (mean of capacity bathing, lifting, walking, and working) and sick day (1 if the member did not miss a day of work due to illness in the last year, 0 otherwise). Time productive labor is minutes spent yesterday spent on bags, agriculture, business, home labor, and animals. Time leisure is minutes spent yesterday on religious activities, social activities, ceremonies, traveling, personal care, and resting. (These time use measures are taken at the two-year mark, not during the bags program.) Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 9: Effects of GUP, SOUP, and UCT on Income and Labor Supply Among Bags Households During Bags Program - No Controls for Imbalanced Variables

		Monthly HH Income + UCT (1)	Weekly Bags Prod. Index (2)	Weekly # of Bags (3)	Weekly Partic. (0/1) (4)	Weekly Bags Earnings (5)	Daily Time Non-Bags Labor (6)
Panel A: Effects of GUP and SOUP							
GUP	ITT	18.66	0.27	1.13	0.12	0.66	-8.76
	SE	(5.31)	(0.12)	(0.50)	(0.05)	(0.35)	(12.39)
	p-val	0.00	0.03	0.03	0.02	0.06	0.48
	q-val	0.05	0.13	0.13	0.13	0.17	0.54
SOUP	ITT	-3.93	-0.17	-0.61	-0.11	-0.29	-8.66
	SE	(2.80)	(0.13)	(0.50)	(0.06)	(0.38)	(12.75)
	p-val	0.17	0.20	0.22	0.07	0.45	0.50
	q-val	0.29	0.32	0.32	0.17	0.53	0.54
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Obs		288	18816	18816	18816	18816	3442
GUP - SOUP	ITT	22.60	0.44	1.74	0.23	0.95	-0.10
GUP - SOUP	p-val	0.00	0.01	0.01	0.00	0.04	0.99
Panel B: Effects of GUP and SOUP for Households Assigned Complex Bags							
		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
GUP-Complex	ITT	10.11	0.26	1.33	0.10	0.53	-6.72
	SE	(3.86)	(0.18)	(0.75)	(0.09)	(0.42)	(18.59)
	p-val	0.02	0.16	0.08	0.27	0.22	0.72
	q-val	0.13	0.29	0.17	0.33	0.32	0.76
SOUP-Complex	ITT	-6.66	-0.29	-1.12	-0.15	-0.64	-1.98
	SE	(4.05)	(0.14)	(0.51)	(0.08)	(0.35)	(14.67)
	p-val	0.12	0.05	0.03	0.06	0.07	0.89
	q-val	0.23	0.16	0.13	0.17	0.17	0.89
Ctrl Mean		14.44	-0.16	3.01	0.54	1.88	541.23
Obs		160	9156	9156	9156	9156	1608
GUP - SOUP	ITT	16.77	0.55	2.45	0.26	1.17	-4.74
GUP - SOUP	p-val	0.00	0.01	0.00	0.01	0.01	0.76
Panel C: Effects of UCT							
		Monthly HH Income + UCT (1)	**Weekly Bags Prod. Index** (2)	**Weekly # of Bags** (3)	**Weekly Partic. (0/1)** (4)	**Weekly Bags Earnings** (5)	**Daily Time Non-Bags Labor** (6)
High UCT	ITT	21.48	0.35	1.46	0.16	0.87	2.59
	SE	(7.39)	(0.16)	(0.67)	(0.07)	(0.49)	(12.61)
	p-val	0.01	0.03	0.03	0.03	0.08	0.84
	q-val	0.13	0.13	0.13	0.13	0.17	0.86
Low UCT	ITT	14.66	0.18	0.73	0.08	0.42	-22.41
	SE	(8.23)	(0.15)	(0.66)	(0.07)	(0.41)	(19.58)
	p-val	0.09	0.25	0.27	0.22	0.31	0.26
	q-val	0.17	0.33	0.33	0.32	0.38	0.33
Ctrl Mean		8.01	0.00	3.76	0.58	2.45	535.59
Obs		288	18816	18816	18816	18816	3442
High - Low	ITT	6.82	0.17	0.73	0.08	0.45	25.00
High - Low	p-val	0.56	0.41	0.41	0.40	0.46	0.24

Treatment effects among Bags households during the Bags program. In all panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP-Bags and SOUP-Bags. Panel B shows effects of GUP-Bags and SOUP-Bags for households assigned complex bags. Panel C shows effects of GUP-Bags separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with more than 30 compounds. We control for stratification variables. For column 1, we use household-level data from the midline survey during bag-making, and we include surveyor fixed effects as well as the baseline value of the outcome when possible. For columns 2-5, we use household-week-level bag production data, and include station-week fixed effects. The bags production index is a standardized index of the variables in columns 3-5, centered around the Control-Bags mean. For column 6, we use household-month-level data from time use surveys during bag-making, and include station-month fixed effects. Daily time non-bags labor is minutes spent yesterday on farming, business operations, animal production and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 10: Effects of GUP, SOUP, and UCT on Inputs Among Bags Households During Bags Program - No Controls for Imbalanced Variables

Panel A: Effects of GUP and SOUP

		Exp. Labor (1)	**Exp. Herbicide** (2)	**Exp. Fertilizer** (3)	**Time Bags** (4)	**Time Field** (5)	**Time Business** (6)	**Time Home** (7)	**Time All Labor** (8)	**Time Leisure** (9)
GUP	ITT	2.71	4.48	21.67	-3.14	-21.67	9.24	1.89	-13.28	12.27
	SE	(3.36)	(2.32)	(14.22)	(11.93)	(9.86)	(4.58)	(11.57)	(12.11)	(7.24)
	p-val	0.43	0.06	0.14	0.79	0.03	0.05	0.87	0.28	0.09
	q-val	0.59	0.23	0.33	0.84	0.19	0.19	0.89	0.43	0.28
	Bsl p-val	0.482	0.268	0.055	.	.	.	.	.	.
SOUP	ITT	-0.32	2.42	20.80	-20.92	-25.86	9.76	6.46	-23.25	15.10
	SE	(2.26)	(2.50)	(21.59)	(12.80)	(10.76)	(4.75)	(10.97)	(11.61)	(9.14)
	p-val	0.89	0.34	0.34	0.11	0.02	0.04	0.56	0.05	0.10
	q-val	0.89	0.50	0.50	0.28	0.19	0.19	0.67	0.19	0.28
	Bsl p-val	0.513	0.903	0.038	.	.	.	.	.	.
Ctrl Mean		4.02	3.83	57.58	112.94	206.61	28.48	299.27	603.49	127.17
Ctrl SD		15.67	10.82	88.11	106.61	185.65	78.23	155.09	168.19	140.45
Obs		272	272	272	1978	3442	3442	3442	3442	3442
GUP - SOUP	ITT	3.03	2.06	0.87	17.78	4.20	-0.52	-4.57	9.97	-2.83
GUP - SOUP	p-val	0.39	0.44	0.97	0.24	0.71	0.93	0.73	0.37	0.76

Panel B: Effects of UCT

		Exp. Labor (1)	**Exp. Herbicide** (2)	**Exp. Fertilizer** (3)	**Time Bags** (4)	**Time Field** (5)	**Time Business** (6)	**Time Home** (7)	**Time All Labor** (8)	**Time Leisure** (9)
High UCT	ITT	-2.24	0.59	27.81	-28.69	-26.43	14.72	13.09	-14.68	11.50
	SE	(3.76)	(1.56)	(14.86)	(12.66)	(12.04)	(5.60)	(11.07)	(11.46)	(9.45)
	p-val	0.56	0.71	0.07	0.03	0.03	0.01	0.24	0.20	0.23
	q-val	0.67	0.77	0.23	0.19	0.19	0.18	0.39	0.39	0.39
	Bsl p-val	0.446	0.271	0.106	.	.	.	.	.	.
Low UCT	ITT	9.50	9.88	13.20	17.18	-15.94	2.65	-11.56	-11.61	13.19
	SE	(7.30)	(3.45)	(23.62)	(14.17)	(12.56)	(5.99)	(17.30)	(18.99)	(9.44)
	p-val	0.20	0.01	0.58	0.23	0.21	0.66	0.51	0.54	0.17
	q-val	0.39	0.18	0.67	0.39	0.39	0.74	0.67	0.67	0.37
	Bsl p-val	0.791	0.606	0.111	.	.	.	.	.	.
Ctrl Mean		4.02	3.83	57.58	112.94	206.61	28.48	299.27	603.49	127.17
Ctrl SD		15.67	10.82	88.11	106.61	185.65	78.23	155.09	168.19	140.45
Obs		272	272	272	1978	3442	3442	3442	3442	3442
High - Low	ITT	-11.75	-9.29	14.61	-45.87	-10.49	12.07	24.65	-3.06	-1.68
High - Low	p-val	0.21	0.03	0.56	0.01	0.49	0.10	0.17	0.87	0.89

Treatment effects on inputs among Bags households during the Bags program. In both panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP-Bags and SOUP-Bags. Panel B shows effects of GUP-Bags separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with more than 30 compounds. We control for stratification variables. Columns 1-3 use household-level data from the midline survey during bag-making, and we include surveyor fixed effects as well as the baseline value of the outcome when possible. Expenditure on labor, herbicide, fertilizer are expenditures in the last 12 months. Columns 4-9 use household-month-level data from time use surveys (minutes spent on each activity yesterday) during bag-making, and include station-month fixed effects. Home labor includes childcare, cleaning, cooking, collecting firewood/water, and shopping. All labor includes time spent on bags and other wage labor, agriculture, businesses, animals (which is minimal, so we do not show separately), and home labor. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 11: Effects of GUP, SOUP, and UCT on Non-Bags Outputs Among Bags Households During Bags Program - No Controls for Imbalanced Variables

Panel A: Effects of GUP and SOUP							
		Monthly Harvest Value (1)	**Monthly Residual Prod.** (2)	**Monthly Business Revenue** (3)	**Monthly Business Income** (4)	**Monthly Wage Income** (5)	**Monthly Non-Bags Income** (6)
GUP	ITT	93.77	2.93	14.44	8.14	-0.86	6.39
	SE	(56.28)	(73.08)	(10.44)	(5.31)	(0.51)	(5.27)
	p-val	0.11	0.97	0.18	0.14	0.10	0.24
	q-val	0.31	1.00	0.31	0.31	0.31	0.38
	Bsl p-val	0.266	.	0.126	$0.037^{(+)}$	0.440	0.873
SOUP	ITT	0.38	-79.72	-9.59	-3.37	-0.31	-3.61
	SE	(72.45)	(58.36)	(5.64)	(2.34)	(0.80)	(2.62)
	p-val	1.00	0.18	0.10	0.16	0.70	0.18
	q-val	1.00	0.31	0.31	0.31	0.80	0.31
	Bsl p-val	0.119	.	0.444	$0.080^{(+)}$	0.009	0.437
Ctrl Mean		408.47	-13.50	13.94	6.54	1.52	8.01
Ctrl SD		461.91	349.39	55.61	24.83	4.66	25.04
Obs		272	266	287	287	288	288
GUP - SOUP	ITT	93.39	82.66	24.03	11.51	-0.55	10.00
GUP - SOUP	p-val	0.15	0.26	0.04	0.05	0.51	0.08

Panel B: Effects of UCT							
		Monthly Harvest Value (1)	**Monthly Residual Prod.** (2)	**Monthly Business Revenue** (3)	**Monthly Business Income** (4)	**Monthly Wage Income** (5)	**Monthly Non-Bags Income** (6)
High UCT	ITT	178.06	112.90	9.50	5.68	-0.59	4.72
	SE	(52.15)	(62.22)	(15.52)	(7.27)	(0.55)	(7.36)
	p-val	0.00	0.08	0.55	0.44	0.29	0.53
	q-val	0.05	0.31	0.65	0.59	0.41	0.65
	Bsl p-val	0.134	.	0.728	$0.029^{(+)}$	0.708	0.776
Low UCT	ITT	-22.99	-149.83	21.57	11.70	-1.22	8.75
	SE	(66.90)	(79.34)	(13.44)	(7.95)	(0.65)	(8.11)
	p-val	0.73	0.07	0.12	0.15	0.07	0.29
	q-val	0.80	0.31	0.31	0.31	0.31	0.41
	Bsl p-val	0.819	.	$0.094^{(+)}$	0.344	0.003	0.977
Ctrl Mean		408.47	-13.50	13.94	6.54	1.52	8.01
Ctrl SD		461.91	349.39	55.61	24.83	4.66	25.04
Obs		272	266	287	287	288	288
High - Low	ITT	201.06	262.72	-12.07	-6.02	0.63	-4.03
High - Low	p-val	0.01	0.00	0.57	0.59	0.34	0.73

Treatment effects on non-bags outputs among Bags households during the Bags program. In both panels, the omitted group is Control-Bags households in any village. Panel A shows average effects of GUP and SOUP. Panel B shows effects of GUP separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with more than 30 compounds. We control for stratification variables. We use household-level data from the midline survey during bag-making, and we include surveyor fixed effects as well as the baseline value of the outcome when possible. Harvest value is the total revenue from harvest, six months after the bag-making period. Residual productivity contains the residuals from a regression of harvest value on input expenditure, acreage, and average time on field. Business revenue and income pertain to all non-agricultural businesses. Wage income does not include earnings from bags. Non-Bags income includes income from agriculture, business, wages, and animals. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 12: Justifying Imputation of Time Spent on Bags

	Monthly Wage Income (USD) (1)	**Time Bags and/or Wage Labor** (2)
GUP-No-Bags	-0.78**	-10.39**
	(0.31)	(4.89)
SOUP-No-Bags	-0.36	-6.26
	(0.65)	(4.61)
Control-Bags	0.21	105.49***
	(0.39)	(6.11)
Any GUP-Bags	-0.92***	98.71***
	(0.27)	(7.86)
SOUP-Bags	-0.25	75.71***
	(0.58)	(7.31)
Observations	864	789
Ctrl Mean	1.130	9.730
Any-GUP-Bags = GUP-No-Bags	0.590	0
SOUP-Bags = SOUP-No-Bags	0.890	0

Treatment effects among all sample households (Bags and No-Bags) on wage income and wage labor during bag-making. The omitted group is Control No-Bags households in any village. The sample is restricted to villages with more than 30 compounds. We control for stratification variables, imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income), and the baseline value of the outcome when possible. For Column 1, we use household-level data from the midline survey during bag-making, and we include surveyor fixed effects. For Column 2, we use household-level averages over the five monthly time use surveys administered during the bags program, and we include surveyor fixed effects. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel. All monetary values are reported in 2014 USD, Purchasing Power Parity (PPP) terms.

Appendix Table 13: Attrition

		Midline During Bags (1)	**Two-Year Survey** (2)	**Time-Use Midline** (3)
Control-Bags	ITT	-0.01	-0.00	0.11
	SE	(0.01)	(0.01)	(0.03)
	p-val	0.31	0.81	0.00
GUP-No-Bags	ITT	0.00	0.01	0.01
	SE	(0.00)	(0.00)	(0.05)
	p-val	0.93	0.10	0.87
Any GUP-Bags	ITT	0.00	0.00	0.14
	SE	(0.00)	(0.01)	(0.03)
	p-val	0.45	0.45	0.00
SOUP-No-Bags	ITT	0.01	0.00	0.04
	SE	(0.01)	(0.01)	(0.05)
	p-val	0.67	0.81	0.46
SOUP-Bags	ITT	0.01	-0.00	0.10
	SE	(0.01)	(0.01)	(0.03)
	p-val	0.37	0.52	0.00
Ctrl Mean		0.99	0.95	0.67
Ctrl SD		0.10	0.21	0.28
Obs		870	2915	1238
GUP-Bags - Ctrl-Bags	p-val	0.20	0.47	0.36
GUP-Bags - SOUP-Bags	p-val	0.45	0.27	0.17

Survey participation on treatments. The omitted group is Control No-Bags households in any village. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income) and include surveyor fixed effects. The outcome in column 1 is a binary indicator for whether the household was surveyed in the midline that took place during bags, conditional on being one of the 1085 households selected to participate. The outcome in column 2 is a binary indicator for whether the household was surveyed at the two-year mark. The outcome in column 3 is the household's average participation rate in the six time use surveys that took place during the bags program, conditional on being one of the 1542 households selected to participate. Standard errors are clustered at the village level.

Appendix Table 14: Effects of GUP, SOUP, and UCT on Fraction of Bags made by Women in Household

Panel A: Impacts of GUP and SOUP

		Made by Women ($\geq$ 15) (1)
GUP	ITT	-0.01
	SE	(0.02)
	p-val	0.58
	q-val	0.77
SOUP	ITT	-0.02
	SE	(0.01)
	p-val	0.11
	q-val	0.45
Ctrl Mean		0.97
Ctrl SD		0.18
Obs		10582
GUP - SOUP	ITT	0.01
GUP - SOUP	p-val	0.48

Panel B: Effects of UCT

		Made by Women ($\geq$ 15) (1)
GUP High UCT	ITT	-0.02
	SE	(0.02)
	p-val	0.27
	q-val	0.55
GUP Low UCT	ITT	0.00
	SE	(0.03)
	p-val	0.85
	q-val	0.85
Ctrl Mean		0.97
Ctrl SD		0.18
Obs		10582
high UCT - low UCT	ITT	-0.03
high UCT - low UCT	p-val	0.35

Treatment effects on the fraction of bags made by women in bag-making households. In both panels, the omitted group is Control-Bags households. Panel A shows average effects of GUP and SOUP. Panel B shows effects of GUP separately for households receiving high ($3.9 weekly) and low ($1.3 weekly) UCT. The sample is restricted to villages with more than 30 compounds. We control for stratification variables and imbalanced variables (average household age, food security index, land area, monthly per capita consumption, and monthly household income). We use household-week-level bag production data, and include station-week fixed effects. The bags production index is a standardized index of bags, participation (0/1), and bags earnings, centered around the Control-Bags mean. Standard errors are clustered at the village level. We use the Benjamini-Hochberg step-up method to compute q-values, considering all tests in each panel.

Appendix Table 15: Wage Elasticity Results

Panel A: Evidence of Responsiveness to Wages Received for Previously Submitted Bags

	IHS(bags) (1)	**IHS(bags)** (2)	**IHS(bags)** (3)	**IHS(bags)** (4)	**IHS(bags)** (5)	**IHS(bags)** (6)	**IHS(bags)** (7)
log(wage)	0.02	-0.00	0.10	-0.02	0.15**	-0.11***	0.19***
	(0.04)	(0.05)	(0.07)	(0.04)	(0.07)	(0.04)	(0.07)
Observations	23,058	14,822	8,236	16,470	6,588	13,146	9,912
consecutive		no	yes				
experience						no	yes
fourth week				no	yes		

Panel B: Elasticity Estimates with respect to the 3-Week Lagged Wage

	IHS(bags) (1)
log(wage(t-3))	0.16***
	(0.05)
Observations	19,764

Effects of the log wage, which was randomly varied over time within villages, on the inverse hyperbolic sine of weekly bag production. In Panel A we use the effective, current wage. "Consecutive" indicates participants who were randomly assigned two consecutive high wage months and two consecutive low wage months. "Fourth week" indicates participant-weeks in the fourth week in the wage month. "Experience" indicates either the fourth week of the month, or for "consecutive" participants, any week in the second consecutive month with the same wage. In Panel B we use the 3-week lagged wage. We include individual and week fixed effects and cluster standard errors at the village level.

Appendix Table 16: Variable Definitions and Construction

Variable	Definition
asset value (USD)	The total number of assets (including livestock, household and productive assets, and stocks), valued using asset prices relative to the price of goats from other countries.
business income (and revenue), monthly (USD)	Monthly business revenues minus expenses. For each business within the household, we ask about how many months in the last year the business was operating, how many months were "normal" (neither higher nor lower than last month), how many months were "high profit," and how many months were "low profit." We ask about sales and profits in the last month, as well as profits in normal, high, and low months. We use the ratio of last month's profits to high and low profits to impute sales in high and low profit months, compute total sales in the last year by summing over sales in normal, high, and low months, and divide by the number of months the business was operating to get monthly revenue. We do the same exercise for expenses.
consumption, monthly per capita (USD)	Total of all food consumption, temptation goods and fuel expenditures, school, clothing, festival expenditures, transportation, rent, medical expenditures, and home improvement spending. Food consumption is (quantity*market price) for each food item. Scaled to per capita monthly values.
expenditure on inputs	We ask about expenditures on manure, fertilizer, labor, herbicide, insecticide, and other inputs in the last year, and then sum over all categories.
financial inclusion index	Standardized index of two variables, centered around the baseline mean. The first variable is the total amount received in loans by the household in the last year. The second variable is the total savings balances at the time of the survey.
food security index	Standardized index of three variables, centered around baseline means. The first two variables equal 0 if the household answered "all year" or "during the lean season only" to the following questions, about adults and kids, respectively: "Did adults/kids ever reduce number of meals per day or reduce portions over the past year?" The third variable equals 0 if the household answered "all year" or "during the lean season only" to the question "Did adults ever skip entire days without eating?"
female empowerment index	Standardized index of five variables, centered around baseline means. Each variable is the answer to the question, "To what extent do you believe yourself able to make your own decisions concerning X?" The categories X are food, school expenses, health expenses, visiting friends, and purchases. They are measured on a scale from 1 to 3.
harvest value	We ask about the quantity of each crop sold in the last year. If the units of harvested crops are the same as the units of sold crops and we have the sale price, then we use this price to compute the sale value of each crop, and then sum over crops. Otherwise, we use the median price for that crop.
income, monthly HH (USD)	The sum of monthly business income, monthly crop income, monthly wage income, and monthly animal revenue. Monthly crop income is harvest value minus expenditure on inputs (annual), divided by twelve.
livestock value (USD)	Sum of values of goats, fowl, pigs, sheep, and cows.
mental health index	Standardized index of three variables, centered around baseline means. The first is economic satisfaction, measured on a scale from 1 to 5. The second is a standardized index of five measures: feeling sad, crying, not eating, not working, and feeling restless, measured on a scale from not at all, hardly ever, some of the time, or most of the time. The third is whether the individual was not worried in the last year (0 if the member experienced a period of worry in the last year, 1 otherwise).
physical health index	Standardized index of two variables, centered around baseline means. The first is the average daily living score, which is the mean of four variables: capacity bathing, capacity lifting, capacity walking, and capacity working (each measured on a scale from 1 being easily done to 4 being unable to do). The second is whether or not they did not take a sick day (1 if the member did not miss a day of work due to illness in the last year, 0 otherwise).
political involvement index	z-score of attendance at village meetings in the last year (1 if the person attended a village meeting in the last 12 months, 0 otherwise), centered around baseline mean.