SAVING FOR MULTIPLE FINANCIAL NEEDS: EVIDENCE FROM LOCKBOXES AND MOBILE MONEY IN MALAWI

Shilpa Aggarwal, Valentina Brailovskaya, and Jonathan Robinson*

Abstract—We test whether the provision of multiple labeled savings accounts affects savings and downstream outcomes in an experiment with 761 microentrepreneurs in urban Malawi. Treatment respondents received one or multiple savings accounts, in the form of lockboxes or mobile money. We find that while providing additional boxes increased savings by 40%, technical issues marred the efficacy of a second mobile money account. Data from novel high-frequency surveys suggest that both types of accounts had impacts on downstream outcomes, including farming decisions and credit extended to customers. We do not detect differential downstream effects by the number or modality of accounts.

I. Introduction

MOST people have multiple concurrent financial goals. For example, it is common for households to be saving up for large indivisible investments such as buying a house or paying for higher education, while also setting aside smaller amounts for day-to-day expenses or for dealing with unforeseen emergencies. How do people save towards multiple goals simultaneously? One potential strategy is to create separate labeled accounts. Research in behavioral economics suggests that once these accounts are created, withdrawals for any purpose other than the labeled one impose a utility cost on the account holder (Ainslie, 2001; Bénabou & Tirole, 2004; Koch & Nafziger, 2016; Thaler, 1999). Previous studies have shown that accounts labeled for a specific purpose can be effective in increasing savings for that purpose (i.e., Brune et al., 2016; Dupas & Robinson, 2013b; Karlan & Linden, 2014), and this finding likely generalizes to having more than one labeled account.

As a practical matter, however, it is not clear how one can accomplish the cognitively challenging task of keeping track of distinct sums of money that have been mentally allocated towards different purposes.[1] We conjecture that the effectiveness of mental accounts will likely be enhanced when accounts are accompanied by the physical separation of money. The practice of physically separating pots of money meant for distinct uses has precedent,[2] although it is not known if this method actually leads to an increase in deposits. A related, policy-relevant question for developing countries with low levels of formal account prevalence, but where mobile money has emerged as a viable alternative to the banking sector, is whether multiple accounts provided via mobile money can help people save towards multiple purposes.

In this paper, we report results from an experiment with 761 micro-entrepreneurs in the city of Blantyre, Malawi, who were randomized into one of several different treatments designed to measure the savings efficacy of single versus multiple labeled accounts, which we provided either via mobile money or through lockboxes. The average respondent had 2.4 savings goals at baseline, so providing multiple labeled accounts could be beneficial in this population. One treatment group was offered simple metal lockboxes in which to save up, while a second group was offered mobile money accounts. Both of the treatment groups were further randomized into receiving either one or more of the savings device in question (the multiple box group were offered up to 3 boxes, while the multiple mobile money group were offered 2 mobile money accounts). A control group was not offered any accounts.

A primary contribution of the paper is to document the effect of the accounts on a set of downstream outcomes. In addition to savings, our analysis focuses on labor supply (in the main business and in the alternative of farming), business success (profits and revenues), and credit (these outcomes are highlighted in a preanalysis plan, Aggarwal et al., 2019). To measure these and other secondary outcomes at a high frequency, all respondents (including the control group) were given cell phones, and half of the sample was called once or twice per week to get survey measures of savings decisions, labor supply, income, expenditures, and transfers. We also conducted two rounds of monitoring surveys with all respondents.

During the study period (of about 9 months), 73% used their mobile money account at least once, 60% used it 5 times, and average deposits were about $9 per month. Usage of the

Received for publication April 13, 2020. Revision accepted for publication May 25, 2021. Editor: Amit K. Khandelwal.

*Aggarwal: Indian School of Business; Brailovskaya: IDinsight; Robinson: University of California, Santa Cruz and NBER.

We are extremely grateful to Calvin Mhango and Gabriella Fleischman for overseeing the project and to Marble Karuu and David Park for research assistance. We thank Carly Farver for her continuous support, and IPA Malawi for implementing the research protocol in Malawi. We thank Airtel Malawi for their collaboration, especially Tione Kafumba, Bigboy Makaloma and Chris Sukasuka. We thank Naresh Kumar and Raychel Schwartz for supporting data collection for the long-term followup. We thank the Financial Inclusion Program at IPA, ISB, and UCSC for financial support. We thank Amit Khandelwal, Jessie Li, two anonymous referees, and seminar participants at IPA, Williams College, the University of Cape Town, the Paris School of Economics, UCSC, and IDinsight for helpful comments in shaping this paper. This research protocol was approved by the National Commission for Science and Technology in Malawi and the IRBs of UCSC and the ISB. This trial is registered in the AEA registry for randomized controlled trials (AEARCTR-0002449). A preanalysis plan is available on the AEA registry website. All errors are our own.

A supplemental appendix is available online at https://doi.org/10.1162/rest_a_01086.

[1] Research on the cognitive costs of scarcity (i.e., Mani et al., 2013) suggests that this task might be even harder for the poor.

[2] For example, see this oft-quoted anecdote cited in previous work (i.e., Zelizer, 1994; Soman & Cheema, 2011), from Bradley (1923): *"Take for instance Mrs. M's system as she told it to* Women's Home Companion *in the early 1920's: "I collected eight little cans, all the same size, and pasted on them the following words, in big letters: groceries, carfare, gas, laundry, rent, tithe, savings, miscellaneous. . . . [W]e speak of those cans now, as the grocery can, carfare can, etc."* In the present day, a number of personal finance advisors recommend maintaining different accounts for money meant for different purposes—popularly known as the "envelope method."

The Review of Economics and Statistics, July 2023, 105(4): 833–851

https://doi.org/10.1162/rest_a_01086

lockbox was even higher: at least 95% used their box at least once, 91% used it at least 5 times, and total deposits averaged $27 per month.[3] These are substantial sums for this context, in which daily business profits are roughly $3.50. We find strong evidence that providing a second account increased savings for lockboxes, by about 40%. This evidence is consistent with the hypothesis that having a second, labeled account is helpful to achieve savings goals. We find no effect of a second account on savings for the mobile money group, but this is likely due to design problems with the technology. In particular, the telecom company ("telco") can offer only one account per SIM card, and so users had to switch between SIM cards to use multiple accounts, and this proved too burdensome despite having access to a dual SIM phone. This suggests that sophisticated financial products will work only when using them is simple enough.

Second, we find strong evidence that the saving accounts (both boxes and mobile money) had effects on prespecified downstream outcomes. In particular, we find strong evidence that they increased the time spent by respondents working on their farms. Our data suggest that this increase in farm hours came from reductions in labor supply in the main business—we find negative point estimates for labor in the own business, though estimates are not always statistically significant. This finding is related to several recent papers that find labor supply effects of an easing of credit constraints, such as Fink et al. (2020), who find that providing credit to smallholder farmers decreases off-farm labor and increases own-farm labor. Our finding that saving accounts cause a reallocation of labor supply is related to Callen et al. (2019), who find that Sri Lankan households who were given access to deposit collection reduce leisure time and increase their hours worked, a finding that is likely explained by the fact that the accounts were interest-bearing. In our case, further analysis points to greater investments in the farm, which suggests that the reallocation of labor supply between occupations may have been driven by higher marginal returns to farming. Specifically, we find that treatment respondents are more likely to buy and rent land, as well as invest more in farm inputs (though these were not specified as primary outcomes). These investments, particularly in land, are lumpy by nature, and were likely facilitated by savings. We also find an increase in credit extended to customers, and some evidence of an increase in total expenditure.

We make four main contributions to the literature. First, our experimental design isolates the effect of an additional labeled savings place.[4] While there have been many recent papers which have studied the effect of providing un(der)banked households with savings devices, there are a variety of pathways by which accounts may increase savings, including that savings accounts provide security, limit liquidity, increase the salience of saving money or because savings accounts encourage the activation of mental accounts for particular goals. Isolating a single channel in this context is challenging. In this experiment, we offer an incremental savings option which is identical to the first, and thus the design isolates the effect of an additional, physically separate savings location, since other pathways such as security or salience are equalized.[5]

Second, we contribute to a nascent literature on the impacts of mobile money accounts. Given the extensive uptake and usage of mobile money in several sub-Saharan African countries, there has been a large amount of research and policy interest in the topic. However, the popularity of mobile money makes identification challenging because of the difficulty of maintaining a control group.[6] Consequently, the seminal studies in this research area are identified from plausibly exogenous regional differences in mobile money roll-out (Jack & Suri, 2014; Suri & Jack, 2016). More recently, several researchers have implemented RCTs, usually by partnering directly with the telcos. These include Batista and Vicente (2020b), who randomize mobile money access at the community level; Lee et al. (2021), who offer mobile banking to rural-urban migrants in Bangladesh (to both the urban migrant and to the sending rural household); Wieser et al. (2019), who randomize the roll-out of mobile money agents in rural Northern Uganda, and our own companion paper (Aggarwal et al. 2020), which studies the impact of mobile money (pooled across both subtreatments).[7] Our study adds to this literature by randomizing basic access at the individual level in a country where mobile money infrastructure already existed but where takeup was still modest, and by measuring a number of outcomes via our survey modules. Moreover, while the vast majority of the existing literature focuses on the impact of mobile money on resilience (via a reduction in the transaction costs of transfers), the effects in our study are driven primarily by savings. In fact, we observe little effect on interpersonal transfers from the mobile money

[3]The figures for the box were measured earlier, about 5 months after account opening.

[4]There is a small literature on a related issue about partitioning consumption items between physically separated places. Soman and Cheema (2008) conduct experiments in which experimental subjects were paid in different numbers of accounts (i.e., money split into multiple envelopes or chocolates split into separate packages) and find that immediate consumption is decreasing in the number of accounts.

[5]Several other studies offer different accounts to different treatment groups (i.e., Brune et al., 2016; Dupas & Robinson, 2013b), or (more rarely) offering a basic account and an incremental account with different features, such as a commitment account (John, 2020). Other studies offer accounts to respondents who already have basic accounts at a bank (such as Ashraf et al., 2006), which arguably isolates the incremental effect of the new account, even though the account itself is not experimental.

[6]For example, see Dalton et al. (2020) who randomized merchants in Kenya into having a smartphone app that allowed them to accept electronic payments. By the endline—about 2 years later—a quarter of the control firms had adopted the technology on their own.

[7]There have been a number of interventions which layer other financial interventions on top of basic mobile money, such as labeled accounts (Aker et al., 2020, Dizon et al., 2020, Lipscomb & Schechter, 2018), automatic deposits (Blumenstock et al., 2018) or interest-bearing accounts (Batista & Vicente, 2020a). de Mel et al. (2022) layer mobile money on top of basic bank accounts to enhance the ease of making deposits, but their intervention is hampered by very low take-up. See Suri (2017) and a research brief by the Gates Foundation (BMGF, 2020) for an overview.

treatment and only two-thirds of the mobile money treatment respondents ever made an interpersonal transfer, while more than 80% reported using their mobile money accounts for long-term savings. This result is likely attributable to the fact that the treatment was at the individual level and so did not change mobile money access for the risk-sharing networks of treated respondents; we believe that interpersonal transfer effects would likely be present in a larger expansion of the network. Nevertheless, by muting the channel of interpersonal transfers, our study shows that mobile money can be effective in mobilizing savings. These effects will likely only be larger as telcos develop more sophisticated financial products.

Third, our experiment was designed to carefully measure effects on downstream outcomes. While there have been a number of recent studies of savings (including several that look at boxes or mobile money specifically),[8] only a few of them find effects on downstream outcomes such as business investment and personal expenditures (Dupas & Robinson, 2013a), health investment (Dupas & Robinson, 2013b), educational expenditures (Prina, 2015; Jack & Habyarimana, 2018), labor supply (Callen et al., 2019), self-reported financial well-being (Kast & Pomeranz, 2018; Prina, 2015), risk coping (Jones & Gong, 2020), and debt (Kast & Pomeranz, 2018; Aker et al., 2020). However, most of these studies find outcomes on only one of several potential outcomes, while other studies find no effects at all.[9] An obvious reason why effects may be hard to detect is that statistical power is hampered by low take-up.[10] For example, a tabulation in Dupas et al. (2018) finds that in many studies only 20–30% of people ever use accounts, and much lower percentages (rarely larger than 20%) "actively" use accounts (usually defined as making more than a few deposits). But in this study, usage is dramatically higher: within the study period, at least 94% of people used the box at least once and at least 91% used it at least 5 times; for mobile money accounts, the figures are 73% and 59%.

Finally, we provide a direct comparison between boxes and mobile money accounts as savings devices, and their subsequent impact on downstream outcomes. The design of these two savings devices could potentially have important ramifications for their efficacy as savings devices, for example, boxes are physically proximate, which would make it easier to make deposits, but would similarly encourage withdrawals; the hassle costs of visiting an agent could deter mobile money deposits, but also discourage withdrawals. In line with this, we find that boxes are clearly used more—the value of deposits was about 3.5–5 times higher in boxes, while withdrawals were also higher. We detect no significant effects on downstream effects by treatment modality. This suggests that both options can be effective in this setting, but the current project was not set up to comment on their relative efficacy in impacting savings or downstream outcomes.

The rest of this paper proceeds as follows. Section II describes the experiment and the data. Section III presents results and section IV discusses threats to validity. Section V concludes.

II. Experimental Design and Data

A. *Context and Sampling*

The experiment took place with a representative sample of small entrepreneurs operating in Blantyre, the second largest city in Malawi. While Blantyre is an urban center with a population just over 1 million, the outskirts of the city contain farmland. Blantyre contains 26 wards and 392 enumeration areas (EAs). To construct a sample with coverage across the city, we randomly selected 77 EAs (3 each in 25 wards, and 2 from a 26th ward that had only 2 EAs), and aimed to census all businesses in these wards.

Market structure is heterogeneous across EAs—the number of businesses ranged from 0 to 1,649 (mean 104, median 48).[11] Because of the high number of businesses in some EAs, it was not logistically possible to census every business in those EAs. We therefore decided to divide EAs between those with more than 100 and those with less than 100 businesses. In the smaller EAs, we censused all businesses; in the larger EAs, we counted all businesses but only censused a randomly selected subset of approximately 40% of businesses.[12] We counted a total of 9,848 businesses and classified 8,078 (82.1%) of these as small businesses.[13] We attempted to conduct a census survey with 3,857 businesses and completed surveys with 2,842 (74%).[14]

After the census, we imposed additional exclusion criteria. First, we excluded any business with more than 2 employees (6% of the census list). Second, we excluded businesses in which the business owner was a mobile money agent (3%) to prevent confounding the mobile money treatment. Third, we excluded businesses in which the owner was not actively involved in running operations (defined as working there at least 5 days per week) since such owners would not be able to

[8] A partial listing of papers that include locked savings boxes include Ashraf et al. (2006); Karlan and Linden (2014); Dupas et al. (2019); Francis (2018); Karlan and Zinman (2018), and Aker et al. (2020), among others. Those looking at mobile money accounts as a savings device include Blumenstock et al. (2018); Jack and Habyarimana (2018), Dizon et al. (2020), and Lipscomb and Schechter (2018).

[9] See table 3 in Prina (2015) and figure 5 in Dupas et al. (2018) for a summary of the effects found in these studies.

[10] This problem is often exacerbated by the fact that different people choose to spend money on different things, leading treatment effects to become diffused.

[11] Two EAs contained no small businesses. One was an industrial area and the other was farmland.

[12] Since we counted all businesses, we have sampling weights for all EAs.

[13] We excluded several classes of businesses in this exercise since they were unlikely to qualify as a small business. This included gas stations, clinics, hospitals, banks, microfinance institutions, manufacturing plants, warehouses, wholesalers, and supermarkets.

[14] Of the 1,012 (26%) businesses that were not censused, 552 (14%) refused to participate (either before or after we were able to explain the study), 346 (9%) were permanently closed, 114 (3%) were not reached (either because the shop was closed after three visits or the owner was under 18 years old).

reliably answer business-related questions (9%). Fourth, we excluded businesses that were planning to shut down within 6 months (before the project was slated to end—16%).[15]

Once we had a sample of businesses that met our criteria, we imposed two other exclusion criteria, using data that had been collected either at the census or prior to the baseline survey. First, we removed all polygamous households, which amounted to 5% of the sample. Second, since we initially planned to collect surveys with paper-and-pencil logbooks (we eventually changed to phone surveys), we excluded business owners who were illiterate (about 20% of the sample) and those whose eyesight prevented them from reading a printed page (about 10% of the sample).

These exclusion criteria left us with approximately 1,640 eligible businesses from which we drew our final sample, stratified by financial access (defined by having either a mobile money or bank account) and self-reported distance to the nearest mobile money agent (defined as above or below the sample median). In drawing the sample, we chose to oversample businesses connected to the electricity grid: while 26% of eligible businesses were connected to the grid, we sampled 35%.[16] We replaced respondents who could not be found (about 6.5%) or refused to participate (another 6.5%) with randomly chosen backups, ultimately yielding a sample of 801 businesses, which we randomized into the various treatment arms.[17]

B. Experimental Design

The experimental design is summarized in web appendix figure A2 and the timeline of project activities is shown in web appendix figure A3. As discussed below, the experimental design cross-cut the provision of savings accounts with the frequency of surveying.

Savings account treatment. We offered two types of savings products: metal lockboxes and mobile money accounts. The lockboxes were similar to those offered in prior studies—they had a deposit slit in the top, and a latch that could be locked. The boxes were produced by a local artisan and cost about $3.40. Respondents were also given a lock and key, worth about $1. While a sizeable minority of people (22%) had lockboxes at baseline, these were of lower quality than the project boxes as they were typically made of wood or cardboard, and either could not be locked or had to be broken to be opened. Finally, as in Dupas and Robinson (2013b), respondents were given a passbook to record withdrawals and deposits, so that they could track the balance without having to open their box(es). Respondents were also encouraged to use the project account(s) to save towards their savings goal.

Those in the mobile money treatment received mobile money accounts with Airtel Malawi, the leading telecom company in Malawi with just over 50% market share. The accounts were identical to those already commercially available, with several important differences. First, we reimbursed withdrawal fees for the duration of the project.[18] Second, in pilot work, we found that knowledge of mobile money was limited. Some respondents were not fully aware of fees or lacked basic knowledge about how to access and use mobile money accounts. To address this, we developed training modules on how to use the accounts, which were administered at the time of account-opening.[19] Third, as with the boxes, we encouraged people to use the accounts to save towards their goals. While we view each of these elements as relatively light-touch actions that could easily be independently implemented by the telco, the combined mobile money intervention does differ from status-quo commercial mobile money accounts available at the time in Malawi.

One of the key features of the project is to measure the effects of having an incremental account. Therefore all respondents in the savings device treatments were further randomized into receiving one or multiple of the savings device in question. For the box respondents, those in the single account treatment group were offered only one box, while those in the multiple account group were offered 3 lockboxes. However, people were allowed to take less than 3 boxes if they wanted, and in fact some people did: 24% took only 1 box, 33% took 2 boxes, and 42% took 3 boxes (1% did not take any boxes).[20] To differentiate the boxes, they were painted different colors—everyone received a silver box, while the second and third boxes were painted black and brown respectively. For those in the multiple box groups, the project passbooks allowed for separate tracking of deposits and withdrawals for each box.

In the mobile money treatment, respondents in the single mobile money group were offered only 1 account, while those in the multiple account group were given the choice to get up to 2 mobile money accounts. Each account had a separate SIM card and associated phone number due to the fact that the existing mobile money product offered exactly one account within a single phone number. To encourage people to use these with minimal hassle, the respondents were provided a dual-sim phone, discussed below. As with the boxes, the mobile money accounts were labeled as silver or black on the mobile money interface of respondents' phones, but during

[15] This high turnover rate is indicative of the level of churn in these types of businesses. Some businesses are seasonal and business closure is common.

[16] This decision was made to improve the power of the related paper, Brailovskaya (2018), which utilizes some of the data in this experiment to calculate the effect of power outages on business outcomes.

[17] See web appendix figure A1 for the geographic distribution and spread of the various treatment arms across the city of Blantyre.

[18] Fees were reimbursed weekly by making transfers equivalent to the fee amount to each account through a batch process. Therefore respondents had to incur fees and then wait to be reimbursed, and so our treatment is not identical to a policy in which withdrawals were actually free. The average fee for transactions observed in our sample would have been about 5%.

[19] This module, along with surveys, can be found on the authors' websites.

[20] It it surprising that some people chose to take less than 3 (since they were free). We can only conjecture that people might have felt guilty about taking boxes that they did not intend to use.

surveying we referred to those accounts by the last four digits of the account phone number.

During baseline, we asked all respondents about their savings goals. At the time when accounts were opened (which was about 2 months after baseline), those in the treatment groups were encouraged to use the project account(s) to save towards their savings goal. Specifically, we asked respondents to write their savings purpose on a piece of paper, which was then attached to the lid on the inside of their project boxes for the box treatment, while those in the mobile money account treatment, were asked to keep it with them. Summary statistics of savings purposes are reported in web appendix table A1. Column 1 shows purposes elicited at baseline for the entire sample, while the next four columns show purposes elicited at account opening. Goals are heterogeneous: at baseline, sizeable fractions listed business reinvestment, general consumption, and emergencies, while others listed a variety of other goals including land, durable goods, and education. Given the heterogeneity in goals, it is likely that accounts would have impacts on a variety of different outcomes.

Phone surveys. A primary goal of this project was to measure the effect of savings devices at high frequency. To do this, we opted to measure outcomes using high-frequency phone surveys (described in more detail in section IIC). To implement this effectively, we gave everyone in our sample a basic feature phone, but with dual SIM capability (worth about $12). Even though 95% of the sample already had a phone at baseline, we decided to give out new phones because we were concerned that the phones people already owned were of low quality and might break or need to be replaced prior to the end of the experiment. Moreover, in order to successfully implement the multiple mobile money accounts treatment, we needed people to have phones to have two slots for SIM cards, which are not commonly used.

Since it is possible that the high-frequency survey itself is a treatment (e.g., by focusing attention on measured outcomes), we randomly split the sample into two groups: only one was administered the high-frequency phone survey (which we call the "HFPS"). To measure the effect of surveying itself, we administered two monitoring surveys to the entire sample, so that we could compare responses between those given high frequency surveys and those who were not. In section IV, we utilize the HFPS randomization to examine the effect of regular surveying on outcomes measured in the monitoring surveys but ultimately find no discernible impact.

C. Data

We utilize information from administrative data as well as a number of different types of surveys to conduct our analysis. At the start of the project, we conducted a baseline survey, which contained questions on household and demographic characteristics, business outcomes, savings, cash flows, and related measures. We also administered a short "intake survey" to all respondents, about 2–3 months after the baseline survey. For those in the treatment groups, survey administration coincided with the handing out of the boxes/opening of the mobile money accounts. This survey included versions of many of the questions that would later constitute our main surveys, recalled over the 3 previous days. These pretreatment values are used as control variables, making all regressions ANCOVA (McKenzie, 2012).

We have two main sources of data to measure treatment effects: high-frequency phone surveys (HFPS) and two monitoring surveys (all surveys are on the authors' webpages). The HFPS measured business outcomes and labor supply at the daily level, and household expenditures, transfers, savings, credit, shocks and related outcomes at the weekly level. The HFPS was conducted in two waves, one in September–October 2017 (covering 8 weeks) and another in February and March 2018 (covering 6 weeks). In Wave 1, respondents were called twice per week, that is, every 3–4 days. In one of the weekly interviews, the respondent was administered a "short" survey which took about 15 minutes and which asked about business outcomes over the past 3 days (day by day). The other "long" survey took about 40 minutes and included all the questions in the short call, but also added a recall module for other outcomes that were expected to be rarer or more memorable and thus could be reliably remembered over a week. These included shocks such as household illness and funerals, expenditures, deposits and withdrawals from various saving sources,[21] and transfers given and received. For example, for deposits, a short survey which was administered on say, a Wednesday, would ask the respondent the amount of deposits they made in each savings place on each of the preceding 3 days, that is, on Sunday, Monday, and Tuesday; while a long survey done on that day would ask the same set of questions for each of the preceding 4 days.[22] Long-survey-only questions, such as expenditure on food, were asked for each of the preceding 7 days. Due to budgetary constraints, respondents were called only once per week in wave 2 and were administered only the "long" version of the survey.

Respondents were randomized into which days they were to be called, and this day remained unchanged throughout the project.[23] Respondents were able to pick the time of day when they wished to be called. To encourage compliance, respondents were given U.S. $2 in airtime for each week of the survey and were also enrolled into a lottery in which there was a 1/4 chance of winning an additional U.S. $3. Airtime was sent directly to the respondents' phones after each phone

[21] We collapse information on savings into 7 categories: (1) formal accounts in banks, MFIs, or SACCOs, (2) savings groups (ROSCAs or VSLAs), (3) nonproject mobile money, (4) nonproject boxes, (5) project mobile money, (6) project boxes, and (7) cash at home. For cash at home, we asked about money in a "secret place at home" since this was the most effective way of asking about this in pretesting. Within each type 1–4 and 7, we only asked about total savings and did not ask about possible multiple accounts.

[22] The specific question for savings was "How much did put in a {source} for saving purposes on {day}?"

[23] Respondents were allowed to change the day of the survey at the outset of the project but only 1 of 401 respondents ultimately changed the day.

call was completed.[24] We control for lottery payments in all HFPS regressions. If a respondent did not answer the phone, she was called at least twice more that day (3 times total). If the respondent was still unreachable, a make-up call was scheduled for the following day. During this call, we asked for information for the preceding 4 days in order to reconstruct the lost day.

The other surveys used to calculate treatment effects are called the "monitoring surveys." These were conducted twice, after each round of the HFPS, in January 2018 and March 2018. These were also done over the phone, but included the entire sample. These surveys took approximately 75 minutes to complete, and included questions similar or identical to the HFPS, though over a longer recall period, for example, outcomes like labor supply were measured at the daily level over a week-long look back period, while deposits and withdrawals were measured at a monthly level for a 2–3 months-long lookback period. Rarer events, such as a loan, were measured cumulatively for the entire period between survey rounds. The surveys also included several other questions, such as those about self-reported financial security, land purchases, and tuition payments. The second monitoring survey (the endline survey) included additional debriefing questions about people's experience with the accounts, pressures to share money and experiences with the surveys themselves. Respondents were compensated $2 USD via airtime per survey round.

As an independent verification of lockbox usage, we conducted an unannounced in-person "lockbox check" at the respondents' businesses with a random subsample in the box groups in December 2017. During this visit, we first asked a number of questions about usage of the project boxes, including cumulative deposits and withdrawals since the start of the project as well as the current balance. After the survey, enumerators requested to visually inspect the box (as in Dupas & Robinson, 2013b) in order to verify the balance. Seventy-nine percent of respondents who we interviewed were willing and able to open at least one of the boxes.[25]

Finally, we implemented a long-term follow-up survey in September 2019, about 18 months after the conclusion of the study, to track long-term usage of the accounts. We contacted a randomly selected subsample of 200 participants drawn from the treatment groups only (the control group was not included) for a short phone survey to measure if they were still using the accounts.

In addition to all these surveys, we have access to administrative data from the telco on all the transactions made on the project mobile money accounts for the duration of the project (July 2017–April 2018) as well as for 15 months after (until July 2019).

D. Attrition

Attrition from our full sample of 801 respondents is shown in web appendix table A2. Columns 1–4 show attrition during the HFPS, with the odd columns showing whether a respondent appears at least once and the even columns showing the percentage of calls that were successfully completed. In round 1, 99% of respondents completed at least one survey and 89% of calls were made; in round 2, survey completion fell to 84% and 74%, respectively. While not unexpected, lower compliance in round 2 points to the problem of conducting phone surveys in general, as people lose their phones, change phone numbers, or become fatigued with the surveys. We do not find any differences in attrition across the various treatment groups and the coefficients on the treatment indicators are not significant for any of the surveys. However, HFPS respondents were 7 percentage points less likely to complete a first monitoring survey, which might be because HFPS respondents were more likely to be fatigued by the surveying process. Survey completion for the second monitoring survey is balanced across HFPS and non-HFPS respondents. In total, 761 of 801 sampled respondents ultimately appear in our analysis sample.

E. Summary Statistics and Randomization Check

Summary statistics and a check of randomization balance are presented in table 1 (from the baseline survey) and web appendix table A3 (from the intake survey). From table 1 panel A, 46% of the sample is male and has 9 years of education. Ninety-three percent of respondents have an iron roof on their homes and the average value of household assets is $873. Panel B shows statistics on business outcomes. Sixty-eight percent of the businesses are in retail, with the remainder predominantly in services, which includes occupations such as barbershops, tailoring, and welding. These businesses are very small: average weekly profits are about $19 per week and the average firm has only $293 in equipment and inventory. Nevertheless, these respondents are better off than the average Malawian.[26]

Panel C shows statistics on savings. Average savings across all sources was $120 at baseline, split across 2.5 savings places.[27] We observe that people already engage in physical separation of cash for different purposes: the average

[24] Respondents were enrolled in the lottery even if they were not reached for a specific call, but they did not receive the U.S. $2 payment for survey completion.

[25] Of the remaining 21%, 5.4% could not open the box because they did not have the key with them, 9.4% refused to open the box or travel home to show the box to the field officer, 2.4% did not have access to the boxes at box checks, 3.3% respondents with boxes were not checked because the respondent could not travel home with the field officer on the day of the survey.

[26] According to the World Bank, GNI per capita was $340 in 2017 in Malawi. In the 2016 Malawi Integrated Household Survey, only about half the households reported having iron sheets as the material of their roof.

[27] We did not ask whether people had more than one savings account, within a certain group. So, if a household held multiple bank accounts or participated in multiple VSLAs, they may have more savings places than we measure.

TABLE 1.—SUMMARY STATISTICS AND RANDOMIZATION CHECK

	(1)	(2)	(3)	(4)
	Full sample			
		p-value for joint equality		
	Control mean	Boxes	Mobile money	All treatments
Panel A. Demographic information and asset ownership				
Male	0.46	0.63	0.90	0.84
Household Farms	0.71	0.97	0.039**	0.10
Years of Education	9.05	0.97	0.32	0.59
	(2.83)			
Land owned (Acres)	0.66	0.75	0.87	0.86
	(1.18)			
Value of durable assets and livestock	872.60	0.36	0.06*	0.14
	(1,547)			
House has iron roof	0.93	0.26	0.79	0.22
Owns a cell phone	0.94	0.28	0.71	0.46
Has mobile money account	0.56	0.048**	0.64	0.19
Distance to closest mobile money agent (minutes)	11.32	0.62	0.44	0.69
	(14.25)			
Panel B: Business				
=1 if retail	0.68	0.82	0.39	0.49
Average weekly revenue	66.37	0.50	0.82	0.78
	(99.00)			
Average weekly profit	18.77	0.94	0.81	0.97
	(23.20)			
Value of equipment and inventory	199.10	0.92	0.32	0.22
	(401.20)			
Panel C: Savings				
Total cash savings (balance)	119.80	0.96	0.17	0.47
	(196.70)			
Saves in:				
Mobile money	0.32	0.38	0.65	0.70
Bank account	0.30	0.36	0.044**	0.059*
VSLA/ROSCA	0.52	0.55	0.97	0.86
Secret place at home	0.82	0.57	0.75	0.79
Savings box	0.22	0.085*	0.58	0.28
Observations	761			

Means are population weighted. Monetary values are winsorized at 1% and expressed in USD. In columns 1 and 4, standard deviations in parentheses; in the other columns, standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

respondent has 2.4 goals and saves up for these goals in different places. In particular, 78% of people have separate saving places for different purposes, and only 35% save for more than 1 purpose in a single savings place. The most common saving place is keeping cash at home, reported by 82% of respondents. Saving groups (VSLAs and ROSCAs) are used by 52% of the sample, and 47% report using mobile money accounts to save. Thirty percent of the sample have access to a bank account and 22% save in saving boxes. Note that a sizable minority of the sample already saves in lockboxes; our take-up rates suggest that the project lockboxes were seen as being of higher quality and were preferred to the original boxes. Sixty-two percent of the sample have access to a mobile money account, defined as either owning one or using someone else's. About half of the respondents have their own accounts at baseline.

In columns 2–4, we present *p*-values from F-tests of joint equality between the two box treatments and control, the two mobile money account treatments and control, and all four groups together. The samples appear to largely be balanced—of the 21 variables in the table, the *p*-value for the joint test across all groups is below 0.1 for just 1 variable—the dummy for saving in a bank account. We also find some evidence of imbalance within the subtreatments. For the box treatments, we find *p*-values below 0.1 between the box and control group for having a mobile money account and saving in a box (the probability of savings in a box is 0.22 in control, 0.23 in the one box group, and 0.12 in the multiple box group). For the mobile money account treatments, we find *p*-values below 0.1 for whether the household farms, value of assets, and a dummy for whether the household saves in a bank account.[28]

[28] We also examine balance on the variables measured in the intake survey in web appendix table A3. The majority of the characteristics suggest comparability of groups.

III. Results

A. Take-Up

Our primary measure of usage is the value of deposits, that is, the inflow of cash into the account, irrespective of the length of time this cash is held in the account. By definition, this is a flow figure, instead of a stock measure like the balance. While either measure will provide a sense of account usage (and indeed results are similar using other measures of usage including withdrawals and balance), we prefer deposits because accumulating a balance, in and of itself, was not the intended use of these accounts. Rather, we envisioned that the accounts could be used to save up for larger purchases, and therefore, we expected people to withdraw money eventually. Our expectation is that benefits should primarily accrue only after these withdrawals were made and were followed by the intended purchases, implying that the balance itself is not very informative—observing a low balance may be because usage was low, or simply because a withdrawal was made recently. For this reason, much of the prior literature has also focused on deposits.[29]

Figure 1 shows CDFs of deposits into the single and multiple devices. While administrative data would be the preferred data source for this analysis, we lack such data for the lockboxes, and we therefore show results for different data sources in different panels. In each panel, we show results separately for boxes and mobile money accounts. Panels A and B show data from the surveys (panel A is the monitoring survey and panel B is the HFPS). In both, we observe higher deposits among the multiple box group than the single box group, but no difference in deposits between those receiving one or multiple mobile money accounts. In panel C2, we use administrative data from the telco for mobile money users, and find no difference in usage (if anything, usage for multiple mobile money accounts appears to be lower, at least in the left tail).

Table 2 presents summary statistics. Panel A uses records from account opening. As expected, nearly all respondents who were offered an account chose to open one. Of those offered multiple boxes, only 76% took more than 1 box and only 42% took 3 (despite the boxes being free). However, for mobile money accounts, take-up of a second account was much lower (only 29%), which helps to explain the results in figure 1, which show no effect of multiple accounts. The telco was only able to link one account to each SIM card, and so using multiple accounts required people to switch between accounts. Even though people had access to dual-SIM phones, many respondents reported feeling that it was cumbersome to use more than one account, and so used only one. We therefore do not expect to find differential effects of these two treatment groups on downstream outcomes, and the multiple mobile money account treatment can be thought of as similar or identical to the single mobile money account treatment.

Panel B1 displays measurement of cumulative usage from the date of receiving the box until the in-person lockbox check in December 2017, about 5 months later. As discussed above, at this visit, we verified balances but had to rely on self-reports for deposit and withdrawal activity. Self-reported usage of boxes was nearly universal: 94%–97% of people used a box at least once and 91%–92% at least 5 times. This level of take-up is far higher than in many prior studies, including several in Malawi with banks (Dupas et al., 2018; Brune et al., 2016) or VSLAs (i.e., Ksoll et al., 2016). Our preferred measure of usage, the value of deposits, is also substantial: mean deposits were \$23 per month in the single box group and \$31 in the multiple box group (about 34% higher than the single box group).

Panel B2 show cumulative usage from administrative data from the telco. The evidence suggests lower usage than the boxes, though figures are still substantial. About 67%–79% of respondents made at least one mobile money deposit, and between 49% and 70% made more than 5. The average value of deposits was about \$9 per month.

In table 3, we regress the value of deposits (measured in different data sources) on an indicator for receiving multiple accounts (as well as other important background covariates). Across the three measures, we consistently observe that deposits in the multiple account group were no different in the case of mobile money, but were statistically significantly higher for boxes, with a magnitude of 24% in the lockbox check, 48% in the HFPS, and 26% in the monitoring surveys. While the difference in point estimates across measures is not particularly meaningful (since deposits were measured over different windows and may reflect seasonal changes in savings behaviors), these results all point to higher savings from the introduction of an additional box.[30] These regressions also confirm that for the mobile money account groups, receiving a second account did not cause an increase in usage.

B. Treatment Effects

The main analysis for this paper is organized around a preanalysis plan.[31] We prespecified the following outcomes: savings, credit, labor supply, household expenditures, and business outcomes (however, the plan does not specifically explicate how these variables are to be defined). For both the HFPS and monitoring surveys, we report results from two

[29] For example, Dupas and Robinson (2013a) and Dupas et al. (2019) report results for the value of deposits, while papers like Prina (2015) and Dupas et al. (2018) show both stocks and flows.

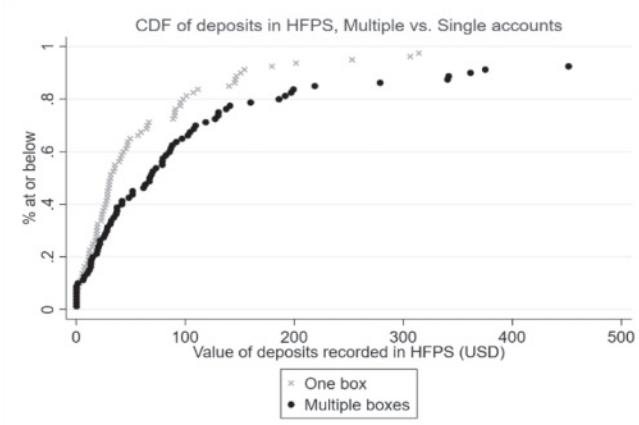
[30] Among background covariates, we find some evidence that people with higher baseline savings or that had bank or mobile money accounts saved more. We find some evidence that people who are more "taxed" by their networks (i.e., people who at baseline reported giving transfers but not receiving them) and people with more assets use boxes more; women use their boxes less. We find that distance to mobile money agents is associated with lower mobile money deposits, which is in line with the hypothesis that transaction costs discourage usage. We do not find any meaningful heterogeneity by other demographic characteristics.

[31] The PAP can be found at https://www.socialscienceregistry.org/trials/2449.

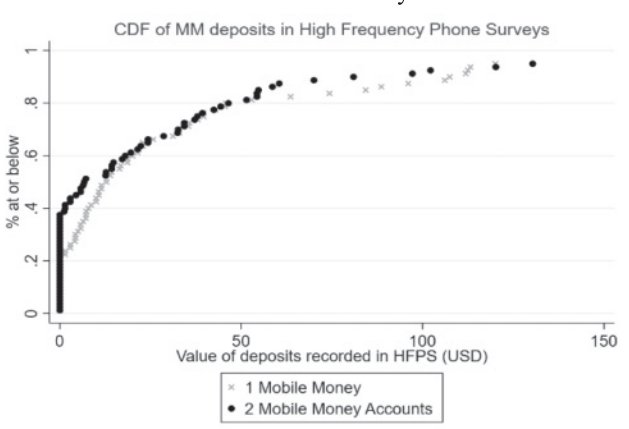
FIGURE 1.—CUMULATIVE DISTRIBUTIONS OF DEPOSITS

Panel A. High Frequency Phone Surveys

A1. Boxes

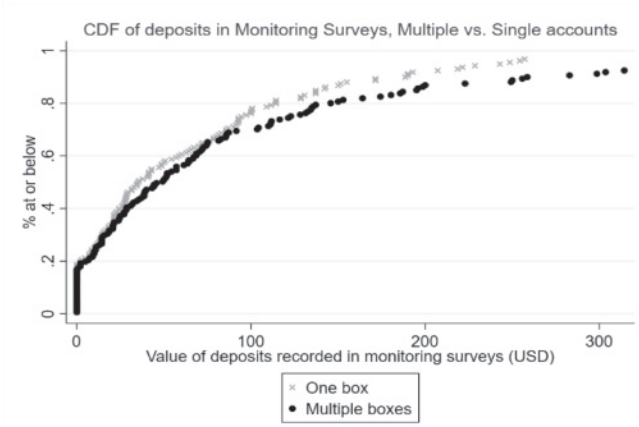


A2. Mobile Money

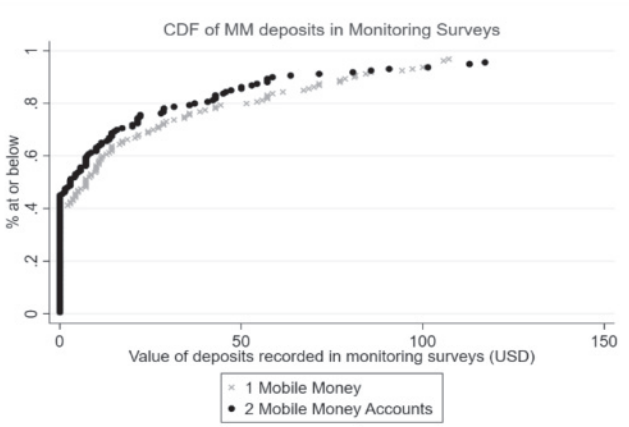


Panel B. Monitoring Surveys

B1. Boxes

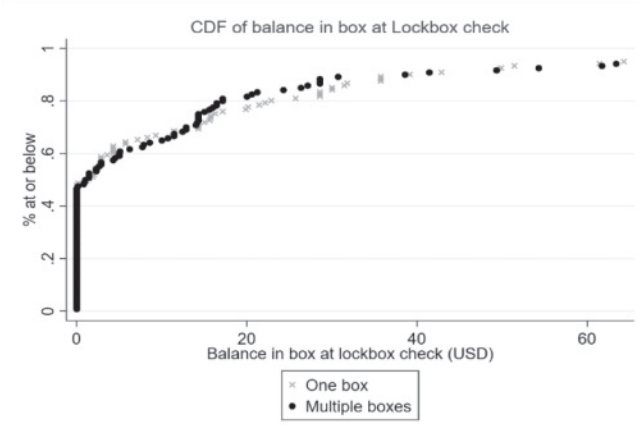


B2. Mobile Money

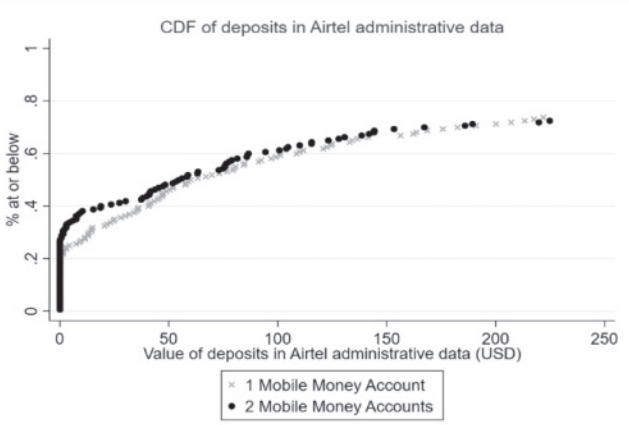


Panel C. Objective Measures

C1. Boxes (Visual Balance Check)



C2. Mobile Money (Telco Data)



Data sources are high frequency phone surveys (panel A), monitoring surveys (panel B), and lockbox check / telco data (panel C). Monetary values are in USD and CDF shows only below 95th percentile. Cumulative usage is calculated over 3.5, 4, 6, and 10 months for panels A1/A2, B1/B2, C1, and C2, respectively.

TABLE 2.—TAKE-UP OF PROJECT ACCOUNTS

	(1)	(2)	(3)	(4)
	Box groups		Mobile money	
	One box	Multiple boxes	One account	Multiple accounts
Panel A. Initial take-up June 2017				
Took at least one account	0.99	0.99	1.00	0.98
If offered multiple accounts: took >1 account	—	0.76	—	0.29
If offered multiple boxes: took 3 boxes	—	0.42	—	—
Observations	160	161	160	160
Panel B1. Lockbox check visit in December 2017				
Reported at least 1 deposit[1]	0.94	0.97		
Reported at least 2 deposits	0.94	0.95		
Reported at least 5 deposits	0.91	0.92		
Reported value of deposits (monthly average)	23.39	31.30		
	(30.58)	(36.18)		
Reported number of deposits (monthly average)	9.83	14.20		
	(10.74)	(13.45)		
Reported value of withdrawals (monthly average)	16.46	23.96		
	(24.95)	(31.37)		
Reported number withdrawals (monthly average)	1.48	2.61		
	(3.52)	(4.32)		
Observations	121	120		
Verified balance	17.98	16.46		
	(28.21)	(27.39)		
Observations	94	99		
Panel B2. Telecom administrative data (July 2017–April 2018)				
Made at least 1 deposit			0.79	0.67
Made at least 2 deposits			0.79	0.56
			0.70	0.49
Total value of deposits (monthly average)			8.72	9.34
			(12.25)	(15.56)
Number of deposits (monthly average)			1.11	1.06
			(1.36)	(1.42)
Total value of withdrawals (monthly average)			9.09	10.34
			(13.55)	(18.13)
Number of withdrawals (monthly average)			1.18	1.19
			(1.49)	(1.75)
Average daily balance			3.21	2.79
			(5.42)	(5.40)
Observations			160	160

See text for discussion of data sources. Means are presented, with standard deviations in parentheses. Verified balance is only reported for respondents who have agreed to show and open their lockboxes. Refer to appendix table C2 for correlations between reported and verified balances and text for explanations/procedures for the lockbox check.

[1] All deposit measures are self-reported.

intent-to-treat specifications. Both of our empirical specifications were prespecified, and are given by equations (1) and (2) below.[32] In equation (1), we analyze the effects of each of our four main treatments (one box, multiple box, one mobile money account, multiple mobile money account) separately:

$$Y_{ist} = \theta_1 + \theta_2 LB_i^1 + \theta_3 LB_i^{\text{mult}} + \theta_4 MM_i^1 + \theta_4 MM_i^{\text{mult}} + \tau L_t + \mu_s + \delta_t + \eta_{\text{HF}} + \lambda X_i + \epsilon_{ist}. \quad (1)$$

In our second specification, we pool the single and multiple account treatments together, to study the effects of having any box or any mobile money account:

$$Y_{ist} = \delta_1 + \delta_2 LB_i + \delta_3 MM_i + \kappa L_t + \mu_s + \delta_t + \eta_{\text{HF}} + \rho X_i + \epsilon_{ist}. \quad (2)$$

In equations (1) and (2), LB_i and MM_i are dummies for the lockbox or mobile money account treatment, while superscripts *1* and mult in equation (1) denote the single and multiple account treatments respectively.

In both equations, Y_{ist} is an outcome for individual i at time t in strata s, L_t is an indicator of airtime lottery wins (measured either daily or weekly), depending on the measurement window of the outcome variable, and μ_s is a strata fixed effect. δ_t is a fixed effect for the date of the interview in the case of monitoring survey regressions, and for the date of the outcome in question for HFPS regressions. η_{HF} is a fixed effect for whether the respondent was sampled for the HFPS surveys, and is therefore estimated only for the monitoring

[32] Note that our specifications are similar but not identical to those written in the PAP. In particular, we add three fixed effects that were not prespecified. They are (1) whether the respondent won the airtime lottery, (2) whether the respondent was sampled for the HFPS (for monitoring survey outcomes only), and (3) the date of the survey (monitoring survey) or outcome (HFPS). The results are not sensitive to these controls but including these fixed effects seem preferable due to questions about whether the lottery or HFPS has an effect, and to control for possible time trends.

TABLE 3.—DETERMINANTS OF ACCOUNT USAGE

	(1)	(2)	(3)	(4)	(5)	(6)
	Lockbox			Mobile money		
	Cumulative deposits measured in:					
	In-person lockbox check (Dec 2017)	HFPS surveys	Monitoring surveys	Telecom administrative data	HFPS surveys	Monitoring surveys
Multiple accounts	37.25*	52.15**	26.01**	9.57	−6.27	−0.17
	(20.54)	(21.11)	(12.09)	(15.36)	(6.89)	(4.15)
Other Covariates						
Time to nearest agent (hours)	12.72	−15.54	−5.12	−65.16*	−16.18	−13.6
	(25.01)	(22.64)	(16.29)	(35.04)	(12.78)	(9.47)
Female	−96.15**	−33.26	−32.19	30.69	−4.68	−4.65
	(45.57)	(46.36)	(26.69)	(33.83)	(14.98)	(9.14)
Married	−54.28	−28.47	−17.7	10.87	3.67	−1.87
	(38.59)	(38.96)	(22.48)	(28.78)	(12.51)	(7.78)
Female * married	47.05	2.02	4.03	−42.73	−3.14	−2.35
	(50.55)	(52.29)	(29.64)	(39.44)	(17.77)	(10.66)
Inverse hyperbolic sine of monetary savings	6.33	13.39*	4.97	5.49	3.88*	3.82***
	(6.38)	(7.10)	(3.91)	(4.36)	(2.11)	(1.18)
Log assets	22.45**	7.51	5.46	7.57	1.89	2.33
	(8.75)	(8.23)	(5.17)	(6.27)	(2.97)	(1.70)
"Taxed" (i.e., gives money but does not receive)	56.06**	16.8	19.74	−9.2	7.78	2.02
	(22.37)	(22.35)	(12.91)	(16.35)	(7.37)	(4.42)
Observations	238	159	317	319	159	319
Mean (1 account group)	116.9	70.11	69.82	87.75	32.5	23.67
Std. dev. (1 account group)	152.90	108.30	88.00	122.70	45.58	34.78

Values are in USD and winsorized at 5%. Columns 1–3 are for the lockbox groups only and Columns 4–6 are for the mobile money groups only. Regressions also control for age, household size, and having a mobile money account, box, or savings account at baseline. See text for discussion of data sources. Deposits are winsorized at 5%. Standard errors in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

survey regressions. X_i is a vector of controls, which includes the mean of the dependent variable (over the day covered by the intake survey), making the specification ANCOVA, as well as a set of other controls that were imbalanced at baseline.[33] Standard errors are clustered at the individual level, and the regressions are estimated with population weights.

All HFPS regressions are at the day level, i.e., we utilize the 3–4 day look-back period in the surveys to convert the data into a daily panel. Monitoring survey outcomes vary in how they are defined across outcomes. We provide detailed explanations for each outcome in the notes for relevant tables. We present two versions of all main tables, one for the HFPS and one for the monitoring surveys. In these tables, we report results from equations (1) and (2) in separate panels A and B. All monetary values are winsorized at 5% and expressed in USD (at 700 MWK to 1 USD).

The panels provide separate sets of information. Panel A tests whether the second account had an effect, by performing a test of equality between one and multiple accounts. This is particularly useful for savings outcomes (tables 4 and 5). However, because each treatment arm is of modest size, the pooled regressions in Panel B provide more power for testing the separate question of whether accounts affected downstream outcomes (whether multiple or single). The main tables are limited to only prespecified outcomes. We present naive p-values in the main tables, and sharpened q-values in web appendix tables A5 and A6 (following Anderson, 2008).[34] We also collected some non-pre-specified outcomes of interest, and show results for these in the appendix.

C. *Savings*

We present effects of treatment accounts on savings in tables 4 (HFPS) and 5 (monitoring surveys), where our dependent variable is daily deposits. In both tables, column 1 shows effects on savings in the project accounts, which are unsurprisingly statistically significant since the control group had zero savings in those accounts by design. In columns 2–5, we estimate the effect of treatment on savings in all boxes (or mobile money accounts), including nonexperimental accounts.

We discuss the lockbox first. In both surveys, we find that providing a single lockbox significantly increases total box savings; that providing multiple lockboxes increases box savings by more, and that the difference between treatment groups is significant. The effects are sizeable: single and multiple boxes increase box savings by \$0.49 and \$1.26 respectively in the HFPS (which is sizeable, even relative to *total* deposits of \$2.43 per day—column 9). In the monitoring surveys, the figures are \$0.47 and \$0.75, respectively,

[33] We include these controls at the suggestion of a referee. Results without these controls are included in appendix E. Results are similar but slightly stronger with controls.

[34] We correct p-values across all 19 prespecified outcomes. We also adjust for the number of regressors. At one extreme, since there are four individual treatment arms, there are 76 hypotheses for the individual regressors. At the other, there are only two tests per regressions for regressions in which we pool observations within box and mobile money account treatments.

TABLE 4.—TREATMENT EFFECTS ON DEPOSITS (HIGH FREQUENCY PHONE SURVEYS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		First Stage				Other savings sources			
	Value of deposits into experimental account	Deposited into any lockbox	Value of deposits	Deposited into any mobile money account	Value of deposits	Cash at home	Bank accounts	Savings groups	Total Value of Deposits
Panel A. Individual accounts									
One lockbox	0.77***	0.49***	0.49***	−0.08**	−0.13**	−0.61***	0.01	−0.13*	−0.19
	(0.12)	(0.05)	(0.13)	(0.04)	(0.06)	(0.18)	(0.10)	(0.07)	(0.34)
Multiple lockboxes	1.58***	0.62***	1.26***	−0.07*	−0.13**	−0.57***	0.03	−0.07	0.84*
	(0.28)	(0.05)	(0.27)	(0.04)	(0.06)	(0.17)	(0.13)	(0.08)	(0.45)
One mobile money account	0.30***	−0.04	−0.14	0.16***	0.19**	0.26	−0.12	−0.09	0.16
	(0.09)	(0.05)	(0.10)	(0.04)	(0.08)	(0.31)	(0.10)	(0.07)	(0.39)
Multiple mobile money accounts	0.31***	−0.03	−0.13	0.11**	0.11	−0.28	−0.03	−0.06	−0.38
	(0.10)	(0.05)	(0.11)	(0.05)	(0.08)	(0.18)	(0.09)	(0.06)	(0.31)
p-value: one box = multiple box	0.006***	0.011**	0.005***	0.78	0.90	0.78	0.90	0.41	0.033**
p-value: one m.m. = multiple m.m.	0.97	0.77	0.89	0.37	0.39	0.045**	0.48	0.69	0.16
Panel B. Pooled lockboxes and mobile money accounts									
Boxes	1.16***	0.55***	0.85***	−0.07**	−0.13**	−0.59***	0.02	−0.10	0.31
	(0.17)	(0.04)	(0.17)	(0.03)	(0.06)	(0.16)	(0.09)	(0.06)	(0.32)
Mobile money	0.29***	−0.04	−0.15	0.14***	0.15**	0.02	−0.08	−0.08	−0.09
	(0.08)	(0.04)	(0.09)	(0.04)	(0.07)	(0.23)	(0.07)	(0.06)	(0.31)
p-value: Lockbox = Mobile money	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.36	0.68	0.22
Observations	4,526	4,526	4,526	4,526	4,526	4,526	4,526	4,526	4,526
Number of businesses	390	390	390	390	390	390	390	390	390
Control mean	0.00	0.25	0.35	0.20	0.24	1.04	0.29	0.45	2.43
Control SD	0.00	0.43	0.82	0.40	0.68	1.80	2.76	0.82	3.93

All results are converted to daily averages. Monetary variables in USD and winsorized at 5%. Column 1: experimental account is the mobile money or box, depending on the treatment group, and equals 0 for the control group. Columns 2–5 include project and nonproject lockboxes/mobile money. Column 8: savings groups include VSLAs and ROSCAs. Column 9: total deposits is the sum of the other columns, as well as other less common types of savings. Standard errors clustered at individual level in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

on a base of $1.53. Turning to the mobile money treatment, we find clear evidence of usage but at a lower level than for the boxes. Total mobile money deposits increase by $0.11–$0.18 in the HFPS and $0.10–$0.12 in the monitoring surveys.

To provide some evidence on crowdout, we show effects on other savings places in columns 6–8 (as well as columns 3 and 5). As expected, in both surveys, we observe a statistically significant decline in cash at home for the lockbox group—in the status quo, many people keep cash at home in a secret location, and in the treatment group they moved some of this money into the box. We also observe declines in some other categories such as mobile money deposits and savings groups. Column 9 shows total deposits, in all savings sources. For the single box group, the treatment effect on total deposits is statistically insignificant in both surveys (and actually negative in the HFPS). For the multiple box group, the treatment effect is $0.84 in the HFPS and $0.49 in the monitoring surveys (significant at 10% and 5%, respectively). In both surveys, we reject equality of the treatments. When we pool the two box treatments (panel B), we see positive point estimates but with t-stats just over 1. For mobile money, we find weaker evidence of an increase in total deposits: because the second mobile money account was not used much, our preferred test is the pooled treatment (panel B), which actually has a negative point estimate in table 4 and a positive, insignificant coefficient in table 5.[35]

While we do not find strong evidence of an increase in total savings, this is not uncommon in the prior literature, largely due to power. Here too, the standard error on total savings (for boxes) is 10%–30% of the control mean, so that confidence interval includes large values. It is common in this literature to not find effects on total deposits, even in cases where there is evidence of downstream effects. Thus the ultimate indicator of usage remains downstream outcomes. Second, even if total deposits did not increase, it is possible that moving money from an insecure place like cash at home to the experimental accounts would be beneficial.[36]

[35] In web appendix B, we measure savings effects on other measures of usage: withdrawals (tables B1 and B2), balance (table B3), and net deposits (tables B4 and B5). As discussed above, we expected effects on both deposits and withdrawals, and we find statistically significant effects for total withdrawals in the monitoring surveys (though not the HFPS) for all treatments.

[36] We show heterogeneity analysis by three prespecified covariates in appendix D, tables D1 and D2. These are (1) pressure to share resources (i.e., if they are "taxed," which we define as giving but not receiving transfers); (2) gender; and (3) displaying hyperbolic preferences in incentivized time preference questions using the convex budget set methodology of Andreoni and Sprenger (2012). Many of these coefficients and F-tests are insignificant so we do not discuss them in the interest of space, and do not perform heterogeneity analysis by these characteristics.

TABLE 5.—TREATMENT EFFECTS ON DEPOSITS (MONITORING SURVEYS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		First stage				Other savings sources			
	Value of deposits into experimental account	Deposited into any lockbox	Value of deposits	Deposited into any mobile money account	Value of deposits	Cash at home	Bank accounts	Savings groups	Total value of deposits
Panel A. Individual accounts									
One lockbox	0.63***	0.61***	0.47***	−0.10*	−0.05	−0.23***	−0.01	−0.08	−0.03
	(0.06)	(0.05)	(0.08)	(0.06)	(0.04)	(0.06)	(0.07)	(0.10)	(0.18)
Multiple lockboxes	0.90***	0.58***	0.75***	−0.07	−0.05	−0.30***	0.17	−0.22**	0.49**
	(0.09)	(0.05)	(0.10)	(0.06)	(0.04)	(0.06)	(0.12)	(0.11)	(0.23)
One mobile money account	0.23***	−0.04	0.01	0.21***	0.11**	(0.01)	0.28**	−0.16	0.24
	(0.04)	(0.05)	(0.06)	(0.06)	(0.04)	(0.07)	(0.12)	(0.10)	(0.19)
Multiple mobile money accounts	0.22***	0.00	−0.03	0.22***	0.11**	−0.07	0.11	−0.10	0.12
	(0.04)	(0.05)	(0.06)	(0.06)	(0.05)	(0.07)	(0.08)	(0.10)	(0.19)
p-value: one box = multiple box	0.008***	0.50	0.019**	0.48	0.93	0.054*	0.11	0.16	0.026**
p-value: one m.m. = multiple m.m.	0.79	0.45	0.45	0.95	0.93	0.40	0.064*	0.48	0.51
Panel B. Pooled lockboxes and mobile money accounts									
Boxes	0.76***	0.60***	0.60***	−0.08*	−0.05	−0.26***	0.08	−0.15	0.22
	(0.06)	(0.04)	(0.07)	(0.05)	(0.04)	(0.06)	(0.08)	(0.09)	(0.17)
Mobile money	0.22***	−0.02	(0.02)	0.21***	0.11***	(0.04)	0.20**	−0.13	0.17
	(0.03)	(0.04)	(0.05)	(0.05)	(0.04)	(0.06)	(0.09)	(0.09)	(0.17)
p-value: Lockbox = mobile money	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.12	0.79	0.75
Observations	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321
Number of businesses	722	722	722	722	722	722	722	722	722
Control mean	0.00	0.30	0.25	0.36	0.18	0.30	0.23	0.54	1.53
Control SD	0.00	0.46	0.62	0.48	0.38	0.55	0.98	0.93	1.80

All results are converted to daily averages. Value of deposits was measured over the past 2 months in the monitoring surveys. Column 1: experimental account is the mobile money or box, depending on the treatment group, and equals 0 for the control group. Columns 2–5 include project and nonproject lockboxes/mobile money. Column 8: savings groups include VSLAs and ROSCAs. Column 9: total deposits is the sum of the other columns, as well as other less common types of savings. Standard errors clustered at individual level in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

D. Labor Supply, Business Outcomes, and Credit

In tables 6 (HFPS) and 7 (monitoring surveys), we examine prespecified outcomes related to labor supply, business success, and credit. Labor supply (columns 1–6) is disaggregated between the main business, secondary occupations, and farming. In both tables, we observe negative point estimates on labor supply in the main business, though the effect is stronger in the HFPS where pooled regressions for mobile money and all accounts show consistent decrease in hours worked. We find strong evidence that hours in farming increased—a finding that is robust even in the subtreatments (implying that the reduction in the main business was put into farming). We also find that the effects on farming hours are consistently larger (by a factor of at least 2), albeit insignificant, for the multiple box group than they are for the single box group, suggesting that greater savings in the multiple box treatment impacted downstream outcomes. Also consistent with the decline in labor supply, we observe negative coefficients on business profits and revenue (columns 7–8). This effect is similar in magnitude across HFPS and monitoring surveys (point estimates are 10%–15% of the control mean), though is only significant for the pooled mobile money treatment in HFPS.

The labor supply and business results suggest an increase in investment in farming, and a reallocation of labor away from the main business, which would be consistent with the marginal return to farming exceeding that of the main business. While we can only speculate as to why this might be, a possibility might be that returns to farming are risky or will only be realized in the future, and so respondents concentrate labor supply in daily business which provides immediate cash at relatively low-risk. Boxes or mobile money accounts may help people build up a buffer that allows them to mitigate this behavior. This finding is related to several recent papers that find labor supply effects of an easing of financial constraints, such as Fink et al. (2020), who find that providing credit to smallholder farmers decreases off-farm labor and increases own-farm labor. Our finding that saving accounts cause a reallocation of labor supply is related to Callen et al. (2019), who find that Sri Lankan households who were given access to deposit collection for interest-bearing accounts decrease their hours of leisure and increase labor supply.

Given that we detect effects on labor supply in farming, we examine a few other farming-related outcomes in appendix table A7. This analysis is restricted to three outcomes which were measured in only the monitoring surveys: renting land, buying land, and expenditures on farm inputs. Point estimates for all treatments are positive, though significant only for the boxes. The effects are large: respondents in the box groups are about 4 percentage points more likely to buy land (against a control mean of 2%) and about 9 percentage points more likely to rent (against a control mean of 5%). For farming

TABLE 6.—TREATMENT EFFECTS ON LABOR SUPPLY AND BUSINESS OUTCOMES (HIGH FREQUENCY PHONE SURVEYS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Labor Supply									
	Main business		Other occupations		Farming		Business outcomes		Credit taken[1]	
	=1 if worked	Hours	=1 if worked	Hours	=1 if farmed	Hours	Profits	Revenues	Took out loan	Value
Panel A. Individual accounts										
One lockbox	−0.08**	−0.78**	−0.02	−0.91	0.02	0.22**	−0.31	−0.48	−0.04	0.49
	(0.03)	(0.40)	(0.04)	(0.93)	(0.02)	(0.11)	(0.32)	(0.99)	(0.05)	(0.44)
Multiple lockboxes	0.00	0.14	0.03	2.33	0.07***	0.46**	−0.32	−0.63	−0.01	0.01
	(0.03)	(0.37)	(0.05)	(2.07)	(0.03)	(0.18)	(0.37)	(1.12)	(0.05)	(0.33)
One mobile money account	−0.05	−0.54	0.02	−1.40*	0.09***	0.52***	−0.49	−0.16	0.03	0.19
	(0.03)	(0.41)	(0.04)	(0.85)	(0.03)	(0.16)	(0.32)	(1.06)	(0.05)	(0.31)
Multiple mobile money accounts	−0.06	0.15	−0.01	−0.59	0.00	0.08	−0.50	0.31	−0.06	−0.14
	(0.04)	(0.56)	(0.04)	(0.85)	(0.02)	(0.10)	(0.35)	(1.14)	(0.05)	(0.32)
p-value: one box = multiple box	0.012**	0.024**	0.29	0.092*	0.055*	0.19	0.97	0.90	0.56	0.31
p-value: one m.m. = multiple m.m.	0.82	0.24	0.51	0.23	0.005***	0.01**	0.98	0.71	0.081*	0.31
Panel B. Pooled lockboxes and mobile money accounts										
Boxes	−0.04	−0.35	0.01	0.71	0.05***	0.33***	−0.32	−0.56	−0.02	0.25
	(0.03)	(0.33)	(0.04)	(1.30)	(0.02)	(0.11)	(0.29)	(0.86)	(0.04)	(0.31)
Mobile money	−0.05*	−0.25	0.01	−1.05	0.04**	0.32***	−0.49*	0.05	−0.01	0.05
	(0.03)	(0.39)	(0.04)	(0.79)	(0.02)	(0.11)	(0.29)	(0.90)	(0.05)	(0.27)
p-value: lockbox = mobile money	0.68	0.78	0.98	0.15	0.88	0.91	0.48	0.50	0.67	0.52
Observations	31,417	31,416	4,597	4,597	4,533	4,533	26,031	26,201	4,536	4,624
Number of businesses	391	391	390	390	390	390	391	391	390	390
Control mean	0.81	8.21	0.18	1.78	0.06	0.20	3.37	11.86	0.33	0.97
Control SD	0.39	4.67	0.38	7.23	0.23	1.20	3.59	15.78	0.47	4.24

The main business outcomes (in columns 1–2 and 7–8) were measured at the daily level. Labor supply in secondary occupations, farming, credit (columns 3–4, 5–6, and 9–10) are presented over a week. There are fewer observations for profits/revenues (compared to main business) because this was only asked for the past 4 days in the second round of HFPS (labor supply was asked for a 7 day recall). Credit includes digital loans, VSLAs, ROSCAs, banks, MFIs, and moneylenders. All monetary variables are expressed in USD and are winsorized at 5%. Standard errors clustered at individual level in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

inputs, we find effects of the order of 25% of the control mean.

Finally, we find evidence of an increase in credit to customers (measured only in the monitoring surveys—table 7, column 11), but no change in loans taken out (columns 9–10 in tables 6 and 7). For credit to customers, we find statistically significant effects for two treatments and borderline significant ones for the other two (t-stats of 1.4), with effect sizes ranging from $0.10–$0.17, equivalent to 43%–73% of the control mean of $0.23. This result suggests that increased liquidity may have been passed on to customers as a way of expanding business in a highly competitive environment. Casaburi and Reed (2019) show a similar finding, though in the different setting of traders buying cocoa from farmers.

E. Expenditures

Tables 8 and 9 show effects on various expenditure categories. All outcomes are expressed in daily terms. Since there are several treatments and several types of expenditures, we start with total expenditures (the last column in both tables). In both surveys, we find sizeable, positive coefficients for boxes. The coefficient is significant at 10% in table 8 but not significant in table 9 (with a t-stat of about 1.2). The mobile money coefficients are also positive, though not significant. We also examine spending on individual categories in the remaining columns. While many of the coefficients are positive, the strongest evidence of an effect is school spending in the HFPS.

F. Effect of Accounts on Ability to Cope With Shocks

One of our prespecified outcomes was to test whether accounts reduced susceptibility to shocks. Appendix tables A8 and A9 show effects of the accounts on interpersonal transfers and loans (please note that these outcomes were not prespecified). While several coefficients are positive, they are modest and largely insignificant. Usage data from the telco (shown in appendix table A10) is consistent with this pattern. While the average respondent deposited close to $120 over the study period, they only sent about $14 and received about $12. Panel B of the table indicates that, by contrast, people used the accounts for savings more than transfers. Web appendix tables A11 and A12 regress our set of prespecified outcomes (table A11) or whether a disease is treated promptly (table A12) on health shocks and an interaction between those shocks and treatment. We find no evidence of an improvement in risk-coping. While this result is in some contrast to much of the existing literature which does find improved resiliency as a result of mobile money, it could be that effects on these types of common health shocks is smaller than on bigger events.

TABLE 7.—TREATMENT EFFECTS ON LABOR SUPPLY AND BUSINESS OUTCOMES (MONITORING SURVEYS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Labor supply										
	Main business		Other occupations		Farming		Business outcomes		Credit taken		
	=1 if worked	Hours	=1 if worked	Hours	=1 if farmed	Hours	Profits	Revenues	Took out loan	Value	Credit to customers
Panel A. Individual accounts											
One lockbox	−0.04	−0.28	−0.02	−0.69	0.05	0.77	−0.47	−2.36*	0.00	1.44	0.14*
	(0.03)	(0.40)	(0.03)	(0.81)	(0.05)	(0.56)	(0.36)	(1.30)	(0.05)	(1.15)	(0.07)
Multiple lockboxes	−0.03	−0.49	0.08*	1.42	0.10*	1.65**	−0.26	−1.17	−0.04	1.49	0.14*
	(0.03)	(0.41)	(0.04)	(1.32)	(0.06)	(0.71)	(0.35)	(1.20)	(0.06)	(1.76)	(0.08)
One mobile money account	−0.07*	−0.95**	0.03	1.31	0.09	1.62***	−0.52	−1.94	0.02	1.18	0.10
	(0.04)	(0.48)	(0.04)	(1.04)	(0.05)	(0.62)	(0.35)	(1.24)	(0.05)	(1.29)	(0.07)
Multiple mobile money accounts	−0.03	−0.44	0.01	0.35	−0.02	0.88	−0.45	−1.31	(0.02)	0.04	0.11
	(0.04)	(0.40)	(0.03)	(0.82)	(0.05)	(0.67)	(0.35)	(1.25)	(0.05)	(1.19)	(0.08)
p-value: one box = multiple box	0.63	0.61	0.02**	0.088*	0.46	0.29	0.58	0.34	0.41	0.98	0.99
p-value: one m.m. = multiple m.m.	0.34	0.29	0.63	0.42	0.036**	0.34	0.83	0.59	0.50	0.37	0.83
Panel B. Pooled lockboxes and mobile money accounts											
Boxes	−0.03	−0.38	0.03	0.31	0.07*	1.18**	−0.37	−1.79*	0.00	1.44	0.14*
	(0.03)	(0.35)	(0.03)	(0.89)	(0.04)	(0.49)	(0.31)	(1.08)	(0.05)	(1.15)	(0.07)
Mobile Money	−0.05	−0.71*	0.02	0.83	0.03	1.27**	−0.49	−1.67	0.01	0.65	0.10*
	(0.03)	(0.37)	(0.03)	(0.73)	(0.04)	(0.51)	(0.31)	(1.10)	(0.04)	(1.07)	(0.06)
p-value: Lockbox = mobile money	0.52	0.28	0.93	0.60	0.35	0.87	0.62	0.89	0.35	0.87	0.93
Observations	9,247	9,247	1,319	1,319	1,321	1,319	1,306	1,315	1,321	1,321	1,315
Number of Businesses	722	722	722	722	722	722	720	721	722	722	718
Control Mean	0.74	7.35	0.13	1.82	0.24	2.05	3.02	10.07	0.67	5.57	0.23
Control SD	0.44	5.08	0.33	8.34	0.43	5.76	3.37	13.51	0.47	10.71	0.54

Labor supply in the main business (columns 1–2) was measured at the daily level (for 7 days prior to the survey date), other variables were measured over the past week. Labor supply in secondary occupations, farming, credit (columns 3–4, 5–6, and 9–10) are presented over a week. Profits and revenues (columns 7–8) are measured at the weekly level, but are converted to daily averages. Credit to customers (column 11) is measured over a month but converted to daily averages. Credit includes digital loans, VSLAs, ROSCAs, banks, MFIs, and moneylenders. See text for regression specification. All monetary variables are expressed in USD and are winsorized at 5%. Standard errors clustered at individual level in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

TABLE 8.—TREATMENT EFFECTS ON EXPENDITURES (HIGH FREQUENCY PHONE SURVEYS)

	(1) Staple foods	(2) Personal expenses	(3) Household expenses	(4) School expenses	(5) Total
Panel A. Individual accounts					
One lockbox	0.11	0.06	0.06	0.14***	0.58*
	(0.12)	(0.06)	(0.08)	(0.05)	(0.34)
Multiple lockboxes	0.03	0.00	0.10	0.15**	0.42
	(0.13)	(0.08)	(0.10)	(0.06)	(0.41)
One mobile money account	0.00	0.02	0.16	0.11*	0.47
	(0.14)	(0.07)	(0.10)	(0.06)	(0.42)
Multiple mobile money accounts	−0.03	−0.02	−0.04	0.02	−0.10
	(0.13)	(0.07)	(0.07)	(0.04)	(0.34)
p-values					
one lockbox = multiple lockbox	0.50	0.42	0.67	0.94	0.71
one m.m. = multiple m.m.	0.80	0.49	0.063*	0.15	0.18
Panel B. Pooled lockboxes and mobile money accounts					
Boxes	0.07	0.03	0.08	0.14***	0.51*
	(0.11)	(0.06)	(0.07)	(0.04)	(0.30)
Mobile money	−0.01	0.00	0.07	0.07*	0.22
	(0.12)	(0.06)	(0.07)	(0.04)	(0.32)
p-values					
Lockbox = mobile money	0.36	0.63	0.90	0.094*	0.34
Observations	4,522	4,536	4,536	4,536	4,536
Number of businesses	390	390	390	390	390
Control mean	1.47	0.50	0.54	0.22	3.93
Control SD	0.90	0.59	0.91	0.58	2.87

Expenditures are measured over the 7 days prior to the survey and are expressed in daily values. Total expenditures include the other columns in addition to other categories not shown here. All regressions control for strata, a measure of the dependent variable during the intake survey (where applicable), imbalanced characteristics at baseline, date fixed effects, and are probability weighted (see in the text for details). All monetary variables are expressed in USD and are winsorized at 5%. Standard errors clustered at individual level in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

TABLE 9.—TREATMENT EFFECTS ON EXPENDITURES (MONITORING SURVEYS)

	(1) Staple foods	(2) Personal expenses	(3) Household expenses	(4) School expenses	(5) Holiday spending	(6) Total
Panel A. Individual accounts						
One lockbox	0.14	0.10	0.36**	0.01	0.01	0.73
	(0.14)	(0.10)	(0.17)	(0.06)	(0.01)	(0.45)
Multiple lockboxes	0.18	0.05	0.07	−0.04	0.00	0.25
	(0.15)	(0.10)	(0.18)	(0.07)	(0.01)	(0.50)
One mobile money account	0.20	0.12	0.21	0.01	0.02	0.69
	(0.16)	(0.10)	(0.18)	(0.07)	(0.01)	(0.49)
Multiple mobile money accounts	−0.01	0.14	0.12	−0.01	0.01	0.30
	(0.14)	(0.10)	(0.17)	(0.06)	(0.01)	(0.48)
p-values						
One lockbox = multiple lockbox	0.76	0.62	0.089*	0.40	0.47	0.30
One m.m. = multiple m.m.	0.17	0.88	0.56	0.67	0.53	0.41
Panel B. Pooled lockboxes and mobile money accounts						
Boxes	0.16	0.08	0.23	−0.01	0.00	0.50
	(0.13)	(0.09)	(0.16)	(0.06)	(0.01)	(0.41)
Mobile Money	0.10	0.13	0.18	0.00	0.01	0.52
	(0.13)	(0.09)	(0.15)	(0.06)	(0.01)	(0.42)
p-values						
Lockbox = mobile money	0.57	0.48	0.68	0.76	0.31	0.96
Observations	1,317	1,320	1,320	1,321	1,321	1,321
Number of businesses	722	722	722	722	722	722
Control mean	1.88	0.72	0.98	0.41	0.06	5.34
Control SD	1.28	0.87	1.46	0.62	0.09	3.73

Expenditures in columns 1–3 are measured over the 7 days prior to the survey and are expressed in daily values. Education and holiday spending (columns 4–5) are measured over a few months prior to the survey and converted to daily values for comparability. Total expenses include the other columns in addition to categories shown here. All regressions control for strata, a measure of the dependent variable during the intake survey (where applicable), imbalanced characterstics at baseline, date of the survey fixed effects, assignment to high frequency group and are probability weighted (see in the text for details). All monetary variables are expressed in USD and are winsorized at 5%. Standard errors clustered at individual level in parentheses. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

G. *Long-Term Usage*

The results in much of the paper are for relatively short-term usage. Do people continue to use the accounts for a longer period? To shed some light on this, we collected a long-term usage survey in September 2019 (26 months after the accounts were first opened; 18 months after the endline survey; and 16 months after the final fee reimbursements were made for the mobile money group). We supplemented the information collected through this survey with long-term usage data (up until August 2019) from the telco. Results are summarized in appendix table A14 (panel A for the survey and panel B for the telco data).

For lockboxes, we find that almost everyone still had access to their boxes. Over 60% of box recipients are still using it, and over 50% made a deposit in the past month. Deposit amounts are sizeable ($14–$19 in the past month). For mobile money accounts, as during the study period, usage is lower than that for the boxes, but meaningful. Across the single and multiple account treatments, 40% reported still using the account and 30% reported having made a deposit in the past month. The reported deposits in the past month are $9, similar to the reported monthly average of deposits during the study period. The telco data corroborates this pattern of sustained usage, with 50% of the respondents having made at least one deposit after the study concluded.

IV. Threats to Validity

A possible concern with our analysis is that our program effects are almost entirely based on survey responses. This is unavoidable in this context, since these small businesses do not have digital records of activity (like barcode scanners) and most businesses this small do not keep detailed financial records (i.e., McKenzie & Woodruff, 2017). Moreover, some outcomes (like expenditures) can typically only be captured in a survey. Thus, while this research would be impossible without relying on surveys, their use does raise some questions which we address in this section. We dedicate a separate empirical appendix C to present all our analyses in this respect.

One type of concern would be experimenter demand effects or social desirability bias, that is, that respondents felt pressure to answer questions in a certain way, because they believed that those responses would be viewed more favorably either in the context of the experiment or because they constitute appropriate behavior in general. Our read of the literature is that existing research suggests that such effects are modest in many settings (i.e., De Quidt et al., 2018; Mummolo & Peterson, 2019; Dhar et al., 2018); moreover, such effects would have to be differentially present by treatment. Even signaling social desirability would be difficult for some outcomes, since the effects on some key outcomes (such as the decline in labor supply or the increase in spending) would have been unlikely to be perceived as desirable.

For a few savings-related outcomes, it is possible for us to explore the possibility of experimenter demand effects, because we have objective measures of actual usage. In particular, for the lockbox groups, we physically verified balances at the lockbox check; for the mobile money account groups, we can observe true usage in the telco's administrative data. We

start with reporting integrity checks for the mobile money account group, which are presented in appendix table C1. Here, we check correlations between the survey and the administrative data for a binary measure of making a deposit over any given period as well as for the amount deposited. The panels are organized in order of time period—panel A is at the day level, panel B at the week level, and panel C is cumulative over the time period. In each regression, the coefficients are statistically significant but point estimates are smaller than one. Point estimates are also monotonically increasing in the time period: the coefficient on total deposits is 0.45 at the day level, 0.65 at the month week level, and 0.82 cumulatively. We interpret these results as suggestive that respondents might have mixed up the exact dates of transactions, but generally truthfully reported on their savings behavior.

At the lockbox check visit, we first asked respondents questions about usage of the lockbox, and then asked to inspect the box right away. In case the box was stored at the respondent's home, we requested them to travel with us immediately to their home to show us the box (so that there was no opportunity to manipulate the amount in the box). There is undeniable selection into participating in this check, for example, respondents who lived far away or who had a high opportunity cost of time would be less likely to travel to their home. In total, out of those who agreed to participate in a survey (without anticipating a lockbox check), 79% of people agreed to show us the box.[37] For those who opened the box, we regress the observed amount on the reported amount in web appendix table C2. Pooling all boxes and individuals together, we find a coefficient very close to 1 (1.07), suggesting that reporting is, on the whole, accurate. Panels B and C run these regressions separately for the groups offered one box and multiple boxes. We find a pooled coefficient of 0.88 for one box and 1.13 in multiple boxes, meaning that for any given amount reported in the survey, people in the multiple box group *actually had more* in the box, potentially suggestive of difficulty remembering balances across multiple places. Thus, while we would view any differences in reporting behavior across treatment groups to be minimal, if anything they would work against finding larger survey measures of savings of multiple boxes goal.[38]

We plot the amount in the box against the amount reported in web appendix figure C1. Panel A shows all values, while panel B focuses only below the 75th percentile (since there are some very large values). As can be seen, the relationship is strong, but many values do not lie on the 45° line. We find that 50% report the exact sum in the box, 37% over-report, and 14% under-report. Thus there is some evidence of over-reporting, but there does not seem to be evidence that this would have substantially biased our results.

A separate concern with our study is that the HFPS itself changed behavior or reporting. We examine this in web appendix table C3 where we examine whether estimated treatment effects on the monitoring surveys differ between HFPS respondents and nonrespondents. In all specifications, a joint test of the significance of the interaction terms yields a p-value well over 0.05.[39]

V. Discussion and Conclusion

People throughout the world save up simultaneously for multiple goals of varying amounts and duration. A simple strategy for saving concurrently for several purposes might be to create multiple physically separated accounts, but this may be challenging in developing countries like Malawi where two thirds of the adult population lacks even a single bank account (see the 2017 Findex, Demirguc-Kunt et al., 2018). In this paper, we experimentally vary the number of accounts that respondents are given, and we find that entrepreneurs who receive multiple lockboxes saved about 30% more. By contrast, providing a second mobile money account had no effect, because of technical challenges. These results strongly suggest that a simple policy of providing multiple accounts with labeled purposes may cost-effectively increase savings, as long as the accounts are simple to use.

In addition, we find robust evidence that getting access to savings accounts had strong impacts on downstream outcomes. We observe that entrepreneurs who received savings accounts invested more in farming (possibly by substituting labor supply away from their small business), increased expenditures, and gave out more credit to customers. However, we do not find differential impacts of an additional box on downstream outcomes (the effect of an additional mobile money account is expected to be minimal based on take-up). While this could be for power reasons, it is also possible that the first box was often used for immediate expenses while the second box was for longer-term goals, and so effects would take longer to manifest. This is purely speculative however.

Our results lend support to the optimism around mobile money, and provide new evidence that mobile money can be effective as a tool for mobilizing savings (above and beyond its value as a transfer system). However, our work does leave some important questions open for future research. In our experiment, we waived withdrawal fees, provided training, and encouraged people to save for their goals. Each of these components was light-touch, and we conjecture that they had

[37] The differences in self-reported balances between those who showed lockboxes and who who did not was about −$11 ($30 for those who showed the box versus $41 for those who did not). The p-value of this difference is 0.37.

[38] It is possible that there exist incremental experimenter demand effects for multiple accounts (i.e., respondents may have felt compelled to report positive usage for each account separately). We do not view this as likely, since we observe no incremental reported effect of the second mobile money account (which we now was not used, from the administrative data).

[39] Another concern is that effects might have been partly driven by the lottery payments. To explore this, we interact each of our treatments with a dummy for having received a lottery payment and run regressions with fixed effects at the individual level. We code the lottery win as being relevant if it was won in the previous period. Results are presented in web appendix table D3. We find modest effects of the lottery in the control group: we see no effect on labor supply, though we observe evidence of an increase in revenues. We see some evidence of a differential labor supply response in the treatment group, but the sign of the treatment effect is not consistent.

a modest effect on usage; however, the evidence on this is not conclusive. For the withdrawal fees, we examine how usage changed in the period after fees were reimbursed (May 2018) in web appendix figure A5. We observe a decline in activity shortly after fees were removed, but a resumption of activity in the following months. The training we provided was very basic and consisted mostly of basic literacy on how to use mobile money accounts. This is something agents are already supposed to be doing, but in practice this training is not provided. Finally, we conjecture that simply encouraging people to save for their goals would not have been effective in isolation, but perhaps the combined effect of encouragement and mobile money was more effective than mobile money alone. Future work might explore these channels more deeply.

REFERENCES

Aggarwal, Shilpa, Valentina Brailovskaya, and Jonathan Robinson, "Daily Income, Labor Supply and Multiple Accounts: A Mobile Money Field Experiment with Micro-Entrepreneurs in Malawi," *AEA RCT Registry*, (2019), 10.1257/rct.2449-2.1.

——— "Cashing In (and Out): Experimental Evidence on the Effects of Mobile Money in Malawi," *AEA Papers and Proceedings* 110 (2020), 599–604. 10.1257/pandp.20201087

Ainslie, George, *Breakdown of Will* (Cambridge, U.K.: Cambridge University Press, 2001).

Aker, Jenny C., Melita Sawyer, Markus Goldstein, Michael O'Sullivan, and Margaret McConnell, "Just a Bit of Cushion: The Role of a Simple Savings Device in Meeting Planned and Unplanned Expenses in Rural Niger," *World Development* 128 (2020), 104772.

Anderson, Michael L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association* 103:484 (2008), 1481–1495. 10.1198/016214508000000841

Andreoni, James, and Charles Sprenger, "Estimating Time Preferences from Convex Budgets," *American Economic Review* 102:7 (2012), 3333–3356. 10.1257/aer.102.7.3333

Ashraf, Nava, Dean Karlan, and Wesley Yin, "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *The Quarterly Journal of Economics* 121:2 (2006), 635–672. 10.1162/qjec.2006.121.2.635

Batista, Catia, and Pedro C. Vicente, "Improving Access to Savings Through Mobile Money: Experimental Evidence from African Smallholder Farmers," *World Development* 129 (2020), 104905.

——— "Is Mobile Money Changing Rural Africa? Evidence from a Field Experiment," unpublished (2020).

Bénabou, Roland, and Jean Tirole, "Willpower and Personal Rules," *Journal of Political Economy* 112:4 (2004), 848–886.

Blumenstock, Joshua, Michael Callen, and Tarek Ghani, "Why do Defaults Affect Behavior? Experimental Evidence from Afghanistan," *American Economic Review* 108:10 (2018), 2868–2901. 10.1257/aer.20171676

BMGF, "The Impact of Mobile Money on Poverty: Research Brief," Technical Report (2020).

Bradley, Alice, *50 Family Budgets* (Woman's Home Companion, 1923).

Brune, Lasse, Xavier Giné, Jessica Goldberg, and Dean Yang, "Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi," *Economic Development and Cultural Change* 64:2 (2016), 187–220. 10.1086/684014

Callen, Michael, Suresh de Mel, Craig McIntosh, and Christopher Woodruff, "What are the Headwaters of Formal Savings? Experimental Evidence from Sri Lanka," *The Review of Economic Studies* 86:6 (2019), 2491–2529. 10.1093/restud/rdz020

Casaburi, Lorenzo, and Tristan Reed, "Interlinked Transactions and Competition: Experimental Evidence from Cocoa Markets," (2019).

Dalton, Patricio, Haki Pamuk, Ravindra Ramrattan, Daan van Soest, and Burak Uras, "E-Payment Technology and Business Finance: A Randomize Control Trial with Mobile Money," (2020).

de Mel, Suresh, Craig McIntosh, Ketki Sheth, and Christopher Woodruff, "Can Mobile-Linked Bank Accounts Bolster Savings? Evidence from a Randomized Controlled Trial in Sri Lanka," this REVIEW, 104:2 (2022), 306–320.

Demirguc-Kunt, Asli, Leora Klapper, Dorothe Singer, Saniya Ansar, and Jake Hess, *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution* (Washington, DC: The World Bank, 2018).

Dhar, Diva, Tarun Jain, and Seema Jayachandran, "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India," Technical Report (2018).

Dizon, Felipe, Erick Gong, and Kelly Jones, "The Effect of Promoting Savings on Informal Risk Sharing Experimental Evidence from Vulnerable Women in Kenya," *Journal of Human Resources* 55:3 (2020), 963–998. 10.3368/jhr.55.3.0917-9077R2

Dupas, Pascaline, and Jonathan Robinson, "Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya," *American Economic Journal: Applied Economics* 5:1 (2013), 163–192. 10.1257/app.5.1.163

——— "Why Don't the Poor Save More? Evidence from Health Savings Experiments," *American Economic Review* 103:4 (2013), 1138–1171. 10.1257/aer.103.4.1138, PubMed: 29533047

Dupas, Pascaline, Anthony Keats, and Jonathan Robinson, "The Effect of Savings Accounts on Interpersonal Financial Relationships: Evidence from a Field Experiment in Rural Kenya," *The Economic Journal* 129:617 (2019), 273–310. 10.1111/ecoj.12553

Dupas, Pascaline, Dean Karlan, Jonathan Robinson, and Diego Ubfal, "Banking the Unbanked? Evidence from Three Countries," *American Economic Journal: Applied Economics* 10:2 (2018), 257–297. 10.1257/app.20160597

Fink, Gunther, B. Kelsey Jack, and Felix Masiye, "Seasonal Liquidity, Rural Labor Markets, and Agricultural Production," *American Economic Review* 110:11 (2020), 3351–3392. 10.1257/aer.20180607

Francis, Eilin, "Paying to Repay? Experimental Evidence on Repayment Commitment," (2018).

Jack, William, and James Habyarimana, "High Hopes: Experimental Evidence on Saving and the Transition to High School in Kenya," (2018).

Jack, William, and Tavneet Suri, "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution," *American Economic Review* 104:1 (2014), 183–223. 10.1257/aer.104.1.183

John, Anett, "When Commitment Fails: Evidence from a Field Experiment," *Management Science* 66:2 (2020), 503–529. 10.1287/mnsc.2018.3236

Jones, Kelly, and Erick Gong, "Precautionary Savings and Shock-Coping Behaviors: The Effects of Promoting Mobile Bank Savings on Transactional Sex in Kenya," unpublished (2020).

Karlan, Dean, and Jonathan Zinman, "Price and Control Elasticities of Demand for Savings," *Journal of Development Economics* 130 (2018), 145–159. 10.1016/j.jdeveco.2017.10.004

Karlan, Dean, and Leigh L. Linden, "Loose Knots: Strong Versus Weak Commitments to Save For Education in Uganda," Technical Report (2014).

Kast, Felipe, and Dina Pomeranz, "Savings Accounts to Borrow Less-Experimental Evidence from Chile," Technical Report (2018).

Koch, Alexander K., and Julia Nafziger, "Goals and Bracketing Under Mental Accounting," *Journal of Economic Theory* 162 (2016), 305–351. 10.1016/j.jet.2016.01.001

Ksoll, Christopher, Helene Bie Lilleør, Jonas Helth Lønborg, and Ole Dahl Rasmussen, "Impact of Village Savings and Loan Associations: Evidence from a Cluster Randomized Trial," *Journal of Development Economics* 120 (2016), 70–85. 10.1016/j.jdeveco.2015.12.003

Lee, Jean N., Jonathan Morduch, Saravana Ravindran, Abu Shonchoy, and Hassan Zaman, "Poverty and Migration in the Digital Age: Experimental Evidence on Mobile Banking in Bangladesh," *American Economic Journal: Applied Economics* 13:1 (January 2021), 38–71. 10.1257/app.20190067

Lipscomb, Molly, and Laura Schechter, "Subsidies Versus Mental Accounting Nudges: Harnessing Mobile Payment Systems to Improve Sanitation," *Journal of Development Economics* 135 (2018), 235–254. 10.1016/j.jdeveco.2018.07.007

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *Science* 341:6149 (2013), 976–980. 10.1126/science.1238041, PubMed: 23990553

McKenzie, David, "Beyond Baseline and Follow-Up: The Case for More T in Experiments," *Journal of Development Economics* 99:2 (2012), 210–221. 10.1016/j.jdeveco.2012.01.002

McKenzie, David, and Christopher Woodruff, "Business Practices in Small Firms in Developing Countries," *Management Science* 63:9 (2017), 2967–2981. 10.1287/mnsc.2016.2492

Mummolo, Jonathan, and Erik Peterson, "Demand Effects in Survey Experiments: An Empirical Assessment," *American Political Science Review* 113:2 (2019), 517–529. 10.1017/S0003055418000837

Prina, Silvia, "Banking the Poor Via Savings Accounts: Evidence from a Field Experiment," *Journal of Development Economics* 115 (2015), 16–31. 10.1016/j.jdeveco.2015.01.004

Quidt, Jonathan de, Johannes Haushofer, and Christopher Roth, "Measuring and Bounding Experimenter Demand," *American Economic Review* 108:11 (2018), 3266–3302. 10.1257/aer.20171330

Soman, Dilip, and Amar Cheema, "The Effect of Partitions on Controlling Consumption," *Journal of Marketing Research* 45:6 (2008), 665–673. 10.1509/jmkr.45.6.665

——— "Earmarking and Partitioning: Increasing Saving by Low-Income Households," *Journal of Marketing Research* 48:SPL (2011), S14–S22. 10.1509/jmkr.48.SPL.S14

Suri, Tavneet, "Mobile Money," *Annual Review of Economics* 9 (2017), 497–520. 10.1146/annurev-economics-063016-103638

Suri, Tavneet, and William Jack, "The Long-Run Poverty and Gender Impacts of Mobile Money," *Science* 354:6317 (2016), 1288–1292. 10.1126/science.aah5309, PubMed: 27940873

Thaler, Richard H., "Mental Accounting Matters," *Journal of Behavioral Decision Making* 12:3 (1999), 183–206. 10.1002/(SICI)1099-0771(199909)12:3⟨183::AID-BDM318⟩3.0.CO;2-F

Wieser, Christina, Miriam Bruhn, Johannes Philipp Kinzinger, Christian Simon Ruckteschler, and Soren Heitmann, "The Impact of Mobile Money on Poor Rural Households: Experimental Evidence from Uganda," *World Bank Policy Research Working Paper* 8913 (2019).

Zelizer, Viviana A., "The Creation of Domestic Currencies," *The American Economic Review* 84:2 (1994), 138–142.