Courant Research Centre

'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 270

Commitment or Concealment? Impacts and Use of a Portable Saving Device: Evidence from a Field Experiment in Urban India

Janina Isabel Steinert, Rucha Vasumati Satish, Felix Stips and Sebastian Vollmer

First Version: April 2020
This Version: September 2021

Platz der Göttinger Sieben 5 · 37073 Goettingen · Germany
Phone: +49-(0)551-3921660 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Commitment or Concealment? Impacts and Use of a Portable Saving Device: Evidence from a Field Experiment in Urban India

Janina Isabel Steinert*[1,2], Rucha Vasumati Satish[3], Felix Stips[4,5], and Sebastian Vollmer[3]

[1]TUM School of Governance, Technical University of Munich, Germany
[2]Department of Social Policy and Intervention, University of Oxford
[3]Chair of Development Economics, University of Goettingen, Germany
[4]Luxembourg Institute of Socio-Economic Research (LISER), Luxembourg
[5]Center for Evaluation (CEval), Germany

Abstract

To quantify the impact of a novel "soft" commitment intervention, we randomly allocate 1525 Indian slum dwellers to receive a zip purse and a lockbox (treatment) or a lockbox only (control). After six months, we document a 19 percent increase in total savings in the treatment arm. The effect is sustained in a subsample of participants we re-interview during the COVID-19 pandemic, twenty months after initial distribution of the devices. While temptation spending was not reduced, additional analyses suggest that the zip purse served as a hiding rather than a self-control device. Our results highlight the importance of considering the role of financial transfers to other household members in future saving promotion programs.

Keywords: Saving, Temptation Spending, Commitment Device, RCT, India
JEL codes: D14, D15, D91, I31, O12, O16
Study pre-registration: Social Science Registry, ID: AEARCTR-0003682

*corresponding author: janina.steinert@tum.de

1 Introduction

Saving has crucial welfare-enhancing functions for the poor. It can smooth consumption, mitigate the impact of adverse financial shocks and make it possible to take advantage of profitable investment opportunities (Hulme et al., 2015; Karlan et al., 2014; Steinert et al., 2018b). Yet, empirical evidence suggests that limited resources and a combination of institutional, social, and psychological barriers hinder low-income individuals from reaching optimal levels of savings (Brune et al., 2021; Dasso and Fernandez, 2015; Goedecke et al., 2018; Houser et al., 2018; Schilbach, 2019). This means that there is much to be gained from identifying effective instruments to promote savings and understand the mechanisms underlying these interventions.

In this paper, we studied the impact of a new saving promotion intervention. The intervention consisted of a portable saving device, a zip purse, and a stationary lockbox. Previous studies found that a lockbox alone could help individuals save more by making withdrawals psychologically costly (for example Aggarwal et al., 2021; Aker et al., 2020; Dupas and Robinson, 2013). The portable saving device was intended to prompt deposits: participants were encouraged to carry the zip purse throughout the day and use it to store money that would otherwise be lost to temptation expenditures. We argue that the zip purse adds value to previous intervention designs in three ways. First, it serves as a daily saving reminder, thus increasing the salience of saving (Akbas et al., 2016; Karlan et al., 2016). Second, it functions as a mental accounting nudge by earmarking money as savings, whereas money held elsewhere - for example in a regular purse - has no such designation (Lipscomb and Schechter, 2018; Shafir and Thaler, 2006; Soman and Zhao, 2011). Third, carrying the portable saving

device may induce feelings of regret (in case money is spent on temptation goods) or gratitude (in case money is redirected into the zip purse for saving purposes) at the point in time when spending decisions are made (Bénabou and Tirole, 2004; Burke et al., 2018; Shefrin and Thaler, 1992).

To test the effectiveness of the device, we conducted a field experiment with 1525 low-income slum dwellers in the city of Pune, India. Half of the participants were randomly assigned to the treatment group and received a lockbox and the portable saving device. The remaining half, the (active) control group, received a lockbox only. Participants in both groups were encouraged to set individual saving goals, aided by community-based program facilitators. We base our analysis on data from endline surveys with 1421 participants, conducted six months after the devices were distributed. We also report results from a COVID-19 follow-up survey with a subgroup of the original sample.

Three main findings emerged. First, we observed a statistically significant higher total savings balance in the treatment group relative to the control group. This corresponds to a 19% increase in the average savings balance using our preferred specification, with a 1% high tail winsorizing of the savings variable. The savings balance includes any money kept in the devices or in a formal bank account, a savings circle, or at home. Twenty months after delivery of the devices and at the peak of the COVID-19 pandemic, we still observed a higher total savings balance in the treatment group relative to the control group. Treatment group participants were also five percentage points more likely to draw on their savings to cope with the economic impact of the pandemic.

Second, the intervention increased female empowerment and reduced debt. However, these treatment effects were not significant at the 10% level after adjusting for multiple hypothesis testing. The former effect could reflect increases in women's financial autonomy, linked to having a saving tool of their own. Against this backdrop, the intervention may have enabled women to make financial decisions without seeking their spouses' approval, and to save money that otherwise would have been spent by their spouse (Fiala, 2017).

Third, we did not find any evidence that more self-control is a transmission channel for the increase in total savings. Temptation spending was generally low, and the treatment and control groups were on par at endline. Nor did we find heterogeneous treatment effects on savings by the baseline level of present bias. Taken together, it seems unlikely that higher savings were due to strengthened self-control and fewer temptation expenditures. This finding contradicts our initial hypothesis that the zip purse would function primarily as a soft commitment device to limit participants' impulsive spending. Instead, additional qualitative and quantitative analyses point to an alternative channel: participants appear to use the portable device to hide private savings from others, most likely from their spouse or other family members. This motivates a reinterpretation of the purpose of the portable saving device as a *hiding* rather than a *commitment* tool.

Findings from this trial feed into three strands of literature. First, they contribute to a large body of literature that assesses the effectiveness of different types of commitment devices in mitigating psychological barriers and increasing savings (for reviews see Bryan et al., 2010; Karlan et al., 2014; Steinert et al., 2018a). These may include

hard commitment devices, which encompass institutionalized flexibility constraints or economic penalties for deviations from a saving plan (Aggarwal et al., 2021, 2018; Aker et al., 2020; Beshears et al., 2020; Brune et al., 2021, 2016; Casaburi and Macchiavello, 2019; Herskowitz, 2021; Karlan et al., 2014), and *soft* commitment devices, which are aimed at manipulating perceived or psychological costs (Bénabou and Tirole, 2004; Breza and Chandrasekhar, 2019; Soman and Zhao, 2011). Our results show that low-income slum dwellers in India exhibited demand for a new type of soft commitment device, a portable (and earmarked) zip purse, as causally expressed by its positive impact on savings, and, qualitatively, expressed by participants during focus group discussions. However, the demand for this device seems to arise not from psychological barriers to saving but rather from a desire to evade social obligations. Our findings add to an evidence base suggesting that avoidance of kin taxes is a crucial component of demand for saving commitments (Baland et al., 2011; Goldberg, 2017; Jakiela and Ozier, 2016).

We also add to a smaller but growing literature that studies the mechanics of temptation spending and commitment (Banerjee and Mullainathan, 2010; Banerjee and Duflo, 2007; Brune et al., 2021; Dasso and Fernandez, 2015; Evans and Popova, 2017; Haushofer and Fehr, 2014; Houser et al., 2018; Mullainathan and Shafir, 2013; Schilbach, 2019; Brune et al., 2021). We introduce a novel measure of temptation spending, which closely mimics the theoretical counterpart of goods that provide utility during consumption but not in anticipation thereof (Banerjee and Mullainathan, 2010). Using this novel measure, we find low levels of temptation spending in our sample. This is in line with previous studies pointing to a minor role of spending on alcohol, tobacco or other temptation goods (Brune et al., 2021, 2017; Evans and

Popova, 2017). Third, we explicitly link temptation spending to saving behavior, finding - in line with the study by Brune et al. (2021) - that temptation spending does not appear to be the main driver of saving increases in the context of our study.

Finally, our results align with a growing body of scholarship that discards a unitary household model, which assumes a common utility function between husband and wife (Anderson et al., 2017). Scholars have argued that prevailing differences in spouses' relative decision-making power are highly relevant for the allocation of household resources (Hoel, 2015). Hence, they can determine how much money is invested into children's futures, business and entrepreneurship activities, and savings (Anderson et al., 2017; Bobonis, 2009; Duflo and Udry, 2004; Duflo, 2003; Iversen et al., 2011). Women with low bargaining power are particularly likely to "lose" potential profits and savings to their husband or partner (Bernhardt et al., 2019; de Mel et al., 2009; Fiala, 2017). This leads to high demand among women for privately held financial tools (Anderson and Baland, 2002; Castilla, 2019; Schaner, 2015). In a similar vein, women in our sample seem to have benefited from the feature of the portable saving device that allowed them to hide money from others and increase their financial autonomy (see also Ashraf et al., 2010). As one of our participants put it: *"I keep some money with me in the purse. If he [husband] wants money he takes it from the box or asks me. I give him the money from the box but he does not know that I have more money with me in my purse."*

In the next section we describe the setting, sample, data, and experimental approach of the trial. In Section 3 we present the main results and heterogeneity analyses. In Section 4 we seek to identify channels that might underlie the treatment effect on

savings, and we discuss caveats and policy implications in the conclusion in Section 5.

2 Experimental design and data collection

2.1 Study setting

The study took place in informal settlements surrounding the cities of Pune and Pimpri-Chinchwad, which are both located in India's western-central state Maharashtra. Maharashtra is the second most populous Indian state and one third of its population lives below the poverty line (World Bank, 2017). Financial inclusion in India is currently at 80% of the adult population (Demirguc-Kunt et al., 2018). While this rate is relatively high compared to other low- and middle-countries, 48% of account holders had not made any deposits or withdrawals in the previous year and account inactivity was thus more prevalent than in other countries (ibid). In addition, rates of financial inclusion were lower among women (only 45% have a bank account of their own) and members of backward casts. Overall, India is still home to 190 million unbanked individuals (Demirguc-Kunt et al., 2018; International Institute for Population Sciences, 2017). We therefore expected demand for alternative saving devices to be high among low-income slum dwellers. India is also one of the lowest-ranked countries on the Gender Inequality Index (World Bank, 2012). Literacy rates are ten percentage points lower for women relative to men, health outcomes for women are substantially worse, and one third of ever married women have experienced physical or sexual violence by a partner (International Institute for Population Sciences, 2017). Potential implications of prevailing gender inequalities were carefully considered for

the intervention design, selection of outcome measures, and interpretation of findings in this study.

2.2 Sampling

The sample of this study consisted of 1525 slum dwellers who were aged 18 years and older.[1] The majority of participants (82%) were female, owed to the fact that women were easier to reach by the study team and more willing to enrol in the trial. Eligibility criteria were defined so as to ensure that participants' financial resources exceeded their subsistence needs and thus gave them the potential to save money. Accordingly, a person was considered eligible for this study if he/she indicated having some income at least once per week or on a monthly basis – either through permanent employment, casual work, remittances or governmental cash transfers.

Enumerators followed a random-walk procedure by starting from a central landmark in each slum or village and then carrying out door-to-door visits to every second household whereby the team split and took both the left and right turn. For each household, enumerators were instructed to recruit one adult participant. While interviews were conducted with only one spouse, the experimental saving devices were delivered to both spouses. Informed consent was obtained from the main respondent

[1] Power calculations for this trial were based on three related RCTs conducted in India, namely (1) (Soman and Zhao, 2011) in which saving was promoted through earmarking money for specific purposes and partitioning savings in envelopes, (2) (Breza and Chandrasekhar, 2019) in which saving was promoted through a peer-based soft commitment intervention, and (3) (Somville and Vandewalle, 2018) in which saving was promoted through a default effect. We used the average of the effect sizes from these studies (δ=0.37) and accounted for the testing of seven hypotheses (i.e. seven primary and secondary outcomes) by using conservative Bonferroni corrections, reducing the alpha level from 0.05 to 0.007. With a minimum detectable effect size of δ=0.37 and an alpha level of 0.007, we would still ensure 80 percent power with a sample size of 1500 participants.

during the first home visit, prior to the baseline interview. We additionally sought informed consent from the participants' spouse prior to delivery of the intervention. We did not provide any monetary incentives for participation in this study.

2.3 The intervention

The intervention, named *"Aaj bachat kara, udya khush raha"* (Marathi for *"Save today, be happy tomorrow"*), was developed with the intention to reduce temptation spending and promote saving. We conducted a preliminary pilot study informed by two focus group discussions to validate and adjust the aim and specific design of the intervention.[2] The core feature of the intervention was a portable commitment device – a zip purse – that participants received in addition to a stationary savings box. Our intervention was built on the guiding hypothesis that the zip purse would serve as a reminder and exert a nudge by penalizing temptation spending with instant feelings of regret (Burke et al., 2018; Lipscomb and Schechter, 2018; Soman and Zhao, 2011). The presumed innovation vis-à-vis existing commitment interventions was that the penalizing force of the device would take effect precisely at the point in time when spending decisions were made.

Active Control Group

The experiment was set up with an active control group. Specifically, participants assigned to the control group received a stationary saving device. The stationary

[2]The focus group discussions (one consisting of twelve women and one consisting of eight men) were conducted in slum locations in Pune. Participants discussed different design options for the portable saving device and the majority favored a zip purse made of traditional fabrics, arguing that these could be easily fixed or pinned to a belt or sari and thus protected from loss.

device was a metal box secured with a padlock. This feature was based on several experiments, which have endorsed lockboxes as effective saving instruments in settings with poorly developed formal savings infrastructure (Aker et al., 2020; Dupas and Robinson, 2013). In this study, each spouse was provided with a key to the padlock to ensure that money retained its liquidity and could be accessed in case of emergencies.

The lockbox and keys were distributed to the participant and his/her spouse during home visits. These were conducted by local community workers who were trained as program facilitators. After explaining the purpose of the lockbox, facilitators asked participants to formulate a savings goal, that is, to specify a target savings amount that they would like to reach over an individually defined timeline. Together with participants, they outlined an individualized saving plan, which specified daily or weekly targets and a timeline to reach the savings goal. The savings goal and implementation plan were visualized on a savings sheet that participants could put on their walls as an additional reminder (see Figure A2). This intervention component was motivated by goal setting theory that predicts a direct link between conscious goals and following actions (Locke and Latham, 2002). Firstly, goal setting helps direct attention towards a specified goal; secondly, increases effort and enthusiasm and may thus help to overcome procrastination; and thirdly, it motivates perseverance (Alan et al., 2019; Karlan et al., 2014). Building on evidence pointing to higher goal attainment when implementation intentions were formed, our intervention combined the savings goals with a concrete implementation strategy (i.e. how much money to deposit each week) (Townsend and Liu, 2012).

Treatment Group

Treatment group participants received the portable saving device, the zip purse, on top of the stationary device (photographs of both devices are shown in Appendix Figure A1). Facilitators instructed participants to carry their empty portable device with them whenever they left their homes, for example when going shopping. Any money that participants would manage to *not* spend - for example on specific temptation goods - should then be moved from their regular wallet or pocket into their portable saving device - so as to keep the money safe for saving purposes. Facilitators further advised participants to move money from their portable into their stationary device on a regular - and ideally daily - basis so as to reduce the risk of loss. Importantly, both spouses received a zip purse of their own. We opted for this joint delivery strategy to alleviate the risk of potential conflicts and power imbalances that may arise from withholding putative intervention benefits from one partner. All other intervention components were kept identical to those of the control group. The full home visit script is provided in the pre-analysis plan (see AEA trial registry).

2.4 Experimental design and timeline

The study randomly assigned 1525 individuals to receive either the lockbox and the portable saving device (treatment group, n= 771) or the lockbox only (control group, n=754). Randomization was performed in Stata and stratified by participant sex, baseline savings, and baseline levels of present bias. The trial and a pre-analysis plan were registered in the American Economic Association's registry for randomized controlled trials (ID: AEARCTR-0003682).

Recruitment and baseline surveys were carried out from November 2018 to January 2019. The saving devices were delivered to treatment and control group participants between February and April 2019. Endline surveys were administered between August and October 2019, with some unanticipated delays due to monsoon-related flooding of several slum locations in Pune.[3] In addition, we held four focus group discussions with program participants in October 2019.

2.5 Data

Baseline and endline data were collected via standardized questionnaires that were administered on mobile tablets. Interviews were conducted with one household member only, either the male or female spouse. We opted for computer-assisted data collection in order to improve data quality by programming built-in skip-patterns, reminders, and consistency checks to prevent item non-response or selection errors, and reduce respondent fatigue through programming visually appealing questionnaires including vignettes and pictures. Questionnaires were available in both English and Marathi, and each item was piloted with the local research team for cultural adequacy and accurate understanding. Enumerators were recruited from local communities and had to be fluent in Marathi. The local research team received a five-day training focused on interview techniques, research ethics, and familiarization with the survey. Individual interviews lasted between 30 and 45 minutes and were typically conducted in participants' homes. The research team made substantial efforts to guarantee a private and comfortable interview atmosphere and avoid the presence of other household

[3]From June to September 2019, India recorded the most extreme level of monsoonal rainfall in the past 25 years. 28.000 people living in Pune had to be evacuated in consequence of flooding.

members (especially of the spouse or parents-in-law) during the interview. Ethical approval for this study was obtained through the University of Göttingen.

The survey captured essential sociodemographic information, including household composition, caste, religion, and a number of wealth indicators, namely education, employment, and income as well as asset ownership. The study's primary outcomes were total savings balances and total temptation expenditures. For the former outcome, we first collected detailed self-reported information on savings held in a formal bank or post office account, savings group or savings held with relatives. For each saving type, we then documented (a) total amounts, (b) past-month deposits, and (c) past-month withdrawals. Subsequently, we hand-counted the money that participants held in their lockbox, and, if applicable, money kept in their portable saving device.[4] Given that participants were not informed about the date and time of their endline interview prior to the visit, we expect this measure to be relatively immune to potential reporting and social desirability biases. We also instructed enumerators to take pictures of the hand-counted money in the zip purse and in the lockbox. This enabled us to verify whether particularly high (or possibly implausible) savings amounts that were recorded in the questionnaires were corresponding to these photographic proofs. To construct our final measure of total savings, we added up the hand-counted savings in the experimental devices and the self-reported savings held in other places. By combining across these saving types, we are able to rule out crowding-out effects

[4]Approximately 70 percent of all respondents were willing to open their lockbox to allow hand-counting of their savings. There was no significant difference in participants' willingness to open the lockbox between experimental groups. In the treatment arm, 62 percent of participants agreed to open their purses for hand-counting. While not all participants agreed to the hand-counting, neither willingness to open the lockbox nor the purse explained any variations in recorded savings amounts. Coefficients for both factors were positive, which suggests that the self-reported savings amounts may - if anything - underestimate rather than overstate the true amounts.

that could result from simply shifting money from one saving mode to another.

To measure temptation spending, we introduced a new measurement approach that has advantages over more conventional measures in that it avoids reliance on *a priori* researcher-defined categories of temptations (for a somewhat similar approach, see Brune et al., 2017). Specifically, the theoretical literature characterizes temptation goods as goods that provide utility when consumed, but not in anticipation of their consumption (Banerjee and Mullainathan, 2010; Bénabou and Tirole, 2004). Using this standard definition, we captured past and desired future consumption of nine food items (sugar, meat, cola/ lemonade, alcohol, fried snacks, cake) and non-food items (gambling, tobacco, toys).[5] The items were only classified as temptation goods if the reported amount of past expenses exceeded the desired future amount. For each respondent, we calculated past-month temptation expenditures by summing the gap between reported and desired expenditures. We used the past month as the reference period to facilitate recall and reduce measurement error relative to longer time frames.

In addition, we included six secondary outcomes in the survey. First, we measured respondents' self-efficacy by drawing on selected items from the *Internality, Powerful Others and Chance (IPC)* scale (Levenson, 1981) and from financial self-efficacy scales used previously by (Lown, 2012; Steinert et al., 2018a) (e.g., *"I am confident that I can plan carefully in advance how to use my money during each week", "My life is controlled by other powerful people".*) Individual items were aggregated into a continuous scale centered around zero by weighting items based on principal component

[5]Our questions were specifically phrased to capture participants' *preferences* (*"How much would you like to spend...?"*) rather than their *expectations* (*"How much do you expect to spend...?"*) on future expenditures.

analysis (PCA) (for more details on the weighting procedure, please see Appendix). We expected increases in self-efficacy, based on previous research suggesting that self-defined goals and implementation plans can instil feelings of self-efficacy and control (Morisano et al., 2010). Second, acknowledging gender biases as a crucial contextual factor in our study population, we assumed that the distribution of saving devices (i.e., the zip purses) for individual usage might affect participants' gender attitudes. In line with the financial inclusion literature (e.g. Duvendack et al., 2014), we included five items on the roles and rights of men and women (boys and girls) in society to elicit participants' gender attitudes (e.g., *"Boys should be fed first and given more food compared to girls", "Daughters should have similar rights to inherited property as sons."*). Again, individual items were aggregated into a continuous scale based on PCA weighting. Third, we included a more direct measure of female empowerment. For this, we only collected data from the women in our sample. Items were drawn from Glennerster et al. (2018) and adapted to match the context of India (the two items were: *"Do you get into trouble for leaving the house without informing your husband or another household member?", "Do you get in trouble for making unescorted outings such as visiting your parents, friends, going to the market?"*). The female empowerment index was aggregated into a continuous scale based on PCA weighting.[6]

Fourth, we measured respondents' financial resilience to potential health or other emergencies. We relied on participants' self-reported capacity to cover the costs for

[6]Note that gender attitudes and female empowerment were listed as one overarching measure (female empowerment index) in the pre-analysis plan. However, factor analysis of individual items pointed to a two-factor solution, indicating two distinct underlying concepts, namely gender attitudes (reported by men and women) and female empowerment (reported by women only). We therefore decided to report these as separate outcomes in the subsequent analyses.

medical treatment and medicine if needed, and if they had experienced an income shock in the previous six months – their ability to financially cope with an emergency (e.g., *"Imagine an emergency would happen tomorrow. How difficult would it be for you and your family to find ten thousand rupees to cope with this emergency?", "Was there a time in the last 4 weeks when you needed to be admitted at the hospital but didn't because you didn't have enough money?"*). The individual items were aggregated based on PCA weighting, yielding a continuous scale ranging from lower to higher levels of financial resilience.

Fifth, assuming that accumulated savings could substitute for loans for investment purposes, we captured respondents' total outstanding debts. Lastly, we collected data on past-month expenditures for six food and seven non-food items (excluding temptation goods). We summed expenditures across these categories into total household expenditures in rupees. This more distal welfare indicator enabled us to examine whether higher savings rates may have beneficial downstream impacts and help alleviate poverty, for instance through more effective protection from economic shocks or returns from business or human capital investments that were realized through accumulated savings (Brune et al., 2016; Dupas and Robinson, 2013; Dupas et al., 2018). All outcome variables are described in further detail in Table A16 in the Appendix.

Lastly, we collected qualitative data to complement our quantitative findings and elucidate possible mechanisms of change. We conducted four focus group discussions with an average of ten participants per group. Three focus groups were conducted with treatment group participants (two gender-segregated and one mixed) and one

with control group participants (mixed). Discussion guides included open-ended questions prompting participants to reflect upon any changes (positive or negative) that they and their families had experienced as a result of receiving the lockbox, and, for treatment group participants, the zip purse. Discussion guides included several questions on intra-household decision-making and probed the dynamics between spouses that shape household financial management. In addition, our enumerators kept written records ("field journals") of their experiences and observations during home visits throughout the implementation and endline phase. Recordings from focus group discussions and enumerator observations were transcribed and translated to English and then coded using thematic analysis (Braun and Clarke, 2006). Qualitative statements that were conceptually similar were categorized into themes and then discussed and validated with a second coder. The qualitative findings can therefore point to prevailing patterns in participants' narratives of changes in their saving behavior.

2.6 Attrition

Sample attrition in this study was 6.8%. We tested for differential attrition using a linear probability model. First, we regressed an attrition dummy on treatment status alone. Then, we repeated this exercise by adding baseline controls. Lastly, we included interactions of treatment and baseline controls. We show that attrition was not associated with treatment status (see Table A1). However, attritors were on average more likely to be female, be employed, and have higher incomes.

2.7 Estimation Strategy

We estimated the average effect of being assigned to the treatment group, the intent-to-treat effect (ITT), on each outcome variable Y with the following regression:

$$Y_i = \alpha + \beta T_i + \gamma Y_{i(t-1)} + \delta S_i' + \epsilon X_i' + \omega_i \quad (1)$$

where T_i was an indicator variable for treatment assignment, equal to 1 if individual i had been assigned to receive the lockbox *and* the mobile saving device, $Y_{i(t-1)}$ the lagged outcome (at baseline), S_i' a vector of stratification variables, X_i' a vector of baseline covariates, and ω_i an error term. For each outcome, we estimated three different specifications: (1) a first specification using only the treatment assignment and stratifying variable as predictors, (2) a second specification including the lagged outcome $Y_{i(t-1)}$ to the previous specification, and (3) a third specification including additional baseline controls X_i' , namely participant age, marital status, educational level, employment, household size, and baseline poverty level. We conditioned on the baseline level of outcomes and additional controls in the ANCOVA specifications (2) and (3) to increase the statistical power and precision of the estimates. Our coefficient of interest was β, which indicated the intent to treat (ITT) effect.

We computed False Discovery Rate (FDR) adjusted q-values to account for multiple hypothesis testing. We utilized the Benjamini-Hochberg method, which is considered less conservative than simple Bonferroni adjustments and particularly suitable when working with a range of outcomes that are likely correlated (see Benjamini and Hochberg, 1995; Benjamini et al., 2006). Adjustments were made across primary and

secondary outcomes separately. In our results section, we report q-values corrected for multiple testing for each pre-specified outcome.

Lastly, we tested for heterogeneous effects based on the following specification:

$$Y_i = \alpha + \beta T_i + \theta TRAIT_i \times T_i + \eta TRAIT_i + \gamma Y_{i(t-1)} + \delta S_i' + \epsilon X_i' + \omega_i \qquad (2)$$

where $TRAIT_i$ was a baseline characteristic for which we hypothesized heterogeneity in the effectiveness of the treatment. These baseline variables were specified in the pre-analysis plan and included (i) participant sex, (ii) female involvement in household financial decision-making, (iii) present bias, and (iv) income levels. The average treatment effect for a subgroup of people with a certain trait (i.e. those below median income) was then given by the sum of the coefficients $\beta + \theta$ for that trait[7].

2.8 COVID-19 phone surveys

Between October and December 2020, we contacted our study sample for a post-hoc phone survey to assess how the COVID-19 pandemic and its socioeconomic repercussions had affected the living conditions of our participants and their families. The phone survey was designed to take less than 15 minutes and focused on measuring participants' total savings balances, while the other trial outcomes were no longer included. We drew on insights from this survey to assess long-term intervention effects - approximately 20 months after the delivery of our saving devices - on total savings balances and households' use of savings to cope with resource shortages induced by the pandemic. We report the findings from this last wave separately from

[7]Data and code are publicly accessible via https://osf.io/p8dj5/.

the pre-specified outcomes and exclude them from FDR adjustments.

In this additional round of data collection, we reached a subsample of 871 participants, representing 57% of our baseline sample. Attrition between baseline and phone surveys was not differential in terms of treatment arm (see Table A1, column (2)). However, participants who participated in the phone interviews were on average younger, more educated, and more likely to be employed and married. In addition, participants who could be reached for a phone interview had a slightly higher baseline savings rate relative to those who could not be reached.

3 Results

3.1 Summary statistics and orthogonality verification of randomization

Baseline characteristics of study participants are summarized in Table 1. Most participants reported being Hindus, 9.7% were Buddhists, and 5.4% were Muslims. More than 80% of our respondents were married and around 40% had not completed any form of primary or secondary schooling.

Almost half of the participants reported belonging to a backward caste or scheduled tribe and thus to a historically marginalized social group. Although discrimination against the so-called "untouchables" is prohibited by the Indian constitution, a person's caste is still a strong predictor of poverty levels today and continues to determine, for example, land ownership, access to public goods, and social capital (Gang

et al., 2008; Kapur Mehta and Shah, 2003; Lastrapes and Rajaram, 2016; Thorat and Madheswaran, 2018). In addition, one third of our respondents were unemployed and the average monthly income (across both study arms) was at 15,589.00 INR (equivalent to approx. 218.00 USD). However, baseline savings balances were already quite high. Almost 80% of our participants indicated holding some form of savings and the average total savings balance reported at baseline was at 7,660.00 INR (equivalent to approx. 96.00 USD). Savings were primarily held in formal bank accounts (approx. 60% of total savings) or post office accounts (approx. 10% of total savings).

At the same time, participants reported low baseline levels of temptation spending. Only 40% of participants reported having purchased a temptation good in the previous month and the average monthly amount spent on these was only 127.00 INR (equivalent to approx. 1.60 USD), i.e. less than 1% of respondents' monthly incomes. This was slightly lower than the rates reported for Uttar Pradesh and Bihar in India by Banerjee and Duflo (2007), who found a share of expenditures on alcohol and tobacco of around 3% of total consumption. However, this difference could be explained by higher average incomes in our sample. Despite this, it is important to note that we relied on participants' self-assessment of temptation spending and that social desirability or self-rationalization effects could have caused some underreporting. Another possible explanation for the relatively low levels of temptation spending was prompted by our enumerators' field notes, documenting that many female participants indicated that their husbands were mainly engaging in temptation spending, which they had little control over. Indeed, gender disparities in temptation spending were illustrated by our quantitative data: total baseline temptation expenditures were higher among male participants (average: 169.35 INR) relative to female

participants (average: 117.50 INR), however, they still represented only a small share of other consumption.

We used a joint orthogonality F-test to assess baseline balance across both study arms and confirmed randomization as effective (F=0.765). The treatment and control group were balanced along all but one characteristic (see Table 1, Column (3)). The only statistically significant difference was a slightly higher rate of married women in the treatment group (p=0.02).

Table 1. Baseline Balance

	Control (N=754)	**Treatment (N=771)**	**t-test Difference (1)-(2)**
Female	0.82	0.81	0.00
	(0.01)	(0.01)	
Age	35.59	36.37	-0.78
	(0.46)	(0.54)	
Scheduled/backward caste or tribe	0.47	0.45	0.02
	(0.02)	(0.02)	
Married	0.81	0.85	-0.05
	(0.01)	(0.01)	
Hindu	0.76	0.78	-0.02
	(0.02)	(0.02)	
Household members	5.01	4.97	0.04
	(0.08)	(0.09)	
Unemployed	0.33	0.32	0.01
	(0.02)	(0.03)	
No education	0.22	0.22	0.00
	(0.02)	(0.02)	
Completed primary education	0.20	0.21	0.00
	(0.02)	(0.01)	
Completed secondary education	0.36	0.39	-0.03
	(0.02)	(0.02)	
Completed tertiary education	0.21	0.19	0.02
	(0.02)	(0.01)	
Past-month income (INR)	17,968.91	13,262.10	4,706.80
	(3,941.90)	(1,754.74)	
Total savings balance (INR)	7,381.29	7,933.16	-551.87
	(1,211.84)	(1,168.38)	
...% Formal savings	*0.71*	*0.70*	
...% Savings at home	*0.10*	*0.11*	
...% Savings in club/group	*0.09*	*0.10*	
...% Savings with relatives	*0.10*	*0.09*	
Past-month temptation spending (INR)	128.06	126.14	1.92
	(23.11)	(16.84)	
Temptation index	1.47	1.44	0.03
	(0.03)	(0.02)	
Self-efficacy index	6.12	6.15	-0.03
	(0.09)	(0.09)	
Female empowerment index	0.02	-0.02	0.04
	(0.04)	(0.04)	
Outstanding debt (INR)	8,810.23	10,651.75	-1,841.52
	(1,994.31)	(2,736.34)	
HH expenditures (selected goods)	5,088.60	5,491.69	-403.09
	(398.14)	(435.67)	
Resilience index	0.27	0.28	-0.01
	(0.02)	(0.02)	
F-test of joint significance (F-stat)			0.765
F-test, number of observations			1,525

Notes: Standard errors in parentheses.

3.2 Impacts on primary outcomes

Results for the study's primary outcomes, total savings balances and total temptation expenditures are reported in Table 2. The ITT effects on total savings balances are shown in Table 2, Column (1) Panel A. In the control group, the average amount of total savings at endline was at 8,400.83 INR, compared to 15,109.09 INR in the treatment group. The treatment effect estimated in specification III of Column (1) corresponds to an 81% increase in total savings and is statistically significant at the 5%-level. This effect remains statistically significant at the 10%-level after adjusting for multiple hypothesis testing. In Panel B, we show that the treatment effect was also robust to 1% high tail winsorizing of the savings variable. However, its magnitude was diminished, corresponding to a 19% higher total savings balance when comparing the treatment to the control group. While this latter result is less sensitive to observations in the upper tail of the distribution of the savings variable, it is important to emphasize that none of the "outliers" seemed to be driven by measurement error. That is, the largest observed savings amounts were mainly composed of savings held in the experimental devices and the exact amount of these was hand-counted and cross-checked against the photographs.

In Table A2, we present estimates for past-month deposits and withdrawals, summed up across all saving sources (e.g. deposits to and withdrawals from experimental devices, bank accounts, home storage, etc.). The results indicate that the treatment effect mostly takes the form of large and statistically significant decreases in withdrawals. Interestingly, the coefficient for deposit rates was also negative (but not

Table 2. ITT Estimates: Primary Outcomes

	(1) Total Savings Balance (INR)			(2) Temptation Expenditures		
	I	II	III	I	II	III
Panel A: Raw Outcome						
ITT: Received Program	6,460.12	6,555.50	6,802.97	1.37	1.34	2.65
	(3084.84)	(3,085.91)	(3,351.22)	(14.91)	(14.91)	(15.25)
	[0.072]	[0.068]	[0.086]	[0.927]	[0.929]	[0.862]
Lagged Outcome		0.09	0.10		-0.00	-0.00
		(0.05)	(0.05)		(0.00)	(0.01)
Mean Control Group	8,400.83	8,400.83	8,400.83	82.27	82.27	82.27
	(20,463.25)	(20,463.25)	(20,463.25)	(262.25)	(262.25)	(262.25)
Panel B: Winsorized Outcome						
ITT: Received Program	1,452.30	1,473.75	1,472.05	-2.95	-2.94	-1.66
	(878.58)	(883.10)	(881.36)	(11.52)	(11.53)	(11.72)
Lagged Outcome		0.27	0.28		-0.00	-0.01
		(0.06)	(0.07)		(0.02)	(0.02)
Mean Control Group	7,886.97	7,886.97	7,886.97	76.59	76.59	76.59
	(14,822.03)	(14,822.03)	(14,822.03)	(219.67)	(219.67)	(219.67)
Stratification variables	yes	yes	yes	yes	yes	yes
Lagged outcome	no	yes	yes	no	yes	yes
Controls	no	no	yes	no	no	yes
Observations	1,421	1,421	1,379	1,421	1,421	1,379

Notes: Multiple hypothesis corrected q-values in square brackets (only for raw outcomes). Robust standard errors in parentheses. Model I includes trial arm and stratification variables, namely participant sex and baseline savings. Model II includes stratification variables (same as Model I) and the lagged outcomes. Model III includes additional controls: participants' age, marital status, educational status, employment, household size, household income. Panel A reports coefficients for the raw outcome. Panel B displays coefficients for 1% high tail winsorizing of the outcome variable. These are, in Column (1) total savings balances, i.e. the total sum of savings (in INR) held in the zip purse (if applicable), lockbox, formal bank account/post office, mobile money account, savings circle, elsewhere at home, and with relatives. The outcome in Column (2) is the total self-reported sum of expenditures (in INR) on individually-defined temptation goods.

statistically significant), suggesting that treatment group participants, when compared to control group participants, put either less frequent and/or smaller amounts of money into their saving devices. We might tentatively interpret this as a stronger binding force associated with the portable saving device. Against this backdrop, participants in the treatment group may ascribe stronger feelings of regret to the idea of making withdrawals from their zip purse (or box) and decide to only deposit money that they do not plan to withdraw.

In Table A3, we assessed whether there was crowding-out from one saving type to another. Overall, we found higher rates of formal savings (i.e. savings held in a bank account, post office, or in the form of mobile money) in the treatment group. In contrast, coefficients for informal savings balances - including savings in the box, at home, or in a savings circle - were negative (although not statistically significant) and the coefficient for savings kept with relatives was imprecisely estimated. Therefore, the decomposition suggests that higher total savings in the treatment group were largely driven by higher savings in formal bank accounts. It appears that treatment group participants were not closely following the facilitator instructions to regularly transfer money from their zip purse into their lockbox but rather deposited the additional money into their bank accounts. We speculate that participants preferred transfers to their bank account over the shared lockbox for reasons of safety or lower visibility and accessibility for other household members. Corroborating this, our collected process data (not shown in table) revealed that 50% of treatment group participants indicated using their zip purse on a regular basis[8] but that only 24% of those had shifted money from their purse to their box on a weekly or more regular basis. While general usage of the lockbox was high in both study groups (91% in the treatment group and 93% in the control group), participants in the treatment group reported less frequent deposits on average. Thus, the percentage of study participants who had not deposited any money into their lockbox in the previous month was five percentage points higher in the treatment group (24% vs. 19%).

Moving to the other primary outcome, namely last month temptation expenditures,

[8]"On a regular basis" was determined by asking participants whether they have carried the purse with them when they last left their house to run some errands.

we found ITT effect estimates close to null. While we observed a general reduction (in the full study sample) in temptation spending between base- and endline, there were no statistically significant differences between study arms at endline. This finding was corroborated by an alternative measurement of temptation spending: in Table A4, we provide ITT effect estimates for a self-reported temptation index[9] and, again, observe coefficients close to zero. Coupled with the descriptive evidence on the overall low prevalence of temptation spending among the study population, these findings put into question our key hypothesis that the device can increase savings by helping to reduce temptation spending.

3.3 Impacts on secondary outcomes

In the next step, we examined whether the portable saving device had positive impacts on any of the pre-specified secondary outcomes. Results are presented in Table 3 and Table 4. First, as shown in Table 3, Column (1), we observed no statistically significant intervention effect on participants' gender attitudes and the difference between study arms was effectively zero. However, albeit non-causal, we observed improvements in gender attitudes over time in both study groups equally. There was a 32 percentage-point improvement on the principal-component-weighted gender attitudes index when comparing baseline to endline measures. To put this into more meaningful terms, at baseline, 88% of participants agreed with the statement that girls and boys should have equal inheritance rights, and at endline 95% of participants

[9]The index was composed of the following three items, rated on a 1-5 Likert scale: *"In the past month, I spent money on things that I did not really need", "...I bought something and later regret that I did", "... I found it difficult to really control how I spend my money".*

agreed with this statement. Similarly, at baseline, 12% had agreed that boys should be given more food relative to their sisters, compared with only 8% at endline. Our study design does not allow us to establish whether these improvements were a direct result of our savings intervention – and its active integration of women into household financial management – or whether they were linked to other co-occurring but unobserved factors.

We further examined whether having the portable saving device improved women's actual empowerment, as captured by women's self-reported autonomy and rights in relation to their husband or partner. For this analysis, we focused exclusively on the sub-sample of female respondents. As documented in Table 3, Column (2), we found statistically significant improvements in the female empowerment index among participants in the treatment group. The treatment effect corresponded to a 17% increase in female empowerment. To put this in more intuitive terms, 26% of women in the control group indicated that they would face negative consequences if they did not inform their husband/partner about leaving the house. In the treatment group, only 20% of women anticipated such consequences. Likewise, in the control group, 78% of women were allowed to leave their homes unescorted, compared to 82% in the treatment group. The overall treatment effect was statistically significant on the 5%-level when using naïve p-values but lost significance (p=0.11) after FDR adjustments.

In Column (3) of Table 3, we present treatment effects on financial self-efficacy. We found no statistically significant differences in financial self-efficacy between study arms, and coefficients were again very small in magnitude. Interestingly, we also did

not find any changes in self-efficacy levels in either of the groups from baseline to endline. These null findings contradict evidence in previous literature (e.g. Steinert et al., 2018a, 2020) - suggesting that financial self-efficacy is a central mediator for changes in financial behavior, such as improvements in savings behavior.

Table 3. ITT Estimates: Secondary Outcomes I

	(1) **Gender Attitudes Index (full sample)**			(2) **Female Empowerment Index (women only)**			(3) **Self-Efficacy Index**		
	I	II	III	I	II	III	I	II	III
ITT: Received Program	-0.00	0.00	-0.02	0.05	0.04	0.04	0.03	0.04	0.04
	(0.12)	(0.12)	(0.12)	(0.02)	(0.02)	(0.02)	(0.13)	(0.12)	(0.12)
	[0.973]	[0.977]	[0.889]	[0.144]	[0.171]	[0.114]	[0.946]	[0.938]	[0.889]
Lagged Outcome		0.09	0.07		0.20	0.17		0.19	0.17
		(0.02)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)
Mean Control Group	-0.24	-0.24	-0.24	-0.23	-0.23	-0.23	6.00	6.00	6.00
	(2.33)	(2.33)	(2.33)	(0.37)	(0.37)	(0.37)	(2.41)	(2.41)	(2.41)
Stratification variables	yes	yes	yes	yes	yes	yes	yes	yes	yes
Lagged outcome	no	yes	yes	no	yes	yes	no	yes	yes
Controls	no	no	yes	no	no	yes	no	no	yes
Observations	1,420	1,417	1,375	1,186	1,176	1,134	1,415	1,409	1,368

Notes: Multiple hypothesis corrected sharpened q-values in square brackets. Robust standard errors in parentheses. Model I includes trial arm and stratification variables, namely participant sex and baseline savings. Model II includes stratification variables (same as Model I) and the lagged outcomes. Model III includes additional controls: participants' age, marital status, educational status, employment, household size, household income. Column (1) displays a principal-component weighted gender attitudes index (item sample: "Boys should be fed first and given more food compared to girls."*), Column (2) displays a principal-component weighted female empowerment index (item sample:* "Do you get in trouble for leaving the house without informing your husband?"*, based on female-only reports, Column (3) displays a principal-component weighted self-efficacy index (item sample:* "When I get what I want, its usually because I worked hard for it").

In Table 4, we report ITT estimates for outcomes that were more distal and thus more reflective of potential changes in participants' economic welfare. First, we assessed participants' financial resilience to unforeseen emergencies such as health shocks. ITT effects were not statistically significant and effectively zero, indicating identical resilience levels in both study arms. Similar to other outcomes, participants in both groups reported substantial improvements in economic resilience from baseline to endline, corresponding to a 59 percentage-point improvement on the resilience index. In focus group discussions and during some of the home visits, participants also reported that they had used savings from their lockbox to manage an emergency. Thus

while there might have been improvements in financial resilience overall, the portable saving device did not seem to have any add-on effect.

We also examined ITT program effects on participants' self-reported total debts from any out-standing loans. We found statistically significant lower levels of debt among participants who had received both the lockbox and the zip purse. Treatment group participants reported a reduction in total outstanding debts of 30% (see Table 4, Column (2), Panel A). This suggests that the demand for loans may have been partly substituted by higher savings rates among treatment group participants, enabling them to make investments or respond to emergencies without resorting to external money lenders. The treatment effect was significant when using naïve p-values (p=0.038) but narrowly failed to reach significance after FDR adjustments (p=0.11).

Lastly, to assess the potential positive downstream impact of higher savings (see Brune et al., 2016; Dupas and Robinson, 2013; Dupas et al., 2018), we estimated program effects on past-month household expenditures (excluding those on temptation goods). The ITT coefficient in Column (3) of Table 4 was positive but not statistically significant and imprecisely estimated with relatively large confidence intervals. Thus, while we could not confirm any substantial downstream impacts of higher savings, we could at least rule out that higher savings rates were achieved by reducing spending on essential consumption goods.

Table 4. ITT Estimates: Secondary Outcomes II

	(1) Resilience Index			(2) Outstanding Debt (INR)			(3) Household Expenditure (INR)		
	I	II	III	I	II	III	I	II	III
Panel A: Raw Outcome									
ITT: Received Program	-0.01	-0.01	-0.01	-498.81	-506.13	-579.65	498.91	458.72	635.62
	(0.02)	(0.02)	(0.02)	(270.40)	(266.18)	(278.87)	(1631.67)	(1653.83)	(1688.83)
	[0.946]	[0.938]	[0.889]	[0.195]	[0.171]	[0.114]	[0.946]	[0.938]	[0.889]
Lagged Outcome		0.22	0.19		0.01	0.01		0.08	0.08
		(0.03)	(0.03)		(0.01)	(0.01)		(0.08)	(0.08)
Mean Control	0.44	0.44	0.44	1,956.51	1,956.51	1,956.51	5,767.05	5,767.05	5,767.05
	(0.44)	(0.44)	(0.44)	(5,319.35)	(5,319.35)	(5,319.35)	(29,496.46)	(29,496.46)	(29,496.46)
Panel B: Winsorized Outcome									
ITT: Received Program	/	/	/	-439.38	-423.29	-509.97	138.46	112.47	115.84
				(187.77)	(186.49)	(190.58)	(281.16)	(282.11)	(280.44)
Lagged Outcome					0.02	0.02		0.09	0.08
					(0.01)	(0.01)		(0.03)	(0.04)
Mean Control				1,754.81	1,754.81	1,754.81	4,041.76	4,041.76	4,041.76
				(3,796.90)	(3,796.90)	(3,796.90)	(5,138.87)	(5,138.87)	(5,138.87)
Stratification variables	yes	yes	yes	yes	yes	yes	yes	yes	yes
Lagged outcome	no	yes	yes	no	yes	yes	no	yes	yes
Controls	no	no	yes	no	no	yes	no	no	yes
Observations	1,421	1,420	1,378	1,421	1,421	1,379	1,421	1,421	1,379

Notes: Multiple hypothesis corrected sharpened q-values in square brackets. Robust standard errors in parentheses. Model I includes trial arm and stratification variables, namely participant sex and baseline savings. Model II includes stratification variables (same as Model I) and the lagged outcomes. Model III includes additional controls: participants' age, marital status, educational status, employment, household size, household income. Panel A reports coefficients for the raw outcome. Panel B displays coefficients for 1% high tail winsorizing of the outcome variable. Column (1) displays a principal-component weighted resilience to economic shocks index (item sample: "Imagine an emergency would happen tomorrow. How difficult would it be for you and your family to find ten thousand INR to cope with this?"*), Column (2) displays the self-reported total sum of debt from any outstanding loans (in INR), Column (3) displays the self-reported total past-month household expenditures on selected food and non-food items (in INR).*

3.4 Heterogeneity in treatment effects

This section examines heterogeneity in treatment effects based on four pre-specified observable characteristics, namely participants' sex, baseline income level, present bias, and women's involvement in household financial decision-making. After adjusting for multiple hypothesis testing, we found no statistically significant evidence for heterogeneity in treatment effects. We elaborate on a few tentative points below and report the regression results in Tables A5-A13.

Treatment effects on primary outcomes did not vary between men and women (see Table A5). The treatment effect on total savings lost statistical significance for participants in the lowest income quantile (see Table A6). It is possible that these indi-

viduals did not have surplus income to save in their portable devices. With regards to present bias, we can cautiously interpret our results as pointing to lower treatment effects on total savings for participants with a higher level of present bias (see Table A7). This finding suggests that our portable saving device was not an effective tool to enhance self-control among more impatient participants. Lastly, there appears to be some heterogeneity in both primary outcomes with regards to women's involvement in financial decisions (Table A8).

As reported in Tables A9-A13, we did not find any substantial heterogeneity in treatment effects on secondary outcomes. The only noteworthy finding was that debts were more substantially reduced for female participants.

3.5 COVID-19 phone survey: Long-term intervention impacts

We reached 871 participants for a COVID-19 follow-up phone survey. Of these, 63% had lost their source of income in the pandemic and 85% reported earning less income than prior to the pandemic. In addition, 10% of the survey participants noted having experienced coronavirus-related symptoms (e.g., cough, fever, loss of sense of taste) and 4% had tested positive for COVID-19. 72% of the phone survey participants indicated having received dry ration kits and 13% indicated having received cash transfers to their Prandhan Mantri Jan Dhan Yojana accounts since the beginning of the pandemic. We added income losses, COVID-19 symptoms, and receipt of emergency support as control variables in the third regression specification of the

following analyses.

Table 5 summarizes the long-term treatment impacts on total saving balances and participants' coping behavior for the phone survey subsample. Corroborating the pre-specified results from above, we still observed significantly higher savings balances among participants in the treatment arm twenty months after distribution of the portable saving device. Specifically, the total savings balance in the treatment group was 45% higher than in the control group. With 1% high tail winsorizing of the outcome variable, the effect coefficient was still positive but failed to reach significance and translated into a smaller effect size of 15% higher total savings in the treatment arm. In addition, we found evidence to suggest that participants were able to use their higher savings to weather the economic shock induced by the COVID-19 pandemic. Specifically, participants in the treatment group were six percentage points more likely ($p<0.10$) to draw on their savings to cope with resource shortages and loss of sources of livelihood. This might have helped them to avoid potentially harmful coping practices such as reducing spending on food, healthcare, and children's education, selling assets, or borrowing money at high interest rates.

4 Discussion of potential mechanisms

Our study showed that the portable saving device was effective in increasing participants' total savings. The intervention, however, did not have any impacts on participants' temptation spending. Similarly, an earlier field experiment in Malawi tested the impact of a commitment intervention based on savings defaults and no

Table 5. ITT Estimates: Long-Term Outcomes (COVID-19 Survey)

	(1) **Total Savings Balance (INR)**			(2) **Savings to cope with COVID-19**		
	I	II	III	I	II	III
Panel A: Raw Outcome						
ITT: Received Program	6,686.42	6,515.92	6,704.43	0.05	0.05	0.06
	(3,637.61)	(3,629.42)	(3,697.56)	(0.03)	(0.03)	(0.03)
Lagged Outcome		0.25	0.25		0.00	0.00
		(0.07)	(0.08)		(0.00)	(0.00)
Mean Control Group	14,943.26	14,943.26	14,943.26	0.65	0.65	0.65
	(8,337.04)	(8,337.04)	(8,337.04)	(0.02)	(0.02)	(0.02)
Panel B: Winsorized Outcome						
ITT: Received Program	1,581.87	1,417.44	1,482.10	/	/	/
	(1,212.35)	(1,167.30)	(1,226.15)			
Lagged Outcome		0.24	0.24			
		(0.05)	(0.05)			
Mean Control Group	9,848.71	9,848.71	9,848.71			
	(2,439.98)	(2,439.98)	(2,439.98)			
Stratification variables	yes	yes	yes	yes	yes	yes
Lagged outcome	no	yes	yes	no	yes	yes
Controls	no	no	yes	no	no	yes
Observations	871	871	831	871	871	831

Notes: (1) Total savings balance includes savings held in experimental devices (box/purse) and savings held in a bank account, savings group, post office, or with relatives. (2) Used Savings to Cope with COVID-19 shock coded 1 if households indicated that they drew on their savings to cope with resource shortages rather than cutting down expenses on health, education, food, selling assets, or taking on loans. Model I includes trial arm and stratification variables, namely participant sex and baseline savings. Model II includes stratification variables (same as Model I) and the lagged outcomes. Model III includes additional controls: participants' age, marital status, educational status, employment, household size, household income, COVID-19 symptoms, income lossess due to the pandemic, and receipt of governmental support due to the pandemic. Panel A reports coefficients for the raw outcome. Panel B displays coefficients for 1% high tail winsorizing of the outcome variable.

changes in temptation spending patterns were observed there either (Brune et al., 2017). Our evaluation of secondary trial outcomes suggests that the higher savings amounts in the treatment group were not driven by increased debt or reductions in general household expenditures either (see Table 4). It thus remains crucial to examine what is driving the treatment effects so as to better understand where the higher savings actually come from. We draw on additional quantitative data on the reported

device usage and related perceptions, and on qualitative data from four focus group discussions and enumerators' field observations. Motivated by prior literature, we explore three channels through which the portable saving device might have worked.

Self-control channel

One possible channel is **enhanced self-control** linked to the (soft) commitment function of the zip purse. If this channel were active, we would expect the portable saving device to foster participants' self-discipline, willpower, and intrinsic motivation, helping them to resist their impulses and temptations (Ashraf et al., 2006; Bénabou and Tirole, 2004; Bryan et al., 2010; Soman and Zhao, 2011). While numerous previous studies have presented evidence on the effectiveness of lockboxes as a commitment tool (Aker et al., 2020), the central question of our research design refers to the add-on effect of the portable device. We argue that the zip purse could have an additional commitment effect in two ways. Firstly, the portable saving device allowed for physical segregation of the liquid cash that participants carried with them and could have amplified mental accounting mechanisms in their day-to-day budgeting decisions. Thaler (1990) argues, based on ethnographic evidence, that money is perceived as less fungible if it is mentally earmarked for a specific purpose (Bénabou and Tirole, 2004). By distributing portable saving devices, we may have helped materialize these mental rules: we allowed participants to keep "money to spend" physically separate from "money earmarked for savings". In line with this logic, treatment group participants may have considered money in their zip purse as explicitly "reserved" for savings and abstained from spending it (Dupas and Robinson, 2013; Karlan et al., 2014; Shafir and Thaler, 2006). Secondly, the portable saving device was designed to

create temporal concurrency between the psychological commitment effect and actual spending decisions. Since the portable device was supposed to be carried during the day, it was likely physically present when most spending decisions occurred. Salience of its commitment function was thus more pronounced in comparison to the lockbox, which was kept at home and thus more distant and abstract. We could therefore assume that any deviations from saving intentions would have *instantly* induced negative emotions and feelings of regret. The perceived psychological costs for a person who had carried the portable saving device would in turn be higher (Shefrin and Thaler, 1992).

We empirically assessed this first channel and found little evidence of a self-control pathway. In our qualitative data, we revealed some narratives that pointed to the self-control function of the zip purse (e.g., *"When I go shopping or some other work, I keep the change in the purse instead of spending it here and there. So I don't buy unnecessary things because of the purse"*, see Table 7, Column (1)). It is also important to note that no participant mentioned any harmful impacts of having received the zip purse - such as increased spending due to greater availability of money. However, it is possible that these accounts were more suggestive of a certain priming or social desirability effect, whereby participants described their usage and perception of the portable device closely in line with the instructions they were given on their intended usage during the intervention delivery. More importantly, the absence of a treatment effect on temptation expenditures may suggest that the zip purse was not primarily used as a commitment device to reduce spending and keep money as savings. Accordingly, in additional exploratory analyses (which remain subject to endogeneity concerns), we show participants who perceived the zip purse as a means of self-control

did not report significantly higher savings amounts (see Table A14, Column (1)).

Reminder channel

A second channel could have been a **reminder function** of the portable saving device. Participants were instructed to carry their portable device during the day to ensure that it was physically present (and possibly visible) whenever spending decisions were made. Accordingly, the zip purse could have been a saving reminder and brought participants' saving intentions "to the top of their mind" (Karlan et al., 2016). Several previous studies have corroborated this argument, demonstrating how reminders have helped to increase savings rates (Kast et al., 2018; Akbas et al., 2016; Karlan et al., 2016). For instance, Karlan et al. (2016) found that participants who received reminder messages were more likely to reach their individual savings goals and held higher savings amounts in their bank accounts at post-test. Similarly, another randomized controlled trial illustrated how feedback text messages that informed participants about their own and their peers' saving performance almost tripled weekly deposit amounts in the treatment arm (Kast et al., 2018).

However, our empirical tests did not reveal a reminder channel. There were very few indications in the qualitative participant accounts of a distinct reminder effect of the portable saving device. Only two of 40 focus group participants made some reference to the device's reminder function and none of the ten enumerators identified this channel in their field observations (see Table 7, Column (2)). Likewise, while we reported higher savings amounts, on average, for treatment group participants who were more inclined to perceive their zip purse as a savings reminder, the coefficient

did not achieve statistical significance (see Table A14, Columns (2) and (4)). This finding is in line with a recent field experiment in Niger in which receipt of a lockbox increased participants' savings amounts while there was no add-on effect for the treatment group who received SMS reminders on top (Aker et al., 2020). While take-up and usage of the box were high, only 20% of SMS recipients even remembered receiving a text message.

Hiding channel

The last channel we tested for was the conversion of the zip purse into a tool to **hide money from others**. This was motivated by our field observations (and hand count data) suggesting that treatment group participants had not always moved the savings from their zip purse into their lockbox – even though the program facilitator had instructed them to do so when delivering the devices. We found that many participants had retained a relatively large share of their savings in their purses instead, which they regularly deposited into a private bank account.[10] The portable saving device may thereby have helped participants to a) reduce the perceived liquidity of cash, b) turn down social requests on any unspent cash, and c) shield transfers and deposits of savings from a spouse or another household member. This interpretation is motivated by a body of literature that points to a high social demand on individuals' disposable incomes, which constrains their capacity to accumulate savings (Jakiela and Ozier, 2016; Riley, 2019; World Bank, 2015). Accordingly, disposable income is often subject

[10]In 2014, the Indian government has launched a large-scale financial inclusion program, which was primarily targeted at female beneficiaries (see Field et al., 2021; Singh et al., 2019). As part of this, more than 400 million Pradhan Mantri Jan-Dhan Yojana (PMJDY) bank accounts have been opened by 2021. The accounts are held by individuals and proof of identity must be shown to access the account.

to social appropriation mechanisms and webs of reciprocal social obligations (Boltz et al., 2019; Fiala, 2017). Individuals experience social pressure to share disposable income with their spouses, other family members or friends in need (Riley, 2019; Ambec and Treich, 2007). Based on a lab experiment in rural Kenya, Jakiela and Ozier (2016) estimated a kin tax of 4% of the observable income. In another experiment, Goldberg (2017) observed that farmers in Malawi who received a windfall payment in public rather than in private had higher expenditures in the week immediately after the experiment, which suggests that money was spent more quickly in order to circumvent social sharing obligations. The consequences of such social pressures are twofold: individuals may either meet their social obligations and simply undersave or spend all liquid cash quickly so that demands from family members, friends or neighbours can be denied (Brune et al., 2021).

Social obligations can also be a defining feature of financial management dynamics between spouses. For instance, Schaner (2015) conducts a field experiment in Kenya to evaluate a model of non-cooperative household savings behavior. She reveals that women are willing to accept economic utility losses (here in the form of lower interest rates) in exchange for holding an account of their own, rather than a joint account with a husband who "will simply withdraw all her savings and spend the funds on current consumption" (ibid, p. 136). Similarly, Anderson and Baland (2002) argue that women tend to join Rotating Savings and Credit Associations (ROSCAs) with the intention of protecting their savings from their husbands' immediate consumption practices. In addition, several lab-in-the-field experiments suggest that participants choose income-hiding over profit maximization when allocating experimental endowments (Almås et al., 2018; Castilla, 2019; Jakiela and Ozier, 2016). While both men

and women may have concealment motives, we can assume that demand for private saving devices is particularly high among women with low hierarchical status and - in consequence - substantially constrained (financial) decision-making power vis-a-vis a spouse (Dupas and Robinson, 2013; Fiala, 2017).

Our qualitative findings are strongly in line with the mechanisms described above. First, enumerators documented ample field observations on the social barriers that many participants faced around saving – and particularly female participants with regards to their spouse. For example, our field enumerators noted: *"She [female participant] said her husband would spend all her savings if he got information about it"*, or *"When it was told that the savings box is for entire family, one female participant asked: 'what if I save the money by hook or by crook and my husband takes it away for drinking alcohol?'"* (see Table 7, Column (3)). In our focus group discussions, we also identified more direct narratives about how women had used their portable saving device to hide money from their husbands. In fact, these hiding motives emerged as the most salient theme throughout the discussion, and were mentioned by two-fifth (21 out of 50) of all focus group participants. Accordingly, one treatment group participant noted: *"I keep some money with me in the purse. If he [husband] wants money, he takes it from the box or asks me. I give him the money from the box but he does not know that I have more money with me in my purse"*(see Table 7, Column (3)).

We conducted additional analyses based on our quantitative data to corroborate the qualitative evidence that hiding motives are at play in our target population. These analyses (see Table 6) revealed that participants in the treatment group reported

lower past-month transfers of money to other household members. More specifically, treatment group participants shared 35% less money than participants in the control group. This is in line with findings from another field experiment in Senegal. In this experiment, individuals were given the opportunity to hide money in order to escape redistributional pressures and decreased the share of their gains to kin by 27% (Boltz et al., 2019). While we did not observe the same reduction with regards to financial transfers to people outside participants' homes, this finding still suggests that the portable saving device had likely helped treatment group participants to keep their saved money for themselves, rather than giving it away to their spouse or to other household members. Further, building on these findings, we conducted additional heterogeneity analyses to examine whether treatment effects on total savings differed between participants who reported having transferred money to other household members at baseline. Our findings suggest that participants who were exposed to such social obligations appeared to have benefited more (although this was not statistically significant) from the zip purse relative to participants who had not reported any transfers to their kin (see Table A15). Lastly, as Table A14, Columns (3)-(4) indicate that higher savings amounts among participants were correlated with characterizations of the zip purse as a hiding tool.

Table 6. ITT Estimates: Responding to Social Demand

	(1) **Past-month transfer to a household member (INR)**			(2) **Past-month transfer to a person outside the household (INR)**		
	I	II	III	I	II	III
Panel A: Raw Outcome						
ITT: Received Program	-395.27	-397.49	-409.16	222.23	222.50	204.26
	(151.36)	(150.76)	(162.34)	(202.28)	(202.56)	(204.59)
Lagged Outcome		0.00	0.00		-0.00	-0.00
		(0.01)	(0.01)		(0.00)	(0.00)
Mean Control	1,185.53	1,185.53	1,185.53	340.98	340.98	340.98
	(3,030.57)	(3,030.57)	(3,030.57)	(2,269.30)	(2,269.30)	(2,269.30)
Panel B: Winsorized Outcome						
ITT: Received Program	-372.68	-360.97	-378.44	39.87	42.49	30.51
	(113.45)	(111.19)	(115.71)	(70.87)	(71.05)	(72.30)
Lagged Outcome		0.20	0.20		0.05	0.04
		(0.04)	(0.04)		(0.04)	(0.04)
Mean Control Group	1,091.78	1,091.78	1,091.78	254.33	254.33	254.33
	(2,445.68)	(2,445.68)	(2,445.68)	(1,242.65)	(1,242.65)	(1,242.65)
Stratification variables	yes	yes	yes	yes	yes	yes
Lagged outcome	no	yes	yes	no	yes	yes
Controls	no	no	yes	no	no	yes
Observations	1421	1421	1379	1421	1421	1379

Notes: Robust standard errors in parentheses. Model I includes trial arm and stratification variables, namely participant sex and baseline savings. Model II includes stratification variables (same as Model I) and the lagged outcomes. Model III includes additional controls: participants' age, marital status, educational status, employment, household size, household income. Outcomes are the past-month transfers of money (in INR). Panel A reports coefficients for the raw outcome. Panel B displays coefficients for 1% high tail winsorizing of the outcome variable.

We return to our above question: Where do the additional savings actually come from? The use of the portable device for concealment purposes is consistent with the treatment effects on primary and secondary outcomes (see section 3.2 and section 3.3). Specifically, we reported substantial increases in total savings at endline, whereas we found neither reductions in temptation expenditures nor in other past-month expenditures or increases in levels of debt. We contend that prior to the intervention, any untapped financial resources of participants, rather than being saved, were spent by spouses or other household members. It follows that our participants have increased

their own savings because the intervention provided them with a private saving device that enabled them to protect their financial resources against social demands. The increased savings are thus a direct function of significant reductions in monetary transfers to other household members (see Table 6).

Table 7. Qualitative Evidence

Self-Control Channel (quoted by 11/50)	**Reminder Channel (quoted by 2/50)**	**Hiding Channel (quoted by 21/50)**
Improved self-control: *"When I go shopping or some other work, I keep the change in the purse instead of spending it here and there. So I don't buy unnecessary things because of the purse."* (female participant, FGD2) *"When I used to go in the market, I couldn't resist myself buying those things. Which were not much useful. But after you told me the importance of saving, I realized that the money I am going to save in the purse you had given us, would turn out to be useful in crunch times."* (female participant, FGD2) *Women used the purse and they benefitted from it also. They told us that they used it to put money in it. Some told us that they fought with their desires because they had decided to save money."* (enumerator field observation) *"I have learnt to resist myself. If I resist myself now, then I can use the same amount of money in the future."* (male participant, FGD4)	***Savings reminder:*** *"Earlier I used to buy vegetables and put the remaining money somewhere. I never saved it. But now I see the purse and put the remaining amount from the purse in the box."* (female participant, FGD4) *"Earlier I used to buy anything I saw. Now I don't because the purse is there."* (female participant, FGD4)	***Hiding tool:*** *"I keep some money with me in the purse. If he [husband] wants money he takes it from the box or asks me. I give him the money from the box but he does not know that I have more money with me in my purse."* (female participant, FGD2) *"I always keep money in the purse so that I can keep it for myself."* (female participant, FGD3) *"My husband can take money from the box. But not from the purse."* (female participant, FGD2) *"Many women saved money without telling it to their family and mostly their husband. They said that if their husbands got to know about their savings, they will spend it."* (enumerator field observation) ***Spousal control:*** *"They also have to take care of not disclosing the amount to their family members so that they won't face any trouble from their family members. That is why not disclosing their savings is one of their priorities."* (enumerator field observation)

Table 7 – continued from previous page

Self-Control Channel	Reminder Channel	Hiding Channel
Mental Accounting Effect: *"It was useful. What I do is, I put small purse inside the big one. When I go out, I put my remaining money in it and when I come back, I put it in the box."* (female participant, FGD4) *"I keep the money aside in it for medical expenses."* (male participant, FGD4) *"I kept the notes in the purse and the coins in the box. So both of them proved to be useful."* (female participant, FGD4)		*"While I was interviewing another household, she came and said she lied about her saving because of her husband's presence. She said her husband would spend all her savings if he got information about it."* (enumerator field observation) *"Some households have so much of patriarchy that women in those households are interested in savings but the son/husband is not ready for it."* (enumerator field observation) *"Lot of men tortures their wives. So when I asked them, they told me that they couldn't keep the boxes at their homes as they live in small place. Their husband might take money from the box."* (enumerator field observation) *"When it was told that the savings box is for entire family, one female participant asked: 'what if I save the money by hook or by crook and my husband takes it away for drinking alcohol?'"* (enumerator field observation)

Notes: Counts based on 40 focus group participants and ten enumerators. Quotes were translated from Marathi into English.

5 Conclusion

We set out to test the effectiveness of a portable saving device, which was distributed to low-income slum dwellers in India's Maharashtra province. Based on a randomized control trial design, we revealed causal impacts on participants' total savings amounts while their temptation expenditures remained unchanged. We can therefore infer that the portable device worked – but that its purpose and usage diverged from our initial hypothesis in our pre-analysis plan. That is, rather than serving as a portable commitment device that helped increase participants' self-control, the device seems to have been used primarily to conceal individual savings and financial resources from other household members.

A limitation is the reliance on self-reported data for measures other than savings held in the lockbox and in the zip purse. This might be particularly problematic for the outcome of temptation spending, which is likely prone to social desirability biases due to the cultural and social stigma attached to alcohol consumption, smoking and gambling. However, while our participants might under-report their actual inclination to give in to temptations, we do not expect any systematic reporting differences between both study arms. Considering that our trial had an "active" control group that was also given a saving device, we contend that participants in both study arms would have been equally exposed to potential interviewer demand or Hawthorne effects.

Further, the final sample size of our study was lower than we had initially planned for. This was mostly because data collection efforts were thwarted by heavy flooding

and temporary evacuations of several treatment locations. As a result, our field experiment was – albeit being sufficiently powered for the main outcome analyses – possibly insufficiently powered to detect heterogeneity in treatment effects (see Karlan et al., 2016; Porter, 2017). This prevents us from drawing conclusions about policies such as targeting strategies or profiling candidates who may benefit the most (Ravallion, 2009).

Another shortcoming is that our outcome measures were only reported by one household member, and in the majority of cases by the female spouse. Therefore, we do not have household-level information on saving amounts and temptation expenditures and are neither able to compare the self-reported financial behavior between husband and wife. In addition, since we did not measure the household's total savings balance, we are unable to fully capture a possible shifting in savings, for example from savings held in a jointly used bank account to savings held in a private saving device. More importantly, following the logic of the identified "hiding channel", we would expect to see declines in the spending patterns of partners or other household members, however, we are unable to confirm these based on the data available.

Overall, our results demonstrate positive effects of receiving a simple portable saving device in the form of a zip purse. Participants in the treatment group reported higher savings amounts and there was some indication of increases in female empowerment – most likely activated by increasing individual decision-making power over how financial resources are used, through the ownership of a private saving device (Field et al., 2021). In addition, we reported decreases in levels of debt, which could have important positive downstream impacts such as greater financial independence

and substantial cost savings on high interest rates. Follow-up interviews collected during the COVID-19 pandemic with a sub-sample of respondents suggest that treatment participants were more frequently able to use their savings for coping with this unprecedented shock.

A broader policy implication of our findings is that the distribution of *private* saving devices can likely help to boost savings rates among low-income individuals, and particularly among women who hold low financial bargaining power relative to their spouses. Our findings also motivate the conclusion that access restrictions to safe and private saving infrastructure appear to be a greater saving barrier than lack of self-control and vulnerability to temptations.

Funding: The study received funding from the Ministry of Science and Culture of Lower Saxony (Program Title: "Reducing Poverty Risk in Developing Countries", the University of Goettingen (Program Title: "Kreativität im Studium"), and the Joachim Herz Foundation.

Acknowledgments: The authors wish to thank Kay Tuschen, Louisa Barzen and Aditi Kuber for their considerable conceptual contributions to the intervention and study design. We further wish to thank Henrike Sternberg, Lena Merkel and Regina Dworschak for substantial contributions to data collection and research design and Nitya Mittal for leading the COVID-19 data collection. The study would not have been possible without the dedication and local expertize of the research team in Pune, including Hrishikesh Patil, Aishwarya Joshi, Bhagyashree Patole, Sagar Ramdas Mankar, Dorthi Robin Joseph, Ravindra Dnyandeo Ambore, Aruna Hitendra Kamble, Sunita Zumbar Salunke, Sujata Sanjay Kale, Sangita Mahesh Kumbhar, Samina Abdul Shaikh, Shrutika Mungi, Pratibha Tenkale, Shilpa Tathe, Ashwini Dawre, Devika Ghule, Shewale Aishwarya Balkrishna, Umesh Ramdas Madane, Anita Dattu Bhalekar, Aditya Mahendra Chandanshive, Mugdha Vijay Narkar, Sarita Goraksh Murtadak, Daingade Vijay Pandurang, Abhijeet Kumar. The authors also wish to thank the SNEH Foundation and Shraddha Deo in Pune and Ashwini Kuber for their logistical support. We are also grateful for feedback from Nianbo Dong and Yulia Shenderovich with regard to the power calculations and Magdalena Krieger and Hannah Prince for their input on the manuscript. Lastly, the authors wish to thank Dean Karlan for very detailed and fruitful advice and the three anonymous referees of the JDE pre-results review process for their valuable comments and the suggestion to collect in-depth usage data.

Author contribution statement: JIS, FS and SV conceptualized the study. JIS wrote the pre-analysis plan, oversaw the trial implementation, and led the data collection. JIS cleaned the data and FS and JIS analysed the data. JIS, FS, and SV interpreted the quantitative findings. RVS led the qualitative data collection, analysis, and interpretation. JIS drafted the first version of the manuscript, which FS and SV revised and RVS provided comments. SV and JIS provided funding for this study.

References

Aggarwal, S., Brailovskaya, V., and Robinson, J. (2021). Saving for Multiple Financial Needs: Evidence from Lockboxes and Mobile Money in Malawi. *The Review of Economics and Statistics*, pages 1–45.

Aggarwal, S., Francis, E., and Robinson, J. (2018). Grain today, gain tomorrow: Evidence from a storage experiment with savings clubs in Kenya. *Journal of Development Economics*, 134:1–15.

Akbas, M., Ariely, D., Robalino, D. A., Weber, M., et al. (2016). How to help poor informal workers to save a bit: Evidence from a field experiment in kenya. *IZA Discussion Papers*, 10024.

Aker, J. C., Sawyer, M., Goldstein, M., O'Sullivan, M., and McConnell, M. (2020). Just a bit of cushion: The role of a simple savings device in meeting planned and unplanned expenses in rural Niger. *World Development*, 128.

Alan, S., Boneva, T., and Ertac, S. (2019). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit. *The Quarterly Journal of Economics*, 134(3):1121–1162.

Almås, I., Armand, A., Attanasio, O., and Carneiro, P. (2018). Measuring and Changing Control: Women's Empowerment and Targeted Transfers. *The Economic Journal*, 128(612):F609–F639.

Ambec, S. and Treich, N. (2007). Roscas as financial agreements to cope with self-control problems. *Journal of Development Economics*, 82(1):120–137.

Anderson, L. C., Reynolds, T. W., and Gugerty, M. K. (2017). Husband and Wife Perspectives on Farm Household Decision-making Authority and Evidence on Intra-household Accord in Rural Tanzania. *World development*, 90:169–183.

Anderson, S. and Baland, J. M. (2002). The economics of roscas and intrahousehold resource allocation. *Quarterly Journal of Economics*, 117(3):963–995.

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quarterly Journal of Economics*, 121(2):635–672.

Ashraf, N., Karlan, D. S., and Yin, W. (2010). Female Empowerment: Impact of a Commitment Savings Product in the Philippines. *World Development*, 38(3):333–344.

Baland, J. M., Guirkinger, C., and Mali, C. (2011). Pretending to be poor: Borrowing to escape forced solidarity in Cameroon. *Economic Development and Cultural Change*, 60(1):1–16.

Banerjee, A. and Mullainathan, S. (2010). The Shape of Temptation: Implications for the Economic Lives of the Poor. *NBER Working Paper Series*, 15973.

Banerjee, A. V. and Duflo, E. (2007). The Economic Lives of the Poor. *Journal of Economic Perspectives*, 21(1):141–168.

Bénabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy*, 112(4):848–886.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

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

Bernhardt, A., Field, E., Pande, R., and Rigol, N. (2019). Household Matters: Revisiting the Returns to Capital among Female Microentrepreneurs. *American Economic Review: Insights*, 1(2):141–60.

Beshears, J., Choi, J. J., Harris, C., Laibson, D., Madrian, B. C., and Sakong, J. (2020). Which early withdrawal penalty attracts the most deposits to a commitment savings account? *Journal of Public Economics*, 183:104144.

Bobonis, G. J. (2009). Is the allocation of resources within the household efficient? New evidence from a randomized experiment. *Journal of Political Economy*, 117(3):453–503.

Boltz, M., Marazyan, K., and Villar, P. (2019). Income hiding and informal redistribution: A lab-in-the-field experiment in Senegal. *Journal of Development Economics*, 137:78–92.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101.

Breza, E. and Chandrasekhar, A. G. (2019). Social Networks, Reputation, and Commitment: Evidence From a Savings Monitors Experiment. *Econometrica*, 87(1):175–216.

Brune, L., Giné, X., Goldberg, J., and Yang, D. (2016). Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi. *Economic Development and Cultural Change*, 64(2):187–220.

Brune, L., Giné, X., Goldberg, J., and Yang, D. (2017). Savings defaults and payment delays for cash transfers: Field experimental evidence from Malawi. *Journal of Development Economics*, 129:1–13.

Brune, L., Kerwin, J. T., and Li, Q. (2021). How Important is Temptation Spending? Maybe Less than We Thought. *World Bank Economic Review*, Forthcomin.

Bryan, G., Karlan, D., and Nelson, S. (2010). Commitment Devices. *Annual Review of Economics*, 2(1):671–698.

Burke, J., Luoto, J., and Perez-Arce, F. (2018). Soft versus Hard Commitments: A Test on Savings Behaviors. *Journal of Consumer Affairs*, 52(3):733–745.

Casaburi, L. and Macchiavello, R. (2019). Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya. *American Economic Review*, 109(2):523–55.

Castilla, C. (2019). What's yours is mine, and what's mine is mine: Field experiment on income concealing between spouses in India. *Journal of Development Economics*, 137:125–140.

Dasso, R. and Fernandez, F. (2015). Temptation Goods and Conditional Cash Transfers in Per. *Mimeo*.

de Mel, S., McKenzie, D., and Woodruff, C. (2009). Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns. *American Economic Journal: Applied Economics*, 1(3):1–32.

Demirguc-Kunt, A., Klapper, L., Singer, D., Ansar, S., and Hess, J. (2018). *Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution*. Washington, DC: World Bank.

Duflo, E. (2003). Grandmothers and Granddaughters : Old-Age Pensions and Intrahousehold Allocation in South Africa. *World Bank Economic Review*, 17(1):1–25.

Duflo, E. and Udry, C. (2004). Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts and Consumption Choices.

Dupas, P., Karlan, D., Robinson, J., and Ubfal, D. (2018). Banking the Unbanked? Evidence from Three Countries. *American Economic Journal: Applied Economics*, 10(2):257–97.

Dupas, P. and Robinson, J. (2013). Why Don't the Poor Save More? Evidence from Health Savings Experiments. *American Economic Review*, 103(4):1138–71.

Duvendack, M., Palmer-Jones, R., and Vaessen, J. (2014). Meta-analysis of the impact of microcredit on women's control over household decisions: methodological issues and substantive findings. *Journal of Development Effectiveness*, 6(2):73–96.

Evans, D. K. and Popova, A. (2017). Cash Transfers and Temptation Goods. *Economic Development and Cultural Change*, 65(2):189–221.

Fiala, N. (2017). Business is Tough, but Family is Worse: Household Bargaining and Investment in Microenterprises in Uganda. *University of Conneticut Department of Economics Working Paper Series*, 2017-05.

Field, E., Pande, R., Rigol, N., Schaner, S., and Troyer Moore, C. (2021). On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms. *American Economic Review*, 111(7):2342–75.

Gang, I. N., Sen, K., and Yun, M.-S. (2008). Poverty in rural India: Caste and Tribe. *Review of Income and Wealth*, 54(1):50–70.

Glennerster, R., Walsh, C., and Diaz-Martin, L. (2018). A Practical Guide to Measuring Women's and Girls' Empowerment in Impact Evaluations. Technical report, The Abdul Latif Jameel Poverty Action Lab.

Goedecke, J., Guérin, I., D'Espallier, B., and Venkatasubramanian, G. (2018). Why do financial inclusion policies fail in mobilizing savings from the poor? Lessons from rural south India. *Development Policy Review*, 36:O201–O219.

Goldberg, J. (2017). The effect of social pressure on expenditures in Malawi. *Journal of Economic Behavior & Organization*, 143:173–185.

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

Herskowitz, S. (2021). Gambling, Saving, and Lumpy Liquidity Needs. *American Economic Journal: Applied Economics*, 13(1):72–104.

Hoel, J. B. (2015). Heterogeneous households: A within-subject test of asymmetric information between spouses in Kenya. *Journal of Economic Behavior and Organization*, 118:123–135.

Houser, D., Schunk, D., Winter, J., and Xiao, E. (2018). Temptation and commitment in the laboratory. *Games and Economic Behavior*, 107:329–344.

Hulme, D., Moore, K., and Barrientos, A. (2015). Assessing the insurance role of microsavings. In Vos, R. and Islam, N., editors, *Financing for Overcoming Economic Insecurity*. Bloomsbudy Publishing.

International Institute for Population Sciences (2017). National Family Health Survey (NFHS-4). Technical report, International Institute for Population Sciences (IIPS), Mumbai, India.

Iversen, V., Jackson, C., Kebede, B., Munro, A., and Verschoor, A. (2011). Do Spouses Realise Cooperative Gains? Experimental Evidence from Rural Uganda. *World Development*, 39(4):569–578.

Jakiela, P. and Ozier, O. (2016). Does Africa Need a Rotten Kin Theorem? Experimental Evidence from Village Economies. *The Review of Economic Studies*, 83(1):231–268.

Kapur Mehta, A. and Shah, A. (2003). Chronic Poverty in India: Incidence, Causes and Policies. *World Development*, 31(3):491–511.

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, 62(12):3393–3411.

Karlan, D., Ratan, A. L., and Zinman, J. (2014). Savings by and for the Poor: A Research Review and Agenda. *Review of Income and Wealth*, 60(1):36–78.

Kast, F., Meier, S., and Pomeranz, D. (2018). Saving more in groups: Field experimental evidence from Chile. *Journal of Development Economics*, 133:275–294.

Lastrapes, W. D. and Rajaram, R. (2016). Gender, caste and poverty in India: evidence from the National Family Health Survey. *Eurasian Economic Review*, 6(2):153–171.

Levenson, H. (1981). Differentiating among internality, powerful others, and chance. In Lefcourt, H. M., editor, *Research with the Locus of Control Construct*, pages 1–15. Academic Press.

Lipscomb, M. and Schechter, L. (2018). Subsidies versus mental accounting nudges: Harnessing mobile payment systems to improve sanitation. *Journal of Development Economics*, 135:235–254.

Locke, E. A. and Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9):705–717.

Lown, J. M. (2012). Development and Validation of a Financial Self-Efficacy Scale. *Journal of a Financial Counseling and Planning*, 22(2).

Morisano, D., Hirsh, J. B., Peterson, J. B., Pihl, R. O., and Shore, B. M. (2010). Set-

ting, elaborating, and reflecting on personal goals improves academic performance. *The Journal of applied psychology*, 95(2):255–264.

Mullainathan, S. and Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. Henry Holt and Company.

Porter, K. E. (2017). Statistical Power in Evaluations That Investigate Effects on Multiple Outcomes: A Guide for Researchers. *Journal of Research on Educational Effectiveness*, 11(2):267–295.

Ravallion, M. (2009). Should the randomistas rule? *Economists' Voice*, 6(2).

Riley, E. (2019). Resisting Social Pressure in the Household Using Mobile Money: Experimental Evidence on Microenterprise Investment in Uganda. *Mimeo*, pages 85, November 9, 2020.

Schaner, S. (2015). Do Opposites Detract? Intrahousehold Preference Heterogeneity and Inefficient Strategic Savings. *American Economic Journal: Applied Economics*, 7(2):135–74.

Schilbach, F. (2019). Alcohol and Self-Control: A Field Experiment in India. *American Economic Review*, 109(4):1290–1322.

Shafir, E. and Thaler, R. H. (2006). Invest now, drink later, spend never: On the mental accounting of delayed consumption. *Journal of Economic Psychology*, 27(5):694–712.

Shefrin, H. M. and Thaler, R. H. (1992). Mental accounting, saving, and self-control. In Loewenstein, G. and Elster, J., editors, *Choice over time*, pages 287–330. Russell Sage Foundation.

Singh, A., Kumar, K., McDougal, L., Silverman, J. G., Atmavilas, Y., Gupta, R., and Raj, A. (2019). Does owning a bank account improve reproductive and maternal health services utilization and behavior in India? Evidence from the National Family Health Survey 2015-16. *SSM - population health*, 7.

Soman, D. and Zhao, M. (2011). The Fewer the Better: Number of Goals and Savings Behavior:. *Journal of Marketing Research*, 48(6):944–957.

Somville, V. and Vandewalle, L. (2018). Saving by Default: Evidence from a Field Experiment in Rural India. *American Economic Journal: Applied Economics*, 10(3):39–66.

Steinert, J. I., Cluver, L. D., Meinck, F., Doubt, J., and Vollmer, S. (2018a). Household economic strengthening through financial and psychosocial programming: Ev-

idence from a field experiment in South Africa. *Journal of Development Economics*, 134:443–466.

Steinert, J. I., Cluver, L. D., Meinck, F., Nzima, D., and Doubt, J. (2020). Opening the Black Box: A Mixed-Methods Investigation of Social and Psychological Mechanisms Underlying Changes in Financial Behaviour. *The Journal of Development Studies*, 56(12):2327–2348.

Steinert, J. I., Zenker, J., Filipiak, U., Movsisyan, A., Cluver, L. D., and Shenderovich, Y. (2018b). Do saving promotion interventions increase household savings, consumption, and investments in Sub-Saharan Africa? A systematic review and meta-analysis. *World Development*, 104:238–256.

Thaler, R. H. (1990). Anomalies: Saving, Fungibility, and Mental Accounts. *Journal of Economic Perspectives*, 4(1):193–205.

Thorat, S. and Madheswaran, S. (2018). Graded Caste Inequality and Poverty: Evidence on Role of Economic Discrimination:. *Journal of Social Inclusion Studies*, 4(1):3–29.

Townsend, C. and Liu, W. (2012). Is planning good for you? the differential impact of planning on self-regulation. *Journal of Consumer Research*, 39(4):688–703.

World Bank (2012). World Development Report - Gender Equality and Development. Technical report, World Bank, Washington D.C.

World Bank (2015). World Development Report - Mind, Society, and Behavior. Technical report, World Bank, Washington D.C.

World Bank (2017). Maharashtra - Poverty, growth, and inequality. Technical report, World Bank, Washington D.C.

Appendix

Tables

Table A1. Sample Attrition

	(1) Not completed endline			(2) Not completed COVID survey		
	I	II	III	I	II	III
Treatment	0.00	0.01	-0.23	0.02	0.02	0.01
	(0.01)	(0.01)	(0.12)	(0.03)	(0.03)	(0.17)
Female		0.10	-0.11		-0.02	-0.08
		(0.02)	(0.03)		(0.04)	(0.05)
Married		-0.03	-0.05		-0.12	-0.10
		(0.02)	(0.03)		(0.03)	(0.05)
Age		-0.00	-0.00		0.00	0.00
		(0.00)	(0.00)		(0.00)	(0.00)
Education		0.00	-0.01		-0.03	-0.03
		(0.00)	(0.01)		(0.01)	(0.01)
Employed		-0.03	-0.02		-0.11	-0.09
		(0.01)	(0.02)		(0.03)	(0.04)
Household Size		-0.00	-0.01		0.01	0.01
		(0.00)	(0.00)		(0.01)	(0.01)
Income		0.00	0.00		0.00	0.00
		(0.00)	(0.00)		(0.00)	(0.00)
Baseline Savings Balance		-0.01	0.01		-0.03	-0.03
		(0.00)	(0.01)		(0.01)	(0.01)
Baseline Time Preference		0.00	-0.01*		-0.01	0.00
		(0.00)	(0.01)		(0.01)	(0.01)
Female x Treatment			0.03			0.12
			(0.05)			(0.07)
Married x Treatment			0.03			-0.06
			(0.04)			(0.07)
Age x Treatment			0.00			0.00
			(0.00)			(0.00)
Education x Treatment			0.01			0.01
			(0.01)			(0.01)
Employment x Treatment			-0.02			-0.05
			(0.03)			(0.06)
Household Size x Treatment			0.01			-0.01
			(0.01)			(0.01)
Income x Treatment			-0.00			-0.00
			(0.00)			(0.00)
Baseline Savings x Treatment			-0.02			0.01
			(0.01)			(0.03)
Baseline Time Preference x Treatment			0.02			-0.01
			(0.01)			(0.02)
Observations	1,525	1,482	1,482	1,525	1,482	1,482
R^2	0.000	0.048	0.063	0.000	0.047	0.052

Note: Robust standard errors in parentheses.

Table A2. ITT Estimates: Impact on Withdrawals and Deposits

	(1) Total Past-Month Withdrawals			(2) Total Past-Month Deposits		
	I	II	III	I	II	III
Panel A: Raw Outcome						
ITT: Received Program	-3,445.25	-3,470.71	-3,542.79	-482.42	-491.12	-512.00
	(1,444.96)	(1,450.53)	(1,525.08)	(332.18)	(332.48)	(342.73)
Lagged Outcome		0.02	0.01		0.06	0.04
		(0.02)	(0.02)		(0.07)	(0.07)
Mean Control	8,765.12	8,765.12	8,765.12	2,315.65	2,315.65	2,315.65
	(34,986.16)	(34,986.16)	(34,986.16)	(7,344.15)	(7,344.15)	(7,344.15)
Panel B: Winsorized Outcome						
ITT: Received Program	-1,887.80	-1,959.47	-1915.99	-258.23	-260.26	-258.67
	(841.59)	(840.89)	(856.03)	(183.93)	(182.22)	(185.02)
Lagged Outcome		0.06	0.04		0.33	0.31
		(0.03)	(0.03)		(0.10)	(0.10)
Mean Control	7,032.62	7,032.62	7,032.62	1,924.86	1,924.86	1,924.86
	(16,870.01)	(16,870.01)	(16,870.01)	(3,790.41)	(3,790.41)	(3,790.41)
Stratification variables	yes	yes	yes	yes	yes	yes
Lagged outcome	no	yes	yes	no	yes	yes
Controls	no	no	yes	no	no	yes
Observations	1,421	1,421	1,379	1,421	1,421	1,379

Notes: Robust standard errors in parentheses. Model I includes trial arm and stratification variables, namely participant sex and baseline savings. Model II includes stratification variables (same as Model I) and the lagged outcomes. Model III includes additional controls: participants' age, marital status, educational status, employment, household size, household income. The outcome variables include past-month withdrawals (in INR) and deposits (in INR) from any savings source (e.g. experimental devices, home savings, formal savings), summed up to the total past-month withdrawal and deposit amount, respectively. Panel A reports coefficients for the raw outcome. Panel B displays coefficients for 1% high tail winsorizing of the outcome variable.

Table A3. ITT Estimates: Disaggregated Savings Balances

	Box Savings	Formal Savings	Home Savings	Savings held in Club	Savings held by Relatives
Panel A: Raw Outcome					
ITT: Received Program	-109.50	2,991.31	-231.02	-21.73	4,162.35
	(133.47)	(1,352.47)	(246.44)	(90.02)	(2,933.37)
Lagged Outcome		0.07	0.05	0.06	0.06
		(0.04)	(0.03)	(0.05)	(0.07)
Mean Control	1,121.82	5,257.97	794.40	653.77	572.87
	(2,728.39)	(15,417.42)	(5,795.69)	(1,717.90)	(4,914.49)
Panel B: Winsorized Outcome					
ITT: Received Program	-66.90	1,271.14	-78.03	6.41	192.03
	(92.97)	(758.86)	(75.83)	(80.83)	(134.79)
Lagged Outcome		0.29	0.18	0.26	0.05
		(0.07)	(0.04)	(0.05)	(0.04)
Mean Control	1,023.01	5,036.38	595.54	623.94	345.60
	(1,723.07)	(12,788.40)	(1,415.41)	(1,494.67)	(2,108.68)
Stratification variables	yes	yes	yes	yes	yes
Lagged Outcome	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes
Observations	1,379	1,379	1,379	1,379	1,379

Notes: Robust standard errors in parentheses. Estimates are for Model III, which includes additional controls: participants' age, marital status, educational status, employment, household size, household income. Outcome variables report absolute savings amounts (INR) in each of the saving sources at the time of the interview. We exclude savings held in the zip purse here given that this variable, by definition, has missing values for all control group participants. Similarly, we do not include the lagged outcome of box savings in the first column as it is by definition zero. Panel A shows estimates on raw outcomes. In Panel B, the outcome (and lagged outcome) is adjusted to reduce the influence of outliers using 1% high-tail winsorization.

Table A4. ITT Estimates: Impact on Self-rated Temptations

	Self-rated Temptations		
	I	II	III
ITT: Received Program	-0.03	-0.03	-0.02
	(0.02)	(0.02)	(0.02)
Lagged Outcome		0.10	0.09
		(0.03)	(0.03)
Mean Control	1.29	1.29	1.29
	(0.46)	(0.46)	(0.46)
Stratification Variables	yes	yes	yes
Lagged Outcome	no	yes	yes
Controls	no	no	yes
Observations	1,417	1,414	1,372

Notes: Robust standard errors in parentheses. Model I includes trial arm and stratification variables, namely participant sex and baseline savings. Model II includes stratification variables (same as Model I) and the lagged outcomes. Model III includes additional controls: participants' age, marital status, educational status, employment, household size, household income. The outcome variable is a principal-component weighted temptation index (sample item: "I am generally willing to give up something today in order to benefit from that in the future.").

Table A5. Heterogeneity in Treatment Effects on Primary Outcomes: Gender

	(1) **Total Savings Balance (INR)**		(2) **Temptation Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome				
ITT Received Program x *Female*	2,136.67 (5,656.85) [0.706]	7,166.78 (3,879.88)	-41.55 (52.88) [0.666]	-4.42 (15.12)
ITT Received Program	5,030.11 (4,381.95)		37.14 (50.67)	
Female	2,120,71 (2,080.42)		-17.33 (41.38)	
Mean (Control x *Female*)	8,530.18 (21,444.17)		82.02 (235.47)	
Panel B: Winsorized Outcome				
ITT Received Program x *Female*	2,334.15 (2,240.21)	1,868.09 (990.38)	-53.00 (35.20)	-10.68 (12.30)
ITT Received Program	-466.06 (1,986.51)		42.32 (33.11)	
Female	1,402.51 (1,449.51)		11.30 (24.13)	
Mean (Control x *Female*)	7,917.03 (14,912.15)		79.81 (221.43)	
Stratification variables	yes		yes	
Lagged Outcome	yes		yes	
Controls	yes		yes	
Observations	1,379		1,379	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (of Panel A). Robust standard errors in parentheses. Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. Female is a binary variable equal to 1 if the gender of the study participant is female. Panel A uses the raw outcome. Panel B uses 1% high-tail winsorization of the outcome (and lagged outcome) to reduce the influence of outliers.

Table A6. Heterogeneity in Treatment Effects on Primary Outcomes: Income

	(1) **Total Savings Balance (INR)**		(2) **Temptation Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome				
ITT Received Program x *Low Income*	6,291.42 (11,444.63) [0.666]	11,600.46 (11,489.29)	-18.55 (33.62) [0.666]	-11.61 (27.96)
ITT Received Program	5,309.04 (2,273.18)		6.95 (18.15)	
Low Income	-4,135.12 (2,272.01)		-1.12 (24.15)	
Mean (Control x *Low Income*)	6,549.67 (14,470.76)		79.68 (259.60)	
Panel B: Winsorized Outcome				
ITT Received Program x *Low Income*	-1,280.81 (1,968.24)	477.04 (1,709.70)	-13.49 (27.46)	-12.03 (23.61)
ITT Received Program	1,757.85 (1,012.92)		1.45 (13.69)	
Low Income	-1,172.91 (1,358.57)		-3.09 (20.44)	
Mean (Control x *Low Income*)	6,549.67 (14,470.76)		74.26 (221.68)	
Stratification variables	yes		yes	
Lagged Outcome	yes		yes	
Controls	yes		yes	
Observations	1,379		1,379	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (of Panel A). Robust standard errors in parentheses. Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. Low income is a binary variable equal to 1 for the lowest income quantile of the sample. Panel A uses the raw outcome. Panel B uses 1% high-tail winsorization of the outcome (and lagged outcome) to reduce the influence of outliers.

Table A7. Heterogeneity in Treatment Effects on Primary Outcomes: Present Bias

	(1) **Total Savings Balance (INR)**		(2) **Temptation Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome				
ITT Received Program x *Present Bias*	-6,148.29 (6,496.33) [0.666]	3,633.43 (2,387.89)	-20.52 (30.03) [0.666]	-7.83 (19.00)
ITT Received Program	9,781.72 (6,081.04)		12.63 (23.44)	
Present Bias	309.07 (1,717.18)		17.95 (24.83)	
Mean (Control x *Present Bias*)	7,768.25 (11,852.48)		84.44 (261.71)	
Panel B: Winsorized Outcome				
ITT Received Program x *Present Bias*	-908.54 (1,758.32)	1,007.82 (1,147.05)	-5.65 (23.33)	-4.50 (15.97)
ITT Received Program	1,916.36 (1,331.69)		1.16 (17.03)	
Present Bias	657.00 (1,106.59)		12.96 (21.55)	
Mean (Control x *Present Bias*)	7,768.25 (11,852.48)		78.67 (211.96)	
Stratification variables	yes		yes	
Lagged Outcome	yes		yes	
Controls	yes		yes	
Observations	1,379		1,379	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (of Panel A). Robust standard errors in parentheses. Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. Present bias is a composed index of four items ("Today is more important than tomorrow", "I am impatient", "I easily give in to my temptations", "It is difficult for me to avoid eating a snack food I enjoy if it is easily available, even if I am not hungry") and centered around zero. Present bias is coded 1 if the index score is greater than 0, thus indicating a higher level of present bias. Panel A uses the raw outcome. Panel B uses 1% high-tail winsorization of the outcome (and lagged outcome) to reduce the influence of outliers.

Table A8. Heterogeneity in Treatment Effects on Primary Outcomes: Female Involvement

	(1) **Total Savings Balance (INR)**		(2) **Temptation Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome				
ITT Received Program x *Female Involvement*	7,819.33 (5,139.55) [0.666]	8,666.22 (4,637.97)	-31.80 (24.25) [0.666]	-11.61 (18.75)
ITT Received Program	846.89 (3,549.10)		20.19 (15.52)	
Female Involvement	569.56 (4,086.75)		75.76 (16.32)	
Mean (Control x *Female Involvement*)	8,651.90 (18,027.15)		98.39 (260.86)	
Panel B: Winsorized Outcome				
ITT Received Program x *Female Involvement*	509.68 (2,303.87)	1,874.51 (1,112.24)	-38.89 (21.64)	-19.65 (15.09)
ITT Received Program	1,364.83 (2,025.36)		19.24 (15.38)	
Female Involvement	399.10 (1,631.94)		71.68 (15.63)	
Mean (Control x *Female Involvement*)	8,338.57 (14,770.60)		96.08 (243.76)	
Stratification variables	yes		yes	
Lagged Outcome	yes		yes	
Controls	yes		yes	
Observations	1,141		1,141	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (of Panel A). Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. Female involvement is a binary variable denoting whether the female spouse (partner) is actively involved in financial decision making processes within the household. Panel A uses raw outcome, while Panel B employs 1% high-tail winsorization to the outcome measure. Robust standard errors in parentheses.

Table A9. Heterogeneity in Treatment Effects on Secondary Outcomes I

	(1) **Gender Attitudes**		(2) **Female Empowerment Index** (women-only sample)		(3) **Self-Efficacy Index**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Gender						
ITT: Received Program x *Female*	-0.29 (0.33) [0.809]	-0.07 (0.13)			0.14 (0.29) [0.944]	0.06 (0.14)
ITT: Received Program	0.22 (0.30)		/	/	-0.08 (0.26)	
Female	0.00 (0.27)				-0.48 (0.22)	
Mean (*Female* x Control)	-0.18 (2.32)				5.88 (2.47)	
Income						
ITT Received Program x *Low Income*	-0.29 (0.31) [0.809]	-0.24 0.27	0.07 (0.05) [0.667]	0.10 (0.04)	0.15 (0.31) [0.944]	0.15 (0.27)
ITT: Received Program	0.05 (0.14)		0.02 (0.02)		-0.00 (0.14)	
Low Income	0.01 (0.21)		-0.07 (0.04)		-0.41 (0.22)	
Mean (*Low Income* x Control)	-0.30 (2.46)		-0.29 (0.39)		5.56 (2.53)	
Present Bias						
ITT Received Program x *Present Bias*	-0.02 (0.24) [0.944]	0.03 (0.17)	0.05 (0.04) [0.809]	0.07 (0.03)	-0.14 (0.24) [0.944]	-0.04 (0.18)
ITT: Received Program	-0.01 (0.17)		0.02 (0.03)		0.11 (0.17)	
Present Bias	0.16 (0.26)		-0.05 (0.04)		0.43 (0.26)	
Mean (*Present Bias* x Control)	0.04 (2.28)		-0.26 (0.39)		5.99 (2.32)	
Female Involvement						
ITT Received Program x *Female Involvement*	0.06 (0.34) [0.944]	-0.05 (0.15)	-0.08 (0.05) [0.667]	0.03 (0.02)	0.51 (0.34) [0.667]	0.17 (0.15)
ITT: Received Program	-0.12 (0.31)		0.11 (0.05)		-0.34 (0.30)	
Female Involvement	0.07 (0.23)		0.09 (0.04)		-0.03 (0.23)	
Mean (*Female Involvement* x Control)	-0.13 (2.29)		-0.20 (0.35)		5.94 (2.49)	
Stratification variables	yes		yes		yes	
Lagged Outcome	yes		yes		yes	
Controls	yes		yes		yes	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (only Panel A). Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. The construction of heterogeneity variables is described in previous tables. Panel A uses raw outcome, while Panel B employs 1% high-tail winsorization to the outcome measure.

Table A10. Heterogeneity in Treatment Effects on Secondary Outcomes II: Gender

	(4) **Resilience Index**		(5) **Outstanding Debt (INR)**		(6) **Household Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome						
ITT: Received Program x *Female*	-0.06 (0.06) [0.809]	-0.02 (0.02)	-2,096.58 (1,043.39) [0.667]	-935.99 (268.22)	-3,124.157 (5,040.90) [0.944]	102.42 (1,861.52)
ITT: Received Program	0.04 (0.05)		1,160.59 (890.81)		3,226.57 (4,576.63)	
Female	-0.01 (0.04)		968.21 (587.98)		866.22 (1,699.74)	
Mean (*Female* x Control)	0.46 (0.44)		2,070.85 (5,560.74)		5,926.08 (32,119.03)	
Panel B: Winsorized Outcome						
ITT: Received Program x *Female*			-1,193.05 (535.71)	-712.69 (205.32)	7.25 (918.24)	117.07 291.32
ITT: Received Program	/	/	480.35 (495.31)		109.82 (864.97)	
Female			773.87 (373.01)		-824.80 (626.32)	
Mean (*Female* x Control)			1,847.12 (3,897.27)		3,867.62 (4,969.67)	
Stratification variables	yes		yes		yes	
Lagged Outcome	yes		yes		yes	
Controls	yes		yes		yes	
Observations	1378		1379		1379	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (only Panel A). Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. The construction of heterogeneity variables is described in previous tables. Panel A uses raw outcome, while Panel B employs 1% high-tail winsorization to the outcome measure.

Table A11. Heterogeneity in Treatment Effects on Secondary Outcomes II: Income

	(4) **Resilience Index**		(5) **Outstanding Debt (INR)**		(6) **Household Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome						
ITT: Received Program x *Low Income*	-0.01 (0.05) [0.677]	-0.02 (0.05)	-1,187.69 (683.03) [0.944]	-1487.71 (571.12)	-692.93 (2,229.76) [0.944]	92.96 (534.48)
ITT: Received Program	-0.01 (0.03)		-300.02 (333.26)		785.89 (2,195.95)	
Low Income	0.05 (0.04)		491.34 (597.49)		-1,585.60 (1,246.97)	
Mean (*Low Income* x Control)	0.48 (0.44)		2,163.64 (6,575.08)		3,091.61 (3,433.58)	
Panel B: Winsorized Outcome						
ITT Received Program x *Low Income*			-821.89 (437.48)	-1139.19 (374.65)	-443.66 (505.27)	-228.37 (373.55)
ITT Received Program	/	/	-317.30 (221.81)		215.29 (346.44)	
Low Income			214.63 (363.36)		-618.84 (372.83)	
Mean (*Low Income* x Control)			1,850.39 (4,088.38)		3,091.61 (3,433.58)	
Stratification variables	yes		yes		yes	
Lagged Outcome	yes		yes		yes	
Controls	yes		yes		yes	
Observations	1378		1379		1379	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (only Panel A). Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. The construction of heterogeneity variables is described in previous tables. Panel A uses raw outcome, while Panel B employs 1% high-tail winsorization to the outcome measure.

Table A12. Heterogeneity in Treatment Effects on Secondary Outcomes II: Present Bias

	(4) **Resilience Index**		(5) **Outstanding Debt (INR)**		(6) **Household Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome						
ITT: Received Program x *Present Bias*	0.00 (0.04) [0.944]	-0.01 (0.03)	-219.11 (524.45) [0.944]	-684.65 (437.93)	-568.64 (3,249.47) [0.944]	343.94 (2,525.69)
ITT: Received Program	-0.01 (0.03)		-465.55 (320.86)		912.58 (2,162.48)	
Present Bias	0.03 (0.05)		1,362.42 (668.22)		334.90 (1,737.12)	
Mean (*Present Bias* x Control)	0.44 (0.43)		2,056.86 (6,270.47)		5,615.63 (32,631.27)	
Panel B: Winsorized Outcome						
ITT Received Program x *Present Bias*			9.78 (383.98)	-500.26 (262.28)	-114.65 (542.21)	60.10 (387.84)
ITT Received Program	/	/	-510.04 (277.19)		174.76 (392.43)	
Present Bias			745.48 (436.01)		643.62 (488.66)	
Mean (*Present Bias* x Control)			1,711.04 (3,746.46)		3,979.71 (4,812.71)	
Stratification variables	yes		yes		yes	
Lagged Outcome	yes		yes		yes	
Controls	yes		yes		yes	
Observations	1378		1379		1379	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (only Panel A). Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. The construction of heterogeneity variables is described in previous tables. Panel A uses raw outcome, while Panel B employs 1% high-tail winsorization to the outcome measure.

Table A13. Heterogeneity in Treatment Effects on Secondary Outcomes II: Female Involvement

	(4) **Resilience Index**		(5) **Outstanding Debt (INR)**		(6) **Household Expenditures (INR)**	
	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome						
ITT: Received Program x *Female Involvement*	-0.08	-0.04	-410.87	-1,020.47	479.26	145.42
	(0.06)	(0.03)	(475.12)	(329.45)	(2,399.30)	(2,399.09)
	[0.736]		[0.809]		[0.944]	
ITT: Received Program	0.04		-609.60		-333.83	
	(0.05)		(338.27)		(662.29)	
Female Involvement	0.01		1,109.15		2,083.28	
	(0.04)		(450.54)		(1,528.41)	
Mean (*Female Involvement* x Control)	0.46		2,315.16		6,758.02	
	(0.44)		(6,062.27)		(36,459.12)	
Panel B: Winsorized Outcome						
ITT Received Program x *Female Involvement*			-137.64	-740.17	117.16	131.57
			(416.27)	(246.88)	(544.72)	(352.70)
ITT Received Program	/	/	-602.54		14.42	
			(329.51)		(420.66)	
Female Involvement			794.16		936.44	
			(360.60)		(411.16)	
Mean (*Female Involvement* x Control)			2,030.07		4,094.67	
			(4,083.37)		(5,349.06)	
Stratification variables	yes		yes		yes	
Lagged Outcome	yes		yes		yes	
Controls	yes		yes		yes	
Observations	1,141		1,141		1,141	

Notes: Multiple hypothesis corrected sharpened q-values in square brackets (only Panel A). Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. The construction of heterogeneity variables is described in previous tables. Panel A uses raw outcome, while Panel B employs 1% high-tail winsorization to the outcome measure.

Table A14. Predictors of total savings amounts in the treatment arm

	Total Savings Balance			
	(1)	**(2)**	**(3)**	**(4)**
Self-Control Purpose	**1,464.05**			**-8,536.85**
	(1,739.50)			**(10,476.08)**
Reminder Purpose		**5,067.72**		**6,754.96**
		(3,344.20)		**(8,128.98)**
Hiding Purpose			**7,239.54**	**8,249.54**
			(4,300.05)	**(5,920.19)**
Female	7,873.88	6,920.00	5,438.45	6,180.76
	(5,420.18)	(4,741.35)	(4,450.42)	(4,845.32)
Baseline Savings	0.15	0.15	0.15	0.15
	(0.09)	(0.08)	(0.09)	(0.09)
Baseline present bias	-2,005.26	-1,918.05	-1,514.32	-1,537.92
	(3,273.91	(3,106.81)	(2,916.18)	(2,941.79)
Married	3,673.98	3,558.81	2,972.54	3,614.99
	(3,584.35)	(3,362.84)	(3,193.08)	(3,542.54)
Age	297.76	400.19	332.12	311.86
	(203.98)	(215.94)	(188.20)	(203.35)
Education level	4,139.32	3,857.95	3,943.22	3,794.57
	(2,583.55)	(2,284.33)	(2,383.67)	(2,280.43)
Employed	-2,583.73	-5,487.98	-2,815.72	-2,106.39
	(8,488.45)	(9,205.93)	(8,873.78)	(8,082.75)
Household size	2,080.45	1,836.47	2,044.71	1,853.90
	(2,797.11)	(2,835.44)	(2,855.22)	(2,673.23)
Baseline income	-0.01	-0.01	-0.02	-0.02
	(0.02)	(0.01)	(0.01)	(0.01)
N	641	645	639	638

Notes: Robust standard errors in parentheses. The outcome variable is the total savings balance. Self-control purpose is captured with the following item: "When I am tempted to buy something I do not really need, the purse helps me to resist my temptations", reminder purpose is captured with: "When I see the purse, it reminds me of the importance to save money", and hiding purpose is captured with: "This purse helps me to keep money for myself and not to give it to other people (my partner, children, friends. . .)". All three items are rated on a 1-5 Likert-scale with higher values indicating higher agreement.

Table A15. Heterogeneity in Treatment Effects by Social Transfers to Kin

	Total Savings Balance (INR)	
	Main & Interaction Effect	Total Effect
Panel A: Raw Outcome		
ITT Received Program x *Social Transfers*	13,467.52	16,431.83
	(9,975.42)	(9,993.26)
ITT Received Program	2,964.31	
	(1,920.40)	
Social Transfers	5,418.25	
	(2,688.61)	
Mean (*Social Transfers* x Control)	10,017.94	
	(21,143.05)	
Panel B: Winsorized Outcome		
ITT Received Program x *Social Transfers*	1,632.51	2,698.60
	(2,115.26)	(1,858.39)
ITT Received Program	1,066.08	
	(984.08)	
Social Transfers	2,874.23	
	(1,282.27)	
Mean (*Social Transfers* x Control)	9,368.49	
	(15,056.11)	
Stratification variables	yes	
Lagged Outcome	yes	
Controls	yes	
Observations	1,379	

Notes: Model I presents results from OLS regressions with socioeconomic controls, stratification variables and the lagged outcome variable (equivalent to Specification III in the main outcome analyses). Model II shows the total effect for participants exhibiting a given trait, which corresponds to the sum of the main effect and the interaction effect as estimated in the previous column and includes corresponding significance levels. The construction of heterogeneity variables is described in previous tables. Panel A uses raw outcome, while Panel B employs 1% high-tail winsorization to the outcome measure.

Table A16. Definition of Outcome Measures

Variable	**Individual Items**	**Response Option**	**Aggregation**
Total Savings	What is the total amount of money that you currently keep in: • your savings box/zip purse (if applicable)? [based on self-report and hand count] • your banke savings account? • your accounts in post offices/ national savings centres? • your mobile phone money account? • cash savings at home (other than those kept in the lockbox)? • cash savings with relatives or friends? • a savings circle, savings group, or ROSCA?	Total amount in rupees	Sum of total savings amounts for each saving method
Temptation Spending	In the past month, how much money did you spend on [sugar, meat, cola/lemonade, alcohol, fried snacks, cake, gambling, tobacco, toys] ? In next month, how much money would you like to spend on this item?	Total amount in rupees	For all items where: past amount > desired future amount, the difference (in rupees) is calculated and added up into a total amount of past-month temptation expenditures
	Self-rated temptation: • In the past month, I spent money on things that I didn't really need. • In the past month, I bought something and later regret that I did. • In the past month, I found it difficult to really control on how I spend my money	5-point Likert scale from never-very often	PCA-weighted index aggregating three individual items

Table A16. ctd. Definition of Outcome Measures

Outcome	Individual Items	Response Option	Aggregation
Self-efficacy	• When I make plans, I am almost certain to make them work • When I get what I want, it's usually because I worked hard for it • My life is controlled by other powerful people • I am confident that I will not run out of money before the next payday • I am confident that I can plan carefully in advance how to use my money during each week	Rated on a 1-10-point Likert scale, ranging from "very much disagree" to "very much agree"	PCA-weighted index aggregating five individual items
Female empowerment	• Boys should not be allowed to get more opportunities and resources for education than girls. • Boys should be fed first and given more food compared to girls. • A husband should be more educated than his wife. • Daughters should have a similar right to inherited property as sons. • It would be a good idea to elect a woman as the President of India again. • Do you get in trouble for leaving the house without informing your husband or another household member? • Do you get in trouble for making unescorted outings such as visiting your parents, friends, going to the market?	Rated on a 1-10-point Likert scale, ranging from "very much disagree" to "very much agree" Last two items are binary and answered by female respondents only	PCA-weighted index aggregating individual items

Table A16. ctd. Definition of Outcome Measures

Outcome	**Individual Items**	**Response Option**	**Aggregation**
Resilience to Emergencies	• Did you experience any kind of emergency in the past six month? • If yes: How difficult was it for you and your family to find enought money to cope with that emergency? • Imagine an emergency would happen tomorrow. How difficult would it be for you and your family to find ten thousand INR to cope with this emergency? • Was there a time in the last 4 weeks when you needed to be admitted at the hospital but didn't because you didn't have enough money? • Was there a time in the last 4 weeks when you needed to buy medicine from a chemist but didn't because you didn't have enough money?	Rated as very difficult, somewhat difficult, not difficult at all	PCA-weighted index aggregating individual items
Household Expenditures	How many Rupees did you spend in the last month for [rice, dal, cleaning utensils, insurances, transport....]	Amount in rupees	Total amount (in rupees) of past-month expenses
Total debt	Are there any outstanding loans that you have to pay back? How much money do you expect to pay for any loan in the next month?	Amount in rupees	Total amount (in rupees) of money owed

Figures

Figure A1. Stationary and portable saving device

Figure A2. Example of individual saving plan

Other

Index Weighting

We will use principal component analysis to determine item weights.

In principal component analysis, variables are expressed as the linear combination of a set of underlying components for each respondent j:

$$a_{1j} = v_{11} \times A_{1j} + v_{12} \times A_{2j} + ... + v_{1N} \times A_{Nj}$$

$$a_{Nj} = v_{N1} \times A_{1j} + v_{N2} \times A_{2j} + ... + v_{NN} \times A_{Nj} \tag{3}$$

where A_N denotes the components and v_N the coefficients on each component for each variable.

Principal component analysis is used to find the linear combination of the individual variables with maximum variance—yielding the first principal component A_{1j} — and then finding a second linear combination with the maximum of the remaining variance, and so forth. The "scoring factors" are then retrieved by inverting the structure of Equation (8), which then produces estimates for the N principal components:

$$A_{1j} = f_{11} \times a_{1j} + f_{12} \times a_{2j} + ... + f_{1N} \times A_{Nj}$$

$$A_{Nj} = f_{N1} \times a_{1j} + f_{N2} \times a_{2j} + ... + f_{NN} \times a_{Nj} \tag{4}$$

Ultimately, the index for each respondent is given by the expression:

$$A_{aj} = f_{11} \times (a*_{1j} - a*_1)/(s*_1) + f_{1N} \times (a*_{Nj} - a*_N)/(s_N) \tag{5}$$

whereby $a*_{1j}$ to $a*_{Nj}$ represent N items for individual j , $a*_1$ the mean of $a*_{1j}$ across respondents and $s*_1$ the standard deviation.