

NBER WORKING PAPER SERIES

RANDOMIZATION FOR CAUSALITY, ETHNOGRAPHY FOR MECHANISMS: ILLIQUID
SAVINGS FOR LIQUOR IN AN AUTARKIC SOCIETY

Ricardo Godoy
Dean Karlan
Jonathan Zinman

Working Paper 29566
<http://www.nber.org/papers/w29566>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021

Research was funded by a grant from the World Bank Research Committee (RFP12503-RESE-BBRSB) and the Bill and Melinda Gates Foundation. We thank Harry Patrinos (World Bank) for his support and patience, Dylan Ramshaw and Innovations for Poverty Action for managing fieldwork, Tomás Huanca and Esther Conde of CBIDSI and CBIDSI surveyors and translators for exceptional fieldwork, and Mike Feng, Zachary Groff, Michael Ioannou, Javier Madrazo, Diego Santa Maria, Glynis Startz, and Jonathan Hy Vayness for excellent research assistance. Institutional Review Board approval from Yale University (0805003895), Northwestern University (STU00208598) and Innovations for Poverty Action (SP201.2010.01.01). This trial is registered in the American Economic Association's registry for randomized controlled trials (ID AEARCTR-0008670). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Ricardo Godoy, Dean Karlan, and Jonathan Zinman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Randomization for Causality, Ethnography for Mechanisms: Illiquid Savings for Liquor in an Autarkic Society

Ricardo Godoy, Dean Karlan, and Jonathan Zinman

NBER Working Paper No. 29566

December 2021

JEL No. D12,I12,O15

ABSTRACT

What should researchers do when confronted with surprising results? Financial access innovations usually leave “temptation” spending unaffected or reduced. However, we found that promotion of savings lockboxes in a largely autarkic society increased alcohol consumption and blood pressure, despite no one reporting intentions to save for alcohol. To probe mechanisms that could explain this pattern, we then used ethnographic methods, including direct observations of drinking (“scans”) and debriefing interviews to discuss the earlier trial results. We learn that sponsoring drinks confers prestige, but the stigma attached to drinking by outsiders likely discouraged reporting intentions to save for it.

Ricardo Godoy
Heller-Brown Building
rgodoy@brandeis.edu

Dean Karlan
Kellogg Global Hub
Northwestern University
2211 Campus Drive
Evanston, IL 60208
and CEPR
and also NBER
dean.karlan@gmail.com

Jonathan Zinman
Department of Economics
Dartmouth College
314 Rockefeller Hall
Hanover, NH 03755
and NBER
jzinman@dartmouth.edu

A data appendix is available at <http://www.nber.org/data-appendix/w29566>

Human behavior is complex and often confounds researchers. What is a researcher to do when confronted with a surprising result? Economists tend to search for alternative theories and probe for evidence of faulty data or identification assumptions. Increasingly, we may use Bayesian estimation and/or attempt to replicate the study.

We take a complementary approach of using ethnographic methods to probe mechanisms. Our ethnography comes after a standard randomized controlled trial (RCT) setup: an intervention sandwiched between a baseline and an endline survey.

We started this project focused on identifying whether and how the introduction of a savings technology would affect economic outcomes and well-being in a relatively autarkic setting. We find some expected results, namely evidence that financial and physical assets increase. We also find some unexpected results: temptation spending and particularly hard alcohol consumption increase. This last result withstands additional statistical scrutiny; most compellingly, we find biometric evidence that blood pressure increases, and specifically on the sub-group (men) in which we observe the largest increase in alcohol consumption. The increase in alcohol consumption pushed against the theory and prior empirical evidence that motivated our initial study design.

Specifically, we study savings among Tsimane', a remote society of horticulturists-foragers in the Bolivian Amazon with infrequent exposure to formal markets at baseline (Ringhofer 2010).¹ We conducted a baseline survey of 1,094 individuals in 61 villages and randomized households to either savings treatment or control. The treatment group received a small wooden savings lockbox provided by a regional NGO.² The control group received a small set of plates, of roughly equal monetary value to the lockbox, from the NGO. After one year, we re-surveyed with a focus on eight outcome measures: financial assets, physical assets, agricultural investments, income sources, expenditures, conflict within the household, consumption of temptation goods, and blood pressure. The first six outcomes were motivated primarily by prior work on savings and other financial interventions, together with baseline data on our subjects' savings intentions. The measurement of temptation goods was motivated primarily by the commitment aspect of the

¹ E.g., in a worldwide comparative study of 13 small-scale rural societies, Tsimane' ranked next to last in market interactions, with only 7% of household calories bought in the market (Henrich et al. 2010).

² Within the treatment group, we independently randomized two further dimensions: (1) whether households were given a key or had to travel to the NGO office in a nearby town to open the lockbox; (2) whether the lockbox was offered to the female or the male household head.

treatment: we hypothesized the lockbox would decrease temptation consumption. Blood pressure measurement was motivated by the expectation of a decrease in temptation consumption and an increase in health investment, the fourth most common savings plan reported at baseline.

We find strong evidence of a positive effect on asset accumulation despite no strong evidence of effects on overall consumption, which is consistent with savings access increasing wealth through savings and investment *per se*. Our estimated treatment effect on intra-household conflict is an imprecise zero, on a low base of conflict (Bauchet et al. 2021).

Turning to the surprising finding, we had little, if any, reason for anticipating a possible increase in alcohol consumption, even with the benefit of hindsight. Nothing in the ethnographic literature had suggested it (n.b. one of the authors is a cultural anthropologist who, at the onset of this study, had already spent 15 years doing research among Tsimane'). Nothing in our baseline data collection suggested it; e.g., we asked those in the treatment groups an open-ended question about their savings goal(s) and not a single person mentioned saving for alcohol.³ Nothing in the extensive literature of randomized trials to encourage savings suggests we should have anticipated wealth effects from any increased savings on the consumption of temptation goods.⁴ And as noted, behavioral economic theory and evidence suggest that any treatment effect of access to commitment savings on temptation consumption might well be *negative* (see e.g., Giné, Karlan, and Zinman 2010; Brune, Kerwin, and Li 2021).⁵ More specific to alcohol but less to commitment savings, Schilbach (2019) finds that a commitment device in India to drink less, in the form of a

³ We see a similar pattern at endline: only one treated person reported using savings box proceeds for alcohol.

⁴ See Table 1 for an overview of 52 studies based on RCTs to encourage savings in developing countries. We find five that estimate treatment effects on temptation consumption, and none of these find treatment effects that are close to statistically significant at conventional thresholds (Abraham et al. 2016; Berry, Karlan, and Pradhan 2018; Brune et al. 2017; Callen et al. 2019; Somville and Vandewalle 2019). Nor does evidence from other financial interventions suggest that increases in temptation spending are common; e.g., Banerjee et al. (2015) find that increasing access to microcredit reduces temptation spending, and studies of cash transfer impacts tend to find null effects (e.g., Brune, Kerwin, and Li 2021).

⁵ Indeed, the first hypothesis we listed in the funding proposal to the World Bank for this project was: "... the saving boxes will reduce... behaviors linked with impulsivity (e.g., smoking, drinking)." The onset of this project pre-dated the widespread use of pre-registration, so we do not have a Pre-Analysis Plan (PAP) and in any case PAPs do not usually predict results (e.g., Coffman and Niederle 2015).

voluntary price increase, reduces daytime drinking and increases saving.⁶

Upon analyzing the endline data and discovering the surprising finding, we decided to use ethnographic methods to probe underlying mechanisms. Specifically, a year after the endline, highly skilled enumerators, including one of the authors, went back to villages and to the central market town to conduct direct observations (“scans”) of Tsimane’ drinking, and to conduct one-on-one open-ended, relaxed, informal interviews with a convenience sample of Tsimane’ that included “debriefing” asking what they thought explained the increase in hard alcohol consumption. Such methods are sometimes used in cultural anthropology and sociology to shed light *ex post* on mechanisms underlying a surprising finding, but they are rarely deployed as such in economics.⁷

We learn commercial alcohol is a prestige good that is consumed socially and visibly, especially when shared, and especially (although not exclusively) among men. This helps explain increased alcohol consumption as an investment, of the social variety, and why respondents apparently had little hesitation reporting their drinking in our baseline and endline surveys. We also learn that alcohol consumption is highly stigmatized by Protestant missionaries proselytizing and living in the study area continuously for the past half century.⁸ This helps explain why respondents did not report intending to save for alcohol using the lockbox at baseline, or having saved for alcohol using the lockbox at endline, since both the surveys and lockboxes were administered by the same NGO, and both interventions and the NGO might have been associated

⁶ See also Ben-David and Bos (2021), where the relaxation of what might be viewed as a collective commitment device leads to more drinking and harmful downstream effects.

⁷ We move beyond casual empiricism about the use of research methods, by examining the universe of 96 papers using RCTs published in the “top five” economics journals from January 2016 through August 2021 (Table 2). We do not find any study taking the approach we do here, with Carneiro et al. (2021) coming closest in their use of “a parallel stream of qualitative analysis” to corroborate quantitative findings and develop hypotheses to test in future work (p. 2545). Beyond that set of papers, the closest economics paper we know of is Bergman et al. (2020), which uses in-depth narrative qualitative interviews to explore mechanisms underlying treatment effects that were not unexpected. See also Fryer (2011) and Allen et al. (2014) in economics, and especially Mannell et al. (2021) in public health, which uses debriefing interviews (but not scans) to probe mechanisms following a diabetes intervention in Bangladesh. Davis et al. (2019) advocate using a subset of the follow-up methods we use here to probe mechanisms as part of a broader framework for more systematically integrating qualitative methods into public health RCT designs.

⁸ This stigmatization has been thoroughly examined in a book written by Protestant missionaries and published after our data collection (Kempf and Kempf 2017).

with missionaries even when no such link existed.⁹

Overall, our work contributes to the literature on savings encouragement described above by studying a novel setting for economists, and to various strands of work on research methods. Most directly, we add to work bringing qualitative data collection and assessment methods from other fields to development economics (e.g., Hirschman 2014 1st ed. 1967; Townsend 1995; Blattman, Jamison, and Sheridan 2017). As documented above, the most novel feature of our work is using ethnographic methods to probe puzzling results *ex-post* (see footnote 7). This approach can help reshape theory and sharpen empirical inferences by uncovering new potential mechanisms and reducing incentives to “file drawer” studies with inexplicable results.

1. Setting, Research Design, and Implementation

A. Overview

Our study took place during 2011-2013 in 61 remote Amazonian villages in Bolivia. In partnership with a local NGO, Centro Boliviano de Investigación Socio Integral (CBIDSI), we conducted a baseline survey and randomization to a gift of a savings lockbox (treatment arms) or plates (control arm) during 2011. We then conducted an endline survey in 2012, roughly 12 months after randomization, also with CBIDSI. The endline survey yielded a puzzling result, which we explore with qualitative data collection by one of the authors and a Tsimane’ enumerator in 2013.

The authors and IPA field staff designed all study protocols in close collaboration with researchers at CBIDSI, including its research field manager—who has a PhD in cultural anthropology based on dissertation research among Tsimane’ and had 16 years of living and doing researching among the Tsimane’ at the onset of our study—and experienced Tsimane’ enumerators.

B. Setting and key actors

Tsimane’ have had centuries of intermittent contact with Westerners, but from the time of first contact with Spaniards in the sixteenth century they have tried to keep missionaries, cattle ranchers, rubber tappers, and traders at arm’s-length by moving farther into remote corners of the rain forest (Godoy 2015). In the 1950s, Protestant missionaries from the USA settled along the

⁹ Our partner NGO has no religious affiliation, but we did promote study participation on the missionary-run radio station. We consider alternative explanations in Section 2-D.

Maniqui River (the homeland of Tsimane') and slowly established the first schools, prepared all textbooks in the Tsimane' language (still used in schools today), translated the Bible, trained villagers to become the first rural teachers, offered medical assistance, and helped create the Tsimane' Council, the official governing body for the group. Their main aims accomplished—translating the Bible, preparing teaching material in Tsimane' for schools, and training a cadre of young Tsimane' preachers—they slowly withdrew their permanent presence from the area starting in the late 1990s, but keep a foothold in the area by running a radio station in Tsimane', a clinic, and periodic refresher courses on religion.

But even at the time of our study Tsimane' were largely autarkic, with most living in small villages of two dozen households that are hours of travel—by foot, canoe, and/or public transport—from the nearest market town. Dwellings are made from local materials such as thatch palm for roofs, with no or rudimentary walls and an open space or lean-to for cooking. Like many native Amazonian societies, Tsimane' follow preferential cross-cousin marriage; e.g., a man is expected to marry the daughter of his mother's father or his father's sister. In a nine-year panel study (2002-2010) we found that the median yearly number of two-headed households was 86%; single-headed households accounted for 11%, and the rest were polygynous households. Decision making power tends to be shared equally among couples (Bauchet et al. 2021).

Most consumption comes from horticulture, fishing, hunting, and livestock,¹⁰ and monetary earnings are only about a dollar per day.¹¹ We know of almost no specialized trades among the Tsimane'—e.g., no carpenters, bakers, or tailors—and most villages lack even a single store. The 2012 census of Bolivia puts the population of Tsimane' at 16,958, most of whom live in the department of Beni (INE 2015).

¹⁰ A study with direct observations of household consumption over five consecutive quarters in 1999-2000 found that only 3-11% of the value of all goods consumed in a household came from the market (Byron 2003, 138; Reyes-Garcia 2001, 39-77).

¹¹ Data from over 50 Tsimane' communities (some of which overlapped with the communities of this study) between 2004 and 2010 produces estimated mean monetary daily earnings per person to be PPP US\$1.07 and mean daily consumption to be PPP US\$0.54, excluding own consumption from forest and farm goods (Godoy 2020b). Men earn income from wage labor in logging camps, cattle ranches, and in the farms of Andean homesteaders who have moved to the lowlands. Women earn income from the sale of thatch palm, and both women and men earn monetary income from selling crops, principally rice and plantains (Vadez et al. 2008; Zycherman 2013).

Tsimane' save in small livestock (e.g., poultry, pigs) and in hardy crops (e.g., manioc, plantains). An annual panel study (2002-2007) with the entire adult population of 13 villages (n=795) found little inter-annual change in the stock of livestock (Undurraga et al. 2014). An estimated 37% of household heads had any cash at baseline in that study, with only 4% keeping money in a bank. Tsimane' have few other marketable assets or possessions besides basic tools and sacred objects. They put the bones of hunted animals in the ceilings of huts as a reverential act to avoid offending animal spirits (Nordenskiöld 1924). Huanca (2006), who managed survey data collection for our study with support from IPA field staff, documents contemporary use of this practice. Over the years, the practice of safekeeping valuables in the roof morphed to include valuable documents and cash wrapped in plastic, often side by side with the sacred bones. The savings box provided to treatment groups in our study built on this tradition of storing valuables.

CBIDSI is a secular NGO that provided the lockboxes and administered the baseline and endline surveys for our study, with the supervision and support of the authors and IPA field staff. CBIDSI was established in San Borja, the central market town of the Tsimane' homeland, along the Maniqui River in 2005 to help foreign researchers carry out academic and applied research among native peoples in the areas. Its key personnel have lived and worked among Tsimane' since the late 1990s.

In our study, participants interacted primarily with a surveyor-translator pair assigned to each village for purposes of administering the randomization and surveys (as detailed in the next two sub-sections). Surveyors were Bolivian Spanish-speaking college graduates who had been working with the NGO in the area for years on a panel study (2002-2010) of Tsimane'. Translators are bilingual Tsimane' who each had a decade or so of surveying experience entering our project.

C. Sampling, randomization, and treatments

The research team, including CBIDSI, worked with the governing council of Tsimane' to select a sample frame of villages. Our best estimate suggested there were about 120 Tsimane' villages with at least eight households at the time of the study.

We ruled out sampling from villages that were too costly to reach, too unsafe, or that contained people from other ethnic groups. This left 61 villages for our sample. Even these villages are far from the nearest market—a median of about 4 hours travel time. They tend to be even farther from the NGO office in the market town of San Borja where savings could be deposited.

All households in a village were eligible for participation in the study. To enhance participation, brief paid advertisements on the missionary radio station promoted the study shortly before the research team would arrive in each study village. The missionary radio station transmits news and announcements three times a day in Tsimane', and can be heard across the study area.

Within each of the 61 villages, CBIDSI and IPA publicly randomly assigned all the households residing in the village at the time of the visit to treatments or to a control group. These random assignments and baseline surveys took place from March through June of 2011. (For more details on the randomization mechanics and baseline survey see Appendix A and Section 1-D.)

All treatment households received a savings lockbox from CBIDSI that is small, portable, wooden, pest-resistant, and has a slit at the top of the box to insert money (Appendix B). N.b. that the lockbox was not linked to any sort of bank account or to any other service.

The main study arms, randomly assigned in roughly equal proportions within each village, are:

Treatment 1 (T1: Lockbox with Key). Savings box *with* a key so recipients could open the savings box at any time.

Treatment 2 (T2: Lockbox without Key). This treatment was the same as Treatment 1, except T2 households did not receive a key. Instead the key was kept in the CBIDSI office in San Borja, which, as noted, was several hours of travel time away from most villages in our sample. Our intent was to provide a strong commitment device to facilitate saving for a lumpy expenditure.

Control Group. The remaining households in a village served as a control and received a placebo intervention of six metal plates. The six plates had about the same monetary value as a savings box (~ PPP US\$11).

We also randomized which household head received the savings box or plates. Heads received the box or plates after completing the baseline survey, which followed immediately after the random assignment. Due to the high prevalence of two-headed households in the population, and our randomization of which head got the box or plates, we did not attempt to survey most single heads and instead gave them a gift after the public randomization. Our analysis sample thus contains only households headed by two adults.

D. Baseline and endline survey content and measures

Baseline and endline surveys covered very similar questions and lasted about 30 minutes on average. Surveyors asked about household composition and outcome variables. They also asked lockbox recipients at baseline about what they planned to save for and about what they had or were saving for at endline. The endline survey added an additional question on intra-household conflict and a few questions on lockbox use. At the end of both the baseline and endline surveys we gave all respondents a gift for the time they had spent in the interview. Women received wool and men received fishhooks and fishing line.¹²

Appendix D details our 8 key outcome measures: financial assets, physical assets, agricultural investments, income sources, expenditures, temptation consumption, blood pressure as measured by a sphygmomanometer during the survey, and conflict within the household. For each of these except for household conflict, we construct a main outcome from multiple component variables. We estimate physical asset values based on survey results on quantity of items combined with village-level asset price data.¹³

E. Baseline survey and sample characteristics

Baseline surveys followed immediately after treatment assignment in each village. We successfully surveyed about 99% of household heads with a random assignment, due to scheduling interviews right after random assignment and promising to deliver the assigned gift of the lockbox or plates after survey completion.

99% of households surveyed at baseline had a female and a male head and nearly all heads are at least 16 years old, and so we drop single-headed households and households headed by minors from our analysis, leaving a baseline sample of 1086 households. Appendix Table 1 summarizes other key demographics, and the seven of eight key outcomes we measure at baseline, and balance checks across treatment arms. As expected, households tend to be large, and heads' school attainment is low; the average respondent years of schooling is just under 3 years. We consider baseline levels of our key outcomes in greater detail below, and for now note that, as expected, our

¹² Appendix C contains more details on survey management and staff training, and data collection protocols, and survey instruments.

¹³ Surveyors collected asset prices by asking 2 or 3 village leaders in each village for the market price of assets they normally sell (e.g., livestock) or buy (e.g., machetes).

sample has low levels of consumption and assets. Regarding balance, we find no more differences across arms that are statistically significant at conventional p-value thresholds than one would expect to find by chance. And later when estimating treatment effects we control for the baseline value of the dependent variable, for the seven of eight key outcomes where we have that data available.

F. Endline survey, attrition, and CBDISI office survey

The endline survey took place during February-May 2012, about one year after the baseline survey. We only attempted to survey the same person as at baseline and succeeded with 918 individuals in our analysis sample, for an 85% retention rate. Appendix Table 1 examines differential attrition across randomization arms. We do not reject equal levels of attrition in three of the four comparisons (Columns 2, 4, and 6). But we do find 8pp (se 3pp) higher attrition among recipients of lockboxes with a key vs. recipients of lockboxes without a key (Column 8), which implies that the comparisons of treatment effects between those two arms should be read with caution.

When boxholders came to the CBDISI office to get their box opened, a CBDISI employee would record the amount saved and ask the holder what they planned to do with the money.

G. Ethnographic follow-up

In May 2013, a year after the endline survey and after preliminary analysis of the data, we returned to the field with standard ethnographic methods designed to explore why savings boxes increased the drinking of commercial alcohol, particularly among men (see Section 2-C for treatment effect estimates). We followed a three-pronged strategy of *debriefing interviews* with study subjects, *scans* of Tsimane drinking in public in towns, and *informal interviews* with liquor store owners.

Debriefing subjects is a common practice in lab experiments and outside of economics, e.g. in ethnography (Gupta and Kelly 2014), and has been done in prior studies among Tsimane' (Reyes-García et al. 2012). We added the additional step of asking subjects about *why* they thought the results turned out as they did.

One of the authors, Godoy, visited seven study villages and conducted about 20 interviews with recipients of savings boxes. Given logistical challenges and budget constraints, we used a

convenience sample of villages and recipients, with Godoy spending about half a day in each of the seven villages, seeking out any lockbox recipients, and starting conversations by asking subject(s) to show us the box, what they had used it for, and how and when they opened it if they had no key. These conversations would include whomever was around, whether an individual or couple, or a group. Other household or community members were welcome to join the conversation/interview and often did. Although the debriefing interview approach here and typically is relaxed and only lightly scripted, the interviewer does engage subjects with objectives and key prompts in mind. In our case, we had three main objectives: (i) to learn what subjects thought about the link between boxes, saving, and increased alcohol consumption (asking something to the effect of “We found that people who got savings boxes ended up drinking more. Can you tell me why?”); (ii) to learn more about how subjects perceived the box and our study (see description of prompts above); (iii) to learn more about how recipients used the box and any challenges they encountered (see description of prompts above).

Scans are another classic method. They are widely used in bio-cultural anthropology for collecting observational data, particularly in rural settings (Bernard 2015). The researcher walks a pre-determined path at a constant pace and notes what people are doing when they first see the subject; after jotting down the information, a researcher might approach the subject to ask questions. We had used scans in prior studies with Tsimane' (Godoy et al., 2009).

Our scans were conducted by an experienced Tsimane' observer under the supervision of Godoy in the market town of San Borja during one Saturday market day in May 2013 from about 7am to 7pm. We chose a market day after consulting with our survey enumerator team about when and where Tsimane' typically congregate in public to drink outside their villages (village drinking is covered in our debriefing interviews).¹⁴

The observer walked the streets of San Borja where Tsimane' typically drink. As the observer found a group drinking, he noted the composition of the group and joined them for a few minutes

¹⁴ Tsimane' come to town on weekends to sell farm and forest products before returning to their villages, typically on Sunday afternoon. A Tsimane' coming to San Borja for a weekend will arrive on Friday afternoon or Saturday morning. Some bring food so they do not have to spend money on food while in town. The poorer, unsalaried Tsimane' who comprise the bulk of our sample will typically sleep in the mission compound on the outskirts of town or with the townspeople who are godparents (*compadres*) of their children. The few salaried Tsimane' (e.g., teachers) also come to town on some weekends to collect their paychecks.

to greet them (but not drink) and find out what they were drinking. Each encounter produces a “vignette” of data from observation and conversation (see Appendix E for examples), and after about two-hours of collecting vignettes, the observer would debrief with Godoy to get guidance on things to look for and questions to ask should the observer encounter the group again.

Godoy conducted informal interviews with liquor store owners while the observer was scanning to find out the price and alcohol content of drinks Tsimane’ bought (see Appendix F).

2. Results

We report monetary values in bolivianos (Bs). PPP US\$1 \approx 3Bs over the course of our study period.

A. Box uses

Table 3 presents statistics from the endline survey on how recipients used the lockbox. Panel A shows that 95% still had the box (Column 1), with no differences across treatment arms (Columns 4 and 7), but that only 77% of those given a key still had it (surveyors verified the presence or absence of the box and key by observation). Panel B shows evidence of active use and that, as intended, those without a key opened the box substantially less often; e.g., a self-reported mean of only 0.66 times over the course of the year (Column 6, se 0.06) vs. 4.26 times among subjects with a key (Column 5, se 0.32). Interestingly, recipients of boxes without a key self-report opening the box much more than we observe in the CBIDS office, e.g., 35% in Column 6 vs. only 13% in the office. This is consistent with some recipients devising ways to open the box without a key (see Section 2-D for more evidence on this). Panel C shows further, and surveyor-verified, evidence of active use, with no statistically significant differences across treatment arms. 46% of recipients have some money in the box at endline (Column 1), and the amounts saved, about 100 Bs (se about 12 Bs), were substantial relative to the baseline of about 200 Bs of cash-on-hand (Appendix Table 1).

B. Self-reported savings plans and uses

Table 4 describes responses to the open-ended savings uses questions we asked of lockbox recipients in the survey, sorted by the frequency of baseline responses to “What are you saving for?” Clothing was the most common response, followed by housewares, transport, medicine &

hygiene, and tools. At endline, in response to “On what did you spend money from the box?”, the most common response was “Did not spend saved money”, followed by medicine & hygiene, food, and clothing.

Most strikingly, given our treatment effect results below, only one person mentioned a temptation good across the two surveys. Nor did anyone mention this in any of the 57 data points, covering 78 reported uses, collected at the CBIDSI office when no-key holders came to get their boxes opened.

Less strikingly but also noteworthy is that the other commonly reported uses at endline are consistent with our treatment effect estimates below. “Did not spend” is quite consistent with our finding below that cash on hand increases substantially. Our treatment effect estimates for various spending categories, including those covering food, medicine & hygiene (“non-food items”), and clothing (“durable items”) tend to be positive but imprecisely estimated.

C. Treatment effect estimates for main outcomes and component outcomes

Table 5 reports OLS intent-to-treat estimates for each of our eight main outcome variables described in Section 1-D. Regression covariates include indicator variables for treatment assignment, the baseline value of the outcome, and an indicator for each village as our stratification variables, since we randomized village-by-village.¹⁵ Each panel-column combination presents estimates from a single regression, with each panel estimating a different treatment effect (“TE”) for that outcome. Panel A presents estimates of the TE for getting a lockbox. Panel B presents estimates separately for female and male lockbox recipients. Panel C presents estimates separately for the key and no-key treatments. For each estimate we also report a q-value showing the minimum false discovery rate (i.e., the expected proportion of rejected null hypotheses that are true) at which the null hypothesis would be rejected for each test, given the other tests run simultaneously within the “family” (Benjamini and Hochberg 1995; Anderson 2008), where we define each panel as a family of tests.

Starting with financial assets, Table 5 Panel A Column 1 suggests a positive treatment effect of 78Bs (se 25Bs, q-value 0.02). This is very large in percentage terms, as the baseline mean is

¹⁵ We set the baseline values of all outcomes to zero if they are missing, and control in each regression for an additional binary variable indicating whether the baseline value of the outcome was missing. For regressions where the outcome is the change in income sources, we use the count of income sources as the baseline outcome variable.

only about 200Bs (Appendix Table 1) and the control group endline mean is only 129Bs. Appendix Table 2 suggests that this increase is driven by a large increase in cash held at home, and a smaller increase—in level terms—in money lent. We do not find an increase in money held at a bank. Appendix Table 3 shows that we find similar results using inverse hyperbolic sine instead of level assets, although pooled treatment effect on money lent now has a p-value closer to 1 than to ≥ 2 . The point estimates in the Panel Bs of these three tables suggest that the effects on financial assets may be larger when the box is given to a female head, but the p-values show that the treatment effect differences between female and male recipients are imprecisely estimated for the most part. The results in the various Panel C's suggest little difference between the key and no-key treatment effects, subject to the qualifiers that we are not powered to detect smaller differences that would still be quite meaningful economically, and that we do see some evidence of differences in bank savings in Appendix Tables 2 and 3, albeit on a base that is small relative to cash at home.

Turning to other potential sources of wealth and income generation, Table 5 Panel A Columns 2-4 suggests positive but imprecisely estimated effects on physical assets, agricultural investments, and change in income sources (see also Appendix Tables 2-4). The point estimate for the TE on physical assets is larger in level terms (213 Bs, se 120 Bs) than the estimated TE on financial assets (78 Bs, se 25 Bs) but much smaller in percentage terms (7% of the control group's endline mean). It is also substantially less strong statistically, with a q-value of 0.15, and a positive but imprecisely estimated effect in the IHS specification (Appendix Table 3). The results for the other two outcomes here imply nontrivial but imprecisely estimated increases. We find little evidence of heterogeneity for Columns 2-4 by the sex of lockbox recipient or by key access (Panels B and C).

We find little evidence of a treatment effect on total household expenditures in Table 5 (Column 5), or on different types of household expenditures in Appendix Tables 5 and 6. The absence of a clear increase in overall spending, coupled with the increase in financial assets documented above, suggests that the lockbox increased net saving and wealth.

We find little evidence of an effect on intra-household conflict over money in Table 5 Column 8, on a low base of conflict (e.g., only 12% of respondents report fighting over money in the last 2 months).

Table 5 Columns 6 and 7 tentatively suggest the surprising inference that the lockbox increases temptation consumption—defined as alcohol and cigarette consumption during the past seven days—and blood pressure, and perhaps more so or only among men (Panel B). The magnitudes

are substantial, with temptation consumption increasing an estimated 0.16 sd (se 0.07 sd), and blood pressure by 0.13 sd (se 0.06 sd). The q-values on these are 0.08 and 0.06, and the p-values on the differences between female and male recipients are 0.21 and 0.08, so these results are more suggestive than definitive, with several additional tables probing further.

Table 6 estimates treatment effects on the component variables of the temptation and blood pressure indices. For blood pressure, Columns 5-7 suggest increases on each component in Panel A: an indicator for high blood pressure, the average of the 3 systolic blood pressure level readings taken at endline, and the average of the 3 diastolic blood pressure readings taken at endline. The point estimates are larger, and statistically stronger, for male recipients in Panel B. Figure 1 plots the endline distributions of systolic and diastolic blood pressure, by treatment status and sex of the recipient of the lockbox, including the blood pressure of subject in the control group. Figure 1 shows that male blood pressure increases a bit throughout the blood pressure distribution, not just at the threshold for high blood pressure or in the right tail of the highest blood pressure readings in our sample (which, n.b., are low relative to U.S. levels). For temptation consumption (Columns 1-4), Panel A suggests that the lockbox increased commercial hard alcohol consumption by 0.05 liters during the past week, on a small base that is also 0.05 liters. Point estimates for treatment effects on the other components are each positive but imprecisely estimated, again each on a low base. Panel B suggests that males may be increasing traditional chicha and commercial alcohol consumption, and that some females may be drinking more as well, in the form of commercial alcohol and/or beer. Because temptation consumption has skewed distributions (n.b. the bottom rows of Table 6), we also examine treatment effects on alternative functional forms of it and find similar results for drinking (Appendix Table 7).

As noted at the outset, this pattern of results suggesting that the lockbox increases alcohol consumption and blood pressure was surprising given prior theory and evidence on commitment and other financial interventions. Nor do they square with respondent self-reports on savings goals and uses in our surveys at baseline, endline, and when getting the box opened at CBIDSI. Yet the blood pressure data, which as noted above comes from multiple cuff measurements, seemed unlikely to produce a false-positive treatment effect. And the results suggesting that the lockbox increased drinking, particularly among men, is of course medically consistent with the results suggesting statistically stronger and clinically larger blood pressure increases for men.

As such we were motivated to probe these results further, using the ethnographic methods

described in Section 1-G.

D. Insights from ethnographic methods

Here we summarize pertinent findings from ethnographic work on Tsimane', both in our follow-ups, and others' work. The aim is to help interpret the quantitative results reported above.

The lockbox is valued as a savings technology. Debriefing interviews, and observations of recipients who came to the CBDISI office to unlock their boxes, confirmed our initial hypotheses that the box provided more security, and mildew-resistant storage, for cash and documents than the baseline savings technology of a cloth or plastic bag. Having said that, many debriefing interview participants mentioned taking additional precautions to protect the box and its content from theft or appropriation.

The lockbox is also valued as a good in itself. Our debriefing interviews suggest that boxes have acquired semi-ritual status, often being placed in the ceiling next to the spirit-warding animal bones described in Section 1-B. Many recipients reported being motivated to save by rumors, unfounded to our knowledge, that CBDISI would confiscate the box if they found recipients were not using it to save money. It is worth noting that CBDISI was still getting requests for new or replacement boxes many years after data collection ended.

Alcohol is a social good that builds connectedness and confers prestige, especially when shared. This was evident from both the debriefing interviews and scans. Almost all drinking is done socially, in public, and with ritualized sharing among anyone who wants to join in “the drink”. The most traditional form is drinking chicha, a home-brewed fermented alcohol prepared by women with one's own staple crops, with rituals and status connotations that pre-date the arrival of Europeans to America (Zycherman 2015; Turner and Klaus 2020, pp. 163-4). Chicha preparation and drinking is strictly gender-normed and -segregated, and chicha is prepared and consumed only in the village. Commercial alcohol drinking is somewhat more gender-fluid, and done both in the village and in-town, but it is still almost always in groups and with ritualized sharing. E.g., our formal scans in the market town of San Borja found Tsimane' wage laborers and horticulturists congregating around a few well-known street corners near a liquor store with

friends, family, and acquaintances (see Appendix E for more details). A “sponsor” passes a bottle or plastic bag of alcohol around and invites passers-by to join, including non- Tsimane’. When the liquor runs out, the sponsor, or a different sponsor from the group, goes to the liquor store to buy more liquor. This tends to continue for several hours. Wealthier Tsimane’, such as those with formal employment (e.g., teachers) tend to sit at bars but also invite friends, acquaintances, and passers-by to join them for drinks. Overall, public drinking ranks in the top third of activities in cultural visibility among Tsimane' (Undurraga et al. 2016).

Commercial alcohol is quite probably closer to the margin of consumption decisions than chicha, since preparing and sharing chicha takes substantial time and crop inputs, relying heavily on a traditional household structure—both heads working near home, substantial manioc crop cultivation—that is now far from universal (Godoy 2020a). Other prestige goods (e.g., owning a motorcycle, TV) are well outside the means for most subjects in our sample.

Commercial alcohol is a lumpy expenditure, requiring effortful savings. Debriefing interviews, conversations emerging from the scans, and data collected from the liquor store in San Borja all suggest that our subjects typically have little cash-on-hand and need to engage in effortful savings to accumulate sufficient cash to purchase alcohol for themselves and sharing. This seems especially so in cases where subjects wish to go to town to drink and must incur additional expenses (e.g., transport, food) to do so.

Given this, and the social investment aspect of drinking, it seems that saving up to purchase commercial alcohol tends to be a deliberate act, not one of succumbing to temptation, as we had initially hypothesized.

Consistent with our quantitative results, we find little evidence that the box created intra-household conflict. Debriefing interviews revealed little if any intra-household tension around savings plans, for alcohol or otherwise.

Consistent with our quantitative results, we find that men are more likely than women to drink commercial alcohol and report using savings proceeds for drinking, although these are not exclusively male activities (as noted above).

Some recipients remove money from the box without a key. One villager demonstrated how to do so using a small plant stem as a makeshift pliers. This could help explain the lack of treatment effect differences between the key vs. no-key arms.

Drinking is highly stigmatized by Protestant missionaries (Kempf and Kempf 2017), and, as noted above, several debriefing interviews mentioned rumors that the missionary radio station had made an announcement that CBIDSI would confiscate boxes from people who were not using it for saving. Thus, it seems likely that lockbox recipients also feared confiscation if they revealed that they were and had been saving for alcohol.¹⁶ Note also that our debriefing interviews were substantially less strongly linked to CBIDSI than the surveys, since they were conducted by one of the authors not by CBIDSI employees.

Alternative explanations for why recipients report alcohol consumption, but not saving for alcohol, in the baseline and endline surveys do not fit the full pattern of evidence.

One possibility is that subjects did not report saving for alcohol in the baseline and endline surveys because they interpreted those questions as regarding something other than the (anticipated) treatment effect of the lockbox.¹⁷ This seems unlikely, given that several people *do* report saving for alcohol in the debriefing interviews that are less strongly linked to CBIDSI, the lockbox provider.

It also seems unlikely that the treatment effects on alcohol consumption and blood pressure are better explained by an increase in stress. E.g., it seems unlikely that stress about saving, or even about losing the lockbox, led to drinking. First, neither we nor others find any evidence that Tsimane' drink to alleviate stress; indeed, there is no word akin to "stress" in Tsimane' and Tsimane' are reported to have some of the lowest levels cortisol and coronary artery diseases in the world (Nyberg 2009; Kaplan et al. 2017). Second, the increase in male blood pressure seems to occur throughout the blood pressure distribution (Figure 1), not primarily in the right tail as one might expect if the most-stressed were turning to drinking to alleviate that stress. Third, as

¹⁶ We did not want to put anyone on the defensive in the debriefing interview by inquiring directly why no one had reported saving for alcohol in our baseline or endline surveys.

¹⁷ See Karlan, Osman, and Zinman (2016) for a discussion of related conceptual and measurement issues, in the context of self-reported uses as compared to estimated causal impacts of microcredit.

discussed above, drinking is typically and traditionally a planned and prestige-enhancing social activity among Tsimane'. Fourth, we find no evidence that increased saving led to increased intra-household conflict that might have created stress.

E. Summing up: Our interpretation of likely mechanisms

Our preferred interpretation of the likely mechanisms follows. We infer that the lockbox led to increased drinking by facilitating saving for commercial alcohol as a social good. We observe these increases primarily among men because men drink substantially more, in levels and on the margin, than women. An increase in blood pressure among men follows from the increase in drinking. Drinking was self-reported in surveys when not linked to saving because it is a prestige good stigmatized only by outsiders, but not self-reported as a use of savings because of fear, fed by rumors, that saving for alcohol would lead to confiscation of the lockbox by the NGO providing it. This fear can be viewed as a nontrivial risk of a very costly outcome—losing a lockbox which is valued as both a savings technology and a good in its own right—that the respondent can mitigate cheaply by fudging to the enumerator or CBIDSI office worker about what they plan to do with their savings. Alternative explanations do not fit the full pattern of evidence.

3. Conclusion

We ran a commitment savings RCT, in a largely autarkic native Amazonian society in the Bolivian rainforest, and found that a savings lockbox increased asset accumulation as expected. But we also found treatment effect estimates suggesting that the lockbox increased alcohol consumption and blood pressure, especially among men. This pushed against our priors, which were based on behavioral theory and the absence of any extant empirical evidence that financial interventions increase temptation consumption in low-income settings. So, we decided to probe these surprising findings using ethnographic methods, specifically scans and debriefing interviews. We learn, among other things, that alcohol is a lumpy purchase that requires effortful saving, and that it is not just a consumption good but an investment in social ties and prestige.

Our main takeaway is methodological, namely that standard ethnographic methods can complement standard approaches in economics for vetting surprising results and shedding light on potential mechanisms. There is no recipe yet for doing so, but our study provides something of a draft template and Davis et al. (2019) is another resource.

REFERENCES

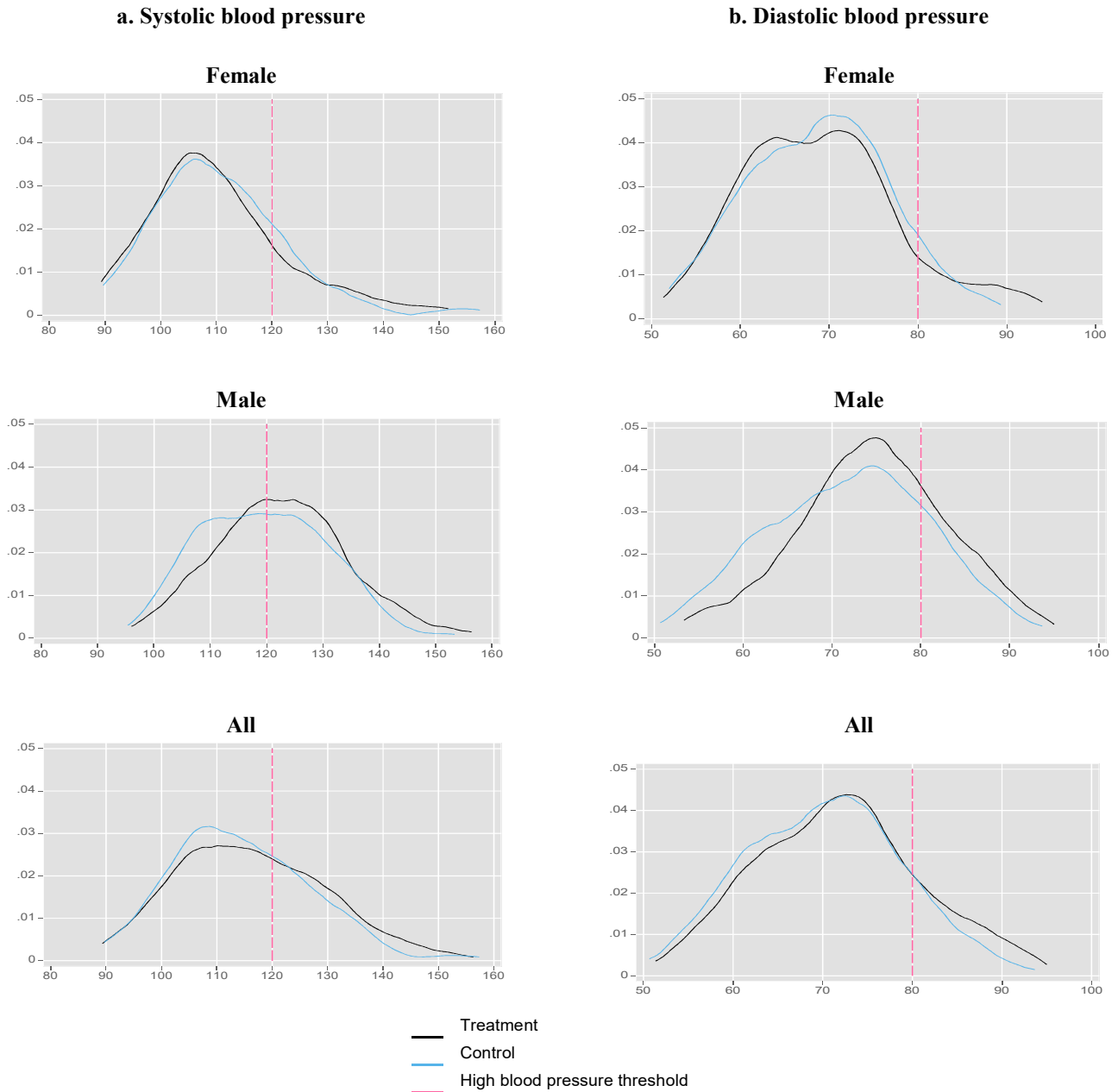
- Abraham, Justin, Merve Akbas, Dan Ariely, and Channing Jang. 2016. "Using Lotteries to Encourage Saving: Experimental Evidence from Kenya." Working Paper. <https://pdfs.semanticscholar.org/cb18/210ce2b42d42bc7c4f4787a607bc98a26794.pdf>.
- Allen, Heidi, Bill J. Wright, and Katherine Baicker. 2014. "New Medicaid Enrollees In Oregon Report Health Care Successes And Challenges." *Health Affairs* 33 (2): 292–99. <https://doi.org/10.1377/hlthaff.2013.1002>.
- Anderson, Michael. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–95. <https://doi.org/10.1198/016214508000000841>.
- Bauchet, Jonathan, Eduardo Undurraga, Ariela Zycherman, Jere Behrman, William Leonard, and Ricardo Godoy. 2021. "The Effect of Gender Targeting of Food Transfers on Child Nutritional Status: Experimental Evidence from the Bolivian Amazon." *Journal of Development Effectiveness* 13 (3): 276–91. <https://doi.org/10.1080/19439342.2021.1924833>.
- Ben-David, Itzhak, and Marieke Bos. 2021. "Impulsive Consumption and Financial Well-Being: Evidence from an Increase in the Availability of Alcohol." *The Review of Financial Studies* 34 (5): 2608–47. <https://doi.org/10.1093/rfs/hhaa094>.
- Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.
- Bergman, Peter, Raj Chetty, Stefanie Deluca, Nathan Hendren, Lawrence Katz, and Christopher Palmer. 2020. "Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice."
- Bernard, H. Russell, ed. 2015. *Handbook of Methods in Cultural Anthropology*. Second edition. Lanham: Rowman & Littlefield.
- Berry, James, Dean Karlan, and Menno Pradhan. 2018. "The Impact of Financial Education for Youth in Ghana." *World Development* 102 (February): 71–89. <https://doi.org/10.1016/j.worlddev.2017.09.011>.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. 2017. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107 (4): 1165–1206. <https://doi.org/10.1257/aer.20150503>.
- Brune, Lasse, Xavier Giné, Jessica Goldberg, and Dean Yang. 2017. "Savings Defaults and Payment Delays for Cash Transfers: Field Experimental Evidence from Malawi." *Journal of Development Economics* 129 (November): 1–13. <https://doi.org/10.1016/j.jdeveco.2017.06.001>.
- Brune, Lasse, Jason Kerwin, and Quingxiao Li. 2021. "How Important Is Temptation Spending? Maybe Less than We Thought." *World Bank Economic Review*.
- Byron, Elizabeth. 2003. "Market Integration and Health: The Impact of Acculturation on the Self-Perceived Morbidity, Diet, and Nutritional Status of the Tsimane' Amerindians of Lowland Bolivia." ProQuest Dissertations Publishing. <https://www.proquest.com/docview/305324335?pq-origsite=primo&accountid=12861>.

- Callen, Michael, Suresh de Mel, Craig McIntosh, and Christopher Woodruff. 2019. "What Are the Headwaters of Formal Savings? Experimental Evidence from Sri Lanka." *The Review of Economic Studies* 86 (6): 2491–2529. <https://doi.org/10.1093/restud/rdz020>.
- Carneiro, Pedro, Lucy Kraftman, Giacomo Mason, Lucie Moore, Imran Rasul, and Molly Scott. 2021. "The Impacts of a Multifaceted Prenatal Intervention on Human Capital Accumulation in Early Life." *American Economic Review* 111 (8): 2506–49. <https://doi.org/10.1257/aer.20191726>.
- Coffman, Lucas C., and Muriel Niederle. 2015. "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible." *Journal of Economic Perspectives* 29 (3): 81–98. <https://doi.org/10.1257/jep.29.3.81>.
- Davis, Katy, Nicole Minckas, Virginia Bond, Cari Jo Clark, Tim Colbourn, Sarah J. Drabble, Therese Hesketh, et al. 2019. "Beyond Interviews and Focus Groups: A Framework for Integrating Innovative Qualitative Methods into Randomised Controlled Trials of Complex Public Health Interventions." *Trials* 20 (1): 329. <https://doi.org/10.1186/s13063-019-3439-8>.
- Fryer, Roland G., Jr. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials *." *The Quarterly Journal of Economics* 126 (4): 1755–98. <https://doi.org/10.1093/qje/qjr045>.
- Giné, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contraction for Smoking Cessation." *American Economic Journal: Applied Economics* 2 (4): 213–35.
- Godoy, Ricardo. 2015. "A Natural Experiment in Westernization: Of Faustian Deals and the Primitive." <https://heller.brandeis.edu/gds/pdfs/working-papers/godoy-2015-9.pdf>.
- . 2020a. "Too Little Too Late: Modernization and the Lifestyle of a Native Amazonian Society in Bolivia - Brandeis University." <https://scholarworks.brandeis.edu/esploro/outputs/book/Too-little-too-late-Modernization-and/9923926194001921>.
- . 2020b. "Tsimane' Expenditures." *Brandeis University Working Paper*. <https://heller.brandeis.edu/gds/resources/working-papers.html>.
- Gupta, Clare, and Alice Bridget Kelly. 2014. "Social Relations of Fieldwork: Giving Back in a Research Setting." *Journal of Research Practice* 10 (2): E2–E2.
- Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, et al. 2010. "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment." *Science (New York, N.Y.)* 327 (5972): 1480–84. <https://doi.org/10.1126/science.1182238>.
- Hirschman, Albert O. 2014. *Development Projects Observed*. Washington: Brookings Institution Press. <https://muse.jhu.edu/book/36694>.
- Huanca, Tomás, Tomás, and SEPHIS (Programme : Netherlands). 2006. "Tsimane' Oral Tradition, Landscape, and Identity in Tropical Forest." *SEPHIS – South-South Exchange Programme for Research on the History of Development*.
- INE. 2015. "Censo de Población y Vivienda 2012, Bolivia. Características de La Población." Instituto Nacional de Estadística. https://bolivia.unfpa.org/sites/default/files/pub-pdf/Caracteristicas_de_Poblacion_2012.pdf.
- Kaplan, Hillard, Randall C. Thompson, Benjamin C. Trumble, L. Samuel Wann, Adel H. Allam, Bret Beheim, Bruno Frohlich, et al. 2017. "Coronary Atherosclerosis in Indigenous South

- American Tsimane: A Cross-Sectional Cohort Study.” *Lancet (London, England)* 389 (10080): 1730–39. [https://doi.org/10.1016/S0140-6736\(17\)30752-3](https://doi.org/10.1016/S0140-6736(17)30752-3).
- Karlan, Dean, Adam Osman, and Jonathan Zinman. 2016. “Follow the Money Not the Cash: Comparing Methods for Identifying Consumption and Investment Responses to a Liquidity Shock.” *Journal of Development Economics* 121 (July): 11–23. <https://doi.org/10.1016/j.jdeveco.2015.10.009>.
- Kempf, Dino, and Elaine Kempf. 2017. *Journey to Chimane Land*. Maitland, FL: Xulon Press.
- Mannell, Jenevieve, Katy Davis, Kohenour Akter, Hannah Jennings, Joanna Morrison, Abul Kuddus, and Edward Fottrell. 2021. “Visual Participatory Analysis: A Qualitative Method for Engaging Participants in Interpreting the Results of Randomized Controlled Trials of Health Interventions.” *Journal of Mixed Methods Research* 15 (1): 18–36. <https://doi.org/10.1177/1558689820914806>.
- Nordenskiöld, Erland. 1924. *Exploraciones y aventuras en Sudamérica*. La Paz, Bolivia: Apoyo para el Campesino Indígena del Oriente Boliviano (APCOB). <http://books.google.com/books?id=4I7jAAAAMAAJ>.
- Nyberg, Colleen. 2009. “Market Integration, Stress, and Health: An Exploration of Hypothalamic-Pituitary-Adrenal Dynamics among the Tsimane’ of the Bolivian Amazon.” *Northwestern University Department of Anthropology Dissertation*.
- Reyes-García, Victoria. 2001. “Indigenous People, Ethnobotanical Knowledge, and Market Economy: A Case Study of the Tsimané Amerindians in Lowland Bolivia.” ProQuest Dissertations Publishing. <https://www.proquest.com/docview/304694064?pq-origsite=primo&accountid=12861>.
- Reyes-García, Victoria, Martí Orta-Martínez, Maximilien Gueze, Ana C. Luz, Jaime Paneque-Gálvez, Manuel J. Macía, and Joan Pino. 2012. “Does Participatory Mapping Increase Conflicts? A Randomized Evaluation in the Bolivian Amazon.” *Applied Geography* 34 (May): 650–58. <https://doi.org/10.1016/j.apgeog.2012.04.007>.
- Ringhofer, Lisa. 2010. *Fishing, Foraging and Farming in the Bolivian Amazon: On a Local Society in Transition*. Springer Netherlands. <https://doi.org/10.1007/978-90-481-3487-8>.
- Schilbach, Frank. 2019. “Alcohol and Self-Control: A Field Experiment in India.” *American Economic Review* 109 (4): 1290–1322. <https://doi.org/10.1257/aer.20170458>.
- Somville, Vincent, and Lore Vandewalle. 2019. “Access to Banking, Savings and Consumption Smoothing in Rural India.” Working Paper HEIDWP09-2019. International Economics Department Working Paper Series. Graduate Institute of International and Development Studies.
- Townsend, R. M. 1995. “Financial Systems in Northern Thai Villages.” *The Quarterly Journal of Economics* 110 (4): 1011–46. <https://doi.org/10.2307/2946647>.
- Turner, Bethany, and Haagen Klaus. 2020. *Diet, Nutrition, and Foodways on the North Coast of Peru: Bioarchaeological Perspectives on Adaptive Transitions*. Bioarchaeology and Social Theory. Springer International Publishing. <https://doi.org/10.1007/978-3-030-42614-9>.
- Undurraga, Eduardo A., Veronica Nica, Rebecca Zhang, Irene C. Mensah, and Ricardo A. Godoy. 2016. “Individual Health and the Visibility of Village Economic Inequality: Longitudinal Evidence from Native Amazonians in Bolivia.” *Economics & Human Biology* 23 (December): 18–26. <https://doi.org/10.1016/j.ehb.2016.06.004>.
- Undurraga, Eduardo A., Ariela Zycherman, Julie Yiu, Taps Bolivia Study Team, and Ricardo A. Godoy. 2014. “Savings at the Periphery of Markets: Evidence from Forager-Farmers in

- the Bolivian Amazon.” *The Journal of Development Studies* 50 (2): 288–301.
<https://doi.org/10.1080/00220388.2013.833322>.
- Vadez, Vincent, Victoria Reyes-Garcia, Tomás Huaca, and William R Leonard. 2008. “Cash Cropping, Farm Technologies, and Deforestation: What Are the Connections? A Model with Empirical Data from the Bolivian Amazon.” *Human Organization* 67 (4): 384–96.
- Zycherman, Ariela. 2013. “The Changing Value of Food: Localizing Modernity among the Tsimané Indians of Lowland Bolivia.” <https://doi.org/10.7916/D8GT5VCT>.
- . 2015. “Shocdye’ as World: Localizing Modernity Through Beer in the Bolivian Amazon.” *Food, Culture & Society* 18 (1): 51–69.
<https://doi.org/10.2752/175174415X14101814953684>.

Figure 1. Distribution of blood pressure measurements at endline, by sex of recipient and treatment arm



Notes: Each plot shows the distribution of the corresponding blood pressure measurement (unit: mm Hg) at endline for the sample indicated in the heading. Sample size at endline is equal to 431 for male recipients (291 in any treatment and 140 in control) and 487 for female recipients (307 in any treatment and 180 in control). Each blood pressure measurement (i.e., systolic and diastolic blood pressure) is the average of three measurements taken by a team of specialists at endline (2012). According to the American Heart Association standards, an individual is considered to have elevated blood pressure if systolic blood pressure is at least 120 mm Hg, and hypertension if systolic blood pressure is at least 130 mm Hg or diastolic blood pressure is at least 80 mm Hg. We define "high blood pressure" as having systolic blood pressure \geq 120 mm Hg or diastolic blood pressure equal to or above 80 mm Hg (see Appendix D).

Table 1. Review of Savings Encouragement RCTs in Developing Countries

Paper	Temptation spending definition, if measured	Estimated average treatment effect (SE) if any	Savings product illiquidity features, if any	Country	Sample description
Abarcar et al. (2019)				Philippines	urban, all genders
Abebe et al. (2018)				Ethiopia	urban, all genders, entrepreneurs
Abraham et al. (2016)	Gambling more than usual, binary [1]	0.06 (0.05)		Kenya	urban, all genders
Aggarwal et al. (2020)				Malawi	urban, all genders, entrepreneurs
Aker et al. (2020)				Niger	rural, all genders
Ashraf et al. (2015)				El Salvador and USA	urban, all genders
Ashraf et al. (2006a)				Philippines	rural, all genders
Ashraf et al. (2006b)				Philippines	rural, all genders
Atkinson et al. (2013)				Guatemala	urban + rural, all genders, entrepreneurs
Attanasio et al. (2019)				Colombia	urban + rural, all genders
Avdeenko et al. (2019)				Ethiopia	rural, all genders, farmers
Banerjee et al. (2020)				Ghana	urban + rural, all genders, households
Bastian et al. (2018)				Tanzania	urban, women, microentrepreneurs
Battista and Vicente (2020)				Mozambique	rural, all genders, farmers
Beaman et al. (2014)				Mali	rural, all genders, households
Berry et al. (2018)	Temptation goods index [2]	-0.027 (0.042)		Ghana	urban + rural, all genders, high school students
Blumenstock et al. (2018)				Afghanistan	urban + rural, all genders
Bruno et al. (2016)				Malawi	rural, all genders, farmers
Bruno et al. (2017)	Alcohol+tobacco, fats+sugars, prepared food [3]	Reported in graph	Withdrawal restriction	Malawi	urban + rural, all genders, households
Bruno et al. (2021)				Malawi	rural, all genders
Callen et al. (2019)	Alcohol+tobacco+gambling [4]	0.64 (65.2)		Sri Lanka	rural, all genders, households
Carter et al. (2016)				Mozambique	rural, all genders, households
Cole et al. (2011)				India, Indonesia	rural (India), urban+rural (Indonesia), all genders, households
de Mel et al. (2020)				Sri Lanka	urban, all genders, households
Dizon et al. (2020)				Kenya	urban + rural, vulnerable women
Dupas and Robinson (2013a)				Kenya	rural, all genders, small-scale businesses

Search methodology: (a) Search for completed trials with keywords "savings" and "account or access or encouragement" in the AEA RCT Registry; (b) Add any RCTs not in (a) but covered in Knowles (2018), a meta-analysis on formal savings accounts; and (c) Add any RCTs not in (a) or (b) but in the Innovations for Poverty Action publications registry (filtering on published papers or working papers and "savings" as the topic). The IPA registry can be found at: <https://www.poverty-action.org/publications>.

Study-specific notes on temptation spending definition and measurement:

[1] Gamble more=1 if respondent reports gambling more than they usually do after the savings program. Paper also report effects on l= gamble less, l=more tempted to gamble. The effect we report above is for the lottery treatment arm.

[2] Standardized index of amount spent on snacks, non-food goods, and entertainment in last 7 days, and amount student would spend on fun if given 5 cedis.

[3] Reported using different definitions of spending in the week following a cash windfall: (1) alcohol and tobacco; (2) fats and sugars + (1); and (3) prepared foods sold by vendors + (2). Treatment effects are only shown graphically and are close to 0 regardless of the definition used: "The difference in spending is always less than USD 1, and accounts for one percent or less of the total transfer."

[4] Spending on alcohol, tobacco, and gambling in the past month, measured by directly asking participants and indirect elicitation. Effect reported in Sri Lankan Rupees (LKR).

Table 1 cont. Review of Savings Encouragement RCTs in Developing Countries

Paper	Temptation spending definition, if measured	Estimated average treatment effect (SE) illiquidity features, if any	Country	Sample description
Dupas and Robinson (2013b)		Savings product	Kenya	rural, all genders, households
Dupas et al. (2012)			Kenya	rural, all genders, households
Dupas et al. (2018)			Uganda, Malawi, Chile	rural, all genders
Flory (2018)			Malawi	rural, all genders, households
Gertler et al. (2017)			Mexico	urban + rural, all genders
Habyarimana and Jack (2018)			Kenya	urban + rural, all genders, students
Horn et al. (2021)			Uganda	urban, all genders, youth clubs
John (2020)	Withdrawal restriction		Philippines	rural, all genders
Karlan and Linden (2014)	Withdrawal restriction		Philippines	peri-urban + rural, all genders
Karlan and Zinman (2018)	Withdrawal restriction		Philippines	peri-urban + rural, all genders
Karlan et al. (2016)				
Site 1: The Philippines	Withdrawal restriction		Philippines	rural + small urban, all genders
Site 2: Peru			Peru	urban + rural, all genders
Site 3: Bolivia			Bolivia	urban, all genders
Kast and Pomeranz (2014)	Lose higher interest rate and insurance		Bolivia	urban, mostly women, microentrepreneurs
Kast et al. (2018)			Chile	
Study 1: Peer Groups			Chile	urban, mostly women, microentrepreneurs
Study 2: Feedback Messages			Chile	urban, mostly women, microentrepreneurs
Laajaj (2017)			Mozambique	rural, all genders, farming households
Lipscomb and Schechter (2018)	Earmarked for sanitation product purchase		Senegal	urban, all genders, households
Prina (2015)			Nepal	urban, women
Salas (2015)	VSLA model, can borrow from fund		Colombia	mostly rural, all genders, households
Schaner (2017)	Withdrawal fees		Kenya	rural, all genders
Schaner (2018)	Withdrawal fees		Kenya	rural, married couples
Somville and Vandewalle (2019)	Goods other than necessities [5]	0.94 (8.75)	India	rural, all genders
Supanantarook et al. (2017)			Uganda	all genders, primary school children

Search methodology: (a) Search for completed trials with keywords "savings" and "account or access or encouragement" in the AEA RCT Registry; (b) Add any RCTs not in (a) but covered in Knowles (2018), a meta-analysis on formal savings accounts; and (c) Add any RCTs not in (a) or (b) but in the Innovations for Poverty Action publications registry (filtering on published papers or working papers and "savings" as the topic). The IPA registry can be found at: <https://www.poverty-action.org/publications>.

Study-specific notes on temptation spending definition and measurement:

[5] Spending on paan (a preparation chewed for stimulant effects), alcohol, tobacco, drinks and snacks from the market, hair oil, lotion and perfumes, measured weekly over 17 weeks. Effect reported in Indian rupees (INR).

Table 2. A search of papers in top-5 journals, covering the last 5+ years, does not find any using qualitative methods to explore explanations for an RCT result

Paper	Journal	Issue	Qualitative follow-up?	Paper	Journal	Issue	Qualitative follow-up?
Carvalho et al. (2016)	AER	Vol 106 No. 2	No	Finkelstein and Notowidigdo (2019)	QJE	Vol. 134 No. 3	No
Deming et al. (2016)	AER	Vol 106 No. 3	No	Coffman et al. (2019)	QJE	Vol. 134 No. 4	No
Chetty et al. (2016)	AER	Vol 106 No. 4	No	Jones et al. (2019)	QJE	Vol. 134 No. 4	No
Kline and Tartari (2016)	AER	Vol 106 No. 4	No	Agan and Starr (2018)	QJE	Vol. 133 No. 1	No
Neumark et al. (2016)	AER	Vol 106 No. 5	No	Ree et al. (2018)	QJE	Vol. 133 No. 2	No
Farber et al. (2016)	AER	Vol 106 No. 5	No	Bursztyn et al. (2018)	QJE	Vol. 133 No. 3	No
Alpert et al. (2016)	AER	Vol 106 No. 5	No	Fowlie et al. (2018)	QJE	Vol. 133 No. 3	No
Bartoš et al. (2016)	AER	Vol 106 No. 6	No	Heller et al. (2017)	QJE	Vol. 132 No. 1	No
Blattman et al. (2017)	AER	Vol 107 No. 4	No	Atkin et al. (2017)	QJE	Vol. 132 No. 2	No
Shimeles et al. (2017)	AER	Vol 107 No. 5	No	Bandiera et al. (2017)	QJE	Vol. 132 No. 2	No
Goldin et al. (2017)	AER	Vol 107 No. 5	No	Khan et al. (2016)	QJE	Vol. 131 No. 1	No
Bowers et al. (2017)	AER	Vol 107 No. 5	No	Araujo et al. (2016)	QJE	Vol. 131 No. 3	No
Pons (2018)	AER	Vol 108 No. 6	No	Haushofer and Shapiro (2016)	QJE	Vol. 131 No. 4	No
Blumenstock et al. (2018)	AER	Vol 108 No. 10	No	Neve et al. (2021)	JPE	Vol. 129 No. 5	No
Casaburi and Willis (2018)	AER	Vol 108 No. 12	No	Bergman (2021)	JPE	Vol. 129 No. 1	No
Khan et al. (2019)	AER	Vol 109 No. 1	No	Brownback and Sadoff (2020)	JPE	Vol. 128 No. 8	No
Casaburi and Macchiavello (2019)	AER	Vol 109 No. 2	No	Cappelen et al. (2020)	JPE	Vol. 128 No. 7	No
Schilbach (2019)	AER	Vol 109 No. 4	No	Gosnell et al. (2020)	JPE	Vol. 128 No. 4	No
Elias et al. (2019)	AER	Vol 109 No. 8	No	Berry et al. (2020)	JPE	Vol. 128 No. 4	No
Bohren et al. (2019)	AER	Vol 109 No. 10	No	Lee et al. (2020)	JPE	Vol. 128 No. 4	No
Kessler et al. (2019)	AER	Vol 109 No. 11	No	Bursztyn et al. (2019)	JPE	Vol. 127 No. 4	No
Alsan et al. (2019)	AER	Vol 109 No. 12	No	Neumark et al. (2019)	JPE	Vol. 127 No. 2	No
Attanasio et al. (2020)	AER	Vol 110 No. 1	No	Banerjee et al. (2019)	JPE	Vol. 127 No. 1	No
Romero et al. (2020)	AER	Vol 110 No. 2	No	Alan and Ertac (2018)	JPE	Vol. 126 No. 5	No
Baranov et al. (2020)	AER	Vol 110 No. 3	No	Banerjee et al. (2018)	JPE	Vol. 126 No. 2	No
Dobbie and Song (2020)	AER	Vol 110 No. 4	No	Bruhñ et al. (2018)	JPE	Vol. 126 No. 2	No
Andrabi et al. (2020)	AER	Vol 110 No. 10	No	Perez-Truglia and Cruces (2017)	JPE	Vol. 125 No. 4	No
Armand et al. (2020)	AER	Vol 110 No. 11	No	Andreoni et al. (2017)	JPE	Vol. 125 No. 3	No
Bergquist and Dinerstein (2020)	AER	Vol 110 No. 12	No	Pallais and Sands (2016)	JPE	Vol. 124 No. 6	No
Levy (2021)	AER	Vol. 111 No. 3	No	Alatas et al. (2016)	JPE	Vol. 124 No. 2	No
Casey et al. (2021)	AER	Vol. 111 No. 5	No	Bó et al. (2021)	EMCA	Vol. 89 No. 2	No
Hjort et al. (2021)	AER	Vol. 111 No. 5	No	Kline and Walters (2021)	EMCA	Vol. 89 No. 2	No
Domurat et al. (2021)	AER	Vol. 111 No. 5	No	Alfonsi et al. (2020)	EMCA	Vol. 88 No. 6	No
Dynarski et al. (2021)	AER	Vol. 111 No. 6	No	Breza and Chandrasekhar (2019)	EMCA	Vol. 87 No. 1	No
Abebe et al. (2021)	AER	Vol. 111 No. 6	No	Duflo et al. (2018)	EMCA	Vol. 86 No. 6	No
Lowe (2021)	AER	Vol. 111 No. 6	No	Sahni and Nair (2020)	REStud	Vol. 87 No. 3	No
Beaman et al. (2021)	AER	Vol. 111 No. 6	No	Sadoff et al. (2020)	REStud	Vol. 87 No. 4	No
Leaver et al. (2021)	AER	Vol. 111 No. 7	No	Armona et al. (2019)	REStud	Vol. 86 No. 4	No
Field et al. (2021)	AER	Vol. 111 No. 7	No	Belot et al. (2019)	REStud	Vol. 86 No. 4	No
Cameiro et al. (2021)	AER	Vol. 111 No. 8	No, but see notes	Karlan and Zinman (2019)	REStud	Vol. 86 No. 4	No
Goldin et al. (2021)	QJE	Vol. 136 No. 1	No	Banerjee et al. (2019)	REStud	Vol. 86 No. 6	No
Bryan et al. (2021)	QJE	Vol. 136 No. 1	No	Callen et al. (2019)	REStud	Vol. 86 No. 6	No
Dahl et al. (2021)	QJE	Vol. 136 No. 2	No	Taubinsky and Rees-Jones (2018)	REStud	Vol. 85 No. 4	No
Burchardi et al. (2019)	QJE	Vol. 134 No. 1	No	Dellavigna et al. (2017)	REStud	Vol. 84 No. 1	No
Burke et al. (2019)	QJE	Vol. 134 No. 2	No	Booij et al. (2017)	REStud	Vol. 84 No. 2	No
Cantoni et al. (2019)	QJE	Vol. 134 No. 2	No	Jakiela and Ozier (2016)	REStud	Vol. 83 No. 1	No
Mbiti et al. (2019)	QJE	Vol. 134 No. 3	No	Beath et al. (2016)	REStud	Vol. 83 No. 3	No
Alan et al. (2019)	QJE	Vol. 134 No. 3	No	Greaney et al. (2016)	REStud	Vol. 83 No. 4	No

Search Methodology: Research assistant looked through the titles of every article in every issue from January 2016 - August 2021 inclusive for the above journals. They skipped any that were clearly not RCTs, and checked the abstracts to confirm that others were actually RCTs before searching through the paper for any qualitative follow-up to explore explanations for a RCT result.

Study specific notes on qualitative methods:

Cameiro et al. do use: "a parallel stream of qualitative analysis" to corroborate quantitative findings and develop hypotheses to test in future work (p. 2545).

Table 3. Savings box usage, as measured in endline survey

	1	2	3	4	5	6	7
	Full sample	Male	Female	Female - Male p-value	Lockbox w/ key (T1)	Lockbox w/o key (T2)	T2 - T1 p-value
Panel A: Still have at endline?							
...box?	0.95 (0.01)	0.95 (0.01)	0.95 (0.01)	0.350	0.96 (0.01)	0.94 (0.01)	0.452
...key? T1	0.77 (0.03)	0.80 (0.03)	0.74 (0.04)	0.275	0.77 (0.03)	-	-
Panel B: Opened box since getting it?							
Yes	0.61 (0.02)	0.62 (0.03)	0.59 (0.03)	0.710	0.87 (0.02)	0.35 (0.03)	0.000
Mean number of times, including zeros	2.46 (0.18)	2.46 (0.26)	2.46 (0.24)	0.732	4.26 (0.32)	0.66 (0.06)	0.000
Panel C: Money in box at endline?							
Yes	0.46 (0.02)	0.46 (0.03)	0.47 (0.03)	0.759	0.47 (0.03)	0.46 (0.03)	0.459
Mean amount, including zeros	103.22 (11.70)	90.47 (15.17)	115.72 (17.77)	0.340	104.43 (15.12)	102.08 (17.74)	0.912
<i>N</i>	598	291	307		292	306	

Notes: Huber-White standard errors in parentheses. The p-values reported in columns 4 and 7 are calculated using ordinary least squares (OLS) regressions that include our randomization strata. Variables in Panel C were verified by the surveyor and set to missing in 23 cases where the respondent balked at verification. The monetary amounts reported in Panel C are in bolivianos (Bs; PPP US\$1 ≈ 3.045Bs in 2012).

Table 4. Responses to open-ended questions on savings uses

	<i>Survey</i>	<i>Baseline</i>	<i>Endline</i>
Use	Count (% of total)		
Clothing		229 (33%)	71 (12%)
Houseware		86 (12%)	26 (4%)
Transport		83 (12%)	22 (4%)
Medicine & hygiene		80 (11%)	111 (19%)
Tools		72 (10%)	36 (6%)
NA / DK		53 (7%)	30 (5%)
Food		43 (6%)	80 (13%)
Livestock		24 (3%)	2 (0%)
Construction materials		17 (2%)	8 (1%)
Others, not including temptation good		15 (2%)	17 (3%)
Temptation good		0 (0%)	1 (0%)
Did not spend saved money		N/A	128 (21%)
Did not use box		N/A	45 (8%)
Stolen/lost box or key		N/A	21 (2%)
	<i>N respondents</i>	702	598

Notes: Questions were asked only to those assigned a lockbox. The baseline question was "What are you saving for?", and the endline question was "On what did you spend money from the box?". Responses are open-ended and were classified by a research assistant. Respondents could name > 1 use, but did so relatively rarely so we include only the first use mentioned. 16.7% of respondents named the same first use in both baseline and endline

Table 5. Estimated lockbox impacts on main outcomes

	1	2	3	4	5	6	7	8
	Value of financial assets (Bs)	Value of physical assets (Bs)	Agricultural investments index	Change in income sources	Total expenditures (Bs)	Temptation consumption index	Blood pressure index	l=Fight over money (past 2 months)
Panel A: Pooled treatment								
Any Treatment	77.75 (24.81) [0.015]	213.00 (120.05) [0.153]	0.08 (0.09) [0.297]	0.10 (0.07) [0.231]	9.76 (31.06) [0.754]	0.16 (0.07) [0.084]	0.13 (0.06) [0.063]	0.02 (0.02) [0.597]
Panel B: By sex of recipient								
Any treatment (T) x Female (F)	96.28 24.30 [0.002]	96.02 136.37 [0.643]	0.05 0.10 [0.690]	0.16 0.09 [0.354]	-28.23 38.70 [0.643]	0.06 0.05 [0.543]	0.04 0.07 [0.690]	0.01 0.03 [0.864]
Any treatment (T) x Male (M)	52.29 (46.07) [0.543]	260.53 (193.02) [0.543]	0.11 (0.10) [0.543]	0.02 (0.11) [0.864]	45.31 (46.23) [0.582]	0.25 (0.14) [0.354]	0.24 (0.09) [0.043]	0.03 (0.03) [0.643]
p-value for T x F = T x M	0.401	0.489	0.694	0.326	0.210	0.212	0.075	0.658
Panel C: By treatment arm								
Lockbox w/ Key (T1)	83.81 (28.89) [0.061]	289.82 (136.71) [0.131]	0.10 (0.08) [0.400]	0.13 (0.08) [0.231]	-1.65 (39.39) [0.967]	0.17 (0.09) [0.193]	0.13 (0.06) [0.131]	0.01 (0.03) [0.751]
Lockbox w/o Key (T2)	72.04 (30.09) [0.131]	140.28 (143.93) [0.526]	0.07 (0.08) [0.535]	0.07 (0.08) [0.526]	20.83 (35.00) [0.631]	0.15 (0.09) [0.193]	0.13 (0.07) [0.131]	0.02 (0.03) [0.580]
p-value for T1 = T2	0.906	0.357	0.693	0.492	0.605	0.847	0.959	0.763
Panel D: Summary information								
# Observations in T1	292	292	292	292	275	292	292	292
# Observations in T2	306	306	306	306	280	306	306	306
# Observations in Control	320	320	320	320	298	320	320	320
Mean of control group at endline	128.90	3010.00	0.00	1.40	341.20	0.00	0.00	0.12
Std. dev. of control group at endline	354.60	2635.00	1.00	1.06	391.80	1.00	1.00	0.33
Share > 0 in control group at endline	0.51	1.00	-	0.80	1.00	-	-	0.12

Notes: OLS intent-to-treat estimates with robust standard errors in parentheses and q-values in brackets. q-values report the minimum false discovery rate (i.e., the expected proportion of rejected null hypotheses that are actually true) at which the null hypothesis would be rejected for each test in each Panel (8 tests in Panel A, 16 tests in Panels B, and 16 tests in Panel C). Each column-panel reports results for a single OLS regression of the dependent variable listed in the column heading on (i) the treatment groups listed in the row headings (control group is omitted), (ii) the baseline value of the dependent variable, (iii) an indicator for whether the baseline value of the dependent variable is missing, and (iv) our randomization strata: a dummy variable for each village. Regressions in Panel B also control for whether the recipient is female. Regressions in column 4 control for the baseline number of sources of income instead of the baseline value of the dependent variable. Regressions in column 7 include controls for whether the recipient is female and the recipient's age. Regressions in column 8 do not control for the baseline value of the dependent variable because it was only collected at endline. All monetary values are in bolivianos (Bs; PPP US\$1 ≈ 3.045Bs in 2012). All indices are standardized with respect to the control group in the same time period. For a detailed description of the outcomes and their components please see Appendix D.