Retrieval Failures and Consumption Smoothing: A Field Experiment on Seasonal Poverty*

Ned Augenblick† B. Kelsey Jack Supreet Kaur Felix Masiye Nicholas Swanson

September 22, 2023

Abstract

Individuals may fail to recall and use information they already know when making decisions. We empirically investigate whether such "retrieval failures" distort consumption smoothing behavior among Zambian farmers, who derive their income from one annual harvest and then spend it down over the course of the year. We document that individuals underestimate upcoming spending by 50%, creating scope for under-saving. In order to improve recall, we randomize an intervention that prompts individuals to think through their future expenses associatively in categories—without providing any external information or guidance. Treated individuals increase "remembered" expenses by 42%; as predicted by the memory literature, effects are concentrated among small, irregular, and stochastic items. Immediate spending drops and, two months after the intervention, treated households hold 15% higher savings. They subsequently enter the "hungry season"—the final months of the year when consumption typically declines sharply—with one additional month of savings, leading to a flatter spending profile over the year. Households use the increased savings to self-finance additional farm investment, resulting in a 9% increase in the next year's crop revenue. We replicate the intervention's impact on beliefs among low-income Americans, suggesting that retrieval failures generalize across settings and populations.

*We thank Abhijit Banerjee, John Conlon, Xavier Gabaix, Muriel Niederle, Matthew Rabin, Andrei Shleifer, Joshua Schwartzstein, Dmitry Taubinsky, and numerous conference and seminar audiences for helpful comments and feedback. We gratefully acknowledge financial support from the National Science Foundation (Kaur CAREER SES 1848452), the Weiss Family Program for Research in Development Economics, Development Innovation Ventures at the United States Agency for International Development (USAID), Feed the Future through USAID, and the International Growth Center. Akanksha Arora, Anderson Benson, Mathieu Chabaud, Harrison Chang, Anne Fogarty, David Lyu and Jingkai Ong provided excellent research assistance. This research was approved by Berkeley IRB (2018-07-11284), and by the UNZABREC (012-09-18). This study was registered on the AEA RCT registry, RCT ID AEARCTR-0004409.

†Augenblick: UC Berkeley (augenblick@berkeley.edu); Jack: UC Santa Barbara and NBER (kelseyjack@ucsb.edu); Kaur: UC Berkeley and NBER (supreet@berkeley.edu); Masiye: University of Zambia (dean-hss@unza.zm); Swanson: UC Berkeley (nicholas.swanson@berkeley.edu).

1 Introduction

Economic models often assume that if an individual knows a piece of information, she will use it when making decisions. However, the limits of human cognition—such as imperfect memory—suggest that this assumption may not hold: knowing something does not necessarily mean it will be retrieved and utilized. This paper empirically examines the possibility that such retrieval failures are consequential for decision-making and behavior, even at high stakes.

We focus on a classic economic problem: deciding how much to spend today and how much to save for tomorrow's expenses. Solving this problem requires that individuals consider myriad future expenses, from large and certain ones (such as a car payment), to those that are small, irregular, and uncertain (such as gas, maintenance, and repair). Even though many of these expenses can be retrieved with thought or external prompting, our core hypothesis is that some of these expenses are not recalled when making decisions. In contrast, we posit that individuals have fewer issues when retrieving information about their inflows, which generally arise from a few large and consistent sources.[1] Using a simple model, we illustrate how this asymmetry causes an individual to overestimate future savings and overspend today, such that she must cut back on future consumption when the neglected expenses (i.e., those she had previously failed to retrieve) become due. Consequently, an intervention that boosts the retrieval of more future expenses causes her to cut back on spending today, which allows her to spend more in the future.

We focus on the savings decisions of Zambian farmers, who face a particularly stark version of the savings problem. In this setting, farmers harvest maize once a year. This harvest accounts for over 90% of annual household income and must cover all expenses until the next harvest. Consumption cycles are pronounced: while 98% of households have ample food right after the harvest, over 50% report difficulty meeting basic needs in the months before the next harvest, a period known as locally as "the hungry season" (Fink et al., 2020). Such cycles—where consumption fluctuates predictably and repeatedly with income flows—are not unique to Zambia: they are ubiquitous in poor countries (e.g., Paxson, 1992; Dercon and Krishnan, 2000; Bryan et al., 2014; Basu and Wong, 2015; Merfeld and Morduch, 2023) and also among low-income individuals in rich countries (Shapiro, 2005; Pew Charitable Trusts, 2016; Kuhn, 2021).[2]

To motivate the potential relevance of retrieval failures, Figure 1 documents large systematic bias in beliefs: farmers overestimate future savings and underestimate future expenses. Early in the agricultural year, we ask farmers to make incentivized forecasts about their future

[1] This is likely to be true even in low-income settings, where income is highly volatile. For example, while a vegetable vendor's income may fluctuate day to day, the sheer array of certain and uncertain expenses will likely be larger and more volatile.

[2] For example, in the US, SNAP recipients decrease calories by 10-15% from the start to the end of the month (Shapiro, 2005).

savings (i.e., maize stock). 78% of individuals think they will have higher savings 3 months later (at the start of the hungry season) than they actually do, with the average participant overestimating future savings by 81%. In addition, we ask individuals to predict their "worst case scenario": how much savings they will have left in the future if "everything that could possibly go wrong does go wrong." Strikingly, more than 60% of farmers end up with less savings than their worst-case savings forecast. They are similarly overoptimistic about savings levels 5 months in the future (the middle of the hungry season).

This overoptimism about savings coincides with substantial under-estimation of future expenses. We specifically elicit predictions about non-food expenses, which are more irregular, stochastic, and less salient than food expenses—and therefore more likely to be forgotten (Mason et al., 2023; Bordalo et al., 2023).[3] On average, farmers actually end up spending twice as much as their initial forecast on non-food expenses. While such beliefs could be explained by various potential explanations, such as naive present focus, note that these expenses are largely comprised of items such as school fees, farm inputs, and medical costs—items that are unlikely to be temptation goods.

To more cleanly test our hypothesis, we design an intervention that helps individuals retrieve information they already "know" (i.e., from their own memory). We draw heavily on the psychology literatures on the planning fallacy (Kahneman and Tversky, 1977; Buehler et al., 2010) and associative memory (Kahana, 2012; Bordalo et al., 2023), which present robust evidence that individuals are more likely to remember items if they are asked to recall them in finer categories. For example, a farmer is more likely to remember a future seed purchase when asked to think about "farm inputs" rather than "expenditures" as a whole. We leverage this idea to design an "expense board" that shows pictures depicting seven broad categories of expenses (e.g., food consumption, farm inputs, household supplies, medical shocks). Early in the agricultural year (about 3 months after harvest), we ask farmers to think through their expenses in each category, and allocate their maize stock (i.e., savings) across the categories.[4] Importantly, we do not provide any assistance, guidance, or normative advice on the allocation. Consequently, the expense board provides individuals with a tool to more readily retrieve and use information from their mind through their own cognitive effort.

We design two complementary experiments: a shorter "mechanism experiment" and a longer "field experiment," which trade off insight into what is happening in participants' minds with the ability to track changes in longer-run behavior. Each experiment is conducted with a separate sample of farmers in Zambia.

[3]In our setting, food consumption is generally the largest expense and occurs every day in the same form of standardized nshima patties. In our preliminary interviews with farmers, food consumption was almost universally discussed first when considering spending.

[4]To promote cognitive engagement with the exercise, we allot farmers thumbtacks equal to the number of maize bags they have in savings, and ask them to stick the thumbtacks into the boxes corresponding to each category to reflect their spending plan.

The mechanism experiment isolates the impact of thinking through finer categories on information retrieval. In the mechanism experiment, participants in both the treatment and control groups allocate their available savings to an expense board, but we vary the level of aggregation of the categories on the board. The control group receives a "placebo" board, with only two categories: food expenses and non-food expenses. The treatment group receives the full board: twelve boxes depicting food consumption in each month and six separate categories of non-food expenses. We predict that the full expense board will lead to increased information retrieval, especially of items that are a priori more likely to be neglected: i.e., non-food expenses, and especially those that are more irregular and less salient.

We randomize 197 farmers to receive either the full or placebo board. Relative to individuals' prior at baseline, the placebo expense board leads to no detectable changes in beliefs about non-food expenses. This suggests that simply going through the motions of planning alone does not change beliefs. Rather, consistent with our hypothesis, the treatment expense board has large impacts: treated farmers expect to spend 42% more on non-food expenses than the control ($p<0.001$). Our model predicts that such changes in beliefs will make individuals feel "poorer," raising the shadow price of money and therefore lowering spending today. Consistent with this prediction, in a real-stakes opportunity to buy a discretionary consumption good (e.g., new clothing), treated individuals' willingness to pay for the good falls by 37% relative to the control group ($p<0.001$).

Finally, to characterize which expenses are being neglected, we undertake a final set of activities with the control group only. We first ask participants to list the specific expenses that comprise their non-food allocation in the two category placebo board. After this, they complete the more detailed full treatment expense board, and then again list the items that now comprise their allocations under the full board. Similar to the between-subjects treatment effect, the control group raises their non-food allocation by 38% ($p<0.001$) after completing the full board. The set of new items listed following completion of the more detailed board are informative of which expenses were previously neglected. Consistent with a memory channel, the previously-neglected expenses are smaller in magnitude, more irregular, and more stochastic.

At the end of the mechanism experiment, the control group has considered many expenses and has effectively been treated. Therefore, it is not possible to test long-term behavioral changes between treatment and control. For that, we turn to the field experiment, where we adopt two core design changes relative to the mechanism experiment. First, in order to avoid contamination of the control, we do not ask the control group to do any retrieval exercise. Second, under our hypothesized mechanism of retrieval failures, a treated farmer may not remember specific items recalled (or specific plans made) many months later; consequently, we offer farmers expense labels (corresponding to the categories on the expense board) two months after the intervention. Farmers can affix these labels to their maize bags in order to visually depict their spending plan. Importantly, we offer these labels to both the treatment and control groups to mitigate the concern that the labels provide a previously-unavailable technology for reminders or soft commitment.[5] We intentionally delayed the implementation of the labels by two months to enable us to test for the effect of the retrieval exercise alone over a substantial time horizon.

The immediate impacts of the longer field experiment match those from the mechanism experiment. First, in the plan developed through the retrieval exercise, treated farmers indicate that they expect to spend 62% more on non-food expenditures relative to their baseline forecast ($p < 0.001$). Second, treatment participants are, on average, willing to pay 34% less for discretionary consumption goods ($p < 0.001$).[6]

Our theory predicts that the change in beliefs should generate a flatter consumption profile over the annual cycle: reduced spending in early months leads to increased savings. Consistent with this prediction, two months after the retrieval exercise—before labels are attached—treated farmers hold 15% more savings than do control farmers (p=0.026).[7] The treatment group continues to hold increased savings in later months. Consequently, treated farmers enter the hungry season with 20% more savings than the control (p=0.018). This effect size corresponds to the amount the average control household spends in one month on total expenses (food and non-food items) during the hungry season. A higher savings stock enables treated farmers to engage in more spending in later months in the cycle, leading to a smoother spending profile over the year.

In our setting, the effects of increased savings have implications not only for welfare, but also for productivity. Half of the control group completely exhausts their initial maize stock before the end of the hungry season. To raise cash to cover immediate consumption needs, households divert labor away from their own farms to do casual wage labor (Fink et al., 2020). Our intervention reduces the need for this behavior: the treatment group is 42% less likely to sell household labor to others during the hungry season (p=0.022). In addition, we see suggestive evidence that treated farmers use their increased savings stock to finance investment in their farms. For example, they have higher spending on farm inputs, including hired labor (p=0.082) and fertilizer and other chemical inputs (p=0.127). Consequently, the treatment group's crop revenue from the following harvest is 9% higher than the control (p=0.095), leading them to enter the following year with a substantively larger pie.

Together, these results provide consistent support for our model of retrieval failures. In

[5]We offered all farmers the choice between the labels or a small compensation (a bag of sugar) at baseline. Moreover, we explicitly told control participants that some individuals find it helpful use the labels to visually record their spending plan for the year. Treated individuals are substantially more likely to take up the labels (80% vs. 29%), consistent with them recognizing greater value in recording their plan than control farmers.

[6]Choices in the WTP exercise are implemented only for a subset of participants so that the baseline distribution of savings among the treatment and control group remains comparable.

[7]These effects are similar if we consider total savings (saved maize plus cash), or only maize (which we directly measure in participants' homes ourselves).

Section 7, we discuss alternative explanations such as soft commitment, intra-household bargaining and experimenter demand, and show that, while some might explain isolated findings, they cannot simultaneously account for the full set of results without relying on a form of retrieval failures. In addition, in Section 8, we complement our core findings by documenting impediments to learning from one's past or learning from others—helping explain why biased beliefs may persist despite experience.

Finally, to examine the external validity of our mechanism, we run a similar intervention in the United States, and discuss the results in Section 9. Specifically, in a survey of around 700 low income participants, we collect prior estimates of upcoming monthly income and expenses. Next, we lead subjects through a categorization-based retrieval exercise similar to the one we implemented in Zambia, separately for income and expenses. We then elicit their posterior estimates. Subjects revise estimates of both income and expenses upward, but by a considerably larger margin for expenses. This is consistent with the idea that expense items are more susceptible to retrieval failures than are income items. These results suggest that retrieval failures may generalize across populations with varied economic circumstances.

Our study advances the literature on how cognitive constraints alter decision-making. A large existing literature focuses on how people respond to *external* information they did not previously know (Chetty and Saez, 2013; Haaland et al., 2023), did not pay attention to (Chetty et al., 2009; Schwartzstein, 2014; Hanna et al., 2014; Gabaix, 2019) or did not seek out (Kling et al., 2012). We highlight the importance of a different dimension: even *internal* information that is already "known" and available to the individual is not always retrieved and used for decision-making. Our findings demonstrate that internal retrieval failures can be large and consequential, affecting behavior even in high stakes environments. This considerably broadens the scope for and relevance of cognitive constraints in economic decision-making.

Relatedly, a burgeoning body of theoretical work in economics models the implications of imperfect information retrieval (Mullainathan, 2002; Gabaix, 2019; Bordalo et al., 2020, 2023; Malmendier and Wachter, 2023). These models are inspired by psychology research on memory (Anderson and Milson, 1989; Kahana, 2012; Wimmer and Shohamy, 2012; Kahana and Wagner, 2023) and prediction and retrieval biases (Kahneman and Tversky, 1973; Tversky and Kahneman, 1974, 1983; Lichtenstein et al., 1978). However, direct field evidence on imperfect information retrieval, and particularly on the relevance of memory, has been limited. A notable exception is work that shows that sending individuals text message reminders to undertake a specific normatively desirable action (i.e., take a pill or save) can immediately increase compliance with that action (Pop-Eleches et al., 2011; Karlan et al., 2016). In contrast, our design does not direct people toward a specific action, but rather lowers the cost of retrieving the various pieces of information that are inside their minds.[8] We find that this

[8] As we discuss below, participants think through their expenses as a whole, and we do not provide partici-

drastically changes individuals' understanding of their overall maximization problem (i.e., through a substantive change in beliefs), with subsequent changes in behavior, including total household spending over multiple months. This offers complementary and naturalistic evidence that retrieval failures can cause large distortions in economic behavior. Moreover, our design enables us to directly test specific predictions of memory models: irregular and stochastic items are more subject to retrieval failures, and cuing finer and more homogeneous categories improves recall (Bordalo et al., 2023).

Our study relates closely to a large literature in psychology on the "planning fallacy": the empirical observation that people exhibit consistent overoptimism in their prediction of how much time it will take them to complete a task (Kahneman and Tversky, 1977; Buehler et al., 2010; Kahneman et al., 2011). We draw heavily on a common debiasing tool in this literature, known as the "segmentation effect": breaking items into sub-categories increases forecasts (Buehler et al., 1994; Forsyth and Burt, 2008).[9] Existing work has applied these ideas to the domain of budgeting, examining and correcting overoptimistic beliefs about future expenses and savings (Peetz and Buehler, 2009; Stilley et al., 2010; Sussman and Alter, 2012; Peetz et al., 2015; Berman et al., 2016).[10] While these streams of work document remarkably robust impacts on elicited beliefs, there is limited evidence that this is consequential for behavior outside of lab-like settings. We build on and extend work on the planning fallacy by demonstrating not only improvements in belief accuracy, but also substantive changes in high-stakes field behavior over significant time horizons in a population of highly experienced agents.

Relatedly, a line of work finds that making a concrete and detailed plan to undertake a specific task—such as voting or getting vaccinated—increases immediate task completion (Nickerson and Rogers, 2010; Milkman et al., 2011, 2013; Abel et al., 2019), though recent studies have argued this approach is less effective at changing repeated behaviors (Carrera et al., 2018). This literature discusses multiple potential mechanisms, from self-control to reference dependence (Beshears et al., 2016). In our mechanism experiment, simply articulating a plan via the placebo board has no apparent impact; it is only when the intervention induces information retrieval (i.e., via the full expense board) do we see effects. Potentially consistent with our findings, this body of work emphasizes the need for plans to be detailed and concrete. To the extent that detailed planning induces retrieval—for example, forcing

pants any guidance on what kind of expense item should increase.

[9]Note that categorization may sometimes lower accuracy. For example, Peetz et al. (2015) find that segmentation can lower forecast accuracy in cases where initial predictions are unbiased. In the associative memory literature, categorization-based cuing can lead to overweighting of rare events or those subject to interference (Bordalo et al., 2023). Our study offers direct evidence of segmentation changing beliefs toward accuracy: in the mechanism experiment, forecasts increase for categories that are ex ante more susceptible to retrieval failures (i.e., non-food items), but not for more regular and salient items (food).

[10]Research on survey design has also documented that finer categories increase measured consumption and expenditures (Deaton et al., 1998), though typically cannot verify whether they also increases accuracy.

individuals to recall other time commitments or obstacles that may otherwise be neglected—retrieval failures may be one (not mutually exclusive) mechanism for the empirical findings in this normative literature.

Finally, our paper contributes to the literature on the presence and causes of consumption smoothing failures. The existing literature has examined several micro-foundations for potential consumption smoothing failures, including missing markets (e.g., Burke et al., 2019; Fink et al., 2020), present bias (e.g., Shapiro, 2005; Ganong and Noel, 2019; Gerard and Naritomi, 2021), and social pressure to share income with others (e.g., Dillon et al., 2021; Carranza et al., 2021; Jakiela and Ozier, 2016).[11] We augment this literature by offering evidence for an additional (and not mutually exclusive) channel: retrieval failures. We document that ameliorating retrieval failures can have sizable consequences for smoothing behavior, indicating first-order relevance as a mechanism.

2 Study Setting

We conduct our study with maize farmers in rural eastern Zambia. In our sample, households harvest their crops once per year. While they may have some supplementary income (e.g. wage earnings from casual labor), the annual harvest comprises over 90% of average household income for the year.

Farmers store harvested maize in their homes or adjacent granaries in 50 kilogram bags, forming their "bank account" for the year. They eat their maize as part of virtually every meal, and also use it to pay for expenditures—either paying in-kind with maize directly, or first selling the maize for cash. Consequently, this setting resembles a simple "eat-the-pie" problem, where income is available upfront and must be smoothed over the rest of the year.

Households face a large array of potential expenses and shocks over the year. Major expected expenditures include food consumption, farm inputs, school fees, and household supplies. Each of these has numerous components, which arrive at different times of the year. For example, farm inputs include a range of specific items (e.g., seeds, fertilizer, herbicide, hired labor) that must be paid for at different times (planting, growing season, harvest). Similarly, school fees involve not just tuition, but also smaller expenditures such as uniforms, pencils and textbooks. Household supplies range from soap to salt to cooking oil—items that are small, numerous, and purchased at differing intervals. In addition, households face unexpected expenditures, for example, due to health shocks, visitors, or contributions to marriages or funerals in the community. At the same time, opportunities to borrow from the future harvest are limited, and borrowing is not common.

This setting is an attractive one for studying consumption smoothing in general, and re-

[11] Note that other work posits that changes in preferences can rationalize behavior such as consumption drops in retirement as optimal (Aguiar and Hurst, 2005).

trieval failures in particular. The farmer's problem is relatively simple and easy to understand—arrival of one income flow that must be allocated over time—while embodying the complexity typical of budget sets—a vast array of expected and possible unexpected expenses. Borrowing is limited and most households fully exhaust their previous harvest income by the subsequent harvest. Our study participants are highly experienced, having solved this annual consumption smoothing problem for decades, and stakes are extremely high.

3 Model

3.1 Model: Introduction

We model our empirical environment using a stylized "eat-the-pie" problem, in which an individual makes decisions over time about how to spend a fixed endowment on a set of expenses. The core assumption is that the individual fails to retrieve some pieces of information that are "known" to her—i.e., available in her memory, but not retrieved and used when solving her problem.

Our setup is motivated by a core asymmetry that we observe in the farmer's budget problem and we believe holds more broadly: income (inflows) is received from a few large and predictable sources, while the sheer number of potential expenditures (outflows) is huge, including many that are small, irregular, rare, and stochastic. Research on memory (as well as introspection) indicates that items that are small, irregular, and stochastic are more likely to be neglected, whereas important, large, certain, and salient items are more likely to be readily retrieved.[12] Consequently, we posit that retrieval failures will be more severe for expenses than income. In the model, we incorporate this idea starkly—by assuming retrieval failures only for expenses—and demonstrate how this asymmetry leads to *systematic* bias in perceptions and behavior.[13]

We present relatively intuitive results in less formal terms. Appendix Section B.1 contains proofs and a more formal discussion, while Appendix Section B.2 illustrates our predictions using a simple numerical example.

[12] Predictions about the types of items most prone to retrieval failures come from different models. For example, Bordalo et al. (2023) predict that more frequent expenses are more likely to be retrieved (absent cuing), but do not have clear predictions about the size of the expense. In a review of associative learning and memory models in psychology, Mason et al. (2023) summarize the factors affecting sampling from memory including extremeness, recency and frequency. See also Wachter and Kahana (2019, 2023).

[13] Time budgeting has a similar asymmetry: while the "inflow" of time is fixed (24 hours a day), the number of potentially unexpected outflows (meetings, conversations, traffic, sickness, etc.) is large. Therefore, our model can be easily modified to make predictions that match the planning fallacy.

3.2 Model: Consumption Smoothing with Retrieval Failures

An individual is endowed with income Y, which she spends over three periods. In each period t, the individual must choose food consumption c_t, which produces utility $u(c_t)$. In addition to food, there are N other possible expenses, which stochastically arise at time t with probability $\pi_{it} \in [0,1]$. That is, some expenses (such as household supplies) arise in every period, some (such as school fees) arise in only one period, and some (such as emergency medical payments) arise stochastically. If expense i arises at time t, the individual chooses an amount e_{it} to spend on the expense and receives utility $v_i(e_{it})$. We assume that the functions $u(\cdot)$ and $v_i(\cdot)$ are increasing and concave, and that $u'(c_t) \to \infty$ as $c_t \to 0$ and $v_i'(e_{it}) \to \infty$ as $e_{it} \to 0$. To isolate the impact of our mechanism, we assume no time discounting, no borrowing, and use a simple-three period model, although modifying these elements does not change the main results.[14]

Our core assumption is that the individual solves the problem using subjective probabilities $\hat{\pi}_{it}$ rather than π_{it}:

Assumption 1. *The individual fails to retrieve some future expenses. That is, for at least one potential expense in each future period, she treats $\pi_{it} > 0$ as $\hat{\pi}_{it} = 0$.*

In other words, the individual solves the budget problem as if some potential expenses will not arise in the future. She remains unaware of these expenses until they arise at period t, at which point $\hat{\pi}_{it} = 1$. In each period, the individual observes which expenses arise for that period, decides on current spending on those expenses and consumption, and creates a state-contingent plan for future consumption and spending that satisfies her (perceived) budget. Then, she enters the next period, observes which expenses arise for that period, and repeats the process given her remaining wealth. The formal maximization problem is written in Appendix Section B.1, and requires additional definitions regarding uncertainty and realizations.[15]

For the most part, we do not explicitly model *which* expenses are subject to retrieval failures because it does not matter for our core results. However, in our setting, a particularly

[14] There are some subtleties. First, all the results hold in a many-period model if expenses are only recalled when they arise. If forgotten expenses can be recalled before arising, the savings result in Initial Prediction 2 can be violated: an individual can *underpredict* future savings in the periods between recall and the due date. Intuitively, she does not realize that her future self will recall this future expense and then save to pay it for it. Second, while adding exponential discounting does not change our results and we see no obvious reason this would not be true for more general discounting functions, we have not proved this: complicated and non-intutive dynamics can arise when the individual is partially sophisticated and strategically manipulating their future self.

[15] Similar models appear in past papers. Karlan et al. (2016) assume that individuals can choose to spend on exactly one non-stochastic, fixed-amount, non-food "expenditure opportunity" in each period, but individuals do not attend to all these opportunities. Bordalo et al. (2023) present a two-period model in which a single fixed-amount expense shock occurs in the second period with probability π, but the shock can arise from many sources. If the sources of the shock are sufficiently heterogeneous, the individual perceives $\hat{\pi} < \pi$.

important example of a large, salient, predictable, and regular expense is food: maize (in the form of nshima patties) is consumed as part of every meal. To capture that food is likely to be recalled in our setting, we explicitly separate consumption c_t in the model and assume that it is not subject to retrieval failures. This generates an additional ex-ante prediction: the individual should particularly under-estimate *non-food* expenditures. We leverage this more specific prediction in constructing our empirical tests (see discussion in Section 4).

The effects of retrieval failures are intuitive. In the initial period, the individual chooses spending without considering the possibility of some future expenses. Consequently, she will consume more than if she fully appreciated these expenses. The individual is then "surprised" in the future when some of these expenses arise and consequently must cut back on planned consumption and other planned expenditures. If the individual experiences our model setup every year, she will experience consumption cycles:

Initial Prediction 1. *[Consumption cycles] In comparison to the rational benchmark, average spending is higher in the first period and then lower in the last period.*

Because the individual is naive about her retrieval failures, she will spend more in the future than she expected and consequently have less savings than expected:

Initial Prediction 2. *[Distorted beliefs] In the first period, the individual will under-predict some future expenditures and have an upward-biased perception of future savings.*

Figure 1 and Appendix Figure A.1 provide empirical support for Initial Predictions 1 and 2. Of course, these patterns are also consistent with a variety of other explanations, such as naive quasi-hyperbolic discounting. Consequently, we use a targeted intervention to create a clean test for the presence of retrieval failures.

3.3 Model: Impact of Increasing Retrieval

Under our hypothesized mechanism, increased recall of future expenses will alter beliefs and behavior—a prediction that distinguishes retrieval failures from other channels. We consequently design an intervention that enables improved recall. To do this, we draw on a robust finding in the psychology literature: thinking through items in categories (in our case, "farm inputs", "household supplies") increases retrieval relative to an aggregate category ("all expenses") (Buehler et al., 2010). Our second assumption is that such an intervention has the intended effect:

Assumption 2. *An individual who is asked to consider expenses in finer categories will retrieve more previously-neglected expenses. That is, the intervention causes some $\hat{\pi}_{it} = 0$ to increase to $\pi_{it} > 0$.*

This assumption can be micro-founded using the more detailed model in Bordalo et al. (2023), which formalizes the idea that, because memory is associative, categories help cue recall of items in that category. In that model, individuals recall stochastic events based on the average *similarity* of the characteristics of the event with all of events in a given cued category.[16] Consequently, retrieval failures will be more pronounced when expense shocks in a cued category are more heterogeneous. Intuitively, "fertilizer" is more likely to be retrieved when considering "farm inputs" than "all expenses" because the items in the former category are more similar.

We now generate additional predictions about the impact of the intervention. Our first prediction is effectively a test that our intervention has the intended effect and causes the individual to retrieve more expenses. This is an implicit test of both Assumptions 1 and 2: retrieval will be boosted only if the intervention mitigates existing retrieval failures.

Intervention Prediction 1. *[Confirming hypothesized impact of the intervention on beliefs] An individual who receives the retrieval intervention will predict higher expected spending on previously-neglected expenses.*

Our second prediction focuses on the immediate impact of the intervention on spending behavior. If the intervention causes an individual to retrieve more future expenses, she will appreciate that she is more financially constrained. Consequently, her immediate desire to spend on a discretionary good will fall (or, more formally, the shadow price of money will rise).

Intervention Prediction 2. *[Changes in perception of budget] The intervention will increase the individual's immediate perceived shadow price of money (the amount of utility gained from a marginal increase in wealth from her plan) and therefore lower her willingness to pay for discretionary goods.*

The third prediction concerns the long-term impact of these changes. Following the second prediction, an individual who appreciates more future expenses will spend and consume less in order to save for these now-retrieved expenses. At some point in the future, these savings will allow her to spend more on consumption and expenses.

[16] While earlier models of associative memory recognize the importance of cues for retrieval, Bordalo et al. (2023) formalize the role of categorization for cuing recall. In their model, there is a similarity function $S(e, H)$ that is assumed to rise as e becomes more similar to the objects in category H; the retrieval of the event $r(e, H)$ is assumed to rise with $S(e, H)$; and the perceived probability of an event $\hat{\pi}(e)$ is assumed to rise with $r(e, H)$. Therefore, the perceived probability of an event (such as "farming equipment breakdown") rises if the individual is cued with a category of similar events (such as "farming expenses") than with broader and more heterogeneous category (such as "all expenses"). Bordalo et al. (2023) is a model of forecasting stochastic events. We provide evidence below that individuals also neglect expenses that are small, irregular and *certain*. To capture this behavior, the Bordalo et al. (2023) model could be modified such that individuals stochastically recall expense-types from a "budget space" rather than stochastic events from a probability space. Here, one important analogous initial assumption would be that individuals retrieve expense-types in some proportion to the expected dollar amount of that type.

Intervention Prediction 3. *[Changes in savings, consumption, expenditures trajectory] The intervention will cause an individual to have a flatter average spending profile (lower initial spending and higher later spending). The individual will have weakly higher average savings throughout the cycle.*

Finally, we note that the results can be extended in intuitive ways. For example, if the individual can use labor to create more income, the intervention will cause her to work more earlier in a cycle (as she now realizes that she needs more money for the future) and will work less later in the cycle (as she has saved more for the previously-unexpected expenses). Similarly, if borrowing is possible but costly, an individual who receive the intervention will borrow less over time as she faces fewer unanticipated shortfalls.

3.4 Retrieval Failures and Quasi-Hyperbolic Discounting

Given that consumption cycles are often explained with quasi-hyperbolic discounting, it is useful to point out the similarities and differences between the two models. In a model with quasi-hyperbolic discounting, the individual sharply discounts future-utility expenses and discontinuously change her discount rate when utility from the good occurs in the present. In our model, she sharply "discounts" neglected expenses and discontinuously change her discount when expenses are recalled, regardless of utility timing. That is, our model predicts that some misprediction in planning comes from failing to retrieve expenses that provide little immediate hedonic benefit (and might be immediately costly), such as paying off past debts or emergency medical spending or farm costs.

One method to distinguish between the models is to examine the types of expenses that are mispredicted. However, this approach requires detailed understanding of the misperception of spending on individual goods and—more importantly—understanding the utility flow of each good over time. While there might some expenses where the timing is obvious (paying a bill for a service already rendered likely causes little immediate gratification), we anticipate that the classification will be difficult and controversial.[17] We consequently focus on the second difference between the models: the style of intervention that will lead to behavior change. Whereas our model predicts that boosting retrieval will change behavior, quasi-hyperbolic discounting (and most standard models) presumes that individuals will not be impacted because they are already fully aware of their expenses.

[17] Is unexpected excess spending on a funeral due to not fully accounting for the likelihood of a funeral or due to overspending on a party to honor the person? Is underestimation of automobile costs due to unappreciated required maintenance or temptation to upgrade something on the car?

4 Intervention and Study Sample

4.1 Intervention

Our key hypothesis is that individuals do not retrieve information about their upcoming expenses that they already "know" (i.e., from their own memory)—leading to consumption smoothing failures. To test this hypothesis, we seek to design an intervention that increases retrieval, without providing any new information, prescriptions, or normative recommendations. We then measure the impact of this intervention on beliefs and behaviors to test Intervention Predictions 1-3 above.

As discussed in Section 3, to construct our intervention, we draw on a well-documented insight in the psychology literature: it is easier to recall items when they are grouped associatively in categories (Kahana, 2012; Bordalo et al., 2023). For example, an individual is more likely to remember that she buys laundry detergent three times a year if she is asked to recall "household supplies" rather than asked to recall "expenditures" as a whole. Consistent with this, the planning fallacy literature robustly documents that thinking in disaggregated categories tends to increase forecasts—a phenomenon referred to as the "segmentation effect" (Kahneman and Tversky, 1977; Buehler et al., 2010).

We leverage these insights in our intervention. We design an "expense board" that prompts individuals to think through their expenses category-by-category (see Figure 2). Using preparatory fieldwork, we identify seven major categories of spending: food (maize allocated to food expenses in each month of the year) and six non-food expense categories (school fees, household supplies, farm inputs, transfers to others, health shocks / other emergencies, and a residual "other" category). In selecting these seven categories, our goal is to design cues that are specific enough to assist with associative recall, but broad and general enough that every household could be expected to have positive expenses within each category. This helps avoid concerns that the categories convey information or normative guidance.[18] In addition, having a relatively small number of categories helps prevent fatigue and keeps the exercise tractable.

The decision to split the board between food and non-food expenses is driven by a specific feature of our setting: adult household members eat maize (in the form of nshima patties) in each meal they consume. Consequently, food expenditures are not only large and salient, but also regular—making them less subject to retrieval failures. Leveraging this feature enables us to make a more specific prediction about *which* expenses will exhibit directional changes in beliefs: thinking through the categories in the expense board will have a disproportionate effect on the retrieval of *non-food expenses* (Intervention Prediction 1).

[18]For example, asking farmers to consider expenses on a specific type of technology might provide new information that this technology exists, or asking them about a normatively-loaded category might cause them to feel obligated to allocate more to that category.

We undertake an exercise with individuals in the treatment group using this physical expense board (Figure 2, Panel A). To promote cognitive engagement, we provide individuals with thumbtacks that equal the number of bags of maize they currently have in savings. We then ask individuals to allocate these thumbtacks across categories on the expense board, in order to depict their spending plan for the coming year.

Note that throughout the retrieval exercise, enumerators do not provide suggestions or make normative statements about how participants should use their maize. They also do not assist participants with doing math. In addition, after the survey is completed, the expense board and thumbtacks are removed from the participant.

For the control group, we face a core tradeoff. On one hand, we would like individuals to complete a similar exercise to the treatment group to gather information on their perceptions of expenses. However, asking control individuals about their upcoming expenses acts as a treatment of sorts and contaminates the control. To solve this tradeoff, we run two separate experiments, with two distinct samples. In the first "mechanism experiment," we extract a large amount of information from the control group in order to precisely pinpoint mechanisms and the characteristics of neglected expenses; this comes at the cost of contamination, making it impossible to examine impacts on longer-run behaviors. We complement this with a "field experiment" in which we ask minimal initial questions to those in the control about their expenses, and track the resultant long-term behavioral differences between the treatment and control groups.

4.2 Sample and Summary Statistics

We provide additional details about each experiment in the subsequent sections. Here, we provide details of the sample that are common across both experiments.

We conduct both experiments in the Eastern Province of Zambia. We draw our sample from villages where most residents derive their income primarily from growing maize, and store their maize in bags after harvest. In both experiments, the sample is comprised of individuals who meet the following criteria: (i) depend on maize as their primary source of income, (ii) store maize in bags, (iii) are not in the upper or lower tail of their village's income distribution (i.e., have little enough maize to report food shortages during the hungry season, but also a sufficiently large maize harvest to make planning worthwhile), and (iv) are not polygamous.[19]

The intervention is always conducted with the head of household alone, with no other family members present. This helps mitigate the concern that treatment effects stem from changes in intra-household bargaining or through information sharing within the household

[19]Note that the above screening criteria are not very restrictive in our setting. For example, 90% of farmers are classified as smallholder farmers in this setting (criteria (i)), and the majority report food shortages in the hungry season (criteria (iii)) (see Fink et al., 2020).

(a concern we discuss in more detail in Section 7.3). All study activities are conducted by enumerators at the participant's home.[20]

The demographic statistics are broadly the same for the mechanism and field experiments, as shown in Appendix Table A.1. Participants are on average around 44 years old, and most grew up on farms—indicating many years of experience in solving the annual "eat-the-pie" problem. Around 80% of our sample is male (female participants are often unmarried heads of household). Participants have on average around 15 bags of maize remaining from their harvest, which is around 50% of the total value of their maize harvest.[21]

5 Mechanism Experiment

5.1 Design

The goal of the mechanism experiment is to isolate the impact of segmentation (i.e., thinking through expenses in finer categories). We construct a design in which all participants allocate their available savings to an expense board, but we vary the level of aggregation of the categories on the board.

Specifically, treatment participants complete the full expense board with 6 non-food categories and 12 monthly food categories as discussed in the previous section and shown in Panel A of Figure 2. Control participants instead use a "placebo" board with only one non-food expense category and one food category, shown in Panel B of Figure 2. This design ensures that both treatment and control participants undertake an exercise with the same mechanics—articulating a spending plan where income equals expenditures—with the difference only in the degree of segmentation, i.e., the extent to which the categories will be useful for cuing retrieval.

Under the null hypothesis that individuals can fully retrieve information from their memory, there should be no difference in the spending plan reflected on the two types of boards. In contrast, if individuals face retrieval failures, then thinking through expenses with finer categories will promote associative recall, leading to increased retrieval—particularly for items that are more subject to retrieval failures. As discussed above, we predict this will lead to a disproportionate treatment effect on forecasted *non-food expenses* compared to forecasted food expenses.

[20] We recruit participants in the same visit that we administer the treatment. We assign treatments during the baseline survey using Survey CTO's randomization tool. Importantly, neither the surveyor nor the participant know the treatment status of the respondent until the retrieval exercise takes place.

[21] Households had already sold a large portion of their maize by the time we undertook our interventions. This suggests that impacts of our intervention could be larger if it were conducted a bit earlier in the year—a sentiment expressed by participants in qualitative debriefs after the study.

5.2 Implementation

Figure 3 displays the timeline of activities for the mechanism experiment. First, we elicit all participants' prior of how much they will spend on non-food expenditures over the coming year.[22] Second, treatment and control farmers undertake the retrieval exercise: allocating their available savings (i.e., maize) using the full expense board or placebo board, depending on their treatment status. Third, we examine short term treatment effects using a willingness to pay exercise. Finally, we have the control group list their recalled expenses both before and after undertaking a second budget exercise with the full treatment board. This offers a within-person comparison of the different expenses that are retrieved using the placebo board and the full expense board. We use this to characterize the features of the items that are most subject to retrieval failures in our setting. Further details are provided in Appendix C.

We undertake the mechanism experiment with 197 farmers in the Fall of 2022, with randomization into treatment and control groups at the individual level. Participants are drawn from 28 villages, with up to 14 participants per village (using the sample selection criteria described in Section 4.2). Appendix Table A.1 shows that the treatment and control groups are balanced on most baseline covariates.

5.3 Results

Intervention Prediction 1: Increase in Perceived Expenses

Intervention Prediction 1 from our model is that individuals in the treatment group will forecast higher expected spending on previously neglected expenses due to improved retrieval. We test this prediction by comparing the number of bags allocated to non-food expenses in the treatment (the sum of all six non-food categories) versus in the control (the single aggregate non-food category).

Results are shown in Panel A of Figure 4. At baseline, the prior is the same on average across the treatment and control groups. Relative to the placebo board, the full treatment board substantially increases expected non-food expenses: the treatment group's allocation to non-food expenses is 38% higher (p<0.001) (see Appendix Table A.3). Furthermore, the distribution of the share of bags allocated to non-food expenses in the treatment group effectively stochastically dominates that of their prior and of the control (Appendix Figure A.3, Panel A).

Note that this result is not purely a mechanical effect of finer categories (or of experimenter demand): the food expense category is also more finely divided for the treatment group than

[22] Specifically, we ask "How many 50kg bags of maize do you expect to sell or use this year for expenses (i.e., not to eat), from the maize that you have remaining from your harvest?" This question is elicited in terms of bags of maize since the expense board exercise also involves allocating maize bags to expenses. This question is embedded in a short baseline survey that is administered to all participants at the start of the session.

the control (12 individual months versus one aggregate category), yet shares of maize allocated to food go down rather than up in the treatment group relative to the control. Note also that the control group's mean expense board allocation is very similar to their prior. This suggests that simply the act of engaging in a budget allocation exercise in itself does not generate meaningful changes in beliefs.[23] Rather, effects only emerge when participants are presented with finer categories in the full budget board—consistent with our hypothesized mechanism.

Intervention Prediction 2: Decrease in Willingness to Pay for Luxury Goods

Intervention Prediction 2 states that the intervention will increase the shadow price of money. To test this prediction, we measure participants' demand for discretionary consumption goods: a chitenge (a cloth wrap), a small solar panel, or a radio. To improve power for this exercise, before the retrieval exercise—at baseline, before participants know their treatment status—we ask each participant to choose which of these three items they would potentially like to purchase at the conclusion of the survey. Then, after the retrieval exercise, we offer to sell the participant the item they had selected earlier, and elicit their willingness to pay for it. We use a Becker-DeGroot-Marschak mechanism: a price card is randomly chosen; if the price is below the individual's stated willingness to pay, the trade is implemented.[24] Note that in-home transactions of these types of goods are not that unusual: households commonly buy goods from "briefcase buyers" who travel to villages after harvest and sell items door-to-door in exchange for maize.

Consistent with our theoretical prediction, the distribution in the control stochastically dominates that in the treatment (Figure 5, Panel A). On average, the willingness to pay in the treatment group is 36% lower than in the control ($p<0.001$) (Appendix Table A.4). This result is robust to including controls for baseline characteristics, item fixed effects, or a Tobit specification. These findings suggest that the intervention changed people's perceptions of their future expenses and made them feel "poorer," at least in the short term.

Types of Expenses Associated with Retrieval Failures

In the final step of the mechanism experiment, we aim to shed light on what types of expenses are most subject to retrieval failures. To achieve this goal, we add a set of steps for the *control group only*. First, control participants are asked to list all individual items they considered when constructing their allocation to the "non-food" category on the placebo expense board. For each item listed, the surveyor asks the amount of the expense, the time when the expense

[23]For example, control individuals decisions might have changed if they were not "adding up" correctly, i.e., if their planned spending did not match their available maize. The expense board, by forcing budget balance, would highlight this contradiction and cause a change in allocation.

[24]Trades are implemented with low probability to avoid generating an imbalance in initial savings between the treatment and control and treatment groups if treatment affects willingness to pay.

is expected to arise, the expected frequency of the expense, and the degree of certainty of the expense. Then, the control group undergoes the full retrieval exercise (i.e., with all 6 categories of non-food expenses). After completing the full expense board, the same follow up questions about items in the non-food categories are asked again.

Completing the full retrieval exercise causes the control group to increase their expected non-food expenses by 37.5% (Appendix Table A.3, column 2), which is very close to the between-subject treatment effects above. By examining which items are added to expenses when going from the placebo board to the full board, we can characterize what kinds of items were initially neglected. Table 1 provides the results of regressing whether an expense was forgotten on different characteristics of that expense. Expenses that are small, infrequent, irregularly-timed, and uncertain are more prone to retrieval failures. These results are broadly consistent with what one might expect under a cognition-based mechanism for retrieval failures, and particularly under imperfect memory (Bordalo et al., 2023).

Note that, by the end of the mechanism experiment, the control has gone through the full retrieval exercise and has therefore been treated. Consequently, we do not expect any long-term behavioral differences between the treatment and control groups. To study these longer-term impacts, we turn to a separate field experiment in which the structure of the control is designed to avoid this type of contamination.

6 Field Experiment

The mechanism experiment provides evidence that our intervention increases forecasted spending by making individuals recall small, irregular and uncertain expenses that are otherwise susceptible to retrieval failures. Consistent with our model, this change in beliefs leads to a reduction in desired expenditures today, as measured by the willingness-to-pay for discretionary consumption. These results highlight the immediate impacts of retrieval failures; the field experiment design allows us to investigate longer-term impacts, namely to test Intervention Prediction 3.

6.1 Experimental Design

The field experiment design includes two substantial changes relative to the mechanism experiment.[25] First, in order to address the concern that even the 2 category placebo board may act as an intervention, we do not conduct any retrieval exercise with the control group in the field experiment. This change to a "pure" control also allows us to estimate the policy-relevant difference between the impact of the retrieval exercise and the status quo.

[25] In addition to the below two changes, we also add a longer prompt by asking treatment individuals to recall their spending in each category last year, before undertaking the exercise for the coming year. We do this to help individuals populate items from memory for each category.

Second, under our hypothesized mechanism of retrieval failures, the treatment group will forget some specific items recalled (or specific plans made) many months later. To address this concern, we introduce a label technology to help participants visually memorialize the results of the retrieval exercise. Each label corresponds to one of the categories on the expense board; we give participants the option to affix a label to each bag of maize according to their expense board allocation. This obviates the need to hold the initial plan in memory, or to undertake the cognitive effort to reformulate it in the future when making spending decisions. To assess demand for this visual representation, we offer all individuals a choice between the labels or a small compensation (a bag of sugar). The treatment group was significantly more likely to take up labels than was the control group, which received no intervention.[26]

To mitigate the concern that the labels introduce confounding mechanisms—such as a previously-unavailable technology for reminders or soft commitment—we incorporate two features into the design. First, we ensure the labels technology is available to *both* treatment and control groups: we offer all participants the labels, and explicitly tell control participants that some individuals find it helpful use the labels to visually record their spending plan for the year.[27] Second, we only provide the labels *2 months after* the retrieval exercise. This enables us to examine the impact of the expense board alone over a substantial time horizon, before labels are introduced.

6.2 Implementation

The field experiment activities were conducted between late August 2019 and Sept 2020.[28] Figure 6 provides an overview of the field experiment timeline and activities. At the start of implementation, most households had just completed shelling their maize. The intervention is embedded in the baseline survey (Visit 1). All participants take a brief baseline survey, which includes information about baseline savings (e.g., maize) and other demographic variables. As in the mechanism experiment, they are asked for their "prior" forecast of non-food expenses. The treatment group completes the retrieval exercise using the full expense board shown in Panel A of Figure 2. Both the treatment and control groups are then offered the choice

[26]Around 80% of treated participants chose to receive labels after completing the retrieval exercise compared to around 30% of the control ($p < 0.01$). Since both groups had the labels explained to them prior to their choice, we interpret this difference as reflecting a higher valuation for the labels in the treatment group. Qualitative responses confirm this: over 95% of treated participants agreed or strongly agreed with the statement "labels are helpful as a reminder of the plan, but you have to have the plan first."

[27]Specifically, all participants across both groups is shown the labels, and each expenditure category (the non-food expenses and consumption for each month of the year) is then explained to the participant. They are then told that if they want, they can attach these labels to their maize bags, to mark what they thought they would spend on each category. Consequently, if there are no retrieval failures but the labels provide soft commitment to sophisticates, then both treatment and control groups should be able to benefit from them equally.

[28]Note that the field experiment was run prior to the mechanism experiment. We discuss the mechanism experiment first to highlight the link with the theory before turning to the more policy-relevant field experiment.

between labels and a bag of sugar. All participants then complete a set of additional survey activities (e.g., questions about child health). Finally, willingness to pay for a discretionary consumption good is elicited. This follows a similar protocol as in the mechanism experiment: early in the baseline survey (before treatment), participants select one good—either a piece of clothing, solar panel, or radio—which they would like a chance to purchase later; we then elicit the willingness to pay at the end of the Visit 1 survey to test Intervention Prediction 2. For the treatment group, the retrieval intervention takes around 45 minutes, and the entire survey (i.e., all Visit 1 activities) takes about 90 minutes.

Our primary outcome, savings, equals the amount of maize in storage plus cash in savings. Because maize comprises the majority of total savings, we can also examine results only on stored maize. This is useful because in each round of data collection, enumerators directly measure the amount of maize in storage, providing an objective measure of savings that does not rely on self-reported data.

Our first follow up is 2 months after the baseline (Visit 2), when we collect data on savings (maize and cash), expenditures, and labor supply. We use the outcomes collected in this visit to test for the impacts of the retrieval exercise on consumption smoothing behavior before labels are provided. For participants that chose to take-up the labels at the end of the baseline survey, the enumerator offers to attach labels to their maize bags; the participant chooses which labels to attach to their remaining maize bags.[29] We complete two additional visits between December and March (Visits 3 and 4), and again measure savings (maize and cash), expenditures, and labor supply.[30] Because planting begins in December, we also collect basic data on farm investment during these two visits. Data collection was paused in March 2020 due to the COVID-19 pandemic.

Finally, we return in September/October—approximately one year after the intervention was conducted—for a final round of data collection (Visit 5). We measure crop yields and revenues and additional farm investment. We also elicit participants' willingness to pay to receive our treatment intervention for the upcoming agricultural year. Finally, to test for persistence, we collect data on expense forecasts for the coming year.

The field experiment was run with 837 farmers. Participants were drawn from 113 villages, sampling up to 14 households per village (using the sample selection criteria described in Section 4.2). We randomize at the individual (rather than village) level in order to improve statistical power. However, this design choice generates some scope for spillovers between participants—for example, control participants may learn about the intervention from treatment households, or may pressure them to share their extra savings during the hungry season. Note that such spillovers would only dampen our measured treatment effects, and only those

[29] Note that this effectively provides individuals an opportunity to revise their spending plan.

[30] Note that in Visit 4, the survey instrument was abbreviated and did not include information on non-food expenditures because we were constrained to finish field activities before data collection shut down in March 2020 due to the COVID-19 pandemic.

collected after our initial interaction with the participant. We mitigate the potential for such spillovers by enrolling no more than 14 participants per village, so that in expectation no more than seven participants per village are treated. The randomization was successful, with balance on baseline covariates (Appendix Table A.1). Appendix C describes the protocols for the field experiment in detail.

6.3 Intervention Predictions 1 and 2: Immediate Effects

The immediate impacts in the field experiment largely match those from the mechanism experiment. This similarity provides additional confidence in external validity given that effects come from different samples in different years.

First, as in the mechanism experiment, the intervention increases participants' forecasted expenses. In the mechanism experiment, we are able to compare forecasts in the treatment and control group. Here, since the control group does not complete a retrieval exercise, we instead rely on a comparison within the treatment group between the stated prior non-food expenses and the sum of the non-food categories in the retrieval exercise. As shown in Panel B of Figure 4, the intervention increases the allocation to non-food expenses by 60% ($p < 0.001$), consistent with Intervention Prediction 1. Furthermore, as in the mechanism experiment, the distribution of forecasted non-food expenses after the intervention stochastically dominates that of the baseline forecasts (i.e., the distribution of priors) (Appendix Figure A.3, Panel B).

Second, the intervention decreases the willingness to pay for a discretionary good. Panel B of Figure 5 qualitatively matches the results from the mechanism experiment. The distribution of the willingness to pay of control individuals effectively stochastically dominates that of the treatment, with an average change of 34% ($p < 0.001$), consistent with Intervention Prediction 2. This result is robust to including controls for baseline characteristics, item fixed effects, or a Tobit specification (Appendix Table A.4).

6.4 Intervention Prediction 3: Long-Term Effects on Saving

Both immediate effects of the retrieval intervention closely mirror those found in the mechanism experiment. We now turn to the longer run results of the field experiment—following the treatment group and measuring impacts relative to the pure control group over the subsequent year.

Empirical Strategy We use repeated household survey data on savings levels to impute expenditures between visits. We also estimate treatment effects in an OLS regression specification. To accommodate time varying treatment effects, we estimate:

$$y_{it} = \sum_{j=1}^{3} \beta_j \mathbb{1}\left(\text{Treatment}_i \times \text{Visit}_{j(t)}\right) + \sigma_t + X_i'\theta + \varepsilon_{it} \quad (1)$$

where y_{it} is the outcome of participant i during time period t. $\text{Treatment}_i \times \text{Visit}_{j(t)}$ is an indicator variable for a participant i that is assigned to the treatment group at baseline, and is responding to survey questions in visit j (at time period t). X_i is a vector of baseline controls for participant i.[31]

The key coefficients of interest are the β_js. We estimate period-specific treatment effects as our preferred measure of impact for two reasons. First, we do not expect the treatment effects to be constant over time. For instance, since the group that we study face an "eat-the-pie" problem, savings should decline throughout the year and differentially between treatment and control, whose levels should converge as they approach the next harvest. Second, outcomes may be non-monotonic. For instance, if the treatment group realizes their available budget is smaller than they thought, they will reduce immediate spending, resulting in more savings available to support higher spending in future periods. In that case, an average treatment effect over the year would mask heterogeneity in the response over time.

Results Intervention Prediction 3 states that the treatment will lead to weakly higher savings, due to lower initial spending (followed by higher spending at later dates). Figure 7 provides evidence consistent with this prediction. The y-axis of the figure measures total spending (on food and non-food expenditures), using a normalized version of the difference between participant savings in each visit. This effect is displayed starting in October, the beginning of Visit 2, and ending in early March, the end of Visit 4.[32] The figure documents that treatment households immediately decrease spending after the intervention, leading to increased savings. As a result, they are able to spend more in later parts of the year, with a crossing point around the end of November, after which the treatment group uses 5-10 kilograms more of maize per week, through the duration of the project. Overall, this results in a flatter spending profile across the year for treated households.

We examine the effect on the evolution of savings stocks directly in Table 2. Our primary specification measures savings as the sum of the amount of unprocessed maize in storage and the value of cash savings (converted to maize equivalents). Treated households held 100 kilograms more than those in the control group at the first follow up (Visit 2), an increase of around 15% relative to the control group (column 1, $p = 0.026$). This first follow up was on

[31] In most specifications, we control for the baseline value of the outcome of interest, however, we also present specifications showing robustness to alternate sets of controls. There was a slight imbalance in the timing of Visit 2 across treatment and control groups; we consequently include week-of-survey fixed effects in all specifications that use the panel data.

[32] Since the outcome variable is based on the difference in savings between each visit, we are unable to show this outcome from the baseline survey. However, the lines should start from the same point, given the balance in the randomization.

average 43 days after our baseline visit, and occurred *prior* to attaching labels to participants' maize bags. Consequently, this 15% treatment effect on savings in the first 43 days reflects the impact of the expense board alone, without labels.

During Visit 3, which coincided with the lead up to and beginning of the hungry season, treated participants had almost 70 kilograms (20%) more in savings than the control group ($p = 0.018$). The magnitude of this treatment effect corresponds to how much the control group spends in total (on food plus non-food) in an entire month on average at this time of the year. During Visit 4, in the middle of the hungry season, treated participants held on average 15 kilograms of maize more than the control group, although this effect is not statistically significantly different from zero ($p = 0.41$).[33]

Our results are not meaningfully affected by the inclusion of baseline controls (column 2). We also see similar effects when we disaggregate savings into maize (column 3) and cash (column 4). Recall that, while cash savings are self reported, the number of maize bags in the households were counted and verified by surveyors, making this measure robust to reporting bias. The magnitudes imply that both savings sources contribute to the total savings effect, consistent with fungibility of these different assets in our context.

We consider two potential concerns regarding the interpretation of our savings results. First, our savings measures require a number of assumptions to aggregate cash and grain savings. Results are robust to alternative assumptions about how to convert cash into kilograms of maize (Appendix Table A.6, columns 1 and 2). Second, the ideal savings measure would reflect all assets held by households that could conceivably be used as savings, since rural agricultural households save in multiple forms (Fafchamps et al., 1998). Our measure of savings, by contrast, reflects cash and unprocessed maize only. While these are the two primary vehicles for saving in this context, we may be missing some substitution of savings across asset classes. To help alleviate this concern, we perform several tests. We first incorporate processed maize into our measure of savings, to ensure that our savings outcome does not just capture substitution of the treatment group from processed to unprocessed maize (Table A.6, column 3). Next, we examine total expenditures by treatment on household assets, including livestock, that could conceivably function as savings (Appendix Table A.7).[34] These robustness checks show that, if anything, we are undercounting the savings effect: treatment participants (insignificantly) increase their net holdings of livestock by selling fewer large livestock during the hungry season.

[33]This pattern of savings suggests that treated participants were also able to delay some maize sales relative to the control. We observe a statistically significant delay of 11 days in the sale of the first maize bag in the treatment group, and a positive but statistically insignificant increase in the price per kilogram at the time of sale. Thus, lower early expenditures may lead to overall higher income from later maize sales, though these effects are too small to be measured in our sample.

[34]These were collected through recall measures of purchases and sales of household assets and livestock during follow up Visit 3.

6.5 Long-Term Effects on Other Outcomes

Empirical Strategy To track other inflows and outflows that both contribute to and diminish savings over the year, we collect data on a number of other outcomes in Visits 2-5. We estimate treatment effects on these outcomes following equation (1). First, we measure food consumption during each follow up visit. Following Fink et al. (2020), we record the number of meals per day consumed by adult members of the household over the past two days.

Second, we measure households' supply and demand for ganyu labor, a form of casual day labor common in rural agricultural markets across Southern Africa. Households that are running out of savings may choose to sell casual day labor (ganyu) to the outside market to increase period-specific income (and consumption) but at the expense of leisure (consumption) and family labor supply on-farm (Fink et al., 2020). We measure total days of wage labor (ganyu) performed by the household and total household ganyu earnings during each survey round. The recall period for these outcomes differs across follow up visits.[35] In addition, we measure whether the household hired outside workers to work on its farm (in the Visit 4 survey round only).

Finally, we measure spending on other inputs (fertilizer, herbicides and pesticides) and agricultural output (total harvest quantity and value) during Visit 5, with questions covering the full agricultural cycle. Since these outcomes are measured only in a single survey, we estimate treatment effects following equation (1), but with one observation per participant. Overall, we report results at the frequency at which they were collected: if we collected a given outcome in multiple survey rounds, we report effects separately by round; if we only collected it once, we report one aggregate effect over the recall period.

Results Consistent with the patterns of expenditure smoothing, Table 3 suggests that treated participants slightly reduced consumption of the staple food at the first follow up survey visit, although this reduction is not statistically significant at conventional levels (column 1). The treatment group has slightly higher consumption of staple meals at the beginning of the hungry season (Visit 3, $p = 0.079$), although this difference disappears toward the end of the hungry season in Visit 4 (consistent with the savings results).[36]

Under-saving has implications not only for consumption—and therefore household welfare—but also for productive investments and future income. In our setting, planting of crops occurs after savings have begun to decline, which may affect both labor and non-labor investments. Table 3 indicates that the increased savings induced by the treatment reduced the likelihood of selling household labor (ganyu): during the hungry season, treated households engaged in

[35] In Visit 2, we use a recall period of one month, while in follow up Visit 3, we reduce the recall period to reduce measurement error and ask about the previous week.

[36] Data on a food security index shows a statistically insignificant reduction in food insecurity in both Visits 3 and 4.

32% fewer days of wage labor on others' farms , relative to the control group mean (column 2) ($p = 0.043$), and we find no corresponding increase earlier in the year. Note that this helps explain some of the muted impacts on hungry season consumption: the increased wage earnings among the control group (column 3) are used to purchase food ($p = 0.008$).

Consistent with less diversion of labor away from their farms, treated participants spend more days working on their own farm during the hungry season (column 4) ($p = 0.087$). Besides household labor investments on their farm, we find additional suggestive evidence for increases in other farm inputs. Treated households purchase more casual labor to work on their own farms during the hungry season (column 5, $p = 0.082$), and report higher expenditures on other farming inputs, fertilizer and herbicide/pesticides (column 6, $p = 0.127$). Increased farming inputs lead to higher crop revenues at the subsequent harvest. Treated households report the value of their entire harvest as being 8.9% higher than control households on average ($p = 0.095$).

Altogether, our treatment induces sizable savings effects, which affect household consumption and production. The magnitudes of the investment and revenue impacts are comparable to those found in Fink et al. (2020), which involved giving households a substantial cash or maize transfer during the hungry season.

7 Alternative Explanations

The mechanism and field experiments demonstrate that the retrieval intervention causes participants to increase their perception of small, irregular, and rare expenses, which leads to an immediate drop in their willingness to pay for discretionary consumption. The field experiment demonstrates two longer-term effects. Prior to the second visit (when labels are attached to participants' bags), treated participants spend less and consequently save more. Following this visit, participants draw down these savings to consume and invest more. These impacts provide support for the predictions of our model, and therefore the relevance of retrieval failures and effectiveness of our intervention for boosting retrieval. In this section, we discuss alternative explanations and argue that it is challenging to account for the constellation of results without retrieval failures playing a primary role.

7.1 Reminders and Increased Salience of the Need to Save

Past literature has demonstrated that simply reminding an individual of a specific action can change behavior by bringing the action to the top of the individual's mind. Therefore, while we interpret our intervention as leading to a genuine change in participants' beliefs about their budget and the labels as a tool to recall that change in beliefs, an alternative explanation is that parts of our intervention—particularly the labels placed on the bags during the second visit—might simply increase the salience of spending choices.

While the salience of spending choices may play some role, it cannot account for our core results. To start, in the mechanism experiment, both the treatment and control groups go through an exercise in which expenses and budgets are very salient, but their perceived expenses and willingness to pay differ starkly. Second, in the field experiment, neither group is given access to the "reminder technology" between the intervention and the second visit, and yet savings differ significantly between the groups.

Determining the precise impact of the labels is more difficult, since they were differentially adopted by the treatment group. The labels were designed to help participants memorialize the results of the retrieval exercise. However, the labels might act as a simple nudge to consider spending decisions more carefully. There are two reasons we believe that the latter explanation is unlikely to play a large role. First, past literature demonstrates that the impact of attentional reminders is typically small (Gabaix, 2019) and short-lived (Carrera et al., 2018). And, very similar to our setting, Burke et al. (2019) cross-randomize a maize bag labeling intervention in Kenya and find little effect on behavior. Second, in the field experiment, a follow up survey documents that treated farmers perceive the impact of the labels as largely tied to the intervention: less than 2% report that the labels would be helpful in the absence of the retrieval exercise (Appendix Figure A.4, Panel A) and participants value the labels with the intervention at nearly 10 times their value for labels alone (Appendix Figure A.4, Panel B).[37]

7.2 Present Bias and Soft Commitment

As we discuss in Section 3.4, models of present bias have a maintained assumption—common to most economic models—that individuals have unconstrained access to information that they "know." Therefore, these models would predict that an intervention that manipulates only retrieval will have no impact on perceptions of expenses or immediate behavior, contrary to our findings.

It is therefore challenging to explain the results of the mechanism experiment with a model of present bias. The only difference between the treatment and control (in the mechanism experiment) was the granularity of the expense categorization. It is unclear why present bias would cause a person to retrieve more (small, irregular, and stochastic) expenses with finer categorization and consequently change their immediate spending behavior.

While the field experiment was designed to target retrieval, the treatment group was differentially exposed to other components, such as spending more time with an interviewer, discussing a consumption plan, and potentially receiving labels. One might argue that some

[37] For the value measurement, we used a BDM to elicit participants' preferences for a new participant in the following year to receive a cash payment, labels, or the retrieval exercise plus labels.

of these additional components could be used as a “soft” commitment device by partially-sophisticated present-biased participants to incentivize their future selves to take actions that better align with current preferences.

While soft commitments may be important for some behavior change, we believe that they are unlikely to play a large role in our results for several reasons. Perhaps the most natural way for our intervention to act as a commitment is through the labels, which might act as a salient signal of a previous mental commitment. However, as noted above, our willingness to pay results and main savings effects occur prior to the application of the labels. In addition, all participants have access to the labels: if there are no retrieval failures but the labels provide soft commitment to sophisticates, then both treatment and control groups should be able to benefit from their use. Furthermore, we find little demand for these labels in the control group, suggesting that people do not perceive them as a previously-unavailable commitment technology. A more subtle commitment effect might arise from simply making a plan in front of another person. This effect requires that the one-time articulation of a plan in front of an outside surveyor (that the individual will likely not see again) can significantly impact very long-term behavior, which seems unlikely given past literature (Carrera et al., 2018). Finally, if planning or telling others are effective ways to self-impose a soft commitment, participants presumably could do both of these things without our intervention.

7.3 Intrahousehold Coordination

Since the retrieval intervention affects household behavior, there is a concern is that it may lead to a shift in the structure of communication within the household, which may drive behavior change. For instance, the labels might serve as a coordination device for spouses in household bargaining.

To partially address this concern, we intentionally ran the intervention with the household head only. The retrieval exercise leads to sharp changes in beliefs about total expenses, and reduces willingness to pay for discretionary consumption before the participant has interacted with anyone else in the family. Therefore, our initial impacts from Visit 1 cannot be driven by intrahousehold coordination. For our later results, it is unclear why the intervention would lead to a change in bargaining that shifts behavior in the direction we observe. Rationalizing these directional changes requires particular asymmetries: for example, one possibility is the household head generally prefers a plan with more savings and the intervention provided the head with a previously-unavailable ability to increase bargaining power. However, if the intervention works because of asymmetric preferences in the home, then we would not expect to see beliefs update in response to the intervention. In addition, participants do not mention changes in intrahousehold coordination in our follow up qualitative surveys.

7.4 Demand Effects

Some of our results may be susceptible to experimenter demand effects. For instance, while our retrieval intervention was constructed carefully to avoid making normative statements or suggesting allocations, treated participants may have perceived some demand to reduce immediate spending, leading to dampened valuations in the willingness to pay exercise. It seems difficult, however, to square this mechanism with other results. For instance, it is unclear why treated participants would perceive experimenter demand to spend more on non-consumption items (Appendix Figure A.3). If anything, the perceived demand in this context would probably be to save more for future food consumption, which would suggest that beliefs would shift in the opposite direction, toward allocating a *smaller* proportion of their maize stock toward future non-food expenses. More specifically, to the extent that finer categorization itself is a signal of experimenter preferences, the full treatment board has 12 food categories (one for each month) and 6 non-food categories. Consequently, in the mechanism experiment, when participants go from the placebo board to the full board, there is a larger increase in food categories; it is therefore unclear why the full expense board would signal demand for lower food consumption.

Alternatively, the additional categories in the full expense board may have indicated some experimenter demand to populate those categories that was absent in the two category board provided to the mechanism experiment's control group. The additional non-food categories were chosen based on extensive piloting to represent expenses identified by the majority of households, and restricting the number of non-food categories to six minimizes priming or otherwise signaling the importance of budgeting for certain items.

Finally, it seems implausible that a demand effect alone would generate substantial (objectively observed) savings increases over a 3-6 month period, as we show in Table 2.

8 Extensions: The Persistence of Biased Beliefs

Figure 1 documents remarkably skewed beliefs among highly experienced agents. Our study focuses on one particular explanation that can generate such bias in beliefs: retrieval failures. The main focus of our study is to test for the presence of retrieval failures, and their resultant impacts on consumption smoothing. In this section, we go beyond this core focus, and present suggestive evidence on additional mechanisms for why biased beliefs may persist despite experience.

To motivate this line of enquiry, note that, even under the presence of retrieval failures, the patterns in Figure 1 still present a puzzle: even if individuals do not remember all their future expenditures, they could realize that they always run out of maize earlier than expected, and debias their beliefs about future savings this way. In line with this, the planning fallacy literature highlights three ways in which individuals could debias their beliefs about the future. The first, analytical forecasting, is thinking through all components of the problem (i.e., all future expenses) and computing the correct forecast. Our paper demonstrates that, due to retrieval failures, people do not do this perfectly.

Second, alternatively, individuals could use recall based forecasting: thinking about their own past history (i.e., when maize has run out in past years) to form a more accurate assessment of the future (i.e. at what point maize is likely to run out this year). We investigate whether individuals learn from past experience in two ways. We first look in the cross-section at whether individuals with more experience—proxied by age—hold more accurate beliefs. 85% of individuals in the lowest age quartile are overoptimistic, compared to 73% in the top age quartile, where the mean age in each quartile group is 28 and 62, respectively (Appendix Figure A.5).[38] Therefore, while there is some suggestive evidence for updating over decades of experience, experience appears insufficient to eliminate the overoptimistic bias in forecasts.

Why aren't people learning from their own experience? We gather additional evidence by returning to our field experiment sample in September 2020, one year after our initial intervention. At this time, we ask participants to recollect their savings forecasts and realized savings in the preceding year. We find evidence for systematically biased memory, where individuals recall the past as being better than it was (Appendix Figure A.6). More than 70% of participants recall having more maize savings than they actually did at Visit 3. This difference is meaningful: the average recall bias can explain 80% of the average forecast error, meaning that if participants actually had the number of bags they recalled, their forecast error at baseline would be only 20% of the size observed.[39] Note that while participants exhibit rosy memory, they do not completely disregard the hungry season: on average, they recall having 52% less maize in Visit 4 (hungry season) than in Visit 3 (early hungry season), while the actual decline in maize was 63.6%. This bias in memory is consistent with psychology literature on the planning fallacy (Roy et al., 2005; Griffin and Buehler, 2005), and may slow learning but is too small to fully explain the beliefs we document.

We also ask participants to recall the forecast that they made at baseline about how much maize they would have at Visit 3. We find recollections of forecasts to be noisy, but the error (recalled forecast relative to actual forecast) to be more or less mean 0.[40] Taking these two findings together, we can compare a participant's recollection of their forecast error, to

[38] Appendix Figure A.5 shows the cumulative distribution function of forecast error—(incentivized) forecasts of future savings minus realized savings—for the bottom and top age quartiles.

[39] Doubling the incentive that we pay from 1 bag to 2 bags of sugar for correct recall does not meaningfully change this response, which is also inconsistent with models of motivated reasoning. Participants in the treatment group are no better able to recollect their past than are control participants.

[40] To be precise: let the participants' actual forecast about future savings at baseline be $\hat{S}_3$, i.e., it is their forecast at baseline of how many bags of maize they would have in storage at Visit 3. We represent realized savings by S_3. Note that the participant's forecast error is $\hat{S}_3 - S_3$. The participant's recall of these items after the following year is $\bar{\hat{S}}_3$ and $\bar{S}_3$. Using these two elements, we construct the recalled forecast error as $\bar{\hat{S}}_3 - \bar{S}_3$.

their actual forecast error. Appendix Figure A.7 shows participants' actual forecast error, as measured by baseline forecasts of future savings less actual savings, and recalled forecast error, as measured by recall of baseline forecast less recalled of savings. This measure shows that participants appear somewhat naive about their overoptimism: while more than 80% of participants exhibit positive forecast errors, fewer than 40% recall that their forecasts were overoptimistic. Together, these pieces of evidence shed light on why experience alone is not enough to debias over-optimism.

The third potential approach to debiasing beliefs is reference-based forecasting: using the experiences of others similar to oneself to form more accurate beliefs. During the same follow-up final survey visit, we ask questions about own forecasted savings and, later in the survey, about the forecasted savings of other households like their own. We find participants place themselves at the extremes of the distribution: more than 50% of participants forecast that they will have more maize than all ten similar households for whom they provided forecasts (Appendix Figure A.8). While this exercise was not incentivized, it is consistent with other experiments that find evidence of asymmetric naivete (Fedyk, 2018), and with literature on the planning fallacy that shows that people underestimate their own task completion time but not that of others (Buehler et al., 1994).

Together, these results are consistent with cognitive frictions that slow the learning process and contribute to the persistence of biased beliefs in an experienced population. Rosy memory, naivete about own forecast errors and overoptimism about oneself compared to others all impede belief updating. Examining these mechanisms further constitutes an interesting diretion for additional research.

9 External Validity: Low Income Households in the U.S.

We have so far demonstrated the importance of retrieval failures in the context of an annual consumption smoothing problem in Zambia. However, the potential relevance of retrieval failures is much broader, extending both to other populations and other classes of problems. In fact, past literature has demonstrated the tendency to underpredict future expenses or overpredict savings in high income countries (Peetz and Buehler, 2009; Stilley et al., 2010; Sussman and Alter, 2012; Peetz et al., 2015; Berman et al., 2016).

To complement our evidence from Zambia, we run a short static online survey among low and middle-income households in the United States to assess the external validity of retrieval failures.Between January and April 2023, we recruited 721 employed adults, whose household income was between $20,000 and $70,000 per year, through Prolific's survey panel. We exclude individuals who are not currently employed or who are living with their parents. This enables us to examine beliefs about not only expenses, but also income.

Before showing results from a segmentation exercise, we first report respondents' perceptions of how income and expenses match their forecasts in Figure 8. When presented with the scenario, "There are months where *more* expenses come up than I had initially expected," only 13% say this happens "rarely or never". In contrast, 59% rarely or never have *less* expenses come up than expected. In other words, when it comes to expenses, there is directional bias in the realization relative to expectations: individuals are generally more likely to be unpleasantly "surprised" (Panel A). In contrast, we see no such substantive asymmetry for income (Panel B). Finally, similar to the pattern for expenses, individuals state they often end up with less savings than they anticipated (Panel C). These patterns roughly match what one would expect under retrieval failures.

We then examine whether engaging in a segmentation exercise—thinking through categories—affects beliefs about income and expenses. First, we collect priors: we ask participants to forecast total expenses for the coming month. We also collect forecasted monthly income. Second, participants undergo a category-based elicitation for expenses and for income. Finally, we ask participants make revised forecasts for both income and expenses.[41] Example screenshots from the survey are shown in Appendix Figure A.9.

Figure 9 shows the resultant impact on beliefs from the segmentation exercise: the CDFs of the posterior, relative to the prior, for expenses and incomes. Consistent with Intervention Prediction 1 of the model, we observe that 58% of the sample revises expenses upward, with a mean change in forecasted expenses of 13%. While the mean update in forecasted income is also positive, it is substantively smaller: the mean change is 4.6%, and the modal change is 0%, with 59% of respondents having no change in their income forecast after the segmentation exercise. The p-value of a test of whether the mean change in income equals the mean change in expenses is <0.01).

These results offer suggestive evidence that (i) overoptimism in forecasted expenses and savings, but not in income, is common among low income Americans, and (ii) fine categories aid retrieval of expenses, and—to a lesser degree—income. Notably, when asked to explain the divergence between their prior and posterior estimates, 70% of respondents say that forgetting expenses was an "important" or "very important" reason for the discrepancy. Tracking longer term outcomes in this population would be more challenging than in our field setting in Zambia, given the diversity of income flows and additional complexity of expenditures. However, these findings suggest that similar retrieval interventions may yield consumption smoothing benefits across a range of settings.[42]

[41] Note that this last step, which measures the debiasing effect of finer category-based elicitation, distinguishes this exercise from research on survey design, which considers how more or less aggregated income, expenditure and consumption questions affect measurement (e.g., Crossley and Winter (2014)).

[42] The labor allocation and productivity impacts that we document in Zambia are less likely to generalize, though better consumption smoothing may, for example, allow households to take fewer payday loans, resulting in overall higher income.

10 Discussion and Conclusion

In this paper, we posit that people do not always retrieve and utilize information that they "know" when making decisions. To test for the empirical importance of this mechanism, we employ a simple intervention designed to help individuals retrieve information about their future expenses. We find that this leads individuals to (i) substantially increase the amount of savings that they believe they need to allocate to non-food expenses, (ii) reduce spending today and therefore (iii) increase savings by 15-20% in the months following the intervention. We find that the additional savings have meaningful consequences: participants reduce their off-farm labor, self-finance increased investments in their farms, leading to an estimated increase in total farm revenues of 8.9%. Although the majority of the paper focuses on the decisions of Zambian farmers, we believe that the basic mechanism generalizes to other populations, and we provide additional survey evidence from low-income individuals in the United States that supports that view.

We see these failures as a consequence of the fundamental limits of the human retrieval system. Although we can imagine a role for willful neglect of the budget problem, our results are somewhat inconsistent with standard models of optimally-chosen motivated beliefs (e.g., Bénabou and Tirole, 2016).[43] Even if individuals are intentionally shifting their beliefs, our results show that the structure of the memory system shapes the form of manipulation. For example, it is unclear why someone who desires to willfully neglect expenses would be impacted by looking at finer categorizations, unless those categorizations somehow force undesired retrieval. Similarly, it is unclear why unconstrained willful manipulation would lead to misperception of small, irregular and uncertain expenses rather than of larger expenses or income. This suggests that any form of willful neglect would involve intentionally avoiding thinking deeply about the problem (Bolte and Raymond, 2022), which leads to specific retrieval failures arising from an imperfect retrieval system. Moreover, if individuals were holding motivated beliefs, then we might expect under-estimation of expenses to be accompanied by over-estimation of income—contrary to our findings in the US sample. Finally, even if individuals are willfully refusing to think about their expenses, our results challenge the notion that the they are optimally trading off the benefits of distorted beliefs with the costs of distorted behavior: although precise welfare statements are naturally challenging, the utility costs of budget distortions are substantial in our context and our evidence suggests that the benefits of more comprehensive retrieval are large.

If individuals are intentionally avoiding thinking about their expenses, they may have preferred to avoid the retrieval intervention. We investigate this in follow-up data collection when we return one year after the initial intervention. We ask treatment farmers their willingness

[43] Brunnermeier et al. (2008) provide an economic model of the planning fallacy that invokes motivated reasoning to explain overoptimistic task completion times.

to pay to receive the intervention again. Over 90% are willing to pay for the intervention, and we find no evidence of a desire to avoid it in other qualitative questions. This suggests that participants find the intervention welfare improving, but also have difficulty replicating or completing it on their own (perhaps due to a range of other issues, including present bias).

Our findings point to policies that may help address "under-saving" in low income populations. Most specifically, in addressing seasonal poverty, existing work in development has largely focused on financial interventions, such credit or incentives for migration during lean seasons (Bryan et al., 2014; Aggarwal, 2018; Burke et al., 2019; Fink et al., 2020). However, this approach leaves open the fundamental question of why seasonal poverty exists in the first place. In these settings, households begin the year with an endowment of wealth post-harvest, and then eat this endowment down over the course of the year. Hungry seasons arise because households exhaust their endowment before the end of the year (i.e., before the next harvest)—suggesting a failure of savings rather than credit. We depart from the bulk of the literature by framing seasonality as a savings problem rather than borrowing problem, and propose simple interventions that focus on beliefs rather than technology. These types of policies may be more cost effective than those that require credit or capital interventions. More generally, incorporating beliefs into savings and smoothing interventions may be relevant across a wide range of settings, as suggested by our evidence from the United States.

Finally, while our main application focuses on intertemporal allocation decisions, retrieval failures likely affect a broad class of decisions that require retrieving many detailed pieces of information to accurately optimize. For example, a microentrepreneur ordering inventory must consider existing stocks, a range of sales scenarios, and the substitutability of different items. Neglecting one of these pieces of information may lead to a sub-optimal order and lower profits. Similarly, a school teacher who is deciding whether to extend the number of days spent teaching a particularly challenging concept must keep in mind all future topics, unforeseen challenges in teaching them, and disruptions to the school schedule. Alternately, an executive deciding on the roll-out date of a new product must consider the various steps that must be completed and potential shocks. Retrieval failures therefore provide broad scope to explain mis-optimization, though testing the range of their explanatory power requires new empirical work. In addition, better understanding what gives rise to retrieval failures in the first place, and the impediments to learning (e.g., why is memory biased?) is important to both assess the generalizability of the phenomenon and to inform interventions to correct it.

References

Abel, Martin, Rulof Burger, Eliana Carranza, and Patrizio Piraino, "Bridging the intention-behavior gap? The effect of plan-making prompts on job search and employment," *American Economic Journal: Applied Economics*, 2019, *11* (2), 284–301.

Aggarwal, Shilpa, "Do rural roads create pathways out of poverty? Evidence from India," *Journal of Development Economics*, 2018, *133*, 375–395.

Aguiar, Mark and Erik Hurst, "Consumption versus expenditure," *Journal of Political Economy*, 2005, *113* (5), 919–948.

Anderson, John R and Robert Milson, "Human memory: An adaptive perspective," *Psychological Review*, 1989, *96* (4), 703.

Basu, Karna and Maisy Wong, "Evaluating seasonal food storage and credit programs in east Indonesia," *Journal of Development Economics*, 2015, *115*, 200–216.

Bénabou, Roland and Jean Tirole, "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–164.

Berman, Jonathan Z, An TK Tran, John G Lynch Jr, and Gal Zauberman, "Expense neglect in forecasting personal finances," *Journal of Marketing Research*, 2016, *53* (4), 535–550.

Beshears, John, Katherine L Milkman, and Joshua Schwartzstein, "Beyond beta-delta: The emerging economics of personal plans," *American Economic Review*, 2016, *106* (5), 430–34.

Bolte, Lukas and Collin Raymond, "Emotional Inattention," Technical Report, Working paper 2022.

Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer, "Memory and probability," *Quarterly Journal of Economics*, 2023, *138* (1), 265–311.

_, Nicola Gennaioli, and Andrei Shleifer, "Memory, attention, and choice," *Quarterly Journal of Economics*, 2020, *135* (3), 1399–1442.

Brunnermeier, Markus K, Filippos Papakonstantinou, and Jonathan A Parker, "An economic model of the planning fallacy," Technical Report, National Bureau of Economic Research 2008.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh," *Econometrica*, 2014, *82* (5), 1671–1748.

Buehler, Roger, Dale Griffin, and Johanna Peetz, "The planning fallacy: Cognitive, motivational, and social origins," in "Advances in experimental social psychology," Vol. 43, Elsevier, 2010, pp. 1–62.

—, —, and Michael Ross, "Exploring the "planning fallacy": Why people underestimate their task completion times," *Journal of Personality and Social Psychology*, 1994, *67* (3), 366.

Burke, Marshall, Lauren Falcao Bergquist, and Edward Miguel, "Sell low and buy high: arbitrage and local price effects in Kenyan markets," *Quarterly Journal of Economics*, 2019, *134* (2), 785–842.

Carranza, Eliana, Aletheia Donald, Florian Grosset, and Supreet Kaur, "The Social Tax: Redistributive Pressure and Labor Supply," 2021.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky, "The limits of simple implementation intentions: Evidence from a field experiment on making plans to exercise," *Journal of Health Economics*, 2018, *62*, 95–104.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and taxation: Theory and evidence," *American Economic Review*, 2009, *99* (4), 1145–77.

— and Emmanuel Saez, "Teaching the tax code: Earnings responses to an experiment with EITC recipients," *American Economic Journal: Applied Economics*, 2013, *5* (1), 1–31.

Crossley, Thomas F and Joachim K Winter, "Asking households about expenditures: what have we learned?," in "Improving the measurement of consumer expenditures," University of Chicago Press, 2014, pp. 23–50.

Deaton, Angus, Margaret Grosh et al., "Consumption," *Designing Household Survey Questionnaires for Developing Countries: lessons from ten years of LSMS experience. Washington, USA. The World Bank*, 1998, pp. 1–78.

Dercon, Stefan and Pramila Krishnan, "Vulnerability, seasonality and poverty in Ethiopia," *Journal of Development Studies*, 2000, *36* (6), 25–53.

Dillon, Brian, Joachim De Weerdt, and Ted OâDonoghue, "Paying more for less: why donât households in Tanzania take advantage of bulk discounts?," *World Bank Economic Review*, 2021, *35* (1), 148–179.

Fafchamps, Marcel, Christopher Udry, and Katherine Czukas, "Drought and saving in West Africa: are livestock a buffer stock?," *Journal of Development Economics*, 1998, *55* (2), 273–305.

Fedyk, Anastassia, "Asymmetric naivete: Beliefs about self-control," *Available at SSRN 2727499*, 2018.

Fink, Gunther, B Kelsey Jack, and Felix Masiye, "Seasonal liquidity, rural labor markets, and agricultural production," *American Economic Review*, 2020, *110* (11), 3351–92.

Forsyth, Darryl K and Christopher DB Burt, "Allocating time to future tasks: The effect of task segmentation on planning fallacy bias," *Memory & Cognition*, 2008, *36* (4), 791–798.

Gabaix, Xavier, "Behavioral inattention," in "Handbook of behavioral economics: Applications and foundations 1," Vol. 2, Elsevier, 2019, pp. 261–343.

Ganong, Peter and Pascal Noel, "Consumer spending during unemployment: Positive and normative implications," *American Economic Review*, 2019, *109* (7), 2383–2424.

Gerard, François and Joana Naritomi, "Job displacement insurance and (the lack of) consumption-smoothing," *American Economic Review*, 2021, *111* (3), 899–942.

Griffin, Dale and Roger Buehler, "Biases and fallacies, memories and predictions: Comment on Roy, Christenfeld, and McKenzie (2005).," 2005.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing information provision experiments," *Journal of Economic Literature*, 2023, *61* (1), 3–40.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, "Learning through noticing: Theory and evidence from a field experiment," *Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

Jakiela, Pamela and Owen Ozier, "Does Africa need a rotten kin theorem? Experimental evidence from village economies," *The Review of Economic Studies*, 2016, *83* (1), 231–268.

Kahana, Michael J. and Anthony D. Wagner, *Oxford Handbook of Human Memory*, Oxford University Press, 2023.

Kahana, Michael Jacob, *Foundations of human memory*, Oxford University Press, 2012.

Kahneman, Daniel and Amos Tversky, "On the psychology of prediction," *Psychological Review*, 1973, *80* (4), 237.

_ and _ , "Intuitive prediction: Biases and corrective procedures," Technical Report, Decisions and Designs Inc Mclean Va 1977.

_ , Dan Lovallo, and Olivier Sibony, "Before you make that big decision," *Harvard Business Review*, 2011, *89* (6), 50–60.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman, "Getting to the top of mind: How reminders increase saving," *Management Science*, 2016, *62* (12), 3393–3411.

Kling, Jeffrey R, Sendhil Mullainathan, Eldar Shafir, Lee C Vermeulen, and Marian V Wrobel, "Comparison friction: Experimental evidence from Medicare drug plans," *Quarterly Journal of Economics*, 2012, *127* (1), 199–235.

Kuhn, Michael A, "Electronic Benefit Transfer and Food Expenditure Cycles," *Journal of Policy Analysis and Management*, 2021, *40* (3), 744–773.

Lichtenstein, Sarah, Paul Slovic, Baruch Fischhoff, Mark Layman, and Barbara Combs, "Judged frequency of lethal events," *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4* (6), 551.

Malmendier, Ulrike and Jessica A Wachter, *Oxford Handbook of Human Memory*, Oxford Univeristy Press, 2023.

Mason, Alice, Elliot Ludvig, and Christopher Madan, *Oxford Handbook of Human Memory*, Oxford Univeristy Press, 2023.

Merfeld, Joshua D and Jonathan Morduch, "Poverty at higher frequency," *Working Paper*, 2023.

Milkman, Katherine L, John Beshears, James J Choi, David Laibson, and Brigitte C Madrian, "Using implementation intentions prompts to enhance influenza vaccination rates," *Proceedings of the National Academy of Sciences*, 2011, *108* (26), 10415–10420.

_ , _ , _ , _ , and _ , "Planning prompts as a means of increasing preventive screening rates," *Preventive Medicine*, 2013, *56* (1), 92–93.

Mullainathan, Sendhil, "A memory-based model of bounded rationality," *Quarterly Journal of Economics*, 2002, *117* (3), 735–774.

Nickerson, David W and Todd Rogers, "Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making," *Psychological Science*, 2010, *21* (2), 194–199.

Paxson, Christina H, "Using weather variability to estimate the response of savings to transitory income in Thailand," *American Economic Review*, 1992, pp. 15–33.

Peetz, Johanna and Roger Buehler, "Is there a budget fallacy? The role of savings goals in the prediction of personal spending," *Personality and Social Psychology Bulletin*, 2009, *35* (12), 1579–1591.

_ , _ , Derek J Koehler, and Ester Moher, "Bigger not better: Unpacking future expenses inflates spending predictions," *Basic and Applied Social Psychology*, 2015, *37* (1), 19–30.

Pew Charitable Trusts, "Barriers to saving and policy opportunities: The role of emergency savings in family financial security (Brief)," 2016.

Pop-Eleches, Cristian, Harsha Thirumurthy, James P Habyarimana, Joshua G Zivin, Markus P Goldstein, Damien De Walque, Leslie Mackeen, Jessica Haberer, Sylvester Kimaiyo, John Sidle et al., "Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: A randomized controlled trial of text message reminders," *AIDS (London, England)*, 2011, *25* (6), 825.

Roy, Michael M, Nicholas JS Christenfeld, and Craig RM McKenzie, "Underestimating the duration of future events: Memory incorrectly used or memory bias?," *Psychological Bulletin*, 2005, *131* (5), 738.

Schwartzstein, Joshua, "Selective attention and learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

Shapiro, Jesse M, "Is there a daily discount rate? Evidence from the food stamp nutrition cycle," *Journal of Public Economics*, 2005, *89* (2-3), 303–325.

Stilley, Karen M, J Jeffrey Inman, and Kirk L Wakefield, "Planning to make unplanned purchases? The role of in-store slack in budget deviation," *Journal of Consumer Research*, 2010, *37* (2), 264–278.

Sussman, Abigail B and Adam L Alter, "The exception is the rule: Underestimating and overspending on exceptional expenses," *Journal of Consumer Research*, 2012, *39* (4), 800–814.

Tversky, Amos and Daniel Kahneman, "Heuristics and biases: Judgement under uncertainty," *Science*, 1974, *185* (4157), 1124–30.

_ and _ , "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment," *Psychological Review*, 1983, *90* (4), 293.

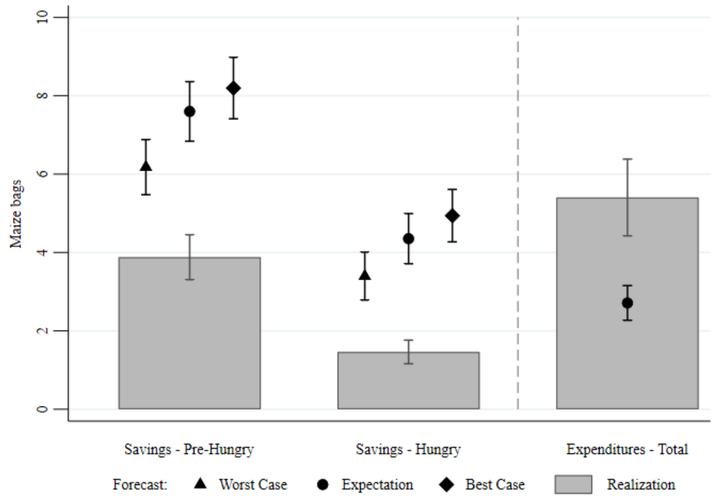
Wachter, Jessica A and Michael J Kahana, "Associative learning and representativeness," *Working Paper*, 2023.

_ and Michael Jacob Kahana, "A retrieved-context theory of financial decisions," Technical Report, National Bureau of Economic Research 2019.

Wimmer, G Elliott and Daphna Shohamy, "Preference by association: How memory mechanisms in the hippocampus bias decisions," *Science*, 2012, *338* (6104), 270–273.

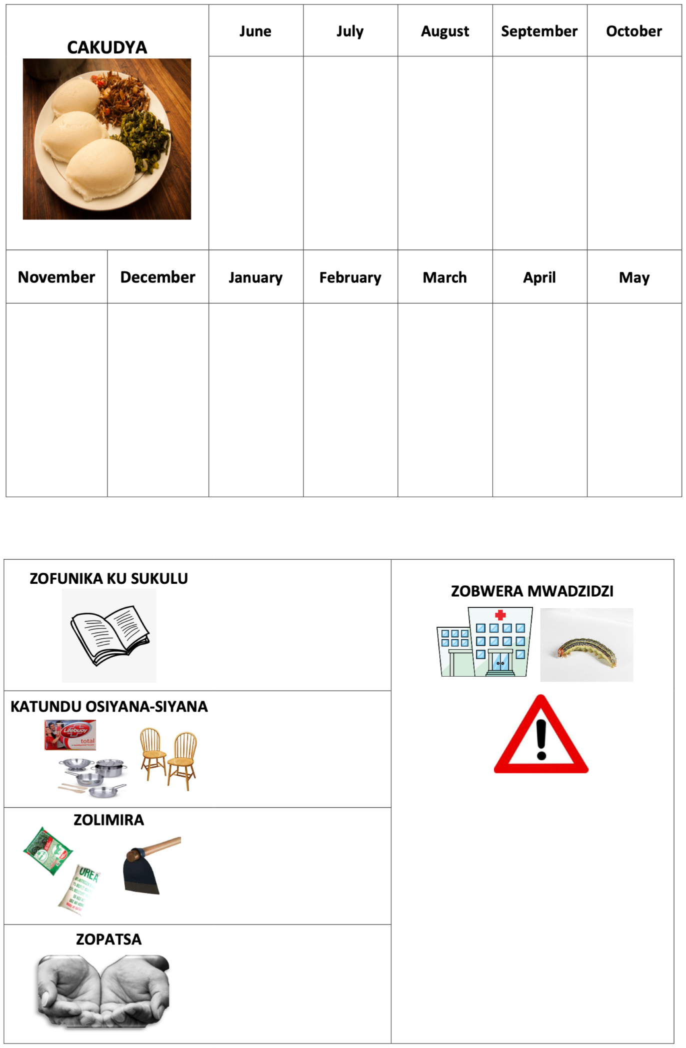
11 Tables and Figures

Figure 1: Overoptimism in savings and expenditure forecasts (field experiment)



Notes: Forecasts (bars) and realizations (dots), for savings (on the left) and non-food expenditures (on the right). Error bars correspond to 95% confidence intervals. Participants were asked to predicted their expected savings (in number of bags of maize) at two future dates, as well as savings in the best and worst-case situations. They were also asked to predict non-food expenses until the next harvest. Realizations were measured during follow-up survey visits. The sample is restricted to participants in the control group, whose forecast was incentivized and who surveyed in follow-up rounds.

Figure 2: Expense board and categories

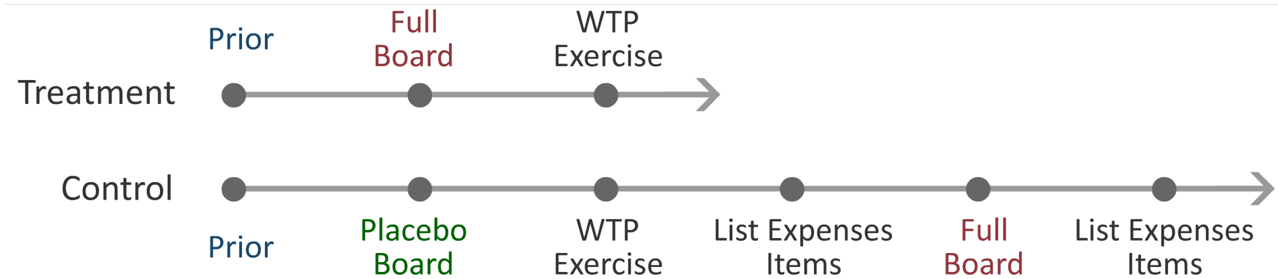


(a) Panel A: Full expense board

(b) Panel B: Two category expense board

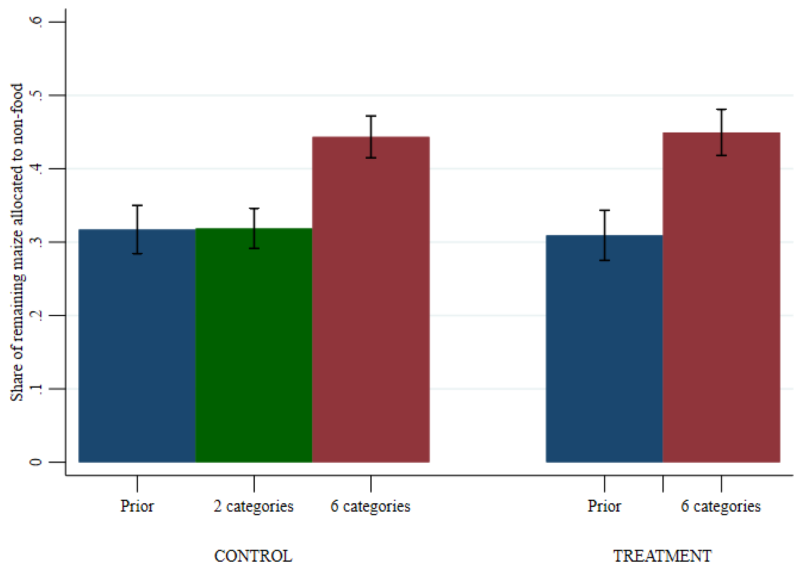
Notes: Panel A: The full expense board used by the treatment group to allocate their savings to expense categories. As part of the treatment, participants receive a set of pins, with each pin representing one bag of maize from their savings. Participants assign pins to food consumption (allocated by month) or to six broad non-food expense categories: school fees, household supplies, farming inputs, transfers, emergencies, and other expenses (for which participants place pins outside of the categories shown). Panel B: The "placebo" expense board used by the control group in the mechanism experiment to allocate their income (maize bags) to expense categories (either in the food or in the non-food category) using pins.

Figure 3: Timeline (mechanism experiment)

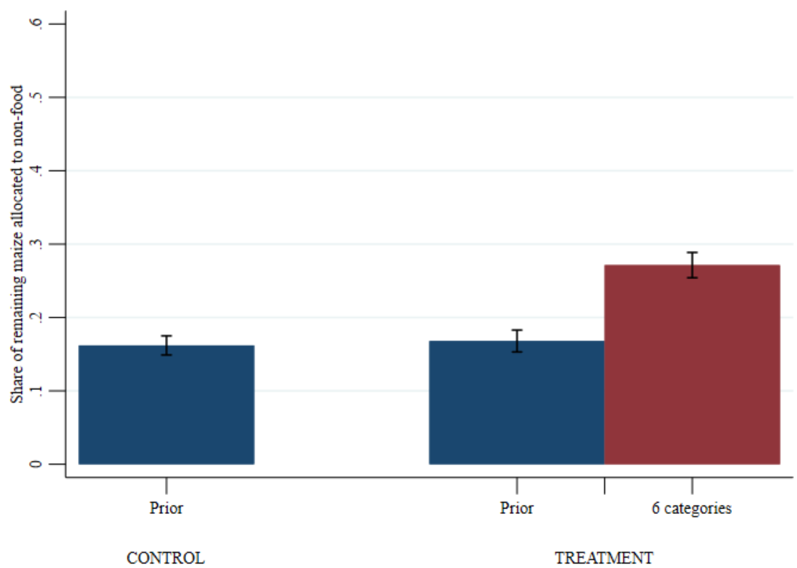


Notes: Mechanism experiment timeline. All participants are asked for their prior of non-food expenses. They are then randomized into treatment and control groups. The treatment group completes the expense board in Panel A of Figure 2 and the control group completes the board in Panel B. Both groups then complete the willingness-to-pay exercise. Finally, the control group goes through an additional set of activities to identify the previously-neglected expense items.

Figure 4: Share of maize bags allocated to non-food expenses



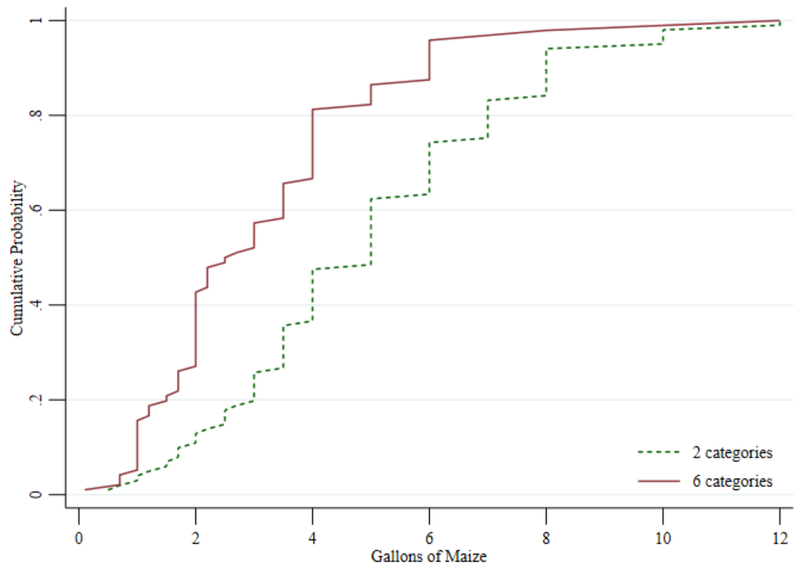
(a) Panel A: Mechanism experiment



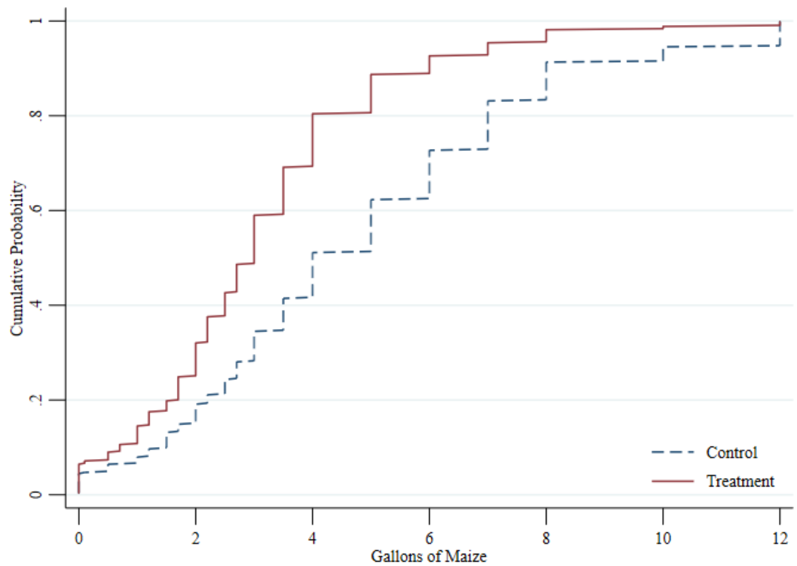
(b) Panel B: Field experiment

Notes: Panel A shows results from the mechanism experiment. Both treatment and control are first asked their estimate ("prior") of non-food expenditures without any retrieval board. The blue bar represents the share of their current maize stock allocated to non-food expenses. The control then completes the simplified two-category retrieval exercise (green bar) while the treatment completes the full six-category exercise (maroon bar). After, the control also completes the six-category exercise (maroon bar). Panel B shows results from the field experiment. Both groups are first asked for their prior (blue bar). Only the treatment completes the six-category exercise (red bar).

Figure 5: Willingness to exchange maize for discretionary consumption



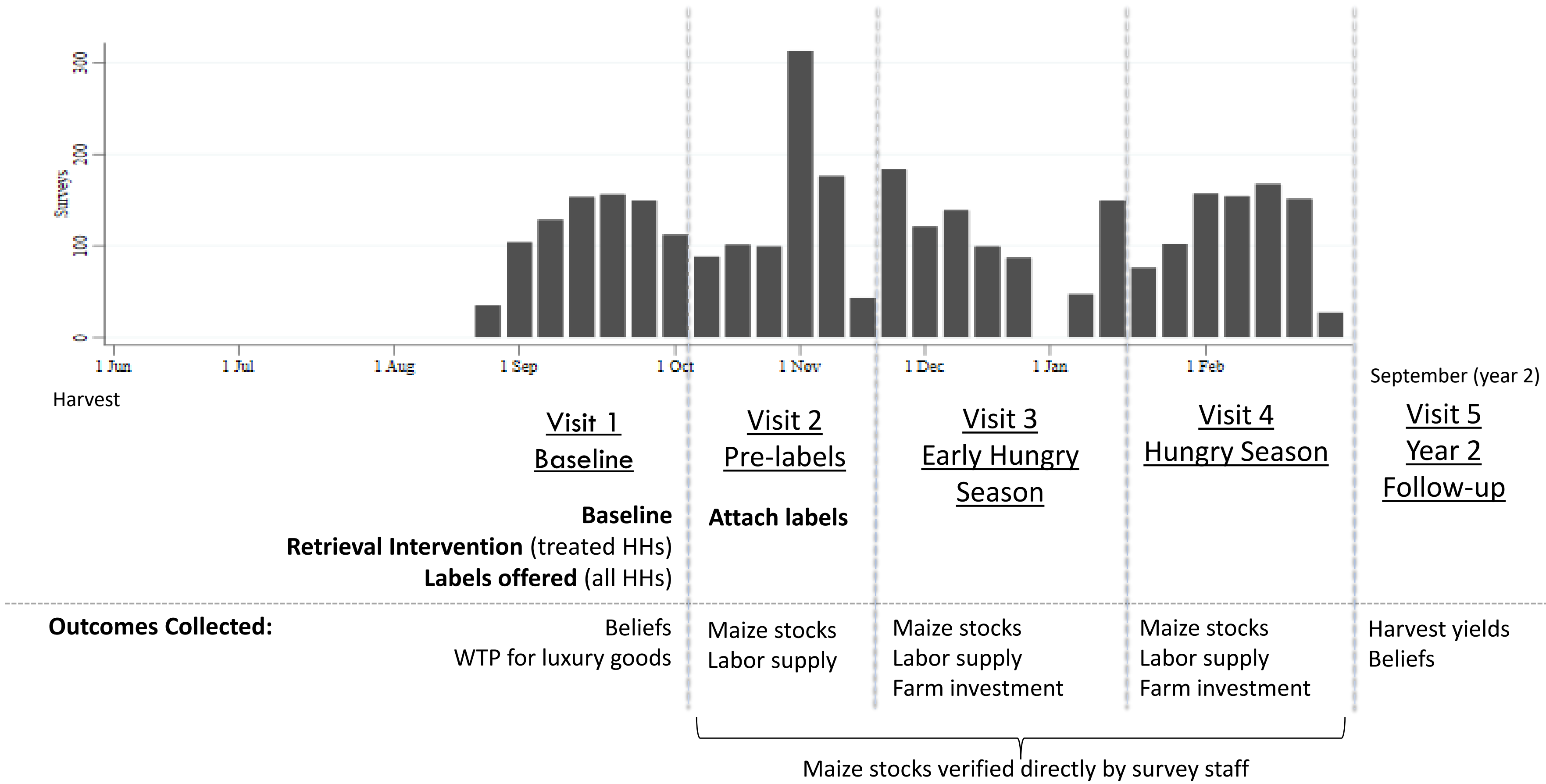
(a) Panel A: Mechanism experiment



(b) Panel B: Field experiment

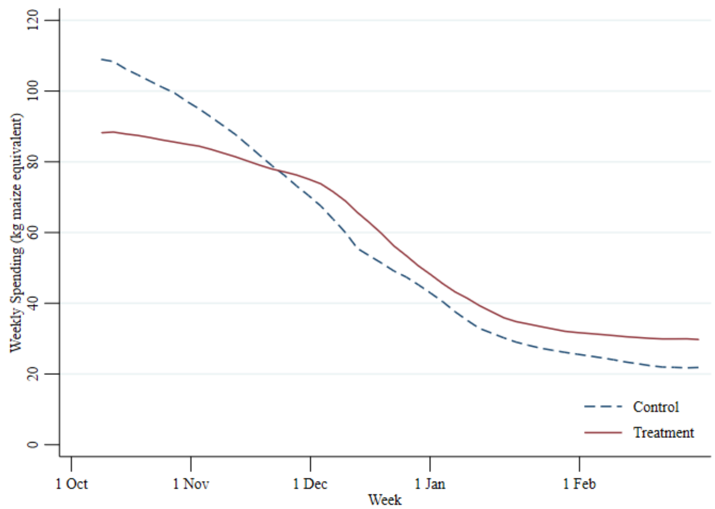
Notes: CDF of the willingness to pay (WTP) for a discretionary consumption items. WTP is elicited using the Becker-DeGroot-Marschak method after the retrieval exercise. In Panel A the green dashed line shows the WTP in the control group (two-category exercise), and the maroon solid line shows the WTP in the treatment group (six-category exercise) for the mechanism experiment. In Panel B the blue dotted line shows the WTP in the control group (no retrieval exercise), and the maroon solid line shows the WTP in the treatment group (six-category exercise) for the field experiment. Valuations are measured in gallons of maize. Maize could be traded for one of three items: 1) a cloth wrap used as clothing, 2) a radio or 3) a solar panel. The preferred item was chosen at the beginning of the baseline survey, prior to the intervention.

Figure 6: Timeline and data collection (field experiment)



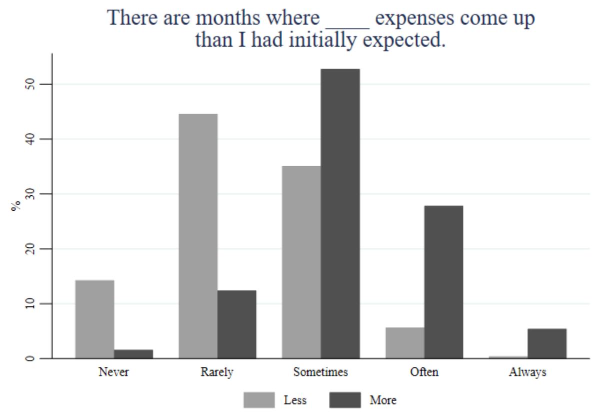
Notes: Field experiment timeline, including the intervention and data collection. Vertical bars represent the number of participants interviewed each week over the course of the study, by data collection rounds. The intervention was administered at the same time as the baseline survey (Visit 1). Labels were provided to participants that took them up in Visit 2. The main outcome measures collected during each round are printed at the bottom of the figure. See text for additional detail on the outcomes.

Figure 7: Implied consumption path based on observed savings (field experiment)

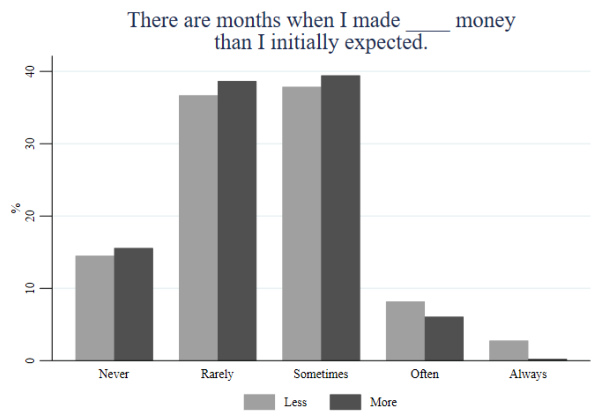


Notes: Smoothed consumption paths for treatment (maroon solid line) and control (blue dotted line) participants in the field experiment. The dependent variable is constructed as the difference in savings (measured as kilograms of maize plus the maize value of cash savings), divided by the number of weeks between survey visits. This approximates "weekly consumption". This is then regressed on dummies for baseline wealth. The residuals are fit with a smoothed local polynomial. The residual series is rescaled so that the starting stock matches the level measured in kilograms of maize.

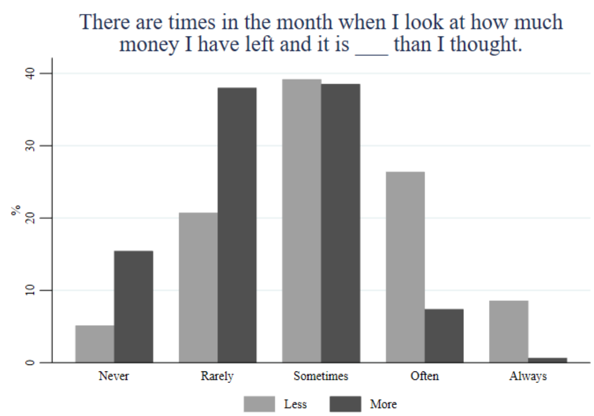
Figure 8: Asymmetry in income and expenses (U.S. sample)



(a) Panel A: Expenses



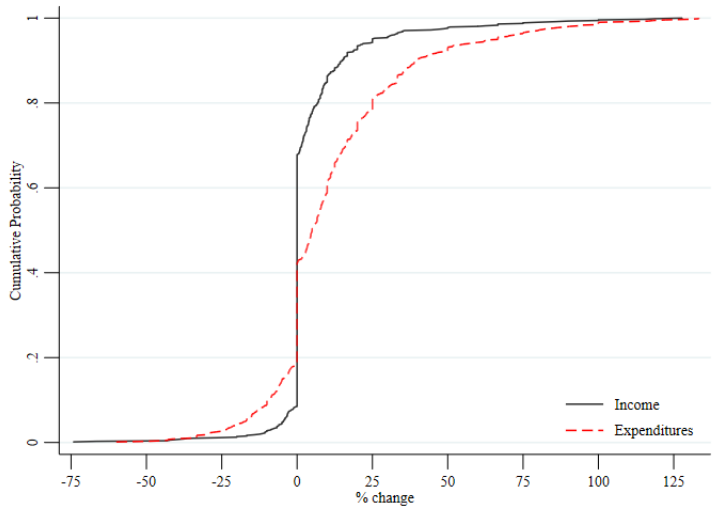
(b) Panel B: Income



(c) Panel C: Savings

Notes: Respondants are show a series of situations about income, expenditures and savings and asked how often they occur. For example, in Panel A, the light bars represent the distribution of responses to "There are months where less expenses come up than I had initially expected" and dark bars to "There are months where more expenses come up than I had initially expected." The sample is restricted to employed individuals not living with their parents (N=721). Respondents see all versions of the questions.

Figure 9: Percentage change before/after retrieval exercise (U.S. sample)



Notes: CDF of the percentage update of income and expenditures of Prolific survey participants after undergoing the retrieval exercise. The grey solid line corresponds to the update in income, the red dashed line corresponds to the update in expenditures. The percentage update is calculated as the percentage difference between the aggregate estimate and the sums of the category-by-category estimates. The sample is restricted to employed individuals not living with their parents (N=721). The 5% of participants who update their income or expenses by more than 400% are excluded from the plot.

Table 1: Characteristics of forgotten expenses (mechanism experiment)

	Forgotten (1)	Forgotten (2)	Forgotten (3)	Forgotten (4)	Forgotten (5)
Expense Size (bags)	-0.05**				-0.05**
	(0.02)				(0.02)
Frequency Uncertain		0.32***			0.18*
		(0.05)			(0.11)
Irregular			0.27***		0.06
			(0.04)		(0.08)
Item Uncertain				0.30***	0.07
				(0.04)	(0.11)
N	467	467	467	467	467
Mean Ref. Category	0.61	0.55	0.54	0.54	0.53
Baseline Controls	Yes	Yes	Yes	Yes	Yes

Notes: Characteristics of items retrieved in the full expense board that were not included when participants used the two category board. "Frequency uncertain" refers to expenses with an uncertain frequency of expenditure. "Time uncertain" refers to expenses with uncertain expenditure timing. "Item uncertain" refers to expenses where the exact spending is uncertain (e.g., in the case of "emergencies"). Expense characteristics were elicited from participants following the full expense board exercise. Baseline controls include: quantity of maize remaining and level of savings. Standard errors are clustered at the individual level.

Table 2: Impact of the retrieval exercise on savings (field experiment)

	Cash & Maize (1)	Cash & Maize (2)	Maize Bags (3)	Cash (ZMW) (4)
Treat x Visit 2 (Pre-Labels)	99.86**	101.45***	0.80**	150.94*
	(44.76)	(37.79)	(0.37)	(87.20)
Treat x Visit 3 (Early Hungry)	68.18**	70.14***	0.57**	102.33**
	(28.70)	(24.13)	(0.27)	(45.12)
Treat x Visit 4 (Hungry)	15.45	15.52	0.09	39.45
	(18.60)	(18.63)	(0.19)	(43.75)
Dependent Variable unit	Kg	Kg	Bags	Kwacha
N	2480	2480	2480	2480
Control Mean Visit 2	660.51	660.51	7.99	426.60
Control Mean Visit 3	335.83	335.83	3.93	277.95
Control Mean Visit 4	156.72	156.72	1.43	234.04
F-test 2v3	0.32	0.32	0.48	0.53
F-test 3v4	0.01	0.01	0.03	0.14
Baseline controls	No	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes

Notes: Treatment effects on savings in maize and cash. The coefficients show the effect by survey visit, in chronological order. The outcome variable in columns 1-2 is total unprocessed maize in storage, plus the maize value of cash savings. Cash savings are converted into maize using the price of maize in Katete market (on the day of the survey visit). The dependent variable in column 3 is the number of bags of maize that the participant had in storage, measured by direct surveyor observation. Column 4 shows self reported cash savings, measured in the local currency (Zambian kwacha). Columns 2-4 control for baseline characteristics, and all specifications control for week of survey fixed effects. Standard errors are clustered at the household level.

Table 3: Impact of the retrieval exercise on consumption and investment (field experiment)

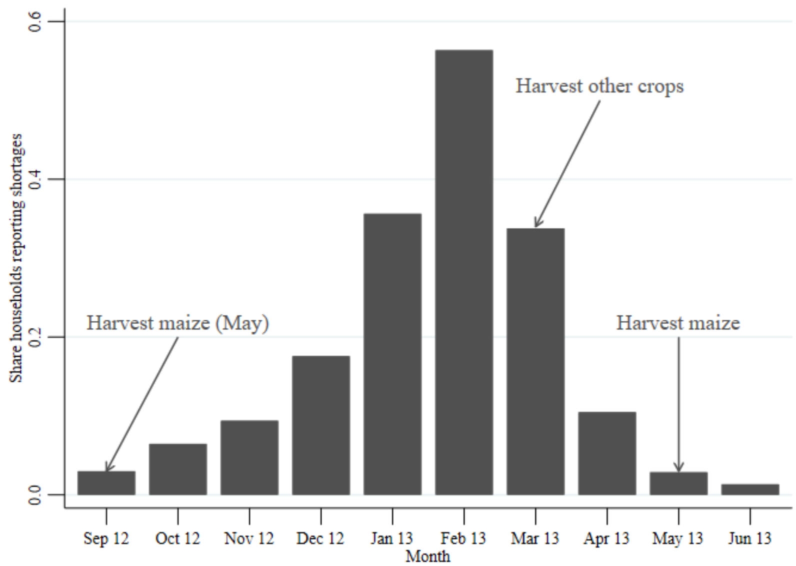
	Meals Per Day (1)	Days Wage Labor (2)	Wage Earnings (3)	Days on Farm (4)	Days Hired Labor (5)	Fertilizer/ Chemical Exp (6)	Total Crop Revenue (7)
Treat						69.16	698.48*
						(46.47)	(392.98)
Treat x Visit 2 (Pre-Labels)	-0.03	0.29	0.38				
	(0.03)	(0.24)	(13.15)				
Treat x Visit 3 (Early Hungry)	0.05*	-0.09	0.43	0.74			
	(0.03)	(0.11)	(4.16)	(0.73)			
Treat x Visit 4 (Hungry)	-0.01	-0.22**	-10.67***	1.23*	0.66*		
	(0.03)	(0.11)	(3.98)	(0.72)	(0.38)		
Dependent Variable unit	# meals	Days	Kwacha	Days	Days	Kwacha	Kwacha
N	2480	2480	2480	1654	823	814	814
Control mean						718.68	7433.35
Control Mean Visit 2	2.11	1.24	56.11				
Control Mean Visit 3	2.01	0.51	17.42	18.21			
Control Mean Visit 4	2.02	0.68	25.58	15.26	1.10		
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Treatment effects on consumption and investment outcomes. The dependent variables are: the self reported number of staple meals consumed yesterday (column 1), the number of person-days household members performed ganyu (wage labor) over the past 4 weeks (Visit 2) or past week (Visit 3 and 4) (column 2), total household earnings from ganyu (column 3), self reported person-days of family labor on the household farm (column 4), number of person-days of hired labor (column 5), annual expenditures on fertilizer and other pesticides/herbicides (column 6), and total harvest value for the agricultural year 2019/2020 (column 7). Data used to estimate columns 1-5 were collected in one or more short-recall survey visits. Data used to estimate columns 6 and 7 were collected after the following harvest. All specifications control for baseline characteristics. Standard errors are clustered at the household level.

A Supplementary Appendix

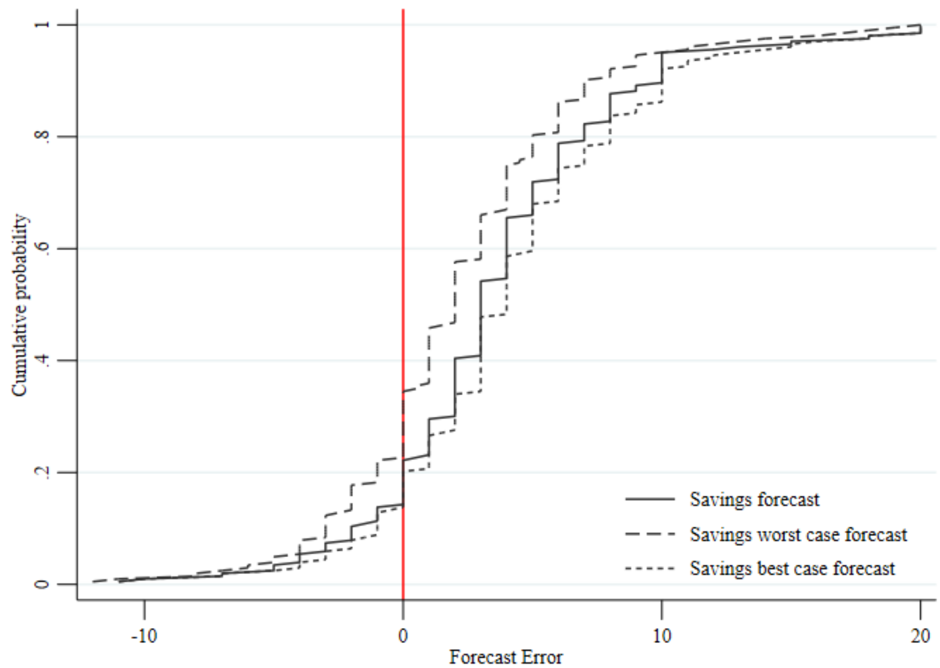
A.1 Appendix Figures

Figure A.1: Reported food shortages by month among Zambian farmers

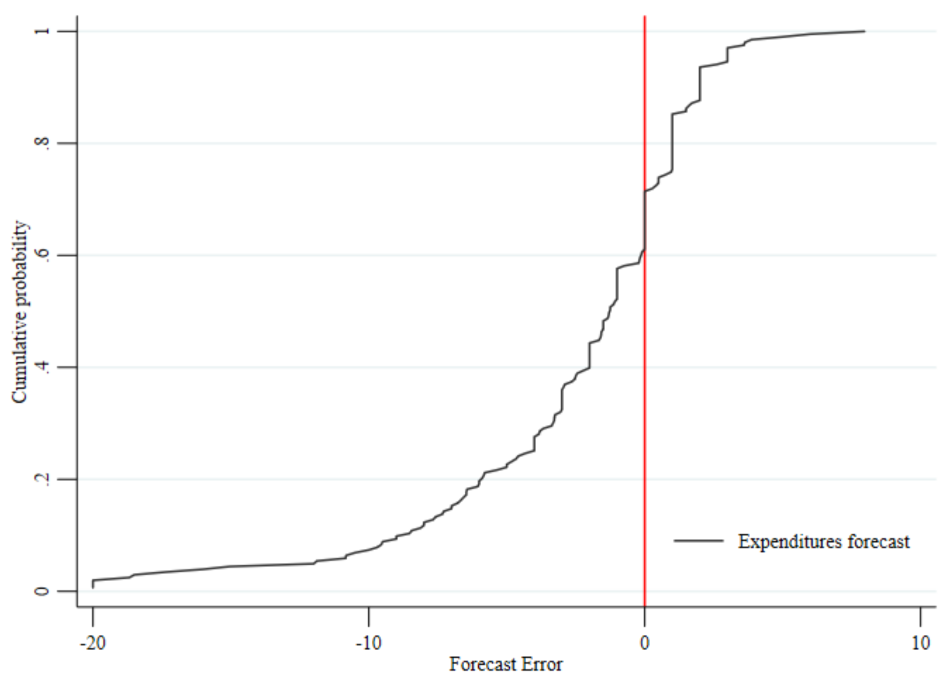


Notes: Proportion of farmers in Eastern Zambia reporting food shortages by month. The data come from Fink et al. (2020). The sample consists of farming households located in Chipata district, Zambia.

Figure A.2: Overoptimism in savings and expenditure forecasts (field experiment)



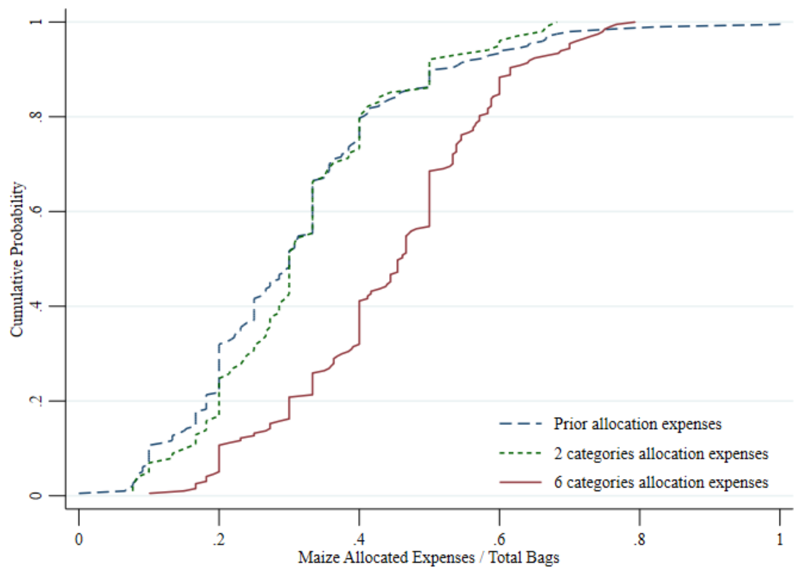
(a) Panel A: Savings forecast



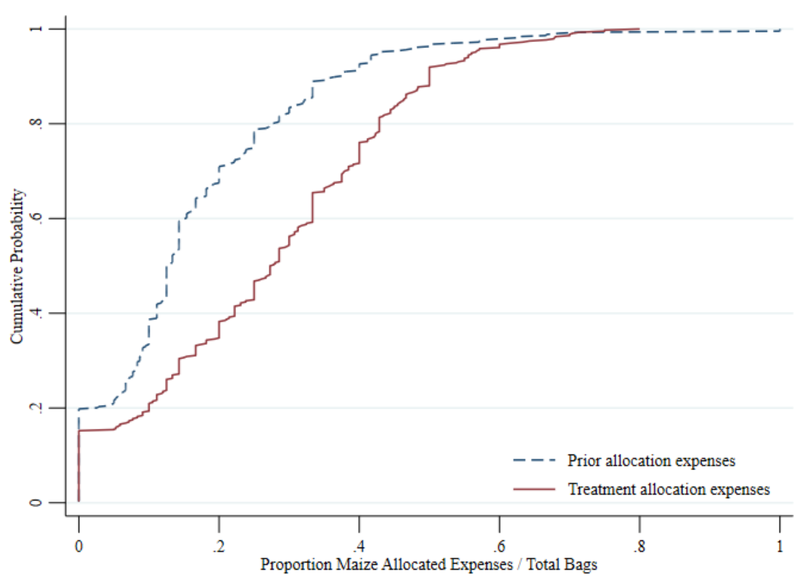
(b) Panel B: Expenditure forecast

Notes: Panel A shows the CDF of participant forecasts of future savings, relative to realizations measured in Visit 3 (Pre-Hungry Season). Panel B shows the CDF of participant forecasts of future expenditures, relative to realizations measured in Visit 3 and 4. We refer to these differences as the participant's forecast error. Participants are asked the number of bags of maize they expected to have in savings by a future survey visit or to have spent during the year. Participants were also asked the best and worst case expected savings.

Figure A.3: Proportion of current maize wealth allocated to non-food expenses



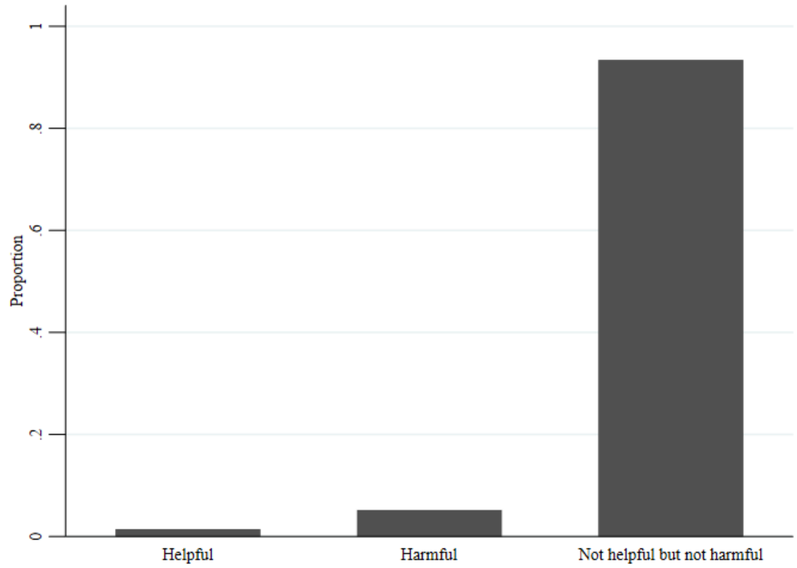
(a) Panel A: Mechanism experiment



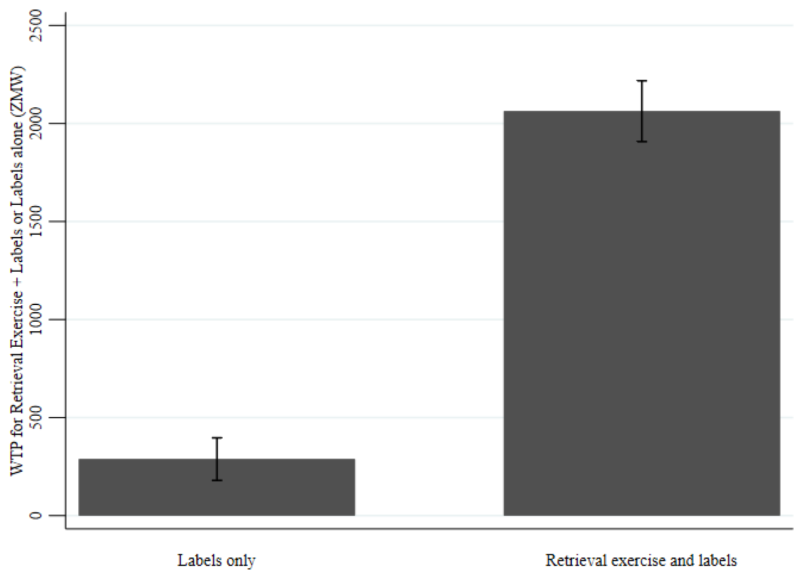
(b) Panel B: Field experiment

Notes: CDF of the proportion of the participant's current stock of maize bags allocated to non-food expenses. The blue dashed line shows the proportion allocated prior to undergoing the retrieval exercise. Data from the participant's expense board are used to construct an updated belief for the participant, post retrieval exercise. Panel A shows results from the mechanism experiment, which compared a simplified two category expense board (control, green dotted line) with the full six category board (treatment, maroon solid line). Panel B shows results from the field experiment, which elicited priors for both treatment and control, then measured the share of bags allocated to non-food expenses in the treatment group during the retrieval exercise. The blue dashed line shows the treatment group prior. The maroon line shows the post-retrieval allocation based on school fees, household supplies, farming inputs, emergencies and transfers.

Figure A.4: Evidence of value of labels (field experiment)



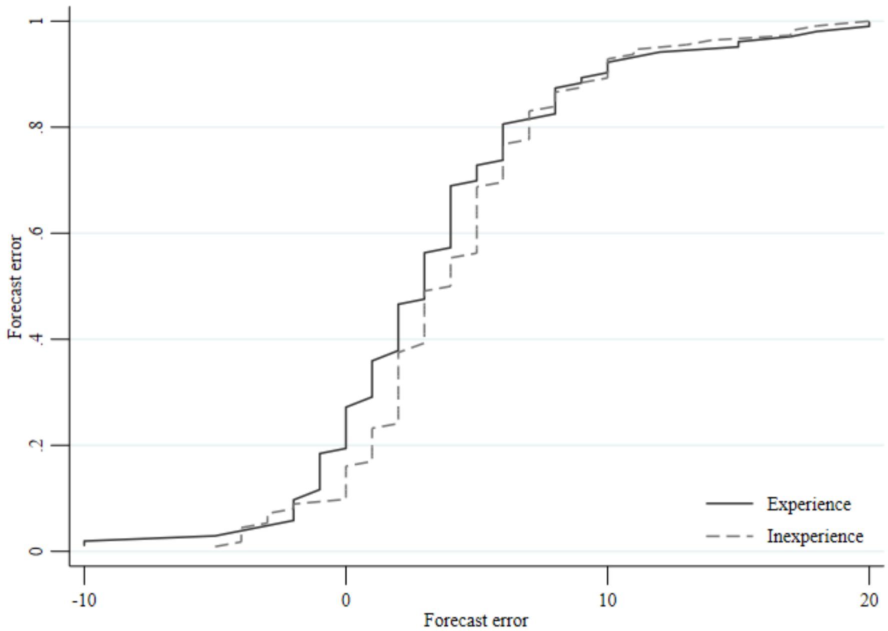
(a) Panel A: Hypothetical effect of labels without retrieval exercise



(b) Panel B: Valuation of retrieval and labels versus labels alone

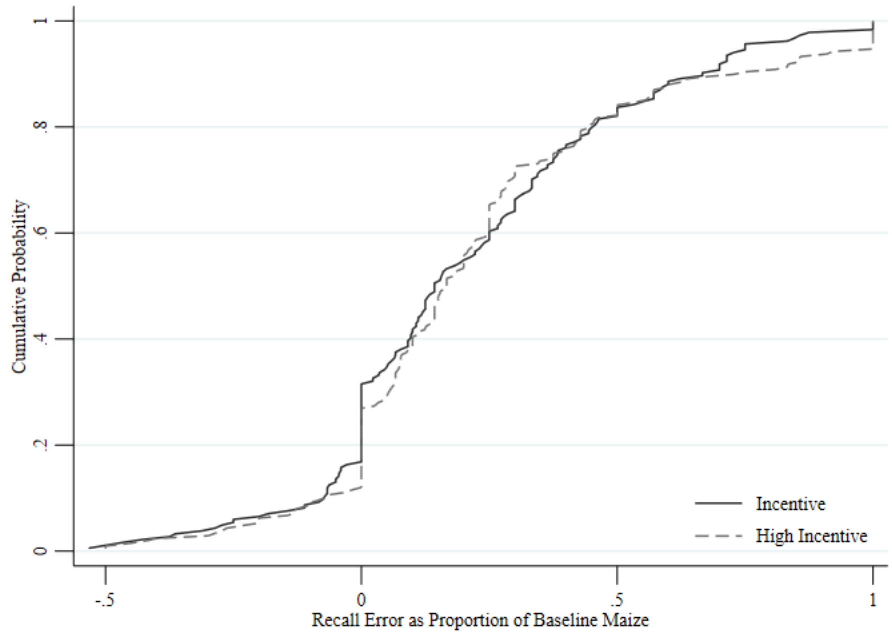
Notes: Panel A shows the proportion of participants who indicated that being given labels alone would have been 1) helpful, 2) harmful or 3) neither helpful nor harmful if labels were not accompanied by the retrieval exercise. The questions was asked during the second year of the project. The sample is restricted to the treatment group. Panel B shows participants' willingness to pay for labels alone or labels bundled with the retrieval exercise. Participants were told to consider a household similar to their own and asked whether they preferred this household to receive 1) labels alone or 2) cash, for varying amounts of cash in a Becker-DeGroot-Marschack procedure. The exercise was then repeated with the same household but the respondent was asked whether they preferred this household to receive 1) the retrieval exercise and labels or 2) cash. The sample is restricted to the treatment group.

Figure A.5: Forecast versus realized maize savings, by experience (field experiment)



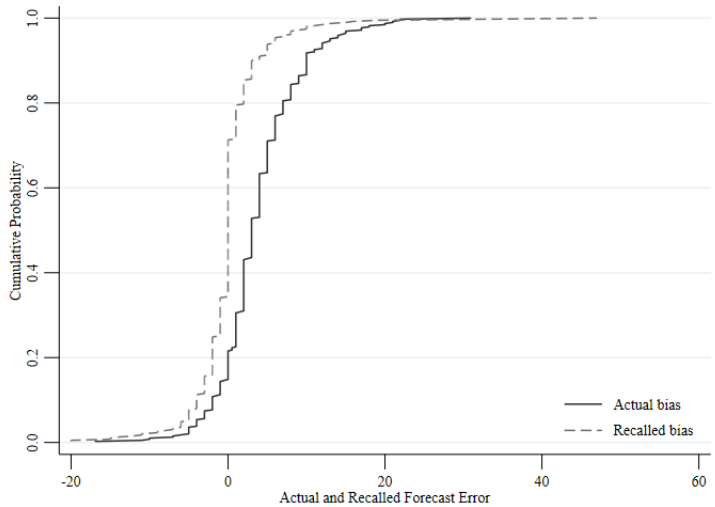
Notes: Forecasts versus realizations of savings. Experience is proxied by age: individuals in the lowest age quartile (mean age 28 - dashed line) are categorized as inexperienced, those in the top age quartile (mean age 62 - solid line) as experienced. The sample is restricted to control participants whose forecast was incentivized at baseline.

Figure A.6: Memory error – Recalled maize savings versus actual savings (field experiment)



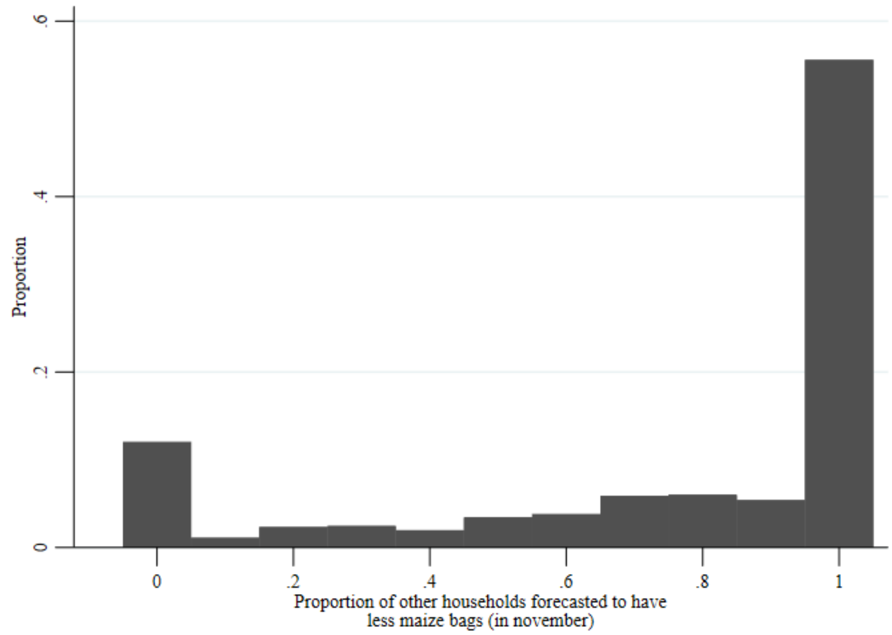
Notes: Recollection of savings in maize bags at the time of a past survey visit minus the number of maize bags measured during that past visit. The solid line shows the low incentive condition for accurate recall; the high incentive condition (dashed line) doubles the payoff for accuracy.

Figure A.7: Sophistication of past overoptimism (field experiment)



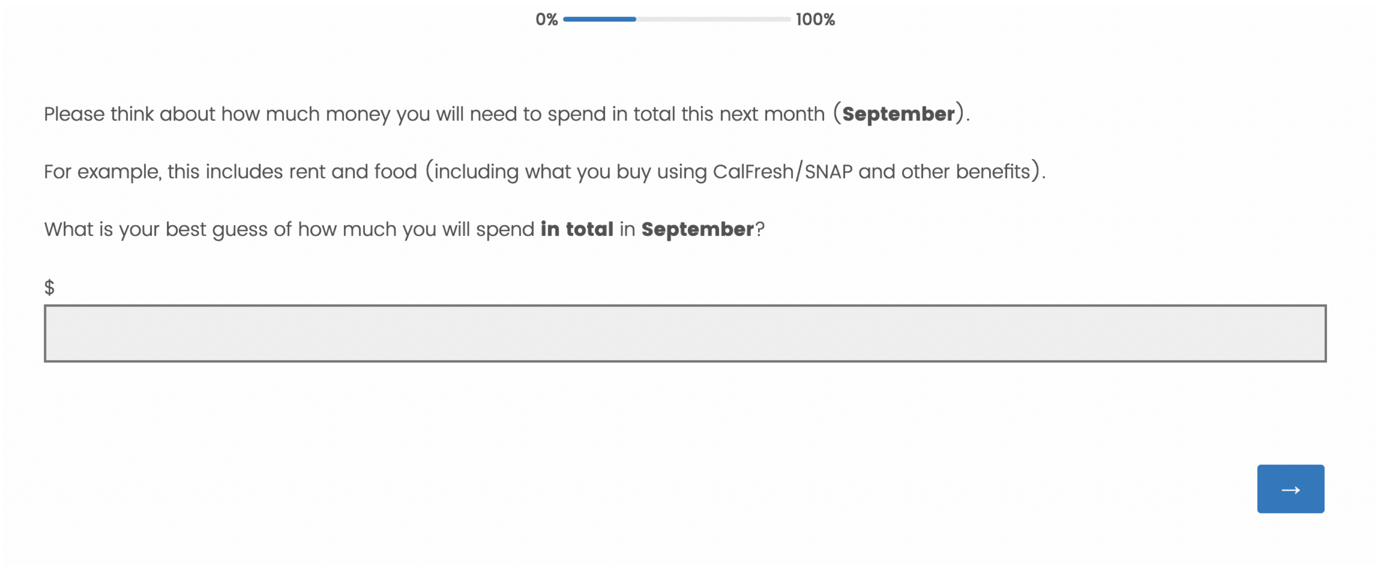
Notes: Actual forecast error and recalled forecast error. Actual forecast error is measured as forecasted savings minus realized savings (dark grey solid line). Recalled forecast error is the household's recalled forecast minus recalled savings (light grey dashed line). The sample is restricted to control households only.

Figure A.8: Own forecast, relative to forecasts for other similar households (field experiment)



Notes: Proportion of other households forecasted to have less maize own forecast. Participants were asked to forecast how many bags of maize they would have in a future survey round. Participants were then asked to think about 10 other households similar to themselves, i.e. households that lived nearby, had similar household sizes, had the same harvest size, and the same amount of that harvest remaining. They were then asked to forecast how many bags of maize these 10 households would have. Participants' forecasts of own savings relative to others is shown in the histogram. A value of one indicates that a participant forecast they would have more bags of maize than all ten other households.

Figure A.9: Prolific Survey Screenshots



0% 100%

Please think about how much money you will need to spend in total this next month (**September**).

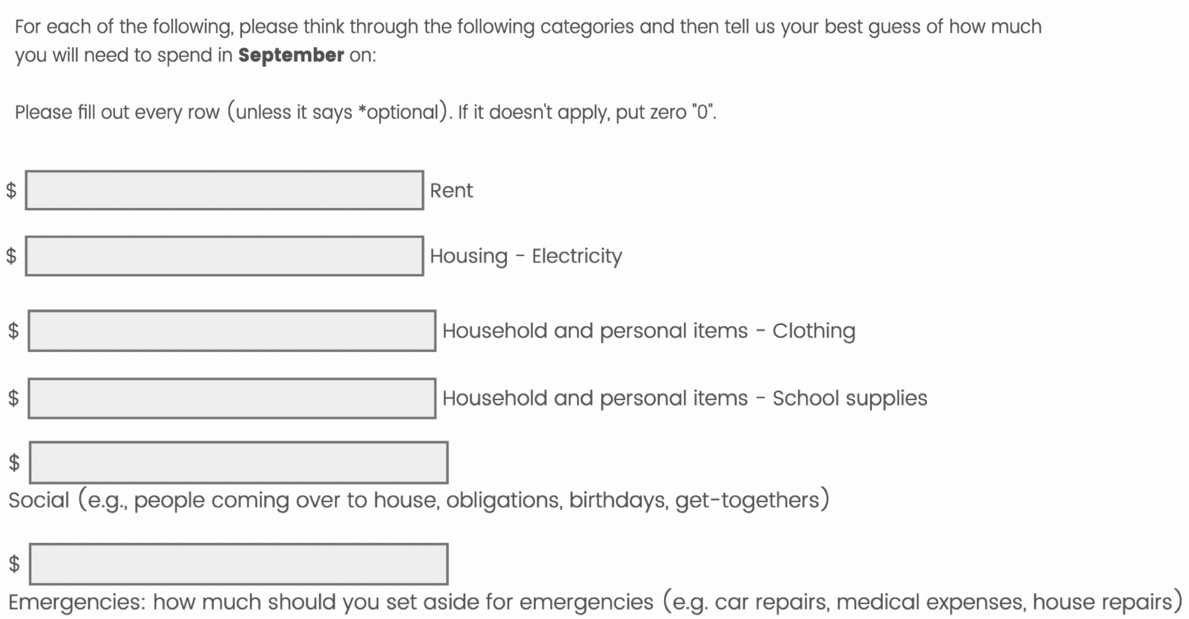
For example, this includes rent and food (including what you buy using CalFresh/SNAP and other benefits).

What is your best guess of how much you will spend **in total** in **September**?

$

→

(a) Panel A: Aggregated Expenses



For each of the following, please think through the following categories and then tell us your best guess of how much you will need to spend in **September** on:

Please fill out every row (unless it says *optional). If it doesn't apply, put zero "0".

$ Rent

$ Housing - Electricity

$ Household and personal items - Clothing

$ Household and personal items - School supplies

$ Social (e.g., people coming over to house, obligations, birthdays, get-togethers)

$ Emergencies: how much should you set aside for emergencies (e.g. car repairs, medical expenses, house repairs)

(b) Panel B: Category-By-Category Expenses

Notes: Screenshots of the Prolific survey used to elicit people's prediction of next month's total expenditures in aggregate (Panel A) and category-by-category (Panel B).

A.2 Appendix Tables

Table A.1: Balance

Variable	Mechanism experiment			Field experiment		
	Mean Control (1)	Mean Treatment (2)	P-value (3)	Mean Control (4)	Mean Treatment (5)	P-value (6)
Demographics						
Male	0.7	0.8	0.10	0.8	0.8	0.10
	(0.5)	(0.4)		(0.4)	(0.4)	
Age	44.8	45.0	0.92	43.6	43.5	0.96
	(12.5)	(11.8)		(13.9)	(13.6)	
Married	0.7	0.8	0.06*	0.8	0.8	0.30
	(0.4)	(0.4)		(0.4)	(0.4)	
Household Size	5.9	6.4	0.14	6.1	6.3	0.33
	(2.2)	(2.5)		(2.5)	(2.3)	
Bags of Maize	15.0	16.7	0.18	15.3	15.2	0.81
	(7.7)	(10.0)		(9.3)	(8.4)	
Savings (kwacha)	1298.0	1671.5	0.32	723.1	764.9	0.69
	(2693.5)	(2593.4)		(1661.1)	(1395.7)	
Farm Acres	4.3	3.9	0.21	4.3	4.2	0.40
	(1.9)	(1.7)		(2.1)	(2.1)	
Meals Yesterday				2.1	2.1	0.43
				(0.4)	(0.4)	
Hired Ganyu				0.5	0.4	0.44
				(0.5)	(0.5)	
Number of People Sold Ganyu				1.7	1.6	0.38
				(1.6)	(1.7)	
Person-days Sold Ganyu				7.4	8.5	0.21
				(10.6)	(15.5)	
F-test of joint significance			0.27			0.34
N	101	96		403	434	

Notes: Baseline participant characteristics for the treatment (columns 1 and 4) and control (columns 2 and 5) groups. Columns 1-3 show balance for the mechanism experiment, and columns 4-6 show balance for the field experiment. Columns 1, 2, 4 and 5 show the standard deviation in parentheses. Columns 3 and 6 shows the p-value of a t-test of equality of means in the control and treatment groups. Ganyu refers to casual labor. F is the p-value from a test of the joint significance of all covariates. N shows the number of observations for each group.

Table A.2: Attrition (field experiment)

	Non-Missing Control (1)	Non-Missing Treatment (2)	P-value (3)
Round 2	0.99	0.98	0.43
Round 3	0.99	0.99	0.93
Round 4	0.98	0.98	0.89
Round 5	0.98	0.97	0.49

Notes: Columns 1 and 2 show the proportion of participants at the baseline survey that are still present in the sample at the next visits, according to the treatment group. Column 3 shows the p-value of a t-test of equality of proportions across treatment groups.

Table A.3: Allocation of bags to non-food expenses according to the exercise (mechanism experiment)

	Number of Bags		Share of Total Bags	
	(1)	(2)	(3)	(4)
Control (2 Categories)	-0.35	0.22	0.01	0.01
	(0.49)	(0.23)	(0.01)	(0.01)
Treat	2.64***	2.04***	0.14***	0.13***
	(0.53)	(0.27)	(0.02)	(0.02)
Control (6 Categories)	1.56***	2.13***	0.13***	0.14***
	(0.55)	(0.26)	(0.02)	(0.02)
N	495	495	495	495
Prior	5.50	5.50	0.31	0.31
P-value Control=Treatment	0.00	0.00	0.00	0.00
Baseline Controls	No	Yes	No	Yes

Notes: The dependent variable is either the number of bags allocated to non-food expenses (columns 1-2) or the share of remaining maize bags allocated to non-food expenses (columns 3-4). Participants in the control group are asked to allocate maize to non-food expenses 3 times: when stating their prior, after the 2 category retrieval exercise and after the 6 category retrieval exercise. Participants in the treatment group are asked to allocate maize to non-food expenses 2 times: when stating their prior and after the 6 category retrieval exercise. The reference (omitted) category is "Prior" so each coefficient can be interpreted relative to prior estimates of non-food expenses. Baseline controls include: quantity of maize remaining and level of savings. Standard errors are clustered at the household level.

Table A.4: Willingness to exchange maize for consumption goods

	OLS (1)	OLS (2)	OLS (3)	Tobit (4)
Panel A: Mechanism Experiment				
Treat	-1.81***	-1.81***	-1.84***	-1.88***
	(0.32)	(0.33)	(0.33)	(0.34)
Radio			-0.43	
			(0.38)	
Solar Panel			-0.80**	
			(0.37)	
N	197	197	197	197
Control Mean	4.94	4.94		4.94
Control Mean - Chitenge			5.50	
Baseline Controls	No	Yes	Yes	Yes
Week FE	No	Yes	Yes	No
Panel B: Field Experiment				
Treat	-1.65***	-1.65***	-1.64***	-1.68***
	(0.18)	(0.18)	(0.18)	(0.19)
Radio			-0.31	
			(0.22)	
Solar Panel			-0.04	
			(0.25)	
N	837	837	827	837
Control Mean	4.81	4.81		4.81
Control Mean - Chitenge			4.80	
Baseline controls	No	Yes	Yes	Yes
Week FE	No	Yes	Yes	No

Notes: Impact of the retrieval exercise on the willingness to exchange maize for discretionary consumption items, after the treatment was administered. Panel A shows results in the mechanism experiment; Panel B shows the results in the field experiment. The dependent variable is the valuation of the item by the participant, in terms of gallons of maize, elicited using the Becker-DeGroot-Marshack method. Participants chose one of three items prior to the elicitation: a radio, solar panel or chitenge (cloth wrap). Columns 1-3 are estimated using OLS; Column 4 is estimated using a tobit model. Columns 2-4 include baseline controls. Column 3 includes item fixed effects; the citenge is the omitted item. Robust standard errors in parentheses.

Table A.5: Allocation of bags to non-food expenditures according to the category of expense and the timing of the question (control group, mechanism experiment)

	Number of Bags (1)	Share of Total Bags (2)	Number of Items (3)
After 6-Categories Exercise	0.35**	0.03***	0.40***
	(0.14)	(0.01)	(0.07)
Farming Inputs	2.53***	0.15***	0.63***
	(0.55)	(0.02)	(0.10)
Household Goods	0.68***	0.06***	0.38***
	(0.18)	(0.01)	(0.07)
Transfers	-0.44**	-0.02**	-0.18***
	(0.18)	(0.01)	(0.06)
Emergencies	-0.38**	-0.02*	-0.09
	(0.18)	(0.01)	(0.07)
After 6-Categories Exercise × Farming Inputs	-0.85***	-0.06***	-0.08
	(0.20)	(0.01)	(0.09)
After 6-Categories Exercise × Household Goods	-0.08	-0.01	0.10
	(0.22)	(0.01)	(0.10)
After 6-Categories Exercise × Transfers	0.40**	0.02**	0.25***
	(0.20)	(0.01)	(0.08)
After 6-Categories Exercise × Emergencies	0.50***	0.03***	0.35***
	(0.19)	(0.01)	(0.10)
N	1005	1005	1005
School (before 6-categories exercise)	0.56	0.03	0.25
Baseline Controls	Yes	Yes	Yes

Notes: The dependent variable is either the number of bags allocated to non-food expenses (column 1) or the share of remaining maize bags allocated to non-food expenses (column 2) or the number of items listed by the respondent (column 3). All specifications include baseline controls. The reference category is "School". Standard errors are clustered at the household level.

Table A.6: Savings results - Robustness checks (field experiment)

	Cash & Maize (1)	Cash & Maize (2)	Cash & Maize (3)	Cash & Maize (4)	Cash & Maize (5)	Cash & Maize (6)
Treat x Visit 2 (Pre-Labels)	108.89***	98.01***	95.65**	99.86***	96.12**	90.43**
	(40.92)	(36.22)	(38.52)	(37.62)	(38.49)	(38.42)
Treat x Visit 3 (Early Hungry)	77.41***	70.04***	76.23***	67.21***	66.84***	67.62***
	(26.28)	(24.07)	(25.31)	(23.71)	(23.88)	(23.97)
Treat x Visit 4 (Hungry)	18.71	15.87	19.17	12.13	13.51	15.97
	(20.97)	(18.56)	(20.82)	(18.54)	(18.95)	(18.63)
Dependent Variable unit	Kg	Kg	Kg	Kg	Kg	Kg
N	2480	2480	2474	2480	2480	2480
Control Mean Visit 2	681.69	650.95	775.49	660.51	660.51	660.51
Control Mean Visit 3	354.96	334.93	495.17	335.83	335.83	335.83
Control Mean Visit 4	173.32	156.46	264.73	156.72	156.72	156.72
Maize price	Lowest	Highest	Current	Current	Current	Current
F-test 2v3	0.36	0.35	0.54	0.30	0.36	0.49
F-test 3v4	0.01	0.01	0.01	0.01	0.01	0.01
Baseline Controls	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	No
Month FE	No	No	No	No	No	Yes
Camp FE	No	No	No	No	Yes	No
Round FE	No	No	No	No	No	Yes

Notes: Robustness of the impact of the retrieval exercise on savings to alternative measurement and specifications. The dependent variable in each column is the number of kilograms of unprocessed maize in storage (except in column 3 where the amount of processed maize in storage is added), plus the value of cash in savings, converted into maize equivalents using market prices. Columns 3-6 convert maize using the prevailing price on the day of the survey visit, columns 1 uses the average monthly price in September 2019 (the lowest average monthly price witnessed during our sample period) and column 2 uses the average monthly price in February 2020 (the highest average monthly price witnessed during our sample period). Column 3 adds the amount of processed maize (meal) that the participant had in storage. This is converted into unprocessed maize at the average processing rate observed in the area (approximately 70%). Column 4 controls for the baseline value of the dependent variable. Column 5 controls for agricultural camp (small geographic unit) fixed effects. Column 6 drops week fixed effects, and instead controls for survey round and survey month fixed effects. All specifications include baseline controls. Standard errors are clustered at the household level.

Table A.7: Impact of retrieval exercise on assets (field experiment)

	Asset Purchase Value (1)	Asset Sale Value (2)	Asset Net Value (3)	Livestock Purchase Value (4)	Livestock Sale Value (5)	Livestock Net Value (6)
Treat	-18.34	0.30	-18.64	7.14	-82.25	89.40
	(120.46)	(2.53)	(120.49)	(80.17)	(62.78)	(97.48)
Dependent Variable unit	Kwacha	Kwacha	Kwacha	Kwacha	Kwacha	Kwacha
N	823	823	823	823	823	823
Control Mean	593.17	1.89	591.28	401.23	326.08	75.15
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Impact of the retrieval exercise on purchases and sales of assets and livestock. Column 1 shows effects on the value of household assets purchased. Column 2 shows effects on the value of assets sold. Column 3 is net purchases (column 1 - column 2). Column 4 shows the value of livestock purchased. Column 5 shows the value of livestock sold. Column 6 shows net purchases (column 4 - column 5). The recall period is the past agricultural season. All specifications include baseline controls. Standard errors are clustered at the household level.

Table A.8: Impact of incentives and retrieval exercise on recall of past savings (field experiment)

	Recall Error (1)	Recall Error (2)	Recall Error (3)
High Incentive	-0.18		
	(0.37)		
Treat		-0.00	
		(0.35)	
Treat x High Inc			-0.18
			(0.52)
Treat x Low Inc			0.30
			(0.50)
Control x High Inc			0.14
			(0.48)
Dependent Variable unit	Bags	Bags	Bags
N	810	810	810
Control Mean	3.09	2.95	2.95
Baseline controls	Yes	Yes	Yes
Week FE	Yes	Yes	Yes

Notes: Impact of incentives and the retrieval exercise on recalled forecast errors in savings. The dependent variable in each column is the recall error: the number of bags of maize that the participant recalled having in a past survey round minus the actual savings. The recall of past savings was elicited in September/October 2020, 8-10 months after savings were measured. Responses were incentivized: Participants were asked five questions (including this one) and were given a payout if they answered one correctly. The incentive was either one bag of sugar (low incentive, monetary value 10-15 Zambian kwacha) or two bags of sugar (high incentive, monetary value 20-30 Zambian kwacha). All specifications include baseline controls. Standard errors are clustered at the household level.

B Appendix B: Model Proofs and Simple Example

B.1 Proofs of Full Model

Definitions

The description of the model in the main text does not formally define some important objects related to uncertainty. As noted in the main text, expense i in time t arises with probability π_{it}. Define the binary random variable that determines whether this expense arises as $\boldsymbol{\theta}_{it}$, which has a realization in $\{0, 1\}$. Define $\boldsymbol{\Theta}$ as the vector of random variables that collects all of these binary variables, with generic realization Θ (the vector is length $3 \cdot N$). For example, $Pr(\Theta = [1, 1, ..., 1])$ is $\pi_{1,1} \cdot \pi_{2,1} \cdot .... \cdot \pi_{N,3}$. We will at times abuse notation and also use $\boldsymbol{\Theta}$ to represent the set of all possible realizations of Θ. Define the information about expenses known in period t as $\mathcal{I}_t$, such that for example $Pr(\Theta|\mathcal{I}_1)$ represents the probability of realization Θ given knowledge of period 1 expenses. Define the subjective probabilities of a person who has retrieval failures as $\hat{\pi}$ and $\hat{Pr}$. Finally, as we are not concerned with prices, we normalize the units of consumption and expenses such that the price of all units is 1.

Maximization Problem

A person making a decision at time 1 observes the realizations of whether each expense arises in time 1, makes a decision about consumption and these expenses in time 1, and makes a state-contingent plan about future consumption and expenses given the uncertainty about future expenses. The full rational person must balance their budget for each future expense contingency:

$$\max_{c_t(\Theta), e_{it}(\Theta)} \sum_{\Theta \in \boldsymbol{\Theta}} Pr(\Theta|\mathcal{I}_1) \Big(\sum_{t=1}^{3} u(c_t(\Theta)) + \sum_{t=1}^{3} \sum_{i=1}^{N} \mathbb{1}(\theta_{it} = 1) \cdot v_i(e_{it}(\Theta)) \Big)$$
$$\text{s.t.} \quad \sum_{t=1}^{3} c_t(\Theta) + \sum_{t=1}^{3} \sum_{i=1}^{N} e_{it}(\Theta) = Y \text{ for all } \Theta \in \boldsymbol{\Theta}$$

A mistaken person who fails to retrieve some expenses solves the same problem, but uses $\hat{\pi}_{it}$ to construct $\hat{Pr}(\Theta)$:

$$\max_{c_t(\Theta), e_{it}(\Theta)} \sum_{\Theta \in \boldsymbol{\Theta}} \hat{Pr}(\Theta|\mathcal{I}_1) \Big(\sum_{t=1}^{3} u(c_t(\Theta)) + \sum_{t=1}^{3} \sum_{i=1}^{N} \mathbb{1}(\theta_{it} = 1) \cdot v_i(e_{it}(\Theta)) \Big)$$
$$\text{s.t.} \quad \sum_{t=1}^{3} c_t(\Theta) + \sum_{t=1}^{3} \sum_{i=1}^{N} e_{it}(\Theta) = Y \text{ for all } \Theta \in \boldsymbol{\Theta}$$

If a person mistakenly perceives a given $\hat{\pi}_{it} = 0$, the person then mistakenly places zero weight on all expense realizations Θ in which $\boldsymbol{\theta_{i,1}} = 1$, such that $\hat{Pr}(\Theta) = 0$ for that realization. Importantly, when the time period arrives, the mistaken person uses the actual expense realizations.

Proof: Solution

We solve the problem using backward induction. First, consider the period-3 problem. The person enters the period with savings s_3 given previous expenditures. The person observes the realizations of expenses and spends money on consumption and expenses. Given that expenses that do not arise do not enter the person's utility function, she must spend zero money on them. Therefore, the person must divide the savings among consumption and the realized expenses in Θ. This is a standard maximization problem in which the Euler equations must hold (recall that the price for each unit is normalized at 1):

$$u'(c_3) = v_i'(e_{i,3}) \equiv MU_3(\Theta, s_3) \text{ for realized expenses in period 3} \tag{2}$$

This equation must be satisfied for all expenses in Θ (i.e. there are no corner solutions) from the assumption that $v_i'(e_{it}) \to \infty$ as $e_{it} \to 0$. We then define the marginal utility in this solution for a given realization of expenses Θ and savings s_3 as $MU_3(\Theta, s_3)$. Note that this function must decrease in s_3 as all additional savings must be spent on consumption and expenses and, given these functions are concave, the marginal utility must drop as spending rises. Next, we show this marginal utility must be rising when a expense realization contains additional expenses. To do this, we first define an ordering: $\Theta_1 >_{(3)} \Theta_2$ if all of the period 3 elements of Θ_1 are weakly higher than those in Θ_2 (with one strictly higher). That is $\Theta_1 >_{(3)} \Theta_2$ if the realization Θ_1 contains additional period-3 expenses than the realization Θ_2. Recall that the person must spend no money on non-realized expenses and at least some money on all realized expenses. Therefore, in comparison to the spending in Θ_2, the person must spend more in Θ_1 on the additional expenses and less on the expenses only realized in Θ_2. Given the concavity of the expense functions, the marginal utility from these expenses must rise, and therefore $MU_3(\Theta, s_3)$ must be rising in Θ.

Now, consider the period 2 problem. In this period, the person arrives with s_2, observes the period-2 realized expenses, and chooses period-2 consumption and expenses (thereby saving s_3). The standard Euler equations then imply that the person's marginal utility from period-2 choices must equal the expected period-3 marginal utility (with the expectation taken over realizations of period-3 expenses):

$$u'(c_2) = v_i'(e_{i,2}) \equiv MU_2(\Theta, s_2) = E_{\Theta|\mathcal{I}_2}[MU_3(\Theta, s_3)] \text{ for realized expenses in period 2} \tag{3}$$

Just as in period 3, we define $MU_2(\Theta, s_2)$ as the marginal utility in period 2 from Equation 3. Just as in period 2, this function must rise in s_2. And, given the analogous definition of $\Theta_1 >_{(2)} \Theta_2$, it must be that $MU_2(\Theta_1, s_2) > MU_2(\Theta_2, s_2)$ if $\Theta_1 >_{(2)} \Theta_2$.

Now, we consider the impact of a mistaken belief that some $\hat{\pi}_{i,3} = 0$ even though $\pi_{i,3} > 0$. This shift causes the probability that the person places on any realization Θ_1 where expense i arises to shift to the comparable realization Θ_2 where expense i does not arise (and all other expense realizations are the same). Note that, given the definition above, $\Theta_1 >_{(3)} \Theta_2$. Given that $MU_3(\Theta, s_3)$ is rising in Θ, it must then be that (holding s_3 fixed):

$$E_{\Theta|\mathcal{I}_2}[MU_3(\Theta, s_3)] > \hat{E}_{\Theta|\mathcal{I}_2}[MU_3(\Theta, s_3)]$$

where $\hat{E}$ represents the expectations given the mistaken beliefs. But, then, the marginal utility from period-2 choices for consumption and expenses in Equation 3 are also larger than $\hat{E}_{\Theta|\mathcal{I}_2}[MU(\Theta, s_3)]$. Therefore, these choices cannot satisfy the Euler equation and therefore cannot be not optimal for the mistaken person. It cannot be optimal for the mistaken person to save more because (1) period-2 marginal utilities must rise given a decrease in period-2 spending, and (2) $\hat{E}_{\Theta|\mathcal{I}_2}[MU(\Theta, s_3)]$ must fall given that $MU(\Theta, s_3)$ is falling in s_2 for all Θ, such that the Euler equation cannot be satisfied. Therefore, relative to a rational person, the mistaken person's savings must fall, such which implies that period-2 consumption and expenses rise, and period-3 consumption and expenses fall.

The same argument holds when considering mistaken beliefs on more than one period-3 expense. Consider comparing a second mistaken person who holds the same beliefs as the original mistaken person, except that she places zero weight on an additional expense. In this case, using the same logic as above, the second mistaken person must have higher period-2 spending, lower savings, and lower period-3 spending in comparison to the first person. The same must then be true for a third person who is mistaken about an additional expense, and so on. Given this chain argument, any mistaken person must have higher period-2 expenses and lower savings in comparison to the rational person.

Therefore, considering the problem in period 1 leads to the same analysis and conclusion given the period-1 Euler equation:

$$u'(c_1) = v_i'(e_{i,1}) \equiv MU_1(\Theta) = E_{\Theta|\mathcal{I}_1}[MU_2(\Theta, s_2)] = E_{\Theta|\mathcal{I}_1}[MU_3(\Theta, s_3)] \text{ for realized expenses in period 1} \tag{4}$$

where $MU_1(\Theta, s_1)$ is defined as above. Given the same arguments as above, it must then be that, relative to a rational person, a person who is mistaken about some period-2 expenses must have a lower $MU_1(\Theta)$, higher period-1 consumption and therefore lower savings s_2. Then, as shown above, a person is mistaken about period-3 expenses will have a lower s_3 than a rational person with the same savings s_3. Therefore, the mistaken person must have lower savings s_3 relative to the rational person. Therefore, they must have lower period-3 spending.[44]

To understand the impact of mistaken beliefs on predictions, consider the impact of a mistaken belief that some $\hat{\pi}_{i,2} = 0$ even though $\pi_{i,2} > 0$. As discussed above, the mistaken person's beliefs about period-2 expenses causes her to shift probability from any realization Θ_1 where period-2 expense i arises to the comparable realization Θ_2 where expense i does not arise (and all other expense realizations are the same). That is, she places zero weight on realization Θ_1 and instead believes that Θ_2 – and the solution consistent with $MU(\Theta_2, s_2)$ – would occur instead. When Θ_1 does occur, she is surprised and instead enacts the solution consistent with $MU(\Theta_1, s_2)$. As $\Theta_1 >_{(2)} \Theta_2$, then $MU(\Theta_1, s_2) > MU(\Theta_2, s_2)$. Therefore, the mistaken person expected higher period-2 consumption and higher period-3 savings for the realizations in which neglected expense i is realized. In period 3, the person enters with lower savings than expected and therefore would spend less on period-3 consumption than expected given no mistaken beliefs about period-3 expenses. Using the same argument as above, mistaken beliefs will cause period-3 consumption to drop even lower. Therefore, period-3 consumption must also be lower than expected. Finally, given that person entered period 2 with savings s_2 and spends less than expected on period-2 and period-3 food consumption, the person must spend more on combined period-2 and period-3 non-food expenses than expected (as all s_2 must be spent on either food consumption or non-food expenses).[45]

[44] The effect of mistaken beliefs on period-2 consumption and expenses is ambiguous given the potential for mistaken beliefs about period-3 expenses. If the mistaken person had no mistaken beliefs about period-3 expenses, lower savings s_2 must lead them to spend less in period 2 in comparison to the rational person. However, (as shown above), mistaken beliefs about period-3 expenses can lead to higher period-2 spending relative to the rational person.

[45] The direction on period-3 non-food expenses is ambiguous. An unanticipated period-2 expense can cause savings to drop

Finally, the impact of the intervention moves the person toward the rational benchmark. As discussed above, all results about the relative comparisons of a mistaken person to a rational person are also applicable to a relative comparison between a person who is mistaken about a set of expenses to a person who is mistaken about a subset of those expenses (i.e. a "less mistaken" person). Therefore, if the intervention only causes people to reduce the set of mistaken expenses, the comparisons above will be applicable. For example, just as a rational person's prediction of total non-food expenses is higher than the mistaken person, a "less-mistaken" person will predict higher non-food expenses than a "more-mistaken" person.

Formally-Stated Predictions

The conclusions from the solution above are collected in the following statements:

Formal Initial Prediction 1. *For every realization of period-1 expenses, the mistaken person will have higher period-1 spending on consumption and expenses than the rational person. Therefore, the mistaken person will have higher average period-1 spending across all realizations. For every realization of expenses, the mistaken person will have lower period-3 spending on consumption and expenses than the rational person. Therefore, the mistaken person will have lower average spending across all realizations.*

Formal Initial Prediction 2. *For all expense realizations in which a neglected period-2 expense arises, the mistaken person's period-1 state-contingent expectation of savings entering period-3* s_3 *will be higher than realized savings; of period-2 and period-3 consumption will be higher than realized consumption; and of combined period-2 and period-3 non-food expenses will be lower than realized expenses. Therefore, these relative statements will be true on average across all realizations.*

Formal Intervention Prediction 1. *A less-mistaken person will have a higher period-1 expectation of period-2 and period-3 non-food expenses.*

Formal Intervention Prediction 2. *A less-mistaken person will have a higher period-1 marginal utility of consumption* $MU(\Theta)$*.*

Formal Intervention Prediction 3. *For all expense realizations, a less-mistaken person will have a lower period-1 spending on consumption and expenses and higher period-3 spending on consumption and expenses. Therefore, these statements will be true on average across expenses. A less-mistaken person will have a higher period-3 savings in all expense realizations that the two people had different predictions about. Therefore, average period-3 savings will be higher across expenses realizations.*

B.2 Simple Model

In this section, we work through a simplified example of our model with no stochasticity and limited expenses to help with intuition. The example is intentionally stark. Harvest income is 1 ($Y = 1$) and expenses all share the same utility function as consumption ($v_i(\cdot) = u(\cdot)$). We assume that a person has two expenses in addition to food (school fees and herbicide) which are only paid in period 2. That is,

relative to period-1 expectations to the point that both consumption and non-food expenses drop in period 3 relative to period-1 expectations.

$\pi_{1,2} = \pi_{2,2} = 1$ and $\pi_{1,1} = \pi_{2,1} = \pi_{1,3} = \pi_{2,3} = 0$. Therefore, a person with perfect retrieval is maximizing five components $\sum_i^5 u(e_i)$ under the constraint that $\sum_i^5 e_i = 1$.

Given that the prices of expenses in the example are assumed equal (at 1) and they share the same (concave) utility functions, the Euler equations demand that the same amount is spent on all expenses. Given the budget constraint, this implies that $e_i = \frac{1}{5}$ for all i. The person at time $t = 1$ fixes the time-1 choice e_1 from this plan and enters time $t = 2$ with savings $s_2 = \frac{4}{5}$. At this point, the person will choose the same plan (she is time-consistent), such that $\frac{1}{5}$ will be spent on all expenses. Note that the person smooths the spending over time for expenses with the same utility function. Also note that the person has rational expectations about the future.

Consider instead a person with retrieval failures who does not account for having to buy herbicide when initially making their plan (that is, $\hat{\pi}_{2,2} = 0$). This person at time $t = 1$ only perceives four expenses and therefore plans on spending $\frac{1}{4}$ on each of these expenses. The person then fixes $c_1 = \frac{1}{4}$ and enters time $t = 2$ with $s_2 = \frac{3}{4}$. At this point, the person realizes that they have neglected to consider herbicide. Taking herbicide into account, they smooth their savings over the four remaining components ($c_2 = e_{1,2} = e_{2,2} = c_3 = \frac{3}{16} < \frac{1}{5}$), fix their time-2 choices, and enact this plan when they arrive in period $t = 3$.

This matches our Initial Predictions 1 and 2. Even though the person does not discount the future and the utility from consumption is constant, they do not smooth spending over food. Furthermore, at time $t = 1$, they would predict (incorrectly) that they will enter period $t = 3$ with savings $s_3 = \frac{1}{5}$ (rather than $s_3 = \frac{3}{16}$).

Our Assumption 2 in this example is again that the intervention (treatment) corrects the retrieval failure, i.e. the treated person recalls herbicide at time $t = 1$ while the control person continues to neglect it. Given this assumption, the treated person will predict that they will spend more on non-food expenses than the control (from $\frac{1}{5}$ to $\frac{3}{8}$). This behavior fits Intervention Prediction 1.

For Intervention Prediction 2, consider the person's marginal utility per dollar from their plan. The treated person recalls all five components and therefore sets $c_1 = \frac{1}{5}$, receiving $u'(\frac{1}{5})$ of marginal utility per dollar (as each good costs $p = 1$) in their initial plan. Meanwhile, the treated person sets $c_1 = \frac{1}{4}$ and therefore receives $u'(\frac{1}{4})$ of marginal utility per dollar in their plan. As $u'(\frac{1}{5}) > u'(\frac{1}{4})$, the treated person has a higher marginal utility per dollar (i.e. a higher shadow price of money).

Finally, the treatment pushes the person to the fully rational solution, so we have already calculated the trajectories. The treated person's initial expenditures are lower ($c_1^{Treat} = \frac{1}{5} < \frac{1}{4} = c_1^{Cont}$), but are higher later ($c_2^{Treat} = e_{1,2}^{Treat} = e_{2,2}^{Treat} = c_3^{Treat} = \frac{1}{5} > \frac{3}{16} = c_2^{Cont} = e_{1,2}^{Cont} = e_{2,2}^{Cont} = c_3^{Cont}$), such that the savings path is higher ($(s_2^{Treat}, s_3^{Treat}) = (\frac{4}{5}, \frac{1}{5}) > (s_2^{Cont}, s_3^{Cont}) = (\frac{3}{4}, \frac{3}{16})$).

C Protocols

C.1 Field experiment

Sample Construction In each of the study districts, to construct a sample of villages, we obtained an agricultural census for the district. To improve accuracy, we supplemented with our own census of villages in several blocks within each district.

Surveyors conducted an initial screening of villages between June and July 2019, interviewing the headman and one additional member of the village. Villages were selected into the study if a large proportion of residents derived their income primarily from farming maize, and stored their maize in bags after harvest. This effectively ruled out, for example, villages that were close to towns, so that many residents obtain income from non-agricultural sources. Of 171 villages we visited, 118 were deemed suitable for the study. We randomly ordered these 118 villages, and conducted our study in the first 113 of these villages to arrive at our desired sample size.

To construct a sample of households, we sampled up to 14 households per village using the following protocol. Two households in the village were selected randomly from the village registry. After these households completed the baseline survey, surveyors followed a "left-hand rule" to approach the next household in the village. Under the left-hand rule, surveyors faced in different directions from their original household, and moved leftward, skipping at least one household before approaching the next household to survey.

Before conducting the baseline survey, the surveyor conducted a screening survey to determine whether the household was eligible to participate in the study. We designed our screening criteria to ensure that households i) were smallholder farmers, ii) stored their maize in bags, iii) reported prior food shortages, so that increased savings might have meaningful consequences, iv) used maize to pay for expenses and did not have alternate means of smoothing consumption, v) had sufficient maize to make planning worthwhile, and vi) were not polygamous.

Retrieval exercise The retrieval intervention was embedded into the baseline survey, which was the only interaction with participants in the mechanism experiment and the first visit to participants in the field experiment. We implement the treatment in September to early October of 2019. At this time, most households have harvested their maize and completed maize shelling (i.e. removing kernels of corn from the cob) in order to prepare it for storage in bags.

Visit 1: Baseline survey During the baseline visit, an adult respondent at the household was asked if they were the household head. If they stated yes, then a screening questionnaire was completed (and if no, then an alternative time to visit the household when the household head would be home was arranged). If the participant was found to be eligible during the screening survey, then the household head completed a baseline survey. All surveys were conducted with the head of the household only, away from other members of the household or other individuals from the village.

The baseline survey included information about baseline savings (e.g. maize) and other demographic variables. Respondents were then asked for their baseline forecast of future expenditures (giving a measure of baseline beliefs as an input to test Prediction 1). The treatment group then undertook the budget exercise. All individuals in both treatment and control groups were then offered the labels. All individuals concluded the baseline survey by undertaking an exercise where we elicited their willingness to pay for a discretionary good (to test Prediction 2). For the treatment group, the budget intervention took about 20-45 minutes, and the entire baseline survey took about 1 hour.

Visit 1: Randomization After the baseline survey questions were completed, participants were randomized using Survey CTO into a treatment or control group, that determined whether they would conduct the retrieval exercise, or not. Importantly, neither the surveyor nor the household knew their treatment status until it was revealed midway through the baseline survey (when it was time to do the budget exercise for treatment households).

We randomized at the household (rather than village) level in order to improve statistical power. However, this design choice generates some scope for spillovers between households—for example, control households may learn about budgeting from treatment households, or may pressure them to share their extra savings during the hungry season. Note that such spillovers would only dampen our measured treatment effects. We mitigated the potential for such spillovers by enrolling no more than 14 households per village in the study, so that in expectation, no more than seven households per village were treated.

Visit 1: Labels introduction After completing the baseline survey, the following information was told to all households, regardless of their treatment status. The household head was told that they would have the option to choose a set of maize labels, or receive a bag of sugar. They were told that the labels had pictures on them, corresponding to common expenditure categories. Each of the label expenditure categories was then explained to the participant. The participant was then shown that there were also labels that corresponded to each month of the year. They were then told that if they wanted, they could attach these labels to their maize bags, to mark what they thought they would spend on each category. Finally, they were told that regardless of their choice, we would return to conduct more surveys with them in a month's time.

Visit 1: Expense board treatment If a participant was randomized into the treatment group, they then completed a retrieval activity, in which the goal was to consider how to allocate the respondent's maize stock to different expenditures over the course of the year.

This retrieval exercise consisted of three steps. First the household head was asked to think about the lean season. They were asked to recollect if there were difficulties at this time of year. They were asked why shortages happened, and if they happened in years where individuals have reasonable harvest sizes. Then, they were told that planning and budgeting was one way that households managed their maize, and could be used to think through and track how they wanted to use their expenditures.

The participant was then asked to think through how they had used their maize in the past year. They were given a sheet of paper that had pictures of consumption of maize in each month. Using preparatory fieldwork, we had identified seven major categories of spending that households in our setting typically engage in: maize consumption for food in each month, school fees, household supplies, farm inputs, transfers to others, health shocks and other emergencies, and a residual âotherâ category.

Participants were then asked how many bags of maize they harvested in the previous year. They were then handed thumb tacks, with one thumb tack representing one bag of maize. They were then asked to allocate how many bags of maize they had spent on a subset of these categories last year: consumption in each month, transfers, and emergencies. These sheets of paper with the thumb tacks in them were left in front of the respondent for the remainder of the planning activity.

After they had shown how they had used their maize in the past year on these expenditure categories, two additional sheets of paper were brought and placed in front of the participant. One sheet of paper had pictures representing five expense categories on them. These expense categories consisted of school fees, payments for farming inputs, payments for household goods, emergencies and transfers. A second sheet of paper showed consumption by month. However, this piece of paper was left face down on the board to begin with.

Households were then given thumb tacks representing the amount of maize that they had currently, with one thumb tack representing one bag of maize. They were then asked to think about how much of their maize they would allocate their maize to each of the pictured expenditure categories. In addition they were told the following. If they want to allocate a bag to any consumption category that they did not see on the board, they could put this on the outside edge of the paper, indicating a category for "other". They were asked to think about how much maize they wanted to allocate to each category. They were told they could start to place thumb tacks into the board if they wanted, but that they also could just think through the allocations in their head. They were also told that at this point the plan was not sticky, and that they could continue moving the thumb tacks as much as they wanted.

After the participant thought through how much maize they wanted to allocate for expenditures, the sheet of paper showing consumption by month was turned over and placed before the participant. The participant was then asked to also think through how much maize they wanted to allocate for consumption. They were also told that if they wanted to use one bag of maize over two months for consumption, they could put it in between two months on the board, indicating that that bag would be split between those two months. Again the household could begin allocating pins to this category, or could just think through how they wanted to allocate the maize in their head.

The participant was then asked to think again on their own about how they wanted to allocate their maize. And they were told that they could continue moving pins if they wanted to in the board. They were then asked to place pins onto the board showing how they planned to use their maize.

After, the participant was asked to repeat their plan back to the enumerator. They were asked to explain how much maize they would use for each category on the board.

Finally participants were asked after they had made their plan if they thought they might need to do additional labor to supplement their income. If they responded affirmatively, they were asked in which month they might want to begin doing this labor.

Throughout the retrieval exercise, enumerators did not provide suggestions or make normative statements about how participants should use their maize. They did not assist participants with doing math. In addition, after the survey was completed, the planning board and thumb tacks were removed from the participant.

Visit 1: Labels Choice All households then were told that they could either receive a set of labels of their choice, or a bag of sugar. They were told that they should choose whichever option they preferred. They were also told that surveyors would return in around one month's time with the labels to help them attach them to their bags (when most households in our sample would have finished shelling their maize). Households were also told that regardless of their choice, a surveyor would return to survey their household in around one month's time.

Visit 2: Labeling Protocol All participants regardless of treatment status and label choice completed a survey at the second visit. In addition, participants that chose the labels during the baseline visit also conducted labeling of their maize bags with a surveyor.

Before starting the labeling exercise, enumerators asked if the participant had shelled any of their maize. This was a prerequisite for doing the labeling, since maize is usually only stored in bags after it has been shelled. If the participant had not yet shelled their maize, the enumerator asked them at what date they thought they might have shelled, and told them that they would return to label the bags at this point.

Before starting the labeling, for participants that were in the treatment group the enumerator began by recapping with the household head the plan they had made previously. To do this, enumerators showed participants the expense board with pictures of different expenditure categories, and placed thumb tacks into the categories showing the allocation that the participant had made previously. The enumerator then asked participants whether they still had the same number of bags, and they still wanted to keep the same allocation. In most cases, the participant said no, and they were encouraged to redo the allocation, showing how they wanted to allocate the maize that they had remaining to different expenditure categories. After redoing the expense board, the surveyor showed the participant the labels that corresponded to their allocation, and then told the participant that they could attach these labels to their maize bags.

Enumerators then went with participants to the area where they kept their maize. They asked all participants if they would keep their maize here for the remainder of the year.

Enumerators then asked all participants which label to place on which bag of maize. Treatment participants could choose any label, but were reminded of the labels they had used as part of their allocation. Control participants could choose any of the labels. After labeling bags, participants were asked if they wanted our enumerators to stack the bags in any particular order.

After all bags were stacked, surveyors asked if there was additional, unshelled maize that the household planned to shell and store in bags. If the participant said yes then the enumerator asked on which date these bags would be shelled and stored in bags. Our enumerators then would conduct up to two additional visits to participants, to attach labels to these bags. At these follow up visits, enumerators would just ask participants (regardless of treatment status) which labels they wanted our enumerators to attach to their bags.

Visits 2-4: Savings measurement We obtained verified measures of participants' savings of maize in visits 2 through 4 as follows. During each survey, we asked the participant if they could take us to the place that they kept their maize, so that we could weigh up to three of their bags of maize. Our enumerators then accompanied the participant head to the area that they kept their bags of maize (usually either a room in the house or a shed). At this time, three bags of maize were selected and weighed using standing scales that our enumerators brought to interviews (fewer bags were weighed if the participant did not have three bags of maize). The participant was then asked how many bags of maize they had remaining. Enumerators verified this quantity. In addition, after this was completed, enumerators asked the participant if they could also weigh their mealie meal (a formed of processed maize).

Visit 5: Endline survey In October 2020, approximately one year after the intervention was conducted, we conduct a final round of household surveys. This includes collecting information on crop yields and revenues and additional farm investment measures. We also elicit households' willingness to pay to receive our treatment intervention for the upcoming agricultural year. Finally, to test for persistence, we collect data on expenditure forecasts for the coming year.

C.2 Mechanism experiment

The mechanism experiment was conducted in November-December 2022 on a different sample of households. The goal of this intervention was to get a better understanding of the mechanisms to rule out alternative explanations of the results.

The sample consists of 197 households, half of them randomly assigned to a treatment group and the other half to a control group. Participants were drawn from 28 villages, with up to 14 participants per village (using the sample selection criteria described in Section 4.2). For the mechanism experiment, due to logistical constraints, most villages were located near a town. Consequently, some of our screening criteria were relaxed: not all households reported food shortages during the hungry season, and households were more likely to have some small alternative sources of income.

The survey was conducted with the head of the household and started by asking some socio-demographic questions (including size of household, size of harvest, number of maize bags remaining, amount of savings). Next, the household head was asked some questions about the frequency of food shortages and the strategies used to deal with them (e.g. working in another farm). The household head was also told that planning and budgeting was one way that households managed their maize, and could be used to think through and track how they wanted to use their expenditures. Then, the respondent was asked how many bags of maize he/she expected to sell or use this year for expenses different from food. This gives a measure of the prior of household concerning future expenses. All these preliminary questions are common to the control and the treated groups.

After this, the intervention differed according to the treatment group:

- Control group (2 category budget board): The respondent was given a number of thumb tacks corresponding to the number of maize bags they had left from their harvest. Then, they were provided with a sheet of paper representing two categories of expenses: food and non-food. The food category was represented by a picture of a plate full of food, and the non-food category by the same picture but crossed in red. The respondent was then asked to allocate the thumb tacks to each of the categories according to its plan on future expenses.

- Treatment group (6 category budget board): The exercise was exactly the same as described in section C.1: the respondent was first provided with a sheet of paper representing five categories of non-food expenses (school, farming inputs, household goods, transfers and emergencies) and then with another sheet of paper representing food consumption per month. The respondent was asked to allocate the thumb tacks to each of the categories according to its plan on future expenses.

Just after the budget exercise, the respondent was asked his or her willingness to pay for a discretionary consumption item, following exactly the same process as in the field experiment Visit 1.

After the willingness to pay exercise, the respondent was asked to list all the items he/she had in mind for the non-food category or categories when choosing how many bags of maize to allocate.[46] For each item volunteered by the respondent, the surveyor asked the amount of the expense,[47] the time when the expense is expected to be realized,[48] and the expected frequency of the expense.[49] and the degree of certainty of the expense.[50] This was done both for the control and the treated groups.

Finally, at the very end of the survey, the control group was asked to do the treatment budget exercise (i.e. with all 5 categories of non-food expenses). This allows to measure within-household variation in the plan according to the number of categories of expenses presented to the respondent. After making the plan, the respondent was asked again to list all the items he/she had in mind during the exercise. For each item, the same questions about amount, time, frequency and degree of certainty are asked. Moreover, the respondent was asked whether he/she included the item during its first plan (i.e. during the 2 category budget exercise) or if it was forgotten.

[46] The exact question was phrased as follows: "When you did your plan previously, you told us that you needed X bags for non-consumption expenditures. Can you explain to us what the expenditures were here that you were thinking about? Please list all the expenditures that were in your mind at that moment.".

[47] The respondent could choose whether to answer in term of maize bags, maize meda (i.e. 1/12 bags) or local currency (kwacha).

[48] The options were: "this week", "later this month", each following months and "don't know".

[49] The options were: "once", "multiple" and "uncertain".

[50] The options were: "certain" and "uncertain".