NBER WORKING PAPER SERIES

RETRIEVAL FAILURES AND CONSUMPTION SMOOTHING: A FIELD EXPERIMENT ON SEASONAL POVERTY

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

Working Paper 35430
http://www.nber.org/papers/w35430

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2026

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, Stefania Elementi, Anne Fogarty, Alan Lee, David Lyu, Jingkai Ong and Richard Nikolai Schulz 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Retrieval Failures and Consumption Smoothing: A Field Experiment on Seasonal Poverty
Ned Augenblick, B. Kelsey Jack, Supreet Kaur, Felix Masiye, and Nicholas Swanson
NBER Working Paper No. 35430
July 2026
JEL No. D14, D15, O12, Q12

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 36-60%; as predicted by the memory literature, effects are concentrated among small, irregular, and stochastic items. Immediate spending drops and, six weeks 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.

Ned Augenblick
University of California, Berkeley
Haas School of Business
ned@haas.berkeley.edu

B. Kelsey Jack
University of California, Berkeley
Haas School of Business
and NBER
kelseyjack@berkeley.edu

Supreet Kaur
University of California, Berkeley
Department of Economics
and NBER
supreet@berkeley.edu

Felix Masiye
University of Zambia
Department of Economics
felix.masiye@unza.ac.zm

Nicholas Swanson
Cornell University
ngs53@cornell.edu

A online appendix is available at http://www.nber.org/data-appendix/w35430
A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/4409

1 Introduction

Economic models typically assume that people use all the knowledge they have when making decisions. This assumption is so foundational that it is rarely stated explicitly. However, due to the limits of human cognition, what a person knows may not always come to mind when making a decision. This paper empirically examines whether such *retrieval failures*—the failure to recall and use relevant information at the moment of choice—can distort decision-making and behavior, even when stakes are high.

We focus on a classic economic problem: deciding how much to spend today versus save for tomorrow. Solving this problem requires individuals to hold in their mind both their expected income as well as the full distribution of future expenses. We argue there is a fundamental asymmetry in how easily these inputs are retrieved. The sheer number of potential future expenses is large in number and often heterogeneous in nature, infrequent or irregular, and stochastic.[1] These are precisely the features that the memory literature identifies as prone to retrieval failure—implying that people are likely to underestimate their future expenses. Income, in contrast, tends to arise from a much smaller set of relatively consistent sources.[2]

We model how this asymmetry—that expenses are more prone to retrieval failures than is income—produces systematic overoptimism about future finances. This, in turn, causes individuals to overconsume today and then suffer shortfalls later, when future expenses are realized. Our theory microfounds information recall as an associative memory process (Kahana, 2012; Bordalo et al., 2023). This implies a simple debiasing intervention: thinking through one's expenditures category by category, which triggers more complete associative recall, will increase the number of retrieved expenses—an approach that has been robustly validated in the "planning fallacy" literature in psychology (Buehler et al., 2010). 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. Consequently, our model predicts that this *retrieval exercise* will increase individuals' forecast of their future expenses. This change in beliefs will then

[1]Future expenses encompass some that are large, regular, and predictable (such as rent or a vehicle payment). However, individuals must also bring to mind the sum of many expenses that are not only small but also heterogeneous (e.g., shampoo, socks, a dental visit, daily morning coffee), arise irregularly (e.g., extra gasoline costs during a future road trip or a birthday present for a friend), and are stochastic (e.g., the possibility of a roof leak or future root canal).

[2]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.

lead to: (i) lower spending today, (ii) weakly higher savings at all points in the future, and (iii) a flatter spending profile over time. In other words, simply helping individuals retrieve what they know from their own minds—without providing any external information—can lead to meaningful behavior change.

We test these ideas in the context of annual consumption cycles in Zambia. 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—setting up a stark and extremely consequential savings problem. Consumption cycles are pronounced: while 98% of households have ample food right after the harvest, over 50% report difficulty meeting basic needs at the end of the agricultural year (i.e., the three months before the next harvest)—a period known locally as the "hungry season" (Fink et al., 2020). Such cycles, where consumption fluctuates predictably and repeatedly with income flows, are ubiquitous in both poor and rich countries.[3] We hypothesize that retrieval failures contribute to consumption cycles in our setting.

To motivate this hypothesis, we document that farmers substantially under-estimate future expenses. On average, their annual expenditure on non-food items is twice as large as their initial forecast.[4] Note that these expenses are largely comprised of items such as school fees, farm inputs, and medical costs—which are unlikely to be temptation goods. The under-estimation of expenditures coincides with high levels of overoptimism about future savings. In incentivized forecasts, 80% of farmers think they will have higher savings 3 months later than they actually do, with the median participant overestimating future savings by 66%. When asked to predict how much savings they will have if "everything that could possibly go wrong does go wrong", a striking 65% of farmers end up with less savings than their worst-case scenario forecast.

To test the remaining predictions of our model, we design a retrieval exercise that prompts individuals to think through their expenses in categories. Specifically, we develop a simple "retrieval board" that is demarcated into seven broad categories of expenses: food consumption in each month and six categories of major non-food expenses

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

[4]In our setting, food consumption is generally the largest expense and occurs every day in the same form of standardized nshima (maize meal) patties, making it quite regular, predictable, and salient. We therefore focus specifically on eliciting predictions about non-food expenses, which are more irregular, stochastic, and less salient—and therefore less likely to be retrieved.

(school fees, household supplies, farm inputs, transfers, health shocks and emergencies, and other). Shortly after harvest, we ask farmers to allocate spending to each category of expenses for the remainder of the agricultural year.[5] Importantly, we do not provide any assistance, guidance, or normative advice on the allocation. Consequently, any changes in beliefs or behavior result from individuals' own cognitive effort.

We leverage the retrieval board to construct two complementary experiments. In the first experiment (the "mechanism experiment"), all participants allocate their available savings to expense categories. However, we vary the level of aggregation—and therefore usefulness—of the categories. The control group receives a "placebo" board, with only two categories: food expenses and non-food expenses. The treatment group receives the full retrieval board, as described above, which has six non-food expense categories rather than one. Our model predicts that the full board will prompt higher retrieval of non-food expenses, especially of items that are a priori more likely to be neglected (i.e. those that are smaller or less salient).

We randomize 197 farmers to receive either the placebo board or the full retrieval board. Relative to individuals' prior at baseline, those who receive the placebo board show no detectable change in beliefs about non-food expenses. This suggests that simply going through the motions of planning alone does not change beliefs. In contrast, consistent with our model, the full retrieval board has large impacts on beliefs: relative to the placebo board control, treated farmers expect to spend 36% more on non-food expenses (p<0.001). In line with memory models, the incremental expense items that are recalled under the full board but not the placebo board are smaller in magnitude, more irregular, and more stochastic. Moreover, the change in beliefs reduces treated individuals' willingness to spend on discretionary consumption. In a real-stakes exercise, treated individuals' willingness to pay for consumption goods—such as new clothing or a radio—falls by 36% relative to the control group (p<0.001).

Our model predicts effects on consumption smoothing behavior across the year—which we test in a second experiment (the "field experiment") with a new sample of 837 farmers. In the field experiment, like in the mechanism experiment, the treatment group receives the full retrieval board shortly after harvest, and is asked to allocate expenses to categories.

We adopt two core design changes relative to the mechanism experiment. First, the

[5]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.

control group is a "pure" control, which is not prompted to think about their expenses.[6] Second, under our hypothesis of imperfect recall, a treated farmer may not remember specific items recalled (or specific plans made) many months later. In contrast, our model predictions require that expenses, once retrieved through the exercise, are recalled in all future periods. To help farmers memorialize their expected expenses over the subsequent months of the year, we offer farmers expense labels (corresponding to the categories on the retrieval board). Farmers can affix these labels to their maize bags in order to visually depict their spending plan. However, importantly, the labels are only delivered over *six weeks* after the retrieval exercise; this enables us to test for the effect of the retrieval board alone over a substantial time horizon before labels are introduced. In addition, 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.[7]

The immediate impacts of the field experiment replicate the findings from the mechanism experiment. First, after undertaking the retrieval exercise, treated farmers increase their forecast of future non-food expenditures by more than 60% relative to their baseline forecast ($p < 0.001$). Second, treatment participants are, on average, willing to pay 34% less for discretionary consumption goods ($p < 0.001$).

Our theory predicts that the change in beliefs should lead to higher savings, leading to smoother consumption over the year. Consistent with this, six weeks after the retrieval exercise—*before* labels are delivered—treated farmers hold 15% more savings than do control farmers (p=0.029).[8] 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=0.015), an effect size roughly equal to one month of total household expenditure in the control group during the hungry season. The higher savings stock enables treated farmers to spend more in later months in the cycle, leading to a smoother spending profile over the year.

[6] Ex ante, it was unclear whether even a placebo board could have some modest treatment effect relative to no intervention at all. To maximize power, the control group in the field experiment sees no expense board at all.

[7] 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.

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

In developing countries, due to high self-employment, increased savings has 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 turn to casual wage labor—diverting labor away from their own farms at a time when its marginal product is quite high (Fink et al., 2020). Our intervention reduces the need for this behavior: treated households engage in 32% fewer days of wage labor during the hungry season (p=0.035). In addition, we see suggestive evidence that treated farmers use their increased savings stock to self-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=0.137). Consequently, relative to the control group, treatment farmers have 9% higher crop revenue in the subsequent harvest (p=0.076)—leading them to enter the following year with a substantively larger pie.

Together, these results provide consistent support for our model of retrieval failures. Moreover, our model's predictions and intervention design distinguish retrieval failures from alternative explanations such as present focus: under present focus alone, there should be no impact of thinking through categories on beliefs or behavior.[9] More generally, 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, 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. Specifically, in a survey of around 700 low income individuals, we collect prior estimates of upcoming monthly income and expenses. Next, participants undertake 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, consistent with the idea that expense items are more susceptible to retrieval failures than are income items.[10] These results

[9]Like neoclassical models, standard models of present focus such as quasi-hyperbolic discounting assume that individuals automatically use all the information available to them when making plans.

[10]In interpreting our findings, we are largely agnostic about whether they stem from a pure memory channel, motivated reasoning, or some other cognitive limitation that prevents retrieval. While only suggestive, these patterns are more consistent with a memory channel than one rooted in motivated

suggest that retrieval failures may generalize across populations with varied economic circumstances.

Our study advances the literature on how cognitive constraints alter decision-making. Existing work 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.

While a burgeoning body of theoretical work explores imperfect information retrieval (Mullainathan, 2002; Gabaix, 2019; Bordalo et al., 2020, 2023; Malmendier and Wachter, 2024),[11] direct field evidence has been limited. A small set of existing studies shows that directing people's attention to a specific action (e.g., taking a pill or saving) through text message reminders can immediately increase compliance with that action (Pop-Eleches et al., 2011; Karlan et al., 2016). Reminders may cause people to retrieve otherwise neglected information, but they may also operate through other channels. Our design offers a more direct test of retrieval failures. Rather than orienting people toward a specific action, our intervention simply prompts them to cognitively engage with the information that is inside their minds. We show that enabling associative recall—without any normative guidance—drastically improves retrieval and subsequent beliefs (via the mechanism experiment), and that this in turn has large impacts on high stakes behaviors over many weeks to months (via the field experiment).

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 subcategories increases forecasts (Buehler et al., 1994; Forsyth and Burt, 2008).[12] Existing work has applied these ideas to budgeting to correct overoptimistic beliefs about future

reasoning, since the latter would lead individuals to over- rather than under-estimate income.

[11]Our model adapts Bordalo et al. (2023) to a budgeting problem, and formalizes the benefits of categorization for retrieval. 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).

[12]Note that categorization may sometimes lower accuracy (e.g. Peetz et al., 2015). For example, it can lead to overweighting of rare events or those subject to interference (Bordalo et al., 2023).

expenses (Peetz and Buehler, 2009; Stilley et al., 2010; Sussman and Alter, 2012; Peetz et al., 2015; Berman et al., 2016; Howard et al., 2022). [13] We build on this work by demonstrating substantive changes in not only beliefs, but also high-stakes field behavior over significant time horizons among highly experienced agents.[14]

Relatedly, a line of research finds that making a plan to undertake a specific task—such as voting or getting vaccinated—increases task completion (Nickerson and Rogers, 2010; Milkman et al., 2011, 2013; Carrera et al., 2018; Abel et al., 2019). 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 retrieval board) that we see effects. Potentially consistent with our findings, this literature emphasizes the need for plans to be detailed and concrete. To the extent that detailed planning induces retrieval—for example, by recalling other time commitments or obstacles that may otherwise be neglected—retrieval failures may be one possible mechanism underlying effective planning interventions.

Finally, existing work in economics 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., Jakiela and Ozier, 2016; Dillon et al., 2021; Carranza et al., 2025). Relatedly, a large literature in development examines reasons for low savings rates among low-income households (for reviews, see Karlan and Morduch, 2010; Karlan et al., 2014; Kremer et al., 2019; Breza and Kaur, forthcoming). Much of this work examines behavioral mechanisms for savings behavior, such as self control problems and commitment (Ashraf et al., 2006; Dupas and Robinson, 2013; Brune et al., 2021), aspirations (Bernard et al., 2026), limited attention (Karlan et al., 2016), default effects (Blumenstock et al., 2018), and mental accounting (Soman and Cheema, 2011; Wicker et al., 2026). We augment these literatures by offering evidence for an additional channel: retrieval failures. Our study indicates that cognitive constraints—by distorting individuals' understanding of their

[13]Research on survey measurement has also documented that finer categories increase measured consumption and expenditures (Deaton et al., 1998), and that small and irregular items are more likely to be neglected (Bee et al., 2015).

[14]The planning fallacy literature is largely agnostic about the underlying mechanism for these empirical regularities—why the planning fallacy arises and why segmentation reduces over-optimism; our model and study design enable a direct link to imperfect memory—offering one (and not mutually exclusive) contributing mechanism.

financial situation—directly impact savings behavior and therefore expenditure profiles over time. We document that ameliorating retrieval failures can have sizable consequences for smoothing behavior, suggesting 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 consume this maize directly as food, and also use it to pay for expenses—either paying in-kind with maize directly, or first selling the maize for cash. Opportunities to borrow from the future harvest are limited, and borrowing is not common. 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 over the year. Food consumption is generally the largest, most regular, and most salient expense. Farmers eat their maize every day, as part of virtually every meal, in the form of standardized nshima (maize meal) patties. Consequently, in our preliminary interviews with farmers, food consumption was almost universally discussed first when considering spending.

In contrast, non-food expenses are more varied, irregular, and unpredictable than food expenses. Major expected non-food expenditures include 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.

This setting is an attractive one for studying consumption smoothing in general, and retrieval 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 Motivating Evidence: Biased Beliefs

To motivate the potential relevance of retrieval failures, we first present suggestive evidence on beliefs. Early in the agricultural year, a few months after the previous harvest, we elicit incentivized forecasts from farmers about their future savings (i.e., maize stock). Specifically, we ask them to forecast how many 50 kg bags of maize they will have 3 months in the future (just before the hungry season) and approximately 5 months in the future (in the middle of the hungry season). We inform participants that they will receive an incentive payment if their forecast is roughly accurate, based on our visual inspection of their maize stock in future visits.[15] We also 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."[16] Note that the agricultural year in which we elicited these forecasts was a typical one—we are not aware of any unusual negative aggregate shocks.

We display the results in Figure I. 80% of individuals think they will have higher savings 3 months later (at the start of the hungry season) than they actually do, with the median participant overestimating future savings by 66%. In addition, 65% of farmers end up with less savings than their worst-case savings forecast. Savings forecasts for 5 months in the future are similarly biased—and if anything more overoptimistic.

This overoptimism about savings coincides with substantial under-estimation of future expenses. Early in the agricultural year, we ask farmers to forecast their upcoming non-food expenses for the year. We limit our elicitation to non-food expenses because they are more irregular, varied, and stochastic than food expenses, making them less salient and likely more prone to retrieval failures. We then compare this with actual non-food expenditures reported by households in surveys for the remainder of the year.

[15] The incentive was a bag of sugar, which corresponds in value to about 30% of a 50 kg bag of maize.

[16] This elicitation approach has been used in the planning fallacy literature (Buehler et al., 2010).

As shown in the final bar of Figure I, realized expenditures are on average twice as high as farmers' initial forecasts.

Under-predicting future spending is potentially consistent with a variety of explanations, such as naive present focus (O'Donoghue and Rabin, 1999).[17] However, in our setting, most of the mispredicted expenses are not temptation goods, but rather items such as school fees, farm inputs, and medical costs. In addition, in Section 6.3 below, we unpack *which* expenses are most likely to be neglected—finding evidence that rare, irregular, and stochastic expenditures are relatively more likely to be forgotten.

4 Model

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 components of the problem that are "known" to her—i.e., available in her memory—but not retrieved when solving her problem. As a result, she solves an incomplete problem and therefore does not choose the optimal action. To capture this idea, we present a two-stage model in which (1) a person retrieves the components of a problem and then (2) solves the problem given those components.

While our model is framed in terms of the problem faced by the farmers in our intervention, it easily maps to more general planning problems with one important caveat. Our core results depend on an asymmetry between inflows and outflows. As with many people in the world, farmers' income is derived from a few large and predictable sources, while outflows come from a many small and unpredictable sources. Given this asymmetry, outflows are more likely to be neglected than inflows, causing people to overspend. We believe that this type of asymmetry is not limited to our setting. For example, in Section 10 below, we offer evidence suggesting a similar asymmetry between income and expenses among low-income Americans. In addition, time budgeting is similar: while the inflow of time is fixed (24 hours a day), the number of potentially unexpected outflows (meetings, conversations, traffic, sickness, etc.) is large.

[17]In this model, people fail to appreciate how strongly their future selves will weigh present consumption, leading them to underpredict their own future consumption of immediate-gratification goods.

4.1 Stage 1: Retrieval

Introduction

In the first stage of our model, the person retrieves the components of the problem. There are several economic models in which agents take only a subset of the relevant variables into account (Sims, 2003; Gabaix, 2014; Sims, 2010; Schwartzstein, 2014; Bordalo et al., 2013; Karlan et al., 2016). In our model, people consider components that come to mind through an *associative* retrieval process (Anderson, 1983; Raaijmakers and Shiffrin, 1981; Bordalo et al., 2023). Specifically, retrieval depends on the associative *similarity* to what is already being considered. For example, a person is more likely to recall "fertilizer" when already considering farming expenses than school expenses. This associative process distorts retrieval away from the optimal set and—crucially for our empirical test—can be nudged by simple manipulations.

Modeling associative retrieval is challenging because the process is inherently stochastic and path-dependent, and likely uses an endogenous stopping rule. Rather than pinning down a fully realistic specification, our goal is to build a tractable model that captures the core prediction that retrieved items are clustered in some associational space and retrieval depends on the cue that initiates search. We therefore adapt Bordalo et al. (2023) (hereafter "BCGKS"), who avoid complex path dependence by assuming each draw depends only on the *initial cue*.[18]

Expenses

The core components of a budget maximization problem are income and expenses. We start by discussing the retrieval of possible expenses, which we index by i and denote $\mathcal{E} = \{e_1, \ldots, e_n\}$. Retrieval of a given expense e_i is presumably influenced by its past financial magnitude, which we capture with an expense-specific weight parameter w_i. However, to isolate the effect of associations, we start with the simplification that each expense has the same weight $w_i = 1$ and then relax this restriction.

[18] BCGKS study probabilistic prediction: to judge the likelihood of a *hypothesis*, people sample *events* depending on (1) the associational similarity to the hypothesis, and (2) the event's objective probability. We port their mechanism to a budget maximization problem with prompted *categories*, where people retrieve *budget components* depending on (1) the associational similarity to the prompted category, and (2) the financial importance of the component.

Setup: Similarity and Categories

Following BCGKS, we let the symmetric similarity parameter

$$S_{ij} \in [0, 1]$$

denote the similarity between e_i and e_j. We assume $S_{ii} = 1$ and $S_{ij} = S_{ji}$. Similarity represents the relative amount that the retrieval of expense i is boosted after considering expense j.

We are interested in how prompting expense categories impacts retrieval. To do so, we modify the logic in BCGKS by defining the similarity between an expense and a non-empty subset (or *category*) of expenses $\mathcal{C} \subseteq \mathcal{E}$ as the average similarity between the expense and all of the expenses in $\mathcal{C}$:

$$S_i(\mathcal{C}) := \frac{1}{n_{\mathcal{C}}} \sum_{j \in \mathcal{C}} S_{ij},$$

where $n_{\mathcal{C}} = |\mathcal{C}|$. Category similarity represents the relative likelihood that expense i will be considered when prompted by category $\mathcal{C}$. This simple definition compresses the richer dynamics of recalling a random chain of expenses—where the next expense depends on the current path—into a tractable single index, while keeping the same comparative statics. We define the *retrieval probability* $r_i(\mathcal{C})$ of expense i when prompted by category $\mathcal{C}$ as depending on similarity (normalized such that $\sum_i r_i(\mathcal{C}) = 1$):

$$r_i(\mathcal{C}) := \frac{S_i(\mathcal{C})}{\sum_j S_j(\mathcal{C})}.$$

That is, intuitively, if a person is prompted with a category, they think of expenses that are similar to those in that category. We assume that the person *draws* $d \geq 1$ expenses *with replacement* given the probabilities $r_i(\mathcal{C})$.[19] Our core interest is the probability that component e_i is retrieved at least once when prompted with category $\mathcal{C}$:

$$p_i(\mathcal{C}; d) \ := \ 1 - \Big(1 - r_i(\mathcal{C})\Big)^d.$$

Since the process is stochastic, retrieval always has a chance of being imperfect given multiple expenses and a finite number of draws d. Another immediate implication is

[19] Drawing with replacement captures the fact that an already-considered item may come to mind again and allows for the possibility that d draws do not always lead to d retrieved distinct components.

that—given a non-degenerate similarity function—prompting a person with different categories will impact the likelihood that a given expense is recalled. In particular, an expense will be recalled more easily when the person is prompted with a category of similar expenses. This structure encodes that "fertilizer" is more likely to be recalled after prompting the category "farm expenses" rather than "school expenses." That is, given associative retrieval, it is possible to use a category prompt to boost a specific expense. Is it also possible to break the entire set of expenses into a partition of *multiple categories* to enhance the retrieval of all expenses?

To answer this question, we consider a person who is presented with m categories $\Gamma = \{\mathcal{C}_1, ...\mathcal{C}_m\}$ that partition the expense space $\mathcal{E}$. In order to not artificially drive retrieval with more categories, we fix the total number of draws at D, such that each category is given D/m draws.[20] Therefore, the total probability of retrieving expense i is

$$p_i^{all}(\Gamma) := 1 - \prod_{c=1}^{m}(1 - p_i(\mathcal{C}_c; D/m)).$$

Given the sampling interpretation, we assume D/m is an integer, although the expression can be naturally extended to accommodate a non-integer. The *total expected recall* over all expenses is then defined as $R(\Gamma) := \sum_i p_i^{all}(\Gamma)$, which captures the expected number of distinct expenses the person retrieves after being prompted with the categories in Γ.

The Benefits of Categories to Recall

The goal of this paper is not to derive an analytically optimal partition of the expense space. Determining the exact optimum for similar problems in the network literature is NP-hard even in relatively simple situations (Brandes et al., 2006; DasGupta and Desai, 2013; Schaeffer, 2007; Fortunato, 2010). Instead, our goal is simply to note that associations drive the benefit of categorization, that categorization can lead to *more* recall than no categorization given associations, and that this benefit is driven by balanced groups with similar expenses.

To start, note that without associations (if $S_{ij} = 1$ for all i,j), the likelihood of

[20] Intuitively, D represents the total search effort, and $d = D/m$ the effort allocated to each category. We fix D to isolate the effect of categories on retrieval, holding constant any independent effect on effort. Allowing D to respond to search productivity would likely reinforce our results: if people stop when additional draws have low expected value or after repeated retrievals, then categories that reduce repeats will increase total search. Similarly, optimal allocation of draws across categories can only improve recall relative to equal allocation.

retrieval is unchanged for any category, as every category is equally associated with all expenses. In this case, categorization will have no impact on total retrieval given a fixed number of draws D:

Proposition 1. Associations are necessary for categorization benefits at fixed search effort. *If $S_{ij} = 1$ for all i, j, then $R(\Gamma) = R(\Gamma')$ for all partitions Γ, Γ'.*

Given a non-constant similarity matrix, categorization can help retrieval because it avoids *repeated draws* of the same expense (thereby allowing those draws to be used on other expenses). When a person continues to think about the same few expenses, these draws are wasted because the person has already recalled that expense. In the language of classic memory models, these commonly-considered expenses are "interfering" with the retrieval of other expenses. Intuitively, when a person is given no categorization, their mind wanders the associative space and often stays in the same "region." Prompting categories continually resets this process in different regions, allowing the person to recall more expenses.

For an extreme example, consider two expenses e_1 and e_2, with $w_1 = w_2 = 1$, and suppose the expenses have no similarity with each other. With two draws from the broad category $\mathcal{C} = \{e_1, e_2\}$, each draw recalls either expense with probability 1/2. Thus, with probability 1/2, the same expense is recalled twice and the other expense is not recalled. By contrast, suppose the person receives one draw from $\mathcal{C}_1 = \{e_1\}$ and one draw from $\mathcal{C}_2 = \{e_2\}$. The person still makes two retrieval attempts, but the attempts are specialized: the first is aimed at e_1 and the second at e_2. Both expenses are therefore recalled with certainty. Categorization helps not by increasing the number of retrieval attempts, but by reducing repeated retrieval.

More generally, every draw given a category distributes one unit of probability mass across the expenses. Therefore, given D draws, the total amount of mass is fixed at D. But, total recall depends on how this mass is arranged. Good categories improve recall by combining two forces. First, they specialize different draws toward different regions of memory, which reduces repeated retrieval. Second, they preserve coverage across the expense space, so that search is not distorted too heavily toward one group of expenses. Therefore, good categories include a set of similar expenses (to specialize the draws) that balance probability across expenses (to keep overall coverage high).[21]

[21] To show why balance is needed, consider a (somewhat perverse) counterexample. Let $\mathcal{E} = C_1 \cup C_2$, with $|C_1| = |C_2| = 50$. Set all similarities between C_1 and C_2 at 0, and all within C_1 be 1. C_2 is more complicated: one hub expense has similarity 1 with every C_2-expense, while all other members

To formalize this intuition, we define a category's *clustering level* $C(\mathcal{C})$ and *broadness* $B(\mathcal{C})$ as

$$C(\mathcal{C}) := \sum_{i \in \mathcal{C}} S_i(\mathcal{C}) - \sum_{i \notin \mathcal{C}} S_i(\mathcal{C}) \quad \text{and} \quad B(\mathcal{C}) := \sum_i S_i(\mathcal{C}).$$

Clustered categories that are reasonably balanced in broadness will always increase recall:

Proposition 2. Reasonably balanced, positively clustered categories increase retrieval *Consider the category of all expenses $\mathcal{E}$ (with $|\mathcal{E}|$ even) versus categorization $\mathcal{C}_1$ and $\mathcal{C}_2$ of equal-sized groups. Suppose that $w_i = 1$ for all i. If categories are*

(i) positively clustered: $C(\mathcal{C}_1) > 0$ and $C(\mathcal{C}_2) > 0$ and

(ii) reasonably balanced:

$$|B(\mathcal{C}_1) - B(\mathcal{C}_2)| < \frac{B(\mathcal{C}_1)+B(\mathcal{C}_2)}{D} \sum_{i=1}^n \left[\left(1 - \frac{r_i(\mathcal{C}_1)+r_i(\mathcal{C}_2)}{2}\right)^D - \left((1 - r_i(\mathcal{C}_1))(1 - r_i(\mathcal{C}_2))\right)^{D/2} \right],$$

then $\mathcal{C}_1$ and $\mathcal{C}_2$ lead to strictly higher total expected recall than $\mathcal{E}$.

That is, although finding the optimal set of categorization may be challenging, it is possible to increase total recall from the broad category by using naturally-clustered balanced categories.

Weighted Recall

In the discussion above, we assumed that each expense was weighted equally. However, retrieval of a given expense is presumably influenced by its expected financial magnitude or other baseline salience features (such as the positive or negative psychological weight associated with the expense).[22] In this case, we can modify the recall probability to be impacted by weights w_i:

$$r_i(\mathcal{C}) := \frac{w_i S_i(\mathcal{C})}{\sum_j w_j S_j(\mathcal{C})}.$$

of C_2 have similarity $\varepsilon = .001$. In this case, categorization lowers total recall given a moderate level of draws. While the categories contain (weakly) similar expenses, C_2 is so misbalanced that it wastefully allocates almost every draw to the hub expense, increasing repeated draws.

[22] Naturally, one might wonder how an expense can be given a retrieval weight relative to its financial magnitude if that financial magnitude depends on its retrieval. While one may imagine a fixed-point model in which the ex-ante expectation of spending based on the realizations of recall determine the likelihoods of recall, here we sidestep the question and appeal to basic intuition (and results from the literature, such as Malmendier and Wachter (2024)) showing that more financially-important goods are more likely to be recalled.

Therefore, removing the effect of associations ($S_{ij} = 1$ for all i,j), people recall expenses in proportion to their weights.

Our interest lies in how weights may influence the types of expenses that are retrieved. Most clearly, the weights suggest that expenses with a very large budget impact are unlikely to be forgotten. For example, in our setting, food is the dominant (and most salient) expense. Therefore, in the categorizations we implement in our field experiments, we place food in its own category and we assume that food will never be forgotten when the person solves their budget problem. We believe that many problems have similar "always-recalled" components: for example, when mapping our model to a time-budgeting problem, it would be analogous to assume that people always consider sleep when allocating time over the 24 hour day.

The model accommodates stochastic expenses by weighing them based on their *expected* financial impact. Given a fixed number of draws D, it is mechanical in our setup that splitting probability across expenses will lead to a smaller weighted total expected recall. Interestingly though, the effect is stronger than intuition might suspect. To see this, suppose that one expense with retrieval probability r is replaced by two rarer expenses with retrieval probabilities $r/2$. In this case, the expected number of draws to recall the original expense is $1/r$. Meanwhile, the expected number of draws to first retrieve *one* of the rarer expenses is $1/r$, but then retrieving the second rarer expense requires $1/(r/2) = 2/r$ more draws, for a total of $3 \cdot 1/r$ draws. That is, even though the probability was split in 2, the number of expected draws is increased by a factor of 3. The same logic implies that breaking a large expense (with weight w) into 2 smaller expenses (with weight $w/2$) requires 3 times more draws.[23] More generally:

Proposition 3. Small and rare expenses require (many) more draws

Consider an expense i with per-draw recall probability r_i that is split into s expenses, each with recall probability r_i/s. The expected number of draws to recall the full set is increased by a factor of $s \cdot H_s$, where $H_s = \sum_{l=1}^{s} 1/l > 1$.

The same result holds when splitting an expense i with weight w_i into s expenses with weight w_i/s (assuming the new expenses have the same similarity structure as the original).

This result is a direct analog of the classical coupon-collector result (Feller, 1968).

[23]This result naturally leads to a question of the specificity of expense categorization in people's minds—if, for example, one recalls very specific expenses (Cheerios on August 15th) rather than broader expenses (groceries for the year), recall will be much lower. In our model, expense categories are given, and the model has little to say about the innate coarseness of categories in people's minds.

H_s is the sth Harmonic number, where $H_s \to \infty$ as $s \to \infty$, such that the multiplier rises without bound. Therefore, in our model—as in many others (Kahana et al., 2024; Malmendier and Wachter, 2024)—small and rare expenses are especially prone to retrieval failure.

Income

Given a large set of income sources, the argument for categorization of income matches that of expenses. However, as discussed above, there is a core asymmetry in the farmer's budget problem: income comes from a few, large sources. Given the small number of income sources, we make little use of categorization of income in our field experiment. Consequently, in our theoretical approach, we assume that income is always retrieved (just as we assume that the main expense of food is always retrieved). Mapping our model to a time-budgeting problem, this is analogous to assuming that people always remember that a day has 24 hours.

4.2 Stage 2: Maximization Given Recalled Components

In stage 2 of our model, the person uses the retrieved components and solves the maximization problem. In doing so, we assume that she is naive about any retrieval failures in stage 1.[24] Given our setup, we model this maximization problem as a simple multi-period "eat-the-pie" problem. As discussed in stage 1 of the model, we assume that people correctly retrieve their income (Y) and the fact that they must eat (choose c_t) in each period. Let $R_t = \{1, 2, ..., n_t\}$ denote the full set of non-food expenses due in period t, and let e_{it} be the amount spent on expense $i \in R_t$ in period t.

For simplicity, we assume that there is no stochasticity in expenses. Stochasticity does not change the core results but requires a notationally-burdensome setup and necessitates that individuals have a full state-by-state contingent plan.[25] To isolate

[24] A person who does not retrieve a specific expense cannot condition her plan on that expense. However, she may suspect that she is "forgetting something" and add a precautionary budgetary buffer based on past shortfalls. The predictions require only that any such buffer is, on average, too small to offset the omitted expenses, as suggested by our motivating facts.

[25] The model in a previous version of the paper was stochastic and the equivalent results were stated in terms of expectations. The problem becomes more complicated if individuals also use associative memory to determine the probabilities of stochastic events conditional on retrieval. Here, individuals may potentially *overestimate* the probability (Bordalo et al., 2023) of retrieved expenses. This leads to more complicated dynamics in which, for example, individuals may over-insure against recalled emergency shocks (because they overestimate the probability when considering that shock) but under-insure overall (because they only consider the few emergency expenses that they can recall).

the impact of our mechanism, we assume no time discounting, no borrowing, and a three-period model.[26]

Therefore, in period 1, the classic maximizer solves the problem:

$$\begin{aligned} \max_{c_t, e_{it}} \sum_t u(c_t) + \sum_t \sum_{i \in R_t} v_{it}(e_{it}) \\ \text{s.t. } \sum_t c_t + \sum_t \sum_{i \in R_t} e_{it} = Y, \end{aligned}$$

where $u(\cdot)$ and $v_{it}(\cdot)$ are utility functions (which we assume satisfy the Inada conditions), and we normalize all units so that prices for all expenses are 1.

Effect of Imperfect Recall

In stage 1, given a finite number of draws, people fail to retrieve some expenses *on average.* Focusing on the period-1 planning problem, let $\tilde{R}_t \subseteq R_t$ denote the subset of period-t non-food expenses that the person recalls while forming her period-1 plan. Rather than take expectations over all possible realizations of draws, we focus on a realization in which recall of future expenses is imperfect:

Realization Restriction 1. *In stage 1, retrieval is imperfect. We focus on a realization in which $\tilde{R}_t \subset R_t$ for each future period.*

Therefore, in period 1, the person instead solves an incomplete version of the problem.

$$\begin{aligned} \max_{c_t, e_{it}} \sum_t u(c_t) + \sum_t \sum_{i \in \tilde{R}_t} v_{it}(e_{it}) \\ \text{s.t. } \sum_t c_t + \sum_t \sum_{i \in \tilde{R}_t} e_{it} = Y, \end{aligned}$$

This is a period-1 planning problem. The person chooses an intended spending path for all three periods, but only period-1 choices (and therefore savings into period 2) are implemented; future choices can be revised when later periods arrive. When a person enters the next time period, we assume that (1) she cannot forget previously-recalled expenses, (2) she recalls all current expenses that are now due, and (3) she cannot

[26] While adding exponential discounting does not change our results, and we see no obvious reason that this would not be true for more general discounting functions, we have not proved this: complicated and non-intuitive dynamics can arise when the individual is partially sophisticated and strategically manipulating their future self.

recall previously-forgotten future expenses.[27] These assumptions are deliberately stark: they provide clean sufficient conditions that isolate the core force in the model, while standing in for the weaker requirement that violations of these restrictions are not large enough, on average, to overwhelm the effect of newly retrieved current expenses. Furthermore, in the case of (1), our experiment uses category labels to help people continue to remember expenses they had already retrieved.

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 (and therefore save less) than in the rational benchmark where she fully appreciated these expenses. The individual is then "surprised" in the future when some of these expenses arise and she consequently must cut back on planned consumption and other planned expenditures. This will generate a consumption cycle, with higher spending early in the cycle and lower spending later in the cycle. Furthermore, because she does not retrieve all her expenses, the individual will spend more in the future than she expected and, consequently, have less savings than expected.

For the formal statements, let $j \in \{R, M\}$ index rational ($j = R$) and mistaken ($j = M$) types, and let c_t^j denote the period-t consumption for type j, $\sum e_t^j$ denote non-food period-t spending, and s_t^j denote the savings entering into period t. Meanwhile, let $\hat{c}_t^j$, $\sum \hat{e}_t^j$ and $\hat{s}_t^j$ denote the respective period-1 predictions of type j. For the rational (time-consistent) type, these predictions always line up with the truth. However, this does not necessarily hold for the mistaken type, as they later appreciate new expenses. Given these definitions:

Proposition 4. Effects of retrieval failures

1. *In comparison to the rational benchmark, total spending is higher in the first period ($c_1^M + \sum e_1^M > c_1^R + \sum e_1^R$) and lower in the last period ($c_3^M + \sum e_3^M < c_3^R + \sum e_3^R$).*

2. *In the initial period, people overpredict future food expenditures ($\hat{c}_2^M > c_2^M$ and*

[27] If (1) is violated, the period-1 self may underpredict food expenditures in period 2. Intuitively, suppose that the period-1 self recalls a large period-3 expense and doesn't appreciate that their period-2 self will forget this expense. The period-2 self will unexpectedly not save for the large period-3 expense. If this new spending desire is large compared to the new need to spend on newly-recalled period-2 expenses, the person may consume more period-2 food than expected. Conversely, if (3) is violated, the period-1 self may underpredict future savings. Intuitively, suppose the period-1 self does not realize that the period-2 self will recall a very large period-3 future expense. After recalling the expense, the period-2 self will save more in 2 than the period-1 self expects. If this new savings desire is large compared to the new need to spend on newly-recalled period-2 expense, the person may save more in period 2 than expected.

$\hat{c}_3^M > c_3^M$), underpredict total future non-food expenditures ($\sum \hat{e}_2^M + \sum \hat{e}_3^M < \sum e_2^M + \sum e_3^M$), and overpredict future savings ($\hat{s}_3^M > s_3^M$).

Although we frame our results using situation-specific objects like "food," the theory broadly predicts that people will underpredict currently-forgotten future expenditures, and that this non-predicted future spending will cause their predictions about currently-recalled future expenditures to be too high. Mapping our model to a time-budgeting problem, an analogous prediction would be that people will underpredict small events like impromptu meetings or transportation time, which cuts into their previously predicted work and sleep time.

Empirically, the misprediction in Figure I matches result (2) in Proposition 4. The result in (1) is more challenging to test as it references behavior of the unobserved rational benchmark—however, under the assumption that spending in the benchmark is reasonably smooth, the food shortages in Appendix Figure A.1 are consistent with this prediction. 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.

The Effects of Categorization on Budgeting

In stage 1, we discuss how categorization changes how a person recalls upcoming expenses. Therefore, we can use changes in categorization to study changes in recall, allowing us to cleanly test if retrieval failures drive beliefs and behavior. To show the effect of retrieval, we compare mistaken types in the control condition (coarser-category) and treatment condition (finer-category), which we index with $k \in \{C, T\}$.

Given reasonable categories, Proposition 2 suggests that finer categorization can boost recall. This result only occurs *on average*, so—as in Restriction 1—we remove the need to take expectations over all possible realizations of recall by comparing realizations in which finer categorization strictly expands future recall. Formally, letting $\tilde{R}_t^k \subseteq R_t$ denote the set of period-t non-food expenses that the person recalls in period 1 given their condition, we assume:

Realization Restriction 2. *Finer categorization expands period-1 recall of future non-food expenses in both periods:*

$$\tilde{R}_2^C \subset \tilde{R}_2^T \subseteq R_2 \quad \textit{and} \quad \tilde{R}_3^C \subset \tilde{R}_3^T \subseteq R_3,$$

Given this restriction, we derive several predictions for how finer categorization affects beliefs and behavior. We define actual and predicted consumption, expenditures, and savings as above, but now indexed by C and T. Next, following our experiment, consider a marginal discretionary purchase available in period 1 that gives a small utility benefit b, and let p^k denote the marginal willingness to pay in condition k.[28]

Proposition 5. Predictions on intervention impacts

An intervention that boosts retrieval leads to:

1. *A first-period increase in the prediction of total future non-food expenditures ($\sum \hat{e}_2^T + \sum \hat{e}_3^T > \sum \hat{e}_2^C + \sum \hat{e}_3^C$).*

2. *A fall in first-period willingness to pay for a small discretionary purchase ($p^T < p^C$).*

3. *A flatter spending profile: lower first period total spending ($c_1^T + \sum e_1^T < c_1^C + \sum e_1^C$) and higher final-period spending ($c_3^T + \sum e_3^T > c_3^C + \sum e_3^C$).*

The first prediction effectively serves as a test of whether the intervention has the intended effect of more complete retrieval of expenses. The 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 new expense will fall. The third prediction concerns the long-term impact of these changes. Following the second prediction, an individual who anticipates future expenses will spend and consume less in order to save for these anticipated expenses. At some point in the future, these savings will allow her to spend more on consumption and expenses.

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, reducing the need to unexpectedly increase labor supply to meet future expenses). Similarly, if borrowing is possible but costly, the intervention will lead to less borrowing over time as she faces fewer unanticipated shortfalls.

[28] b must be small enough that income effects are second order. Formally, the proof shows that the second part of the Proposition is true for small enough b.

4.3 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 expenses whose utility benefits arrive in the future and discontinuously changes her discount rate when utility from the good occurs in the present. In our model, she sharply "discounts" neglected expenses and discontinuously changes her discount when expenses are retrieved, 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 expenses and—more importantly—understanding the utility flow of each good over time. While there might be some expenses where the timing is obvious (e.g. paying a bill for a service already rendered likely causes little immediate gratification), we anticipate that the classification will be difficult and controversial.[29] 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) predicts no effect of an intervention to boost retrieval if individuals already fully account for their expenses.

5 Intervention and Study Sample

5.1 Intervention

The key insight from our model is that individuals do not retrieve information about upcoming expenses that they already "know" (i.e., from their own memory)—leading to consumption smoothing failures. To test the empirical relevance of the model, 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

[29]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?

intervention on beliefs and behavior to test Predictions 1-3 of Proposition 5.

Proposition 2 indicates that finer and positively clustered categories increase retrieval. 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 a "retrieval board" that prompts individuals to think through their expenses category-by-category (see Figure II). 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 and other emergencies, and a residual "other" category). In selecting these seven categories, our goal is to cluster expenses into similar groups—with cues that are specific enough to assist with associative recall—while also keeping categories 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.[30] In addition, having a relatively small number of categories helps prevent fatigue and keeps the exercise tractable.

As discussed in the model, we consider food to be an "always recalled" expense. In our setting, adult household members eat maize (in the form of nshima patties) in each meal they consume (see Section 2). Consequently, food expenditures are not only large and salient, but also regular—making them less subject to retrieval failures. This gives rise to a prediction about *which* expenses will exhibit directional changes in beliefs: thinking through the categories in the retrieval board will lead to higher predicted *non-food spending* (Proposition 5, Prediction 1).

We undertake an exercise with individuals in the treatment group using this physical retrieval board (Figure II, 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 retrieval 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. After the survey is completed, the

[30]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.

retrieval board and thumbtacks are taken back by the enumerator (so that participants cannot hold onto them).

For the control group, we face a core tradeoff: we would like individuals to complete a similar exercise to the treatment group to gather information on their perceptions of expenses, but this in itself may act as a treatment of sorts. 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; however, 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 longer-term behavioral differences between the treatment and control groups.

5.2 Sample and Summary Statistics

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

We conduct both experiments in the Eastern Province of Zambia. We draw our samples 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. These criteria are not very restrictive; for example, 90% of farmers are classified as smallholder farmers in this setting and the majority report food shortages in the hungry season (Fink et al., 2020).

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 (see Section 8.2). All study activities are conducted by enumerators at the participant's home.[31]

[31]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.

The demographic statistics are broadly the same for samples in 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 study participants are male (female participants are often unmarried heads of household). At baseline, 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.[32]

6 Mechanism Experiment

6.1 Design

The goal of the mechanism experiment is to isolate the impact of categorization on retrieval, and on subsequent beliefs and behavior. We construct a design in which all participants allocate their available savings to categories on a board; however, we vary the level of aggregation—and therefore usefulness—of these expense categories.

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

6.2 Implementation

Figure III displays the timeline of activities for the mechanism experiment. First, we elicit all participants' priors of how much they will spend on non-food expenditures over the coming year.[33] Second, treatment and control farmers undertake the retrieval

[32]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.

[33]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 retrieval exercise also involves allocating maize bags to expenses. This question is embedded in a short baseline survey that is administered to all participants

exercise: allocating their available savings (i.e., maize) using the full retrieval board or placebo board, depending on their treatment status. Third, we examine short term treatment effects using a willingness to pay exercise. Finally, we ask respondents to characterize their recalled expenses in light of models of memory such as Bordalo et al. (2023): for each expense item listed, the surveyor asks its amount, when it is expected to arise, its expected frequency, and its likelihood of arising. Control group participants do this following the placebo board, and treated participants following the full retrieval exercise. The control group then undertakes a second retrieval exercise, using the full retrieval board, after which we again ask them to list and characterize their recalled expenses. We also collect their self-reports on whether each expense was remembered or forgotten during the placebo board exercise. 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 5.2). Appendix Table A.1 shows that the treatment and control groups are balanced on most baseline covariates.

6.3 Results

Prediction 1: Increase in Predicted Future Expenses

Applied to the mechanism experiment, Prediction 1 of Proposition 5 implies that individuals in the treatment group will predict higher future spending on previously neglected expenses due to improved retrieval. We test this prediction by comparing the number of bags of maize 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 IV. At baseline, the prior is the same on average across the treatment and control groups. Relative to the placebo board control, the full retrieval board substantially increases expected non-food expenses: the treatment group's allocation of bags of maize to non-food expenses is 36% higher than the control (p<0.001) (see Appendix Table A.3, column 2). 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

at the start of the session.

Figure A.6, 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 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. This is consistent with the full board helping with the retrieval of previously neglected expenses, as per Prediction 1 (Proposition 5). Note also that the control group's mean forecast of non-food expenses is, after completing the placebo board, 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. [34] Rather, effects only emerge when participants are presented with associative categories in the full retrieval board—consistent with our hypothesized mechanism.

Prediction 2: Decrease in Willingness to Pay for Immediate Expenses

Prediction 2 (Proposition 5) states that the intervention will decrease willingness to pay for a small, immediate purchase. To test this prediction, we measure participants' demand for a discretionary consumption good: 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.[35] Note that in-home transactions of these types of goods are not unusual: households commonly buy goods from "briefcase buyers" who sell items door-to-door in villages in exchange for maize.

Consistent with our theoretical prediction, the distribution in the control stochastically dominates that in the treatment (Figure V, 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.5). This result is robust to including controls for baseline characteristics, item

[34]For example, control individuals decisions might have changed their allocation if their planned spending did not match their available maize. The retrieval board, by forcing budget balance, would highlight this contradiction and cause a change in maize allocation.

[35]Trades are implemented with low probability.

fixed effects, or a Tobit specification (columns 2-4 of Appendix Table A.5). 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. Proposition 3 states that small and rare (and uncertain) expenses require many more draws before they are retrieved.

Completing the full retrieval exercise causes the control group to increase their expected non-food expenses by 37% (Appendix Table A.3, column 2), which is very close to the between-subject treatment effects. We use control group reports of the expenses they included at each stage, along with the reported characteristics of the expenses to assess Proposition 3.

Appendix Figures A.2 and A.3 show the frequency with which expenses are assigned to different characteristics, as well as the correlation among characteristics. Participants indicate that 80% of expenses are certain to occur, while 20% are not, and another 20% have uncertain timing. These measures of uncertainty are highly correlated with one another.

Appendix Figure A.4 provides a list of the most frequently remembered and forgotten expenses, according to control group self reports. Appendix Table A.4 shows the descriptive statistics associated with these expenses. Expenses that are uncertain and irregular, like transfers and emergencies, are frequently forgotten. By contrast, expenses that are large and certain to occur, like household goods and fertilizer, are much less likely to be forgotten. Table I shows the results of regressing whether an expense was forgotten on different characteristics of that expense. Consistent with Proposition 3 in our model, expenses that are small, infrequent, irregularly-timed, and uncertain are more prone to retrieval failures.[36]

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. We next turn to a separate field experiment, in which the structure of the control is

[36]Numerous other features of an expense might make it focal and therefore assigned a higher weight, as discussed in our model. We focused on (expected) magnitude because this could be characterized easily across all expenses and all participants. Future work on what determines weights could help inform the design of more effective retrieval exercises.

designed to avoid this type of contamination.

7 Field Experiment

We design a field experiment to investigate longer-term impacts—specifically, effects on savings and household spending profiles (Proposition 5, Prediction 3).

7.1 Experimental Design

The field experiment design includes two substantial changes relative to the mechanism experiment.[37] First, because even the 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.

Second, we introduce a tool to aid the durability of recall. In the model, we assume that once an expense is recalled, it will be recalled in all future periods. This, of course, is potentially unrealistic if the consumption cycle lasts an entire year. People may eventually forget expenses that were recalled at the time of the retrieval exercise, leading to too much spending in intervening periods (see footnote 27). 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 retrieval board; we give participants the option to affix a label to each bag of maize according to their retrieval 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 in both treatment and control groups the labels. All participants across both groups are 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 think they will spend on each category. We also explicitly tell control participants that some individuals find it helpful to use the labels to delineate their spending plan for the year.

[37] In addition to these 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.

All participants are then shown a choice between the labels or a small compensation (a bag of sugar). We find that the treatment group is significantly more likely to take up labels than the control group.[38]

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. Second, we provide the labels over *six weeks* after the retrieval exercise. This enables us to examine the impact of the retrieval board alone over a substantial time horizon, before labels are introduced. We further discuss the effect of the labels in Section 8.

7.2 Implementation

The field experiment activities were conducted between late August 2019 and Sept 2020.[39] Figure VI provides an overview of the field experiment timeline and activities. At the start of implementation, most households had just completed shelling their maize. The retrieval exercise 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 retrieval board shown in Panel A of Figure II. Both the treatment and control groups are then offered the choice 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

[38]Around 80% of treated participants choose 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."

[39]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.

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 replicate the test of Proposition 5, Prediction 2. Note that trades are implemented with low probability to avoid generating an imbalance in initial savings between the treatment and control groups if treatment affects willingness to pay. For the treatment group, the retrieval exercise 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 at each survey visit, 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 around six weeks 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 at the end of the Visit 2 survey; the participant chooses which labels to attach to their remaining maize bags.[40] We complete two additional visits between December and March (Visits 3 and 4), and again measure savings (maize and cash), expenditures, and labor supply. 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; this also meant that the Visit 4 survey instrument was abbreviated and did not contain data on non-food expenditures.

We return in September/October—approximately one year after the intervention was conducted—for a final survey visit (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.

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 5.2). We randomize at the individual (rather than village) level in

[40] Before applying labels to their bags, treated participants were shown their original retrieval board and given the opportunity to undertake the retrieval exercise again with their remaining maize (since the total bags of maize would have changed since the original exercise). The full protocol is provided in Appendix C.

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 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.

7.3 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 (Figure IV; results for the mechanism experiment are shown in Panel A and for the field experiment in Panel B). 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. Panel B of Figure IV shows that the intervention increases the allocation to non-food expenses by more than 60% ($p < 0.001$), consistent with Prediction 1 (Proposition 5). Furthermore, 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.6, mechanism experiment in Panel A and field experiment in Panel B).

Second, the intervention decreases the willingness to pay for immediate spending. Panel B of Figure V qualitatively matches the results from the mechanism experiment (shown in Panel A of the same figure). 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 Prediction 2 (Proposition 5). As before, this result is robust to including controls for baseline characteristics, item fixed effects, or a Tobit specification (columns 2-4 of Appendix Table A.5).

7.4 Prediction 3: Effects on the Time Profile of Savings Drawdown

We now turn to the longer run results of the field experiment—following the treatment group and measuring impacts relative to the control group over the subsequent year.

Empirical Strategy We use repeated household survey data on savings levels to construct the net drawdown of measured savings between visits. We estimate treatment effects in an OLS regression specification. Because our model predicts that treatment effects will vary over time, we estimate:

$$y_{ij} = \sum_{v=2}^{4} \beta_v \left(T_i \times \mathbb{1}\{j = v\}\right) + \lambda_j + \delta_{\text{time}_{ij}} + \sum_{v=2}^{4} \mathbb{1}\{j = v\} X_i' \theta_v + \varepsilon_{ij}. \tag{1}$$

where y_{ij} is the outcome of participant i during Visit (survey round) j. T_i is an indicator for assignment to the budgeting treatment at baseline, and $\mathbb{1}\{j = v\}$ is an indicator for survey-visit v (with Visits 2-4 being the post baseline visits). The coefficients β_v estimate treatment effects separately by Visit. X_i is a vector of baseline controls for participant i, which we allow to enter flexibly by survey visit. We include survey-visit fixed effects, λ_j, and calendar-time fixed effects, $\delta_{\text{time}_{ij}}$, where time_{ij} denotes the calendar week or month in which participant i's visit-j survey was completed. The calendar-time fixed effects account for the fact that survey timing varies across participants and that survey visits may overlap across calendar weeks.[41] We estimate period-specific treatment effects, β_v, as our preferred measure of impact.

Results Prediction 3 (Proposition 5) states that the treatment will lead to a flatter average spending profile over the consumption cycle (lower first period spending and higher later period spending). A direct implication of this change in spending is that treated households will have weakly higher savings, particularly at the start of the consumption cycle. Figure VII provides evidence consistent with this prediction. The y-axis of the figure measures shows the net drawdown of measured savings, which captures the difference in savings quantities between survey visits, divided by the weeks

[41] In most specifications, we control for the baseline value of the outcome of interest; we also present specifications showing robustness to alternative 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. When treatment is interacted with Visit, we also include survey-visit fixed effects.

between each visit. This effect is displayed starting in October, the beginning of Visit 2, and ending in early March, the end of Visit 4.[42] The figure documents that treatment households draw down measured savings more slowly immediately after the intervention, leading to increased savings. As a result, they are able to draw down their savings 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 drawdown of savings across the year for treated households.

We examine the effect on the evolution of savings stocks directly in Table II. 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 almost 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, $p = 0.029$). This first follow up was on 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 retrieval 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 ($p = 0.015$). 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 12 kilograms of maize more than the control group, although this effect is not statistically significantly different from zero ($p = 0.51$).[43]

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

[42] Since net drawdown is computed as savings in the previous visit minus savings in the current 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.

[43] This pattern of savings suggests that relative to the control group, treated participants were also able to delay selling some some of their maize until later in the year. 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.

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.

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.7, 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 capture potential substitution from processed to unprocessed maize (Table A.7, column 3). Next, we examine total expenditures by treatment on household assets, including livestock, that could conceivably function as savings (Appendix Table A.8).[44] 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 livestock.

7.5 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 ask respondents to report the number of meals consumed by adult household members on each of 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) in order to buy maize (for consumption), but at the expense of leisure and family labor on their own farms (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

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

recall period for these outcomes differs across follow up visits.[45] In addition, we measure whether the household hired outside workers to work on its farm (in Visit 4 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), using the cross section. Throughout, we report results at the frequency at which they were collected: if we collected a given outcome in multiple survey visits, we report effects separately by survey visit; if we only collected it once, we report one effect over the recall period.

Results Consistent with the evidence on timing of net savings drawdown, Table III suggests that treated participants slightly reduced consumption of maize (the staple food) between the baseline and the first follow up survey visit (Visit 2), although this reduction is not statistically significant (column 1). The treatment group has slightly higher consumption of staple meals at the beginning of the hungry season (Visit 3, $p = 0.090$), although this difference disappears toward the end of the hungry season in Visit 4.[46]

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 III 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 32% fewer days of wage labor on others' farms, relative to the control group mean (column 2) ($p = 0.035$), 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 higher wage earnings among the control group (column 3) are used to purchase food.

Consistent with less diversion of labor away from their farms, we find suggestive evidence that treated participants spend more days working on their own farm during the hungry season (column 4) ($p = 0.1467$). We find additional suggestive evidence for increases in other farm inputs. Treated households purchase more casual labor

[45] 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.

[46] Note that our very crude measures of food consumption (e.g. meals per day) may be too inelastic to detect impacts on consumption. Data on a food security index shows a statistically insignificant reduction in food insecurity in both Visits 3 and 4.

to work on their farms during the hungry season (column 5, $p = 0.082$), and report higher expenditures on other inputs such as fertilizer and herbicide/pesticides (column 6, $p = 0.137$). Increased farm inputs lead to higher crop revenues at the subsequent harvest. Treated households report the value of their entire harvest as being 9.4% higher than control households on average ($p = 0.076$).

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.

8 Alternative Explanations

The mechanism and field experiments demonstrate that the retrieval exercise causes participants to increase their recall 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. 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. We begin with a discussion of the role of labels, which touches on alternative mechanisms, namely salience and soft commitment.

8.1 The Role of Labels

Reminders and Increased Salience of the Need to Save. We interpret our mechanism as retrieval leading to a genuine change in participants' beliefs about their budget, which in turn alters spending decisions. 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 without enabling more general retrieval.

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. Labels play no role in the mechanism experiment. Second, in the field experiment, neither group is given access to the "reminder technology" (labels) between the intervention and Visit 2, and yet savings differ significantly between the groups. This time periods leading up to the application of labels is key for household consumption smoothing, since it includes the months when expenditures are highest and the opportunity for saving for the hungry season is greatest.

Determining the precise impact of the labels applied in Visit 2 is more difficult, since labels 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.[47] 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.7, Panel A) and participants value the labels with the intervention at nearly 10 times their value for labels alone (Appendix Figure A.7, Panel B).[48]

Present Bias and Soft Commitment As we discuss in Section 4.3, models of present bias have a maintained assumption—common to most economic models—that individuals have unconstrained access to information that they "know." Therefore, under present bias alone, an intervention that manipulates only retrieval should not have an effect. It is therefore challenging to explain the results of the mechanism experiment (i.e. Predictions 1 and 2) with present bias. In addition, 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.

One potential concern is that labels may act as a "soft commitment", serving as

[47]The process of formulating or updating a spending plan based on the household's remaining maize stocks at Visit 2 may have also cued additional retrieval of upcoming expenses.

[48]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.

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, as we describe in Section 7.1, all participants have access to the labels; we even explicitly tell the control group that they can use the labels to mark their spending plan for the year. 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 their use. Furthermore, since we find little demand for these labels in the control group, this suggests 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.[49]

8.2 Intrahousehold Coordination

Since the retrieval exercise affects household behavior, 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 has immediate effects: it 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

[49]If individuals are naive about their present bias, they may choose not to self-impose a soft commitment device. Similarly, if farmers are overly optimistic about their future savings because of retrieval failures, they will not engage in debiasing on their own.

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.

8.3 Demand Effects

Some of our results may be susceptible to experimenter demand effects. For instance, while our retrieval exercise 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-food items (Appendix Figure A.6). If anything, the perceived demand in this context would probably be to save more for future food consumption, which would suggest that beliefs should 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 retrieval board would signal demand for lower food consumption. Moreover, 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 II.

9 Extensions: The Persistence of Biased Beliefs

Figure I 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 further, and present suggestive evidence on additional mechanisms for why biased beliefs may persist despite experience.

To motivate this line of inquiry, note that, even in the presence of retrieval failures, the patterns in Figure I still present a puzzle: even if individuals do not remember all their future expenditures, they could realize that they run out of maize earlier than expected each year, and debias their beliefs about future savings this way.[50] A full explanation of why farmers do not do this is beyond the scope of this paper. Instead, we organize our analysis around the three ways that the planning fallacy literature has shown that individuals could debias their beliefs about the future—we show that farmers appear not to engage in any of them in our setting. 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. Almost 85% of individuals in the lowest age quartile are overoptimistic, compared to 75% in the top age quartile, where the mean age in each quartile group is 28 and 62, respectively (Appendix Figure A.8).[51] 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.9). 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

[50] A natural question is why biased forecasts persist despite repeated experience. One explanation is that learning is limited by experimentation or feedback (Gittins and Jones, 1974; Fudenberg and Levine, 1993). In our setting, however, persistence is more naturally understood as an attention or retrieval failure: households may have experienced many relevant expenses, but fail to retrieve them when budgeting. This is more closely related to accounts in which misspecified models persist because they shape which information agents attend to (Gagnon-Bartsch et al., 2023).

[51] Appendix Figure A.8 shows the cumulative distribution function of forecast error—(incentivized) forecasts of future savings minus realized savings—for the bottom and top age quartiles.

of bags they recalled, their forecast error at baseline would be only 20% of the size observed.[52] Note that while participants exhibit rosy memory, they do not completely disregard the hungry season: on average, they recall having 50% 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.[53] Taking these two findings together, we can compare a participant's recollection of their forecast error, to their actual forecast error. Appendix Figure A.10 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 almost 80% of participants exhibit positive forecast errors, fewer than 30% recall that their forecasts were overoptimistic. In other words, together, these pieces of evidence suggest that many people remember the past as better than it was, and they do not remember having made a forecast error in the past. Both these will limit the amount of learning that is possible from recall-based forecasting.

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.12). While this exercise was not incentivized, it is consistent with other experiments that find evidence

[52] 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.

[53] 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 $\bar{\hat{S}}_3 - \bar{S}_3$.

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 direction for additional research.

10 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. For example, 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 755 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 retrieval exercise, we first report respondents' perceptions of how income and expenses match their forecasts in Figure VIII. 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 our model's description of retrieval failures.

We then examine whether engaging in a retrieval exercise—thinking through categories—affects beliefs about future 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.[54] Example screenshots from the survey are shown in Appendix Figure A.13.

Figure IX shows the resultant impact on beliefs from the retrieval exercise: the CDFs of the posterior, relative to the prior, for expenses and incomes. Consistent with Prediction 1 of Proposition 5, we observe that 57.5% of the sample revises expenses upward, with a mean change in forecasted expenses of 12%. While the mean update in forecasted income is also positive, it is substantively smaller: the mean change is 4%, 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. These findings suggest that similar retrieval exercises may yield consumption smoothing benefits across a range of settings.

11 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

[54] 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)).

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 9.4%. 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).[55] 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 exercise. We investigate this in follow-up data collection when we return one year after the initial intervention. We ask treatment farmers their willingness 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

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

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 of what gives rise to retrieval failures in the first place, and the impediments to learning (e.g., why is memory biased?) that allow them to persist 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, 11(2019), 284–301.

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

Anderson, John R. "A Spreading Activation Theory of Memory." Journal of Verbal Learning and Verbal Behavior, 22(1983), 261–295.

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

Ashraf, Nava, Dean Karlan, and Wesley Yin. "Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines." The Quarterly Journal of Economics, 121(2006), 635–672.

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

Bee, Adam, Bruce D. Meyer, and James X. Sullivan. "The Validity of Consumption Data: Are the Consumer Expenditure Interview and Diary Surveys Informative?" In Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus, editors, "Improving the Measurement of Consumer Expenditures," University of Chicago Press (2015). 204–240.

Bénabou, Roland and Jean Tirole. "Mindful economics: The production, consumption, and value of beliefs." Journal of Economic Perspectives, 30(2016), 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, 53(2016), 535–550.

Bernard, Tanguy, Stefan Dercon, Kate Orkin, Giulio Schinaia, and Alemayehu Seyoum Taffesse. "The Future in Mind: Aspirations and Long-Term Outcomes in Rural Ethiopia." The Quarterly Journal of Economics, 141(2026), 1383–1447.

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

Blumenstock, Joshua, Michael Callen, and Tarek Ghani. "Why do defaults affect behavior? Experimental evidence from Afghanistan." American Economic Review, 108(2018), 2868–2901.

Bolte, Lukas and Collin Raymond. “Emotional Inattention.” Technical report, Working paper (2022).

Bordalo, Pedro, John J. Conlon, Nicola Gennaioli, Spencer Yongwook Kwon, and Andrei Shleifer. “Memory and Probability.” Quarterly Journal of Economics, 138(2023), 265–311.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. “Salience and Consumer Choice.” Journal of Political Economy, 121(2013), 803–843.

———. “Memory, attention, and choice.” Quarterly Journal of Economics, 135(2020), 1399–1442.

Brandes, Ulrik, Daniel Delling, Marco Gaertler, Robert Görke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. “Maximizing Modularity is Hard.” arXiv preprint physics/0608255 (2006).

Breza, Emily and Supreet Kaur. “Psychology and Development.” In Pascaline Dupas, Pinelopi Goldberg, and Rohini Pande, editors, “Handbook of Development Economics,” Elsevier (forthcoming).

Brune, Lasse, Eric Chyn, and Jason Kerwin. “Pay me later: Savings constraints and the demand for deferred payments.” American Economic Review, 111(2021), 2179–2212.

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, 82(2014), 1671–1748.

Buehler, Roger, Dale Griffin, and Johanna Peetz. “The planning fallacy: Cognitive, motivational, and social origins.” In “Advances in experimental social psychology,” volume 43. Elsevier (2010). 1–62.

Buehler, Roger, Dale Griffin, and Michael Ross. “Exploring the "planning fallacy": Why people underestimate their task completion times.” Journal of Personality and Social Psychology, 67(1994), 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, 134(2019), 785–842.

Carranza, Eliana, Aletheia Donald, Florian Grosset-Touba, and Supreet Kaur. “The social tax: Redistributive pressure and labor supply.” Econometrica, 93(2025), 2273–2308.

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, 62(2018), 95–104.

Cattaneo, Matias D, Richard K Crump, Max H Farrell, and Yingjie Feng. "On binscatter." American Economic Review, 114(2024), 1488–1514.

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

Chetty, Raj and Emmanuel Saez. "Teaching the tax code: Earnings responses to an experiment with EITC recipients." American Economic Journal: Applied Economics, 5(2013), 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). 23–50.

DasGupta, Bhaskar and Devendra Desai. "On the Complexity of Newman's Community Finding Approach for Biological and Social Networks." Journal of Computer and System Sciences, 79(2013), 50–67.

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), 1–78.

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

Dillon, Brian, Joachim De Weerdt, and Ted O'Donoghue. "Paying more for less: why don't households in Tanzania take advantage of bulk discounts?" World Bank Economic Review, 35(2021), 148–179.

Dupas, Pascaline and Jonathan Robinson. "Why don't the poor save more? Evidence from health savings experiments." American Economic Review, 103(2013), 1138–1171.

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

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

Feller, William. An Introduction to Probability Theory and Its Applications, volume 1. John Wiley & Sons, New York, 3rd edition (1968).

Fink, Gunther, B Kelsey Jack, and Felix Masiye. "Seasonal liquidity, rural labor markets, and agricultural production." American Economic Review, 110(2020), 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, 36(2008), 791–798.

Fortunato, Santo. "Community Detection in Graphs." Physics Reports, 486(2010), 75–174.

Fudenberg, Drew and David K Levine. "Self-confirming equilibrium." Econometrica: Journal of the Econometric Society, (1993), 523–545.

Gabaix, Xavier. "A Sparsity-Based Model of Bounded Rationality." The Quarterly Journal of Economics, 129(2014), 1661–1710.

———. "Behavioral inattention." In "Handbook of behavioral economics: Applications and foundations 1," volume 2. Elsevier (2019). 261–343.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein. "Channeled Attention and Stable Errors." (2023). Working Paper.

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

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

Gittins, J. and D. Jones. "A dynamic allocation index for the sequential design of experiments." Progress in Statistics, (1974), 241–266.

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

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

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

Howard, Ray Charles "Chuck", David J Hardisty, Abigail B Sussman, and Marcel F Lukas. "Understanding and neutralizing the expense prediction bias: The role of accessibility, typicality, and skewness." Journal of Marketing Research, 59(2022), 435–452.

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

Kahana, Michael J., James D. Paron, and Jessica A. Wachter. "Associative Learning and Representativeness." (2024). Working Paper, University of Pennsylvania and Stanford GSB.

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

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

———. "Intuitive prediction: Biases and corrective procedures." Technical report, Decisions and Designs Inc Mclean Va (1977).

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

Karlan, Dean and Jonathan Morduch. "Access to finance." In "Handbook of development economics," volume 5. Elsevier (2010). 4703–4784.

Karlan, Dean, Aishwarya Lakshmi Ratan, and Jonathan Zinman. "Savings by and for the Poor: A Research Review and Agenda." Review of Income and Wealth, 60(2014), 36–78.

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, 127(2012), 199–235.

Kremer, Michael, Gautam Rao, and Frank Schilbach. "Behavioral development economics." In "Handbook of behavioral economics: applications and foundations 1," volume 2. Elsevier (2019). 345–458.

Kuhn, Michael A. "Electronic Benefit Transfer and Food Expenditure Cycles." Journal of Policy Analysis and Management, 40(2021), 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, 4(1978), 551.

Malmendier, Ulrike and Jessica A. Wachter. "Memory of Past Experiences and Economic Decisions." In Michael J. Kahana and Anthony D. Wagner, editors, "The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications," chapter 78. Oxford University Press (2024). 2228–2266.

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, 108(2011), 10415–10420.

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

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

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

O'Donoghue, Ted and Matthew Rabin. "Doing it now or later." American economic review, 89(1999), 103–124.

Paxson, Christina H. "Using weather variability to estimate the response of savings to transitory income in Thailand." American Economic Review, (1992), 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, 35(2009), 1579–1591.

Peetz, Johanna, Roger Buehler, Derek J Koehler, and Ester Moher. "Bigger not better: Unpacking future expenses inflates spending predictions." Basic and Applied Social Psychology, 37(2015), 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), 25(2011), 825.

Raaijmakers, Jeroen G. W. and Richard M. Shiffrin. "Search of Associative Memory." Psychological Review, 88(1981), 93–134.

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

Schaeffer, Satu Elisa. "Graph Clustering." Computer Science Review, 1(2007), 27–64.

Schwartzstein, Joshua. "Selective Attention and Learning." Journal of the European Economic Association, 12(2014), 1423–1452.

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

Sims, Christopher A. "Implications of Rational Inattention." Journal of Monetary Economics, 50(2003), 665–690.

———. "Rational Inattention and Monetary Economics." Handbook of Monetary Economics, 3(2010), 155–181.

Soman, Dilip and Amar Cheema. "Earmarking and partitioning: Increasing saving by low-income households." Journal of Marketing Research, 48(2011), S14–S22.

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, 37(2010), 264–278.

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

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

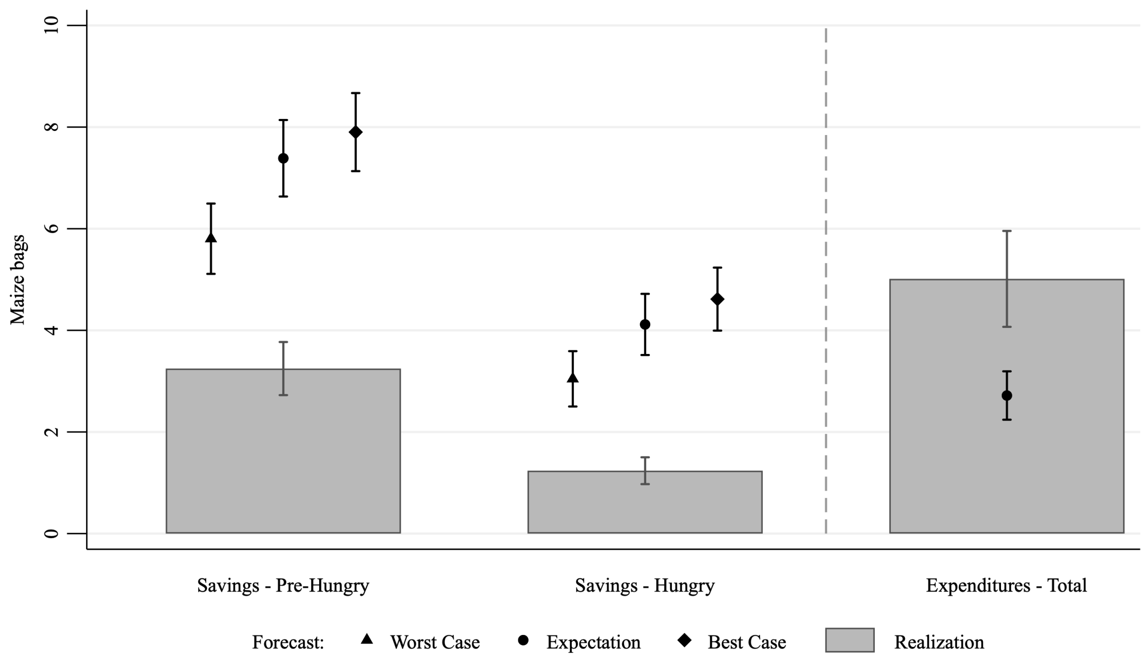
———. "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." Psychological Review, 90(1983), 293.

Wicker, Till, Patricio Dalton, and Daan van Soest. "Mental Accounting and Cash Transfers: Experimental Evidence from a Humanitarian Setting." Working Paper, (2026).

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

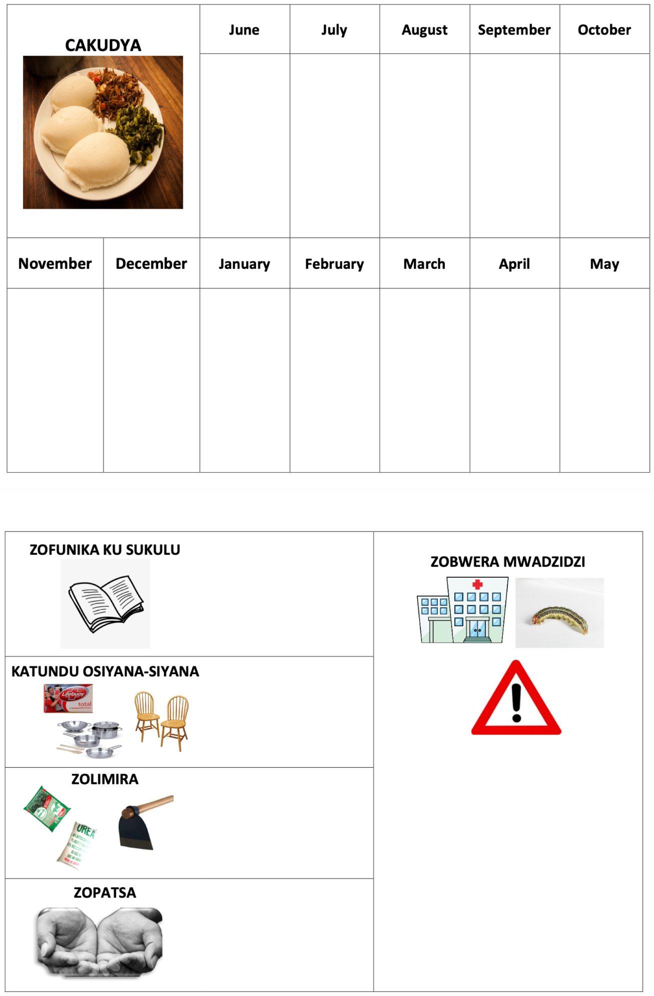
12 Tables and Figures

Figure I: 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 were surveyed in follow-up visits.

Figure II: Retrieval board and categories

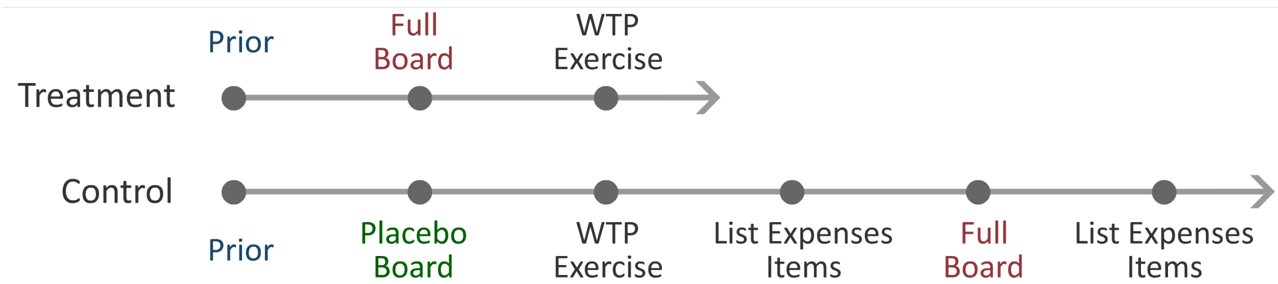


(a) Panel A: Full retrieval board

(b) Panel B: Placebo board

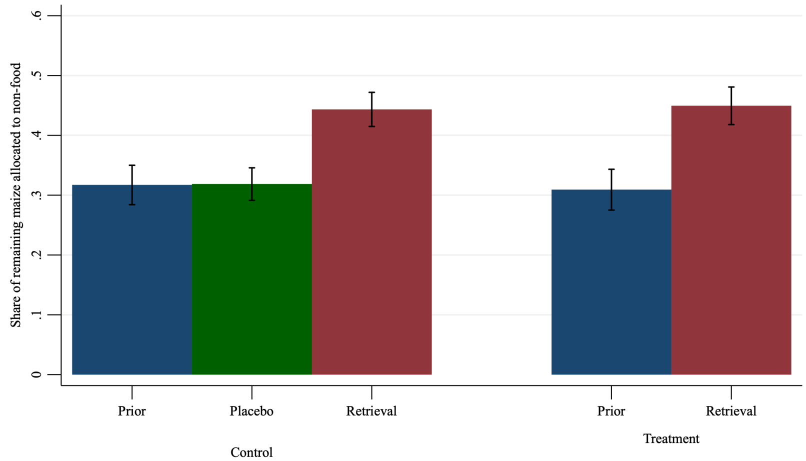
Notes: Panel A: The full retrieval board used by the treatment group to allocate their savings to expense categories. As part of the treatment, participants receive a set of pins, with each pin representing one bag of maize from their savings. Participants assign pins to food consumption (allocated by 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" 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 III: Timeline (mechanism experiment)

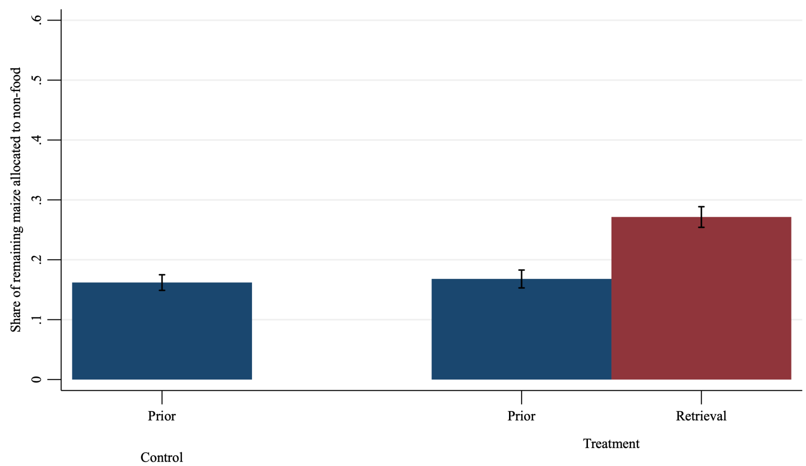


Notes: Mechanism experiment timeline. All participants are asked for their prior of non-food expenses. They are then randomized into treatment and control groups. The treatment group completes the retrieval board in Panel A of Figure II and the control group completes the board in Panel B. Both groups 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 IV: Share of maize bags allocated to non-food expenses



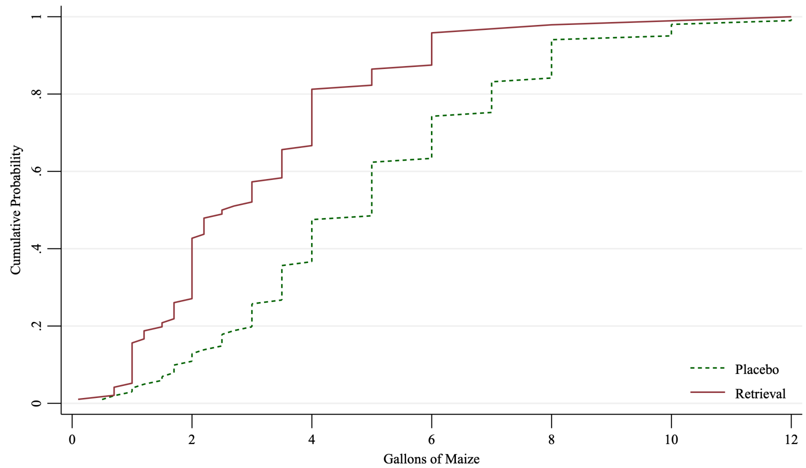
(a) Panel A: Mechanism experiment



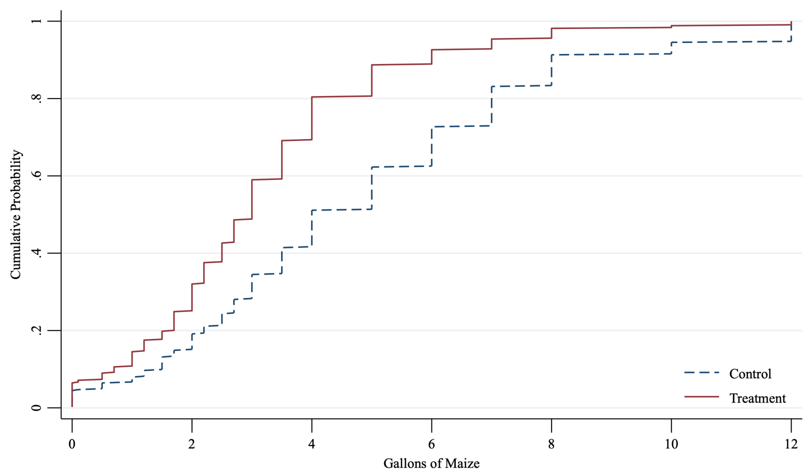
(b) Panel B: Field experiment

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 prior bar reports allocations before any retrieval board. The placebo bar reports allocations after the simplified board. The full-board bars report allocations after the six-category retrieval board. Panel B shows results from the field experiment. Both groups are first asked for their prior. Only the treatment group then completes the full six-category retrieval board.

Figure V: Willingness to exchange maize for discretionary consumption



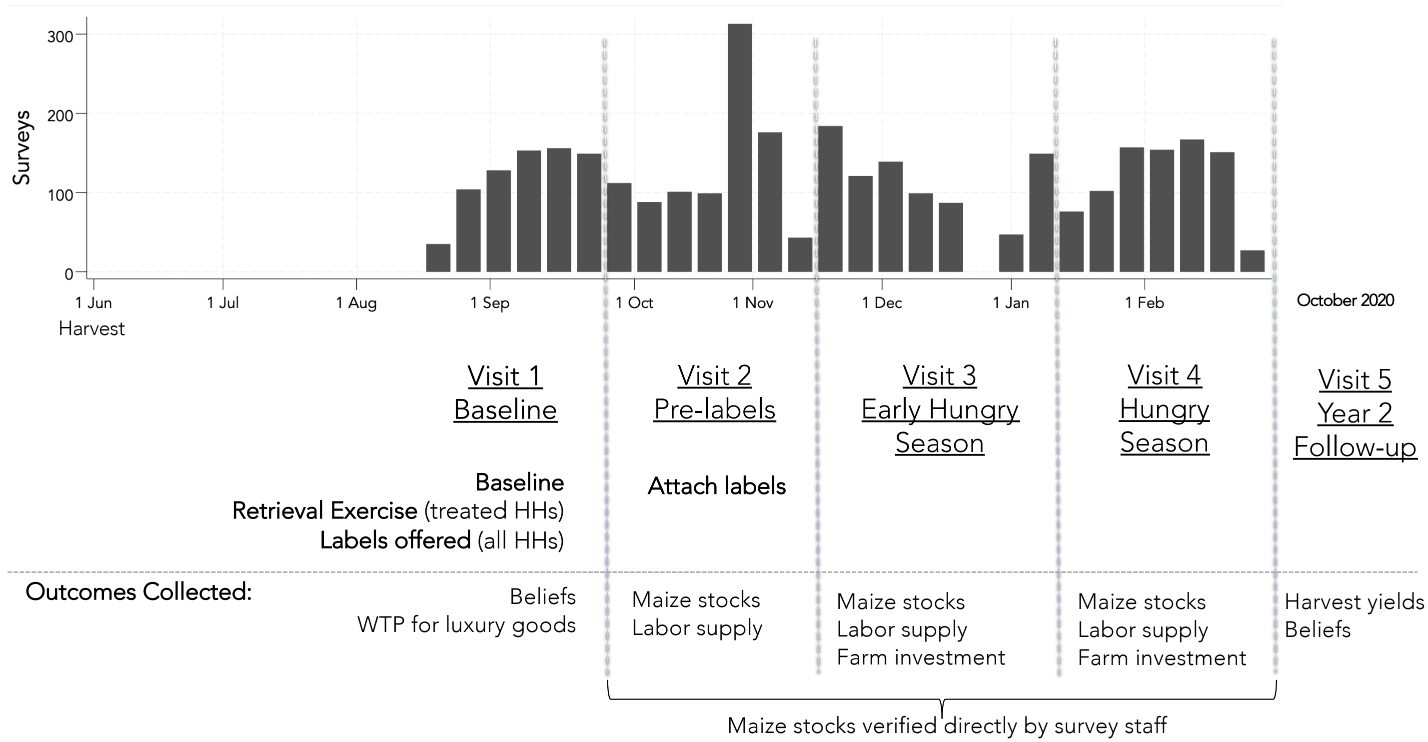
(a) Panel A: Mechanism experiment



(b) Panel B: Field experiment

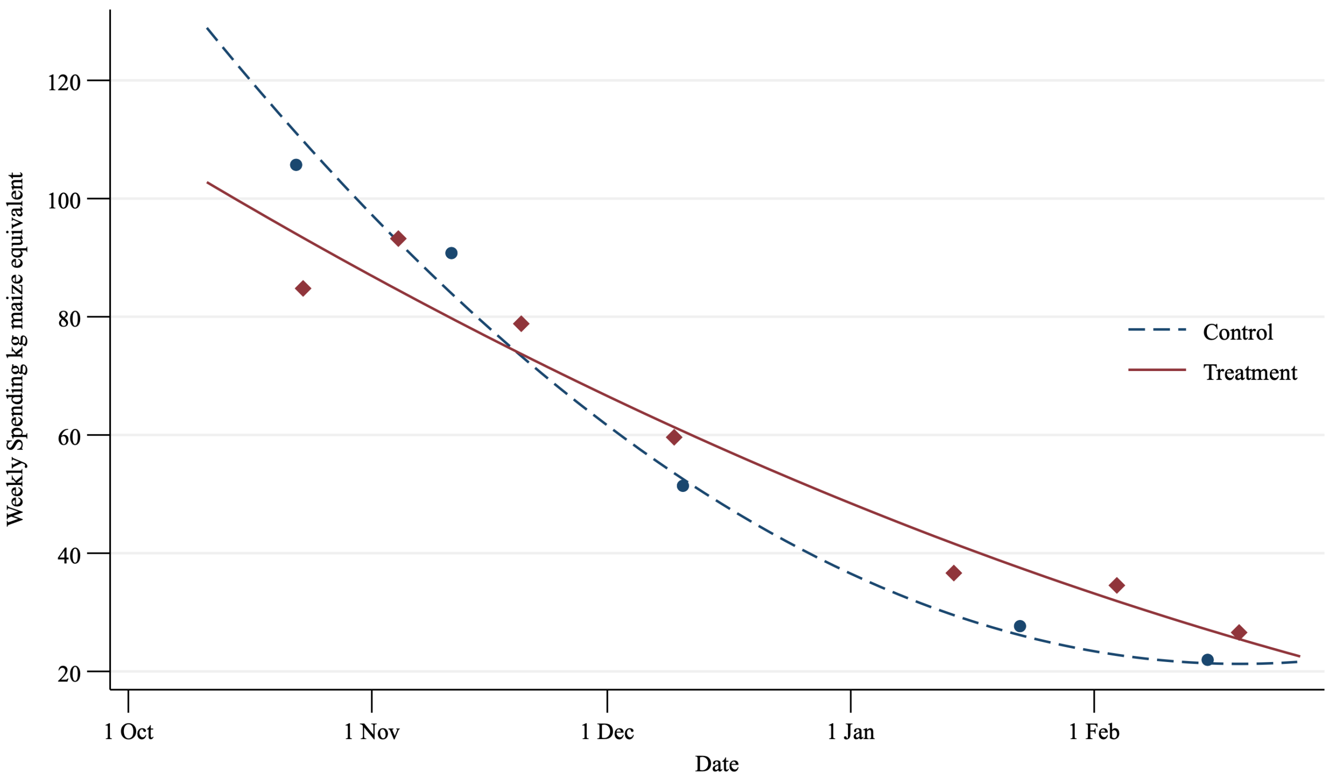
Notes: CDF of the willingness to pay (WTP) for discretionary consumption items. WTP is elicited using the Becker-DeGroot-Marschak method after the retrieval exercise. In Panel A, the dashed line shows the WTP in the control group (placebo board), and the solid line shows the WTP in the treatment group (full retrieval board) for the mechanism experiment. In Panel B, the dotted line shows the WTP in the control group (no retrieval exercise), and the solid line shows the WTP in the treatment group (full retrieval board) 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 VI: Timeline and data collection (field experiment)



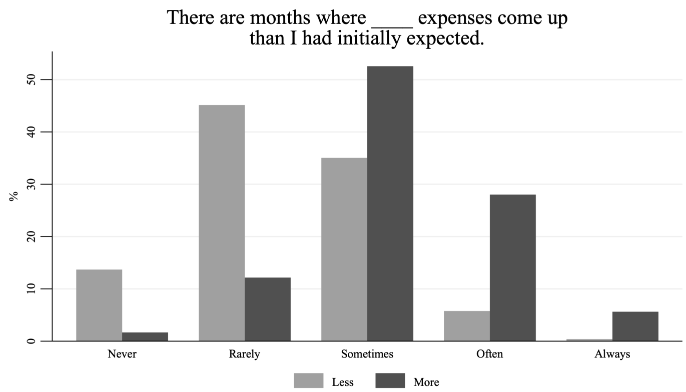
Notes: Field experiment timeline, including the intervention and data collection. Vertical bars represent the number of participants interviewed each week over the course of the study, by survey visits. The intervention was administered at the same time as the baseline survey (Visit 1). Labels were provided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

Figure VII: Net drawdown of measured savings (field experiment)

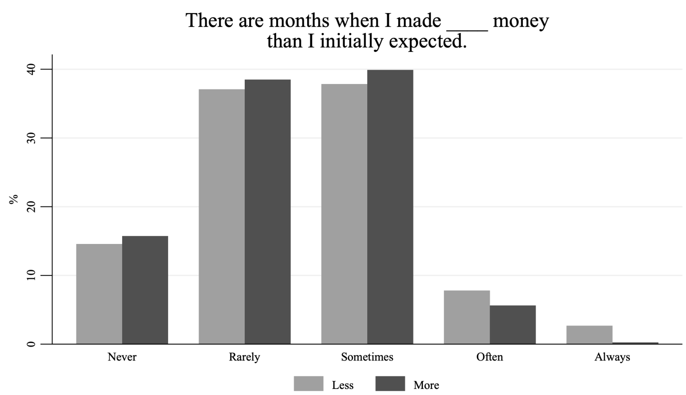


Notes: Smoothed net drawdown of measured savings for treatment (solid line) and control (dotted line) participants in the field experiment. The dependent variable is constructed as the difference in savings (measured as kilograms of maize plus the maize value of cash savings), divided by the number of weeks between survey visits. This approximates "weekly consumption". This is then plotted using a binscatter controlling for baseline cash and maize savings using the binscatter methods of Cattaneo et al. (2024). The binscatter is fit using a polynomial of order 2. Higher orders of polynomial do not meaningfully change the fit (see Appendix Figure A.11).

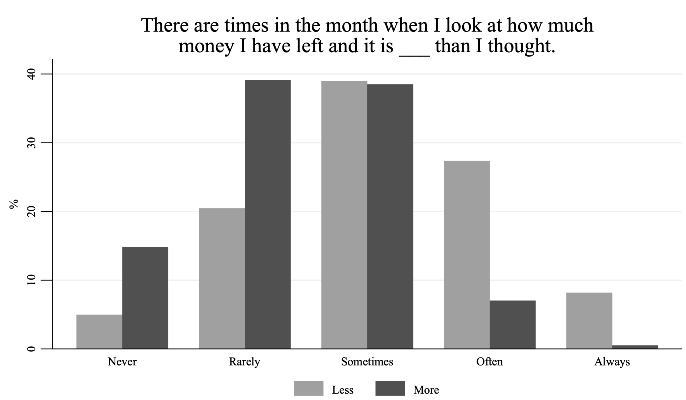
Figure VIII: Asymmetry in income and expenses (U.S. sample)



(a) Panel A: Expenses



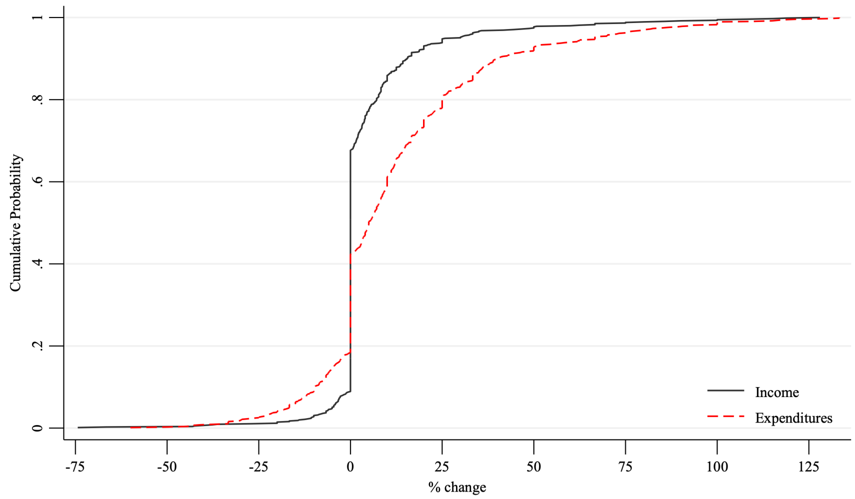
(b) Panel B: Income



(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=751). Respondents see all versions of the questions.

Figure IX: 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 solid line corresponds to the update in income, and the dashed line corresponds to the update in expenditures. The percentage update is calculated as the percentage difference between the prior aggregate estimate and the revised aggregate estimate following the retrieval exercise. The sample is restricted to employed individuals not living with their parents (N=751). In addition, we retain only the individuals whose aggregate income and expenditure revisions are within -75% to 150%, and whose initial aggregate guesses are within -75% to 300% of the implied totals obtained by summing categories. These restrictions trim an additional 5% of observations.

Table I: 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: This table reports regressions of an indicator for whether respondents report that an expense item was forgotten in their original placebo board plan on characteristics of the expense item. The sample consists of non-food expense items listed by control group respondents after completing the full retrieval board, excluding items classified in the residual "Other" category. "Frequency uncertain" refers to expenses with an uncertain frequency of expenditure. "Irregular" 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"). Baseline controls include: quantity of maize remaining and level of savings. Standard errors are clustered at the individual level.

Table II: 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)	99.67**	99.77***	0.75**	159.16*
	(45.67)	(37.69)	(0.36)	(89.50)
Treat x Visit 3 (Early Hungry)	69.47**	69.84***	0.60**	91.93**
	(28.69)	(24.02)	(0.27)	(44.96)
Treat x Visit 4 (Hungry)	12.41	13.38	0.09	41.95
	(18.47)	(16.63)	(0.16)	(42.66)
Dependent Variable unit	Kg	Kg	Bags	Kwacha
N	2480	2480	2480	2480
Control Mean Visit 2	660.87	660.87	7.99	426.60
Control Mean Visit 3	335.83	335.83	3.93	277.95
Control Mean Visit 4	159.16	159.16	1.43	234.04
F-test 2v3	0.36	0.34	0.61	0.40
F-test 3v4	0.01	0.00	0.01	0.23
RI p-value Treat x Visit 2	0.019	0.009	0.039	0.084
RI p-value Treat x Visit 3	0.019	0.004	0.023	0.062
RI p-value Treat x Visit 4	0.461	0.413	0.606	0.302
Baseline controls	No	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Visit 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 and survey visit fixed effects. Baseline characteristic controls are interacted with survey visit indicator variables. Standard errors are clustered at the household level.

Table III: 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 × Visit 2 (Pre-Labels)	-0.04	0.25	-1.80				
	(0.03)	(0.25)	(13.51)				
Treat × Visit 3 (Early Hungry)	0.05*	-0.05	2.80	1.03			
	(0.03)	(0.12)	(4.01)	(0.73)			
Treat × Visit 4 (Hungry)	-0.00	-0.23**	-11.31***	1.04	0.66*		
	(0.03)	(0.11)	(4.14)	(0.71)	(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
Visit FE	Yes	Yes	Yes	Yes	No	No	No

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. Columns 1 through 4 include Baseline characteristic controls interacted with survey visit indicator variables. All columns include week of survey fixed effects, and columns 1-4 include survey visit fixed effects. Standard errors are clustered at the household level.