UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Essays in Development Economics

Permalink

https://escholarship.org/uc/item/09d8n5dv

Author

Swanson, Nicholas

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Essays in Development Economics

By

Nicholas G. Swanson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Supreet Kaur, Co-chair
Professor Edward Andrew Miguel, Co-chair
Professor Ned Augenblick
Professor Frederico Finan

Spring 2024

Essays in Development Economics

Copyright 2024
by
Nicholas G. Swanson

Abstract

Essays in Development Economics

by

Nicholas G. Swanson

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Supreet Kaur, Co-chair

Professor Edward Andrew Miguel, Co-chair

This dissertation examines how market failures in low-income countries contribute to low firm and worker productivity, and interact with psychological biases. Markets in credit, risk and contracting, among others, are often incomplete in low-income countries, and it is important to understand which markets are incomplete, and the implications of this for firm productivity.

Chapter I of this dissertation, joint with Luisa Cefalà, Pedro Naso and Michel Ndayikeza, investigates the existence of a missing market for training in general human capital, and measures the distortions that arise from this market being missing. In low-income countries, contracts in labor markets are often short-term. Because of this, employers might underinvest in training workers in productivity enhancing general skills because the fact that there is weak labor market attachments mean they cannot guarantee that after training the worker they will capture returns from the training. This notion, that employers may have insufficient incentives to train employees, is an old idea in the labor economics literature, going back to Arthur Pigou and Gary Becker, but there is limited empirical evidence suggesting that this mechanism actually leads to meaningful distortions in the economy. We conduct two field experiments with agricultural employers in Burundi who can train casual laborers in improved, labor-intensive, agricultural techniques. In the first experiment, we offer employers in some local labor markets (villages) incentives to train workers. After training, employers who trained workers find it harder to retain them, with many of the trained workers subsequently working for other employers in the local labor market. The spillovers onto the

other employers (who did not train) are large, leading to improved farm outcomes and higher profits. In the second experiment, we randomize employers into a condition that increases the likelihood that the worker will return to work for the employer in the future. Employers receiving this guarantee are 50 percentage points more likely to train the worker. These findings suggest that the wedge between private and social returns to training meaningfully impedes on-the-job-learning, diminishing worker productivity and output.

Chapter II of this dissertation explores the implications of other missing markets on employers' decisions of who to hire in their business. Specifically, this chapter tests whether high unemployment and limited social safety nets in low-income countries generates "pressure" within families for family members who own businesses to hire members of their family as employees, even if privately they would prefer not to hire these relatives. The modal firm in low- and middle-income countries has no employees that are not family members. While this is often attributed to informational or contractual frictions, an alternate view is that pressure to offer financial assistance to extended family in the form of employment may distort employers' hiring decisions. In this chapter I conduct field experiments with employers in Zambia to test whether pressure impacts firm hiring and examine its productivity implications. A sample of urban firms are offered the chance to receive a 3-month subsidy for hiring a full-time permanent employee. A subset of firms is then randomized to receive plausible deniability in their hiring decision: receiving a poster that suggests the firm may not have been eligible for the subsidy if it hired a relative. This increases the probability of choosing to hire a non-related employee, rather than a related one. In the second experiment, I show that it is socially very costly for employers to hire a non-relative: when they have plausible deniability, the subsidy required to get a firm to choose a non-relative rather than a relative falls substantially. These findings suggest that social pressure to hire relatives may distort the composition of employment as well as productivity in developing countries.

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

To my family

Contents

List of Figures

List of Tables

Acknowledgments

I owe a huge debt of gratitude to my advisors for making the Ph.D. experience what it was.

Supreet, you took me under your wing even before I started the Ph.D. and have taught me the importance of using fieldwork to find out what is interesting and true about the real world, and then merge this with economic theory and empirics to try and explain this to others in a way that people take seriously. Working on our joint project was a major part of my learning during this process, and it's only been through your advice and "laser-focused" feedback that I've been able to refine my projects into what they are today, and structure and organise my thinking to become a more serious economist.

Ted, I really appreciated your constant encouragement, and the push that you always gave me to search for the big ideas and major open questions in development, and not settle for anything less. I think that everyone in the Berkeley development group sees your career, merging important research that pushes the research frontier with policy relevant questions, as an inspiration, and I'm no exception.

Ned, I really appreciate your willingness to spend long-meetings talking through the logic of my (and our) projects. Even though many of these meetings took longer than I expected because we had to go through every chain of the logic of an economic problem step by step, that clarity of thought made a lasting mark on me and if I can manage to think through problems half as well as you do I'll consider that a success.

Fred, I regret not talking to you more earlier in the Ph.D., because your advice on my projects this past year has been invaluable. I really appreciated your data driven questions, your willingness to say what you liked and didn't like honestly, and your willingness to go back and forth and push my work as far as it could go.

Beyond my advisors I owe a major debt of gratitude to additional faculty at Berkeley. In the development group, Ben Faber, Jeremy Magruder and Marco Gonzalez-Navarro particularly provided substantial feedback and advice on my projects. I owe a massive debt of gratitude to the labor economics faculty for engaging with me on the first chapter of my dissertation, and particularly to Pat Kline, Jesse Rothstein and Chris Walters who provided a huge amount of feedback that helped to clarify my thought and the exposition of the chapter. For chapters 2 and 3 of my dissertation feedback, I wish to thank also the behavioral group at Berkeley, and particularly Stefano DellaVigna and Dmitry Taubinsky for their comments and feedback. Finally, I want to thank Nick Tsivandis, who was willing to take a chance on me as an RA and has engaged and helped me with all of these projects over time.

I also thank those faculty from before Berkeley who supported me, introduced me to the best of development economics, without whom I wouldn't have reached this position. In particular, I thank Nava Ashraf, Heather Schofield and Frank Schilbach for their support and mentorship.

I also wish to thank my collaborators. Pedro and Michel - you have both been incredible co-authors and without you there is no way I would be working in Burundi today. I really appreciate you both as peers and friends and look forward to the next project. It was also a pleasure working on Chapter 3 of my dissertation with Kelsey and Felix who both really introduced me to the Zambian context, and whom I look forward to continue to collaborate with. I am also indebted to the faculty and staff at the University of Burundi and at CURDES, and in particular to Alexis Bizimungu, Gilbert Niyongabo and Redempteur Ntawiratsa who have been instrumental in facilitating our work in the country. The first chapter of my dissertation would not have been possible without the collaboration of the Burundian country office of One Acre Fund. I really appreciate the assistance of the entire staff, but particularly want to acknowledge Julia Darcey, Ryan Martin, and Marie Karleskind. I also want to thank those who have manged the field project, in particular Franck Irakoze, Berthacy Remesha and Fabrice Kimararungu. In Zambia, I want to thank Inez Dawoodjee for her constant support and collaboration. I also want to thank the field staff, and particularly William Phiri and Sarah Tembo for their leadership over multiple projects. Finally, many of these projects would not have been possible without the support and assistance of CEGA and many of its staff members.

I have fortunate to form strong bonds among my cohort, and in the development group across cohorts, and have had many amazing mentors and folks I could learn from. I particularly want to thank David and Maddie for their support this year.

I thank Luisa who has been with me throughout the Ph.D., who was always willing to push to create something new, and who shared the vision, curiosity and drive to push this research. All of the best parts of this Ph.D. have involved you.

I also thank my family: my father for his encouragement to do economics and his unwillingness to let me settle, my mother for her unending backing and support, and my siblings, for sharing this journey with me.

Chapter 1

Under-training by Employers in Informal Labor Markets: Evidence from Burundi

This chapter is coauthored with Luisa Cefalà, Pedro Naso and Michel Ndayikeza.

1.1 Introduction

The majority of employment in low- and middle-income countries is traded through short-term, informal labor market contracts. In rural economies, there is a prevalent trade of labor through spot labor markets, with minimal long-term, formal contracting (Rosenzweig, 1988; Behrman, 1999; Kaur, 2019).[1] Urban employers exhibit higher rates of turnover as compared to similar employers in high-income countries (Blattman and Dercon, 2018; Adhvaryu et al., 2021; Donovan et al., 2023).

In this paper, we study one distortion that might arise from short-term labor market contracts: employers may find it unprofitable to invest in training workers in productive general skills,[2] since the employer may not capture the future returns from training. Understanding the scope for distortions to firm investments is particularly important in low-income countries since in such settings on-the-job learning is a major source of skill acquisition (Ma et al., 2024), making limited firm investment an important potential explanation for the observed low level and limited growth of labor productivity (Lagakos et al., 2018). While a

[1] For example, 98% of agricultural employment in India is through casual labor contracts (Kaur, 2019).
[2] Skills that workers can use at multiple employers.

long, primarily theoretical, labor economics literature suggests that firms might limit training investments because returns accrue to workers (Pigou, 1912; Becker, 1964; Acemoglu and Pischke, 1999a,b) or future employers (Stevens, 1994; Acemoglu, 1997), there is limited empirical evidence in *any* setting that this mechanism leads to meaningful economic distortions.

We explore whether this mechanism limits employers' training investments in rural casual labor markets, an important setting given that the vast majority of the global extreme poor trade labor in such markets (Merfeld and Morduch, 2023). Specifically, we investigate employers' decisions in Burundian agricultural labor markets of whether to train casual workers in new agricultural technologies. Like many Sub-Saharan African countries, agricultural productivity in this setting is low (Gollin et al., 2014; Dercon and Gollin, 2014), and increasing productivity requires the adoption of high-yielding but labor-intensive agricultural techniques.[3] We focus on one such technique, row planting, that is widely promoted by international organizations and governments in Sub-Saharan Africa (Vandercasteelen et al., 2014) and increases yields by 30-70% (Dusabumuremyi et al., 2014). At baseline, few farmers use row-planting on all of their fields, and a majority of employers cite a lack of casual labor trained in row-planting as a constraint to further adoption. Despite the stated returns to more trained labor, only a thin market for worker training exists: 18% of employers reporting having ever trained casual laborers despite 64% stating they are capable of doing so (Figure 1.2a).

While establishing a market for training could be solved by employer or worker investments, we focus particularly on the constraints facing employers. In our context, direct worker investments in training are not feasible: the average worker is severely liquidity constrained, and perceives firm moral hazard in training effort.[4] Given the inability of workers to directly invest in training, we focus on what limits employers' training investments.[5] We find that the majority state that the nature of short-term contracts means they cannot guarantee they will benefit if they train workers, as this would primarily benefit other employers hiring the worker, or the worker themselves (Figure 1.2b).[6]

Motivated by this evidence, this paper tests whether employers do not train these workers because the structure of labor market means that they cannot "appropriate" sufficient returns

[3]Examples of such technologies include row-planting, manual transplanting (Emerick et al., 2016), demi-lunes (Aker and Jack, 2021) and pit planting techniques (BenYishay and Mobarak, 2019; Beaman et al., 2021).

[4]These features of firm training decisions have also been found in other contexts (Brown et al., 2022).

[5]Recent work also suggests that employers are more likely to finance training costs (Ma et al. 2023).

[6]Because planting occurs at short notice once the rains arrive, training would need to be done in advance of the agricultural season, and employers mention it is difficult to guarantee that a trained laborer would return once planting starts.

from training. We use a simple conceptual framework to develop two predictions associated with a labor market where employers underinvest in training because they do not capture the returns from the investment. The first prediction of this framework is that if employers do train, they are unable to capture all of the returns to training, and specifically that training generates externalities for future *non-training* employers. The second prediction of this framework is that if this externality is limited, by making workers more likely to return to work for the employer who trains them, employers may become more willing to invest in training. Each prediction relies on observing the labor market in a different state: prediction one requires measuring the incidence of training under the labor market status quo and prediction two requires observing employer training decisions in an altered labor market in which workers become more likely to return. Because of this, we design two separate field experiments to test each prediction.

In the first field experiment, the "spillover experiment", we test the first prediction: that when employers train this generates returns for workers and *non-training* employers. We conduct this experiment among Burundian farming communities, in collaboration with the country office of the NGO One Acre Fund (1AF). In these communities farmers can be categorized into two groups that align with the classical training framework: large farmers, who cultivate and regularly employ labor, and laborers, subsistence farmers who also work on other farmers' fields. This setting offers a major advantage in measuring the spillovers from training to other employers. Because villages are isolated and transport costs are high, each village in this setting constitutes a local labor market, allowing us to identify non-training employers who might benefit from "spillovers" from training.

We implement the "spillover experiment" in 80 villages in rural Burundi, sampling more than 3,600 farmers. Within each village, we incentivize some employers (trainer-employers) to invite a laborer (trainee) several months prior to the planting season to a training event that follows a similar structure to training events held by 1AF. We then randomize villages into one of two conditions that generate different incentives for trainer-employers to train the identified worker. In the control condition, enumerators suggest that the trainer-employer could train the trainee in row planting,[7] before giving the trainer-employer an unconditional financial transfer. In the *T1-Financial incentives* condition the payment that is given to trainer-employers is made conditional on training the trainee for at least half a day. Several months after the training event, during the next planting season, we measure labor market and farm outcomes of trainer-employers, trainees, and *spillover-employers* –employers uninvolved in the training, to measure who captures the returns from training.

We find that monetary incentives to train increase the stock of skilled-labor in treated villages. Trainer-employers in T1-Financial incentives villages are almost 80 percentage

[7]The training also involved fertilizer microdosage, a complementary agricultural practice.

points more likely to train their trainees than their counterparts in control villages. This training is effective with trainees in treated villages working on average 3.4 more days for employers doing row planting tasks (p¡0.01).

A large proportion of the trained labor is hired by *non-training* employers. While trainer-employers in treated villages hire 46% more days of labor to do row planting (p¡0.01), which enables them to adopt row planting on 19% more fields (p¡0.01), they also become more likely to state that they tried but were unable to hire the worker they invited to the training event. Consistent with poaching by other employers, the likelihood of trainer-employers rehiring the trained worker in treatment villages is lower than would be predicted by regular labor market churn.[8] Hiring data from spillover-employers in treated villages corroborates this, with these employers hiring 55% more days of labor to do row planting (p¡0.01), which enables them to adopt row planting on 24% more fields (p¡0.01). As a result of being trained, trainees earn 8.2% higher wages during the agricultural season (p=0.02) than trainees in control villages, evidence that some of the returns accrue to workers. We also find that training changes trainees behavior on their own farm, with trainees in treated villages planting 1.3 more of their own fields using row planting in treated villages (p¡0.01).

The returns generated by training for workers and non-training employers are economically meaningful.[9] Farm profitability increases by 10.8% (p=0.05) in treated villages: with the profits of spillover-employers and trainees increasing by 9.6 and 14.2 percent respectively (p=0.10 and 0.04).[10] Aggregating the increase in earnings in treatment villages and comparing it to the cost of training, we estimate a benefit-cost ratio of 2.4, suggesting that a dollar of training investment generates 2.4 dollars of returns. Almost half of this surplus, however, accrues to *spillover-employers*. These results suggest that firms' unwillingness to train generates meaningful distortions in these economies, and that the wedge between private and social returns potentially dilutes training investment incentives.[11]

We turn to our second experiment, the "labor guarantee" experiment, to test whether farmers would train workers if they could guarantee the ability to rehire the worker.[12] In addition to the large measured spillovers from training, additional survey evidence suggests

[8] We augment survey measures with field visits to randomly audited fields and find they are consistent, suggesting that the results do not reflect demand effects or reporting biases.

[9] The increase in earnings includes both labor market earnings and farm profitability. We measure farm profitability directly using farm revenues and subtract all input costs and labor costs. More details are provided in the relevant sections of the paper.

[10] The effect for trainees is driven in part by increased on-farm labor supply. Total labor market earnings for trainees in treated villages also increase by 20% (p=0.02).

[11] This is driven by the fact that non-training employers increase farm profitability and are a large share of the sample.

[12] This is also a test of whether farmers perceive positive returns to training conditional on capturing sufficient surplus.

that employers investments might be hindered by a reliance upon short term contracts. Employers state that most workers exhibit limited commitment in contracting: more than half of employers in control villages report that at least one worker reneged on an agreement to work for the employer during the past season. Contracting training investments is complicated by the fact that employers report it requires a long-term contract, with training needing to occur sometime before the planting-season (see Figure 1.2b). This makes it difficult to specify contract terms: since the exacting timing and quantity of labor required is hard to specify far in advance.

Given this motivating evidence we use the "labor guarantee experiment" to test whether employers become more likely to train workers if they believe they will be able to rehire them during planting. We recruit a second sample of trainer-employers and trainees in different villages using the same protocol as the spillover experiment. In a control group, trainer-employers and trainees are told that the trainee will receive an unconditional cash transfer during the planting season. In the labor guarantee treatment group, trainer-employers and trainees are told that the trainee will receive a cash transfer during the planting season only if the trainee returns to work for the farmer during the planting season.[13] We take several steps to make the treatment credible: we leverage our relationship with 1AF to build trust that the subsidy would be paid and screen workers by asking laborers to choose between the conditional payment during planting or a smaller amount of cash today. In addition, we facilitate contracting during the planting season by using our team of enumerators to facilitate and verify that the worker returns to work during the planting season, prior to any payment being made to the worker. After farmers are assigned to their treatment conditions, we then offer a training event in which trainer-employers in both conditions are invited to attend to train their workers, but we provide no other incentives to train, to measure how an increased likelihood of retaining the worker following training changes farmers' willingness to train.

We find that farmers receiving the labor guarantee treatment are substantially more willing to train workers. Trainer-employers in the labor guarantee treatment group are more than 50 percentage points more likely to attend the training event for half a day or more (a relevant benchmark as used to measure training in the spillover experiment). This effect is particularly concentrated among employers who state that they find it hard to get workers to show up reliably, and employers who believe it will be hard to get a worker to show up reliably after being trained. Training is a meaningful decision, with more than 90% of employers who train the worker hiring that worker and taking up the contract during the

[13]The amount of money offered per day –a top-up of around 70% of the daily wage– is designed to be not so large that it distorts hiring decisions meaningfully but large enough to credibly increase the probability that a worker will work for a particular employer.

planting season.

We explore several alternative explanations for the findings in both experiments. We show that the effects in the spillover experiment are unlikely to reflect salience or social learning, revelation of unobserved worker heterogeneity, or heightened signaling of the technology caused by the training event. Additionally, we find limited evidence that the treatment effects in the labor guarantee experiment can be attributed to demand effects, trainee investments in training due to eased liquidity constraints, or changes in wage bargaining dynamics.

Finally, we discuss why relational contracts do not emerge as a solution to the contracting problem. Relational contracts require the ability to punish the second party for defecting on any contract, but in this context we find that employers and employees have little social capital (likely reflecting their different social statuses), limiting the scope for punishments. Moreover, scarcity of skilled labor in villages creates competition among employers to hire trained workers, creating incentives for workers to defect on agreements, consistent with prior literature documenting that competition can narrow the contracting space (Macchiavello and Morjaria, 2023). The failure of farmers and workers in a village setting to contract these high return investments suggests that relying on decentralized transmission of new skills may be infeasible in environments with weak institutions and a lack of generalized trust. This points to a potentially important role for coordinated or centralized policies that lead to skill transmission, such as coordinated investments in training or, more generally, policies that incentivize individuals who have skills or the know-how to utilize new technologies to disseminate these further among the population.

The remainder of the paper proceeds as follows. Section 1.2 discusses the geographic and economic context of our project and describes our sampling frame. Section 1.3 provides a simple conceptual framework for our treatments and describes the main empirical predictions we test. Section 1.4 discusses the experimental design for the spillover experiment. Section 1.5 outlines the data and our empirical strategy for the spillover experiment. Sections 1.6 shows the paper's core results for the spillover experiment. Section 1.7 describes the design and results of the labor guarantee experiment. Section 2.8 discusses alternate explanations for our findings. Section 3.10 concludes.

1.1.1 Related Literature

This paper contributes to a literature on labor markets in developing countries. A key feature of these markets is that they are often organized via short-term contracts. Past work has highlighted potential advantages of short-term contracts–for example, the ability of the spot labor markets to flexibly respond in the face of shocks (Rosenzweig, 1988; Breza et al., 2021). However, there has been little empirical work studying possible distortions that may arise from short-term informal contracting.We highlight that this feature of developing country labor markets could be consequential for contributing to low labor productivity through a mechanism of low on-the-job learning by workers.

We contribute to a long theoretical and empirical literature on training investments in general skills and firm training impacts. While a substantial literature has long argued that firms are likely to underinvest in training their workers in general skills (Pigou, 1912; Becker, 1964; Acemoglu, 1997; Acemoglu and Pischke, 1998, 1999a,b), some empirical papers have shown that firms invest in general skills beyond a level that is offset by lower wages (Loewenstein and Spletzer, 1999; Autor, 2001).[14] A related empirical literature documents the returns to training interventions and highlights the barriers to firm investments in training in these environments (Card et al., 2011, 2018; McKenzie, 2017). Recent contributions to this literature document that firms in numerous low-income country contexts appear unwilling to invest in training employees (Alfonsi et al., 2020; Caicedo et al., 2022), and two recent papers suggest that worker-firm separation may limit training investments (Brown et al., 2022; Adhvaryu et al., 2023).[15] We contribute to this literature by constructing a clean test for worker turnover limiting firm training, while also demonstrating that training investment maybe too low because of returns captured by future employers.

This paper also speaks to a long literature on the barriers farmers face in adopting improved agricultural technologies in low-income countries. Failures in credit (Jack, 2013), risk (Karlan et al., 2014), and information (Foster and Rosenzweig, 1995) markets, as well as behavioral frictions (Duflo et al., 2011; Bridle et al., 2020) have been proposed as limits to adoption (see De Janvry et al., 2017, for a review). A well-documented literature has also explored the returns to training interventions of new technologies (Kondylis et al., 2017; Aker and Jack, 2021; Barrett et al., 2022; Islam and Beg, 2021). Only recently, however, have labor market frictions been proposed as a meaningful constraint to adoption (Jones et al., 2022).[16] We contribute to this literature by showing clean evidence for one labor market friction meaningfully distorting farmers' adoption decisions.

Finally, we contribute to a literature on social learning and information diffusion in agriculture (Griliches, 1957; Conley and Udry, 2010; Foster and Rosenzweig, 1995), as well

[14] Another strand of literature has shown theoretically and empirically the distinct mechanism that firms may underinvest in hiring novices because there is worker heterogeneity in type that is revealed through hiring. This generates an externality when firms hire workers and their type is revealed that also limits firms' incentives to hire (Pallais, 2014; Terviö, 2009).

[15] Brown et al. (2022) show that conditional financial incentives for firms can lead them to train more and generate large returns for trainees and suggests that worker separation from the firm may limit firm investments in training ex-ante. Adhvaryu et al. (2023) finds in an experiment that managers do not select to train employees with the highest returns and argues that this is because the employees with the largest returns are more likely to separate from the firm after training.

[16] An important related paper in a non-agricultural setting suggests another important labor market constraint limiting the adoption of profitable technologies in manufacturing (Atkin et al., 2017).

as a recent strand of literature that has considered which types of farmers are best able to diffuse agricultural information and under which conditions this diffusion is successful (BenYishay and Mobarak, 2019; Behaghel et al., 2020; Beaman et al., 2021; Chandrasekhar et al., 2022). We contribute to this literature by proposing a novel mechanism that might limit those who hold information related to new technologies from diffusing it more widely.

1.2 Context

In this section we provide a brief overview of the context of the experiment. Additional context is provided in Appendix A.3.

1.2.1 Agriculture in Burundi

We conduct our field experiment with farming households in Muramvya and Gitega provinces, Burundi, and collaborate with the NGO 1AF, which is active in these provinces.[17] The vast majority of the Burundian agriculture is rain-fed and, although the climate is favourable for production, yields are low.[18] Our experiment follows farmers over the course of one of the Burundi's agricultural seasons, over the course of which farmers prepare, plant, weed and harvest their fields (see Appendix Figure A.1.2 for a timeline). The task of planting is particularly time sensitive and labor intensive. Farmers plant their fields at the onset of rains, believing that there is a short window[19] –typically of around one to two weeks following the onset of rains– during which they must finish planting since the season is short.[20] Planting is also relatively time intensive as compared to other tasks required during the season, such as field preparation and weeding. Interviews with farmers and 1AF field officers suggest that they expect around 50% of total labor input for the season to be required during the relatively narrow window for this task.

1.2.2 Row Planting and Fertilizer Microdosage

We study the decision of farmers to train workers in and adopt two planting techniques: row planting and fertilizer microdosage. Row planting requires farmers to till the land and

[17]See Appendix Figure A.1.1 for a map.

[18]In 2018, average maize and bean yields were equal to approximately 1.53 tonnes and 0.66 tonnes per hectare, which are among the lowest yields in the world for these two crops (Ritchie and Roser, 2013).

[19]This belief that delays to planting limit yields is consistent with a long agronomic literature finding that planting delays reduce yields (Howard et al., 2003; Kruger, 2016).

[20]This message is reinforced by the government who sends “moniteurs agricoles” to villages giving farmers windows of time of typically around two weeks that they must complete planting by.

construct well ordered seedbeds, and then sow in parallel lines spaced by the same distance throughout the field, as well as to adjust how they apply complementary inputs, such as fertilizer and compost, to their fields. The microdosage of fertilizer is a complementary technology that requires farmers to apply fertilizer in a particular quantity and order, rather than broadcasting the fertilizer (Vandercasteelen et al., 2016). Traditionally, most planting is done by broadcasting, which is characterized by the semi-random broadcasting of seeds and inputs on farmers' fields.

Adoption of row planting rather than the broadcasting of seed has been found to substantially increase farm yields: agronomic studies in Rwanda, a similar context, find yield increases of 30-70% from spacing alone (Dusabumuremyi et al., 2014), while studies in other contexts find yield increases of 70-100% for other crops (Vandercasteelen et al., 2014).[21] Yield gains are thought to derive primarily from reduced plant competition for water and nutrients, increasing germination rates and chances of survival post-germination, reduced weeding requirements later in the season (Vandercasteelen et al., 2014; Mansingh J and Deressa Bayissa, 2018) and by increasing the yield response of the crop to other inputs, such as fertilizer (Vandercasteelen et al., 2020).

Despite the potential yield gains we find only partial adoption of row planting and fertilizer microdosage in our sample at baseline. Administrative data from agricultural season 2019B shows that 40% of their members' audited climbing bean fields bean fields were planted using row planting, while between 40-50% of fields applied fertilizer incorrectly. For other crops, non-adoption is even more acute.[22]. Moreover, for many farmers adoption does not appear to be constrained by knowledge: only 20.8% of trainer-employers in the spillover experiment control group use row-planting and microdosage on all of their beans plots, despite almost 80% planting using these techniques on at least one field.

1.2.3 Agricultural Labor Markets

Planting is labor intensive and while farmers rely primarily on family labor, around 20% of total labor input is hired for larger farmers.[23] Hired labor is hired from decentralized and informal labor markets, similar to casual labor markets in other LMICs (Breza et al., 2021). Similar to other African settings, high transport costs and distances between villages

[21] Non-experimental estimates of yield returns vary but are generally positive. Monitoring, Evaluation and Learning (MEL) data from 1AF in Burundi finds row planting increases bean yields by 40%, and by 60% when conducted in conjunction with microdosage (1AF, 2016).

[22] This sample includes both 1AF clients and non-clients, and it is representative at the agro-ecological zone level. These adoption rates are lower than in Rwanda, which has similar agroecological conditions

[23] Farmers in the spillover-employer control group in the experiment utilize on average 50 person-days of labor over the course of the agricultural season, see Table A.2.9.

mean that each village in our context comprise a local labor market, as has been found in other African contexts (Jeong, 2021; Fink et al., 2020).[24] Contracting is arranged bilaterally between employers and laborers, often with the employers visiting the households of various laborers, or with laborers visiting employers requesting jobs.[25] In the vast majority of cases, employers attempt to contact laborers in person 1-2 days prior to requiring their labor, and contract labor for just a few days. This style of search offers scope for workers to signal their skills to prospective employers, either by demonstrating their technique on their own fields near their house, or by showing how fields close to their households have been planted (if sufficient time since the onset of rains has passed and these fields were planted sufficiently quickly). After contact with the worker, employers and employees appear to bargain over wages, which depend on a variety of features including the task assigned and size of land required to prepare (Fink et al., 2020).

Despite the village setting, labor markets exhibit churn between seasons. Trainees in the control group report only supplying 30% of the days of labor that they provided in a given year to the same employer they provided labor to the prior year, while 35% of employers report hiring a worker who they had never hired previously each season.

1.2.4 Evidence for Missing Training Markets

At baseline only a thin market that exists for training casual laborers. Figure 1.2 documents that while more than 60% of employers stating that they are capable of training workers, only a small fraction of farmers report ever having done so.

Establishing a training market could be solved by employers or workers bearing the cost of training. In our setting, direct worker investments appear infeasible. Individuals working regularly as laborers are severely liquidity constrained: more than 70% at baseline have less than one day of wages cash on hand. Moreover, some workers also perceive firm moral hazard in training, stating that incomplete contracting means that they might invest in training but only be trained poorly by employers.

Given the constraints to worker investments, which are similar to those found in other settings, we focus instead on what limits firm's incentives to invest in training. A majority of employers state that training is limited by a lack of enforceable long-term contracts, meaning that after training the employer cannot guarantee that the worker would return. Farmers report that it is infeasible to train workers during the planting season, due to time constraints associated with planting quickly after the onset of rains. Training workers *before* the planting season, instead, appears to be limited by the perception that workers will not

[24]Within each village, more than half of households in engage in labor either as an employer, laborer or both, a proportion similar to other rural African contexts (Jeong, 2021).

[25]This nature of contracting is the same as found in Jeong (2021).

return after being trained, because they would work for another employer or would spend more time working on their own fields (see Figure 1.2b).

1.3 Conceptual Framework

This section provides a simple framework to illustrate the returns to training for training firms, workers and non-training firms in our environment. Given our finding that employers and workers bargain over wages (see Section 1.2.3) we base our framework on the bargaining model used by Acemoglu and Pischke (1999b), who model training investment incentives in a frictional labor market. We first illustrate that the level of training in such an environment maybe lower than the first best. We then use this framework to illustrate the effects of two shocks that we use to motivate our experimental treatments. First, we show a shock to the level of training changes returns for workers, training and non-training firms. Second, we show how a change to the separation rate in this environment changes firms' willingness to invest in training.

1.3.1 A stylized model of General Skills Training

We use the model of general skills investments with labor market frictions, such as those in Acemoglu (1997); Acemoglu and Pischke (1998, 1999a,b). These models demonstrate that, because of a compressed wage structure, firms might invest in training their workers in general skills. However the level of training firms invest in maybe less than the social optimum because some of the returns to training accrue to workers, or other firms.

The model consists of two periods. We assume there are three actors: a firm i who trains, a firm j who does not train and a worker. We assume that the productivity of the worker a function of their level of training, τ, and can be written as $f(\tau)$, with $f'(\cdot) > 0$ and $f''(\cdot) < 0$. For firms, we assume there is a cost of training, where $c'(\cdot) > 0$ and $c''(\cdot) > 0$. In period 1 production is normalized to 0 and the worker earns wage W. Firms choose to invest in a worker training level τ at a cost $c(\tau)$.

We assume that in period 2, there is a probability q that the worker is separated from the firm. In period 2, the worker can also quit and search for another job. If the worker quits or is separated, they are able to match with another firm with probability p_w. If the worker does not match with another firm, we assume that they earn nothing.

To match the stylized facts in this environment, we assume in this version of the model that only firms will make investments in general skills training (see Section 1.2.4).This assumption could be microfounded by assuming that the worker is severely liquidity constrained, or by assuming that in the environment contracts are incomplete and cannot enforce transfers from the worker to the firm in the event they leave the firm after being trained.

Given the search friction in our environment, there is positive surplus from a match and we assume that wages are determined with firms by nash bargaining. The parameter β captures the portion of rents that workers capture. Therefore the outside option of the worker in period 2 at the inside firm is $v(\tau) = p_w \beta f(\tau)$. There is wage compression if $p_w < 1$ or $\beta < 1$.

One can then show that the level of training chosen by the social planner and firm respectively satisfy:

$$(1 - q(1 - p_w))f'(\tau^1) = c'(\tau^1) \tag{1.1}$$
$$(1 - q)(1 - \beta)\left(f'(\tau^*) - (p_w \beta f'(\tau^*))\right) = c'(\tau^*) \tag{1.2}$$

Specifically training is lower than the first best because: i) workers capture some of the returns from training through an improved outside option $v(\tau)$, ii) workers capture some returns from bargaining over rents β and iii) other firms may capture some returns when the worker is separated from the firm (with probability q).

1.3.2 Spillovers from Training

We use the framework above to generate simple predictions regarding the incidence of the returns from training in response to the offer of a subsidy paid to firms to train workers. Specifically, consider a subsidy S paid to a firm conditional on offering a level of training τ_1 that is (slightly) higher than the market equilibrium level τ^*. We introduce the notation $\Delta f(\tau_1) = f(\tau_1) - f(\tau^*)$ to simplify notation going forward.

Firms train if

$$(1 - q)(1 - \beta)(\Delta f(\tau_1) - \Delta v(\tau_1)) + S \geq \Delta c(\tau_1)$$

and the total returns from an increase in training can be written as

$$= \underbrace{(1 - q)(1 - \beta)\Big(\Delta f(\tau_1) - \Delta v(\tau_1))\Big) - \Delta c(\tau_1)}_{\text{A - Profits of training firm}} + \underbrace{q p_w (1 - \beta)\Delta f(\tau_1)}_{\text{B - Profits of non-training firm}}$$
$$+ \underbrace{(1 - q)\big(\Delta v(\tau_1) + \beta\big(\Delta f(\tau_1) - \Delta v(\tau_1)\big)\big) + q p_w \beta \Delta f(\tau_1)}_{\text{C - Earnings of Trained Workers}}$$

Implying that the initial level of training was inefficient if:

$$= \underbrace{(1-q)(1-\beta)\Big(\Delta f(\tau_1) - \Delta v(\tau_1))\Big)}_{\text{A - Profits of training firm (excluding training cost)}} + \underbrace{qp_w(1-\beta)\Delta f(\tau_1)}_{\text{B - Profits of non-training firm}}$$
$$+ \underbrace{(1-q)\big(\Delta v(\tau_1) + \beta\big(\Delta f(\tau_1) - \Delta v(\tau_1)\big)\big) + qp_w\beta\Delta f(\tau_1)}_{\text{C - Earnings of Trained Workers}} \geq \underbrace{\Delta c(\tau_1)}_{\text{D - Cost of Training}}$$

This leads to our first two predictions:

Prediction 1. *Training generates spillovers to those not incurring the training cost.*

Prediction 2. *Training is underprovided.*

Specifically, the earnings of trained workers rise as the wage rises, and the combination of a positive separation rate and wage compression entail that other firms capture some of the returns of a given firm's training when they hire these. Moreover if these returns, plus the returns of the training firm, are larger than the cost of training, then the level of training in the economy is inefficiently low.

Finally, the implicit function theorem can be used to generate the following prediction:

Prediction 3. *An increase (reduction) in the separation rate decreases (increases) firm training investments.*

In a frictional labor market in which employers capture returns from training when matched with a skilled worker, reducing the probability of separation increases the expected surplus from training.

1.4 Spillover Experiment: Design and Implementation

1.4.1 Design Overview

Two of the three experimental predictions require observing the returns to training under the labor market status quo. To test these predictions we use our first experiment to measure whether, consistent with a labor market where training is underprovided, we observe that when employers train workers this leads to returns captured by others (the worker or future employers, as per Prediction 1), and whether these returns are economically meaningful, meaning that training is underprovided (as per Prediction 2). Our conceptual test requires us to induce some employers to train in some labor markets, and not others, and then measure who in that labor market captures the returns, among the employer who trains, the worker who is trained, and future, non-training employers.

The primary variation that we use to test this is to provide financial incentives to employers in some labor markets, and not in others, to train workers. To the extent that this training is effective, and generates an exogenous increase in the amount of skilled labor in the village, and then measure the outcome for all labor market participants in treated and untreated labor markets to measure the incidence returns.

1.4.2 Implementation

To observe employers' training decisions, we hold training events in a set of Burundian villages. In each village, the training event broadly follows the protocol of a 1AF training event, which are held regularly to train 1AF members in the community. We recruit two sets of farmers to attend these training events (for more details see 1.7.4 below). The first set of farmers are trainer-employers, who are individuals in the village who regularly hire labor, know how to row-plant and do microdosage, and were willing to attend the event for a known financial incentive. These trainer-employers are then asked to identify a trainee, an individual in the village who works regularly as a laborer and who they would be willing to hire, who does not know row-planting and who is not related to the employer. Farmer and trainee pairs are asked to attend the training event, which consists of the provision of parcels of land and equipment to train. The key difference between our training event and 1AF led events is that in our event we provide incentives for employers to train trainees (rather than the training be a led by a 1AF staff member) away from other pairs.

During the first day of the event, the employer receives a fixed financial incentive to attend the event, and train the worker on an unrelated technology. Employers in different labor markets are then randomized to receive different incentives to return the next day and train the worker in row-planting:

Control - Selected employers in control villages receive an unconditional financial transfer of around 1.5 days wage. Field staff suggests that these farmers that they could train a laborer in row-planting and micro-dosage techniques.

T1—Financial Incentive Treatment - Selected employers in treatment villages receive the same information as in the control group. They are then told that they will receive a financial incentive of around 1.5 days wage conditional on training a laborers in row-planting and micro-dosage techniques.[26]

Interpretation: The spillover experiment increases the returns to training for farmers by offering financial incentives to do so. If the training is effective, this generates an increase in the stock of skilled labor in the village. It can then be measured who captures the returns from this training in this labor market, among the employer who trains, the worker who is trained, and other (non-training) employers in treated villages.

After treatment conditions are explained, farmers are free to continue training or leave. Importantly, during the course of the event staff do not explain how to train, or train participants themselves.[27]

1.4.3 Sampling

We conduct the experiment in 80 villages ("sous-collines", in Burundi) in two provinces in Burundi.[28] We use villages as our unit of randomization because geographic distances and high transport costs mean that each village defines a local labor-market. Because villages typically have 200-300 households, this offers a major advantage for measuring the returns from training, since we can identify both training employers, as well as non-training employers who might feasibly be impacted by treatment, allowing us to quantify the externalities to other employers from training.

Village selection. To obtain a sample of villages for the spillover experiment, we utilize 1AF administrative data to enumerate the villages in two provinces near to 1AF headquarters.[29] We then screen out villages that were unreachable by vehicle during the planting season, villages where 1AF had fewer than 20 clients or the village was deemed to be particularly small, villages where individuals did not primarily derive their livelihood from farming or the farming of beans in season B in particular.[30] In total this left 120 villages eligible for the study. We randomly order these villages and enroll villages based on this ordering. Within each village, we sample four groups of individuals: i) trainer-employers, ii) trainees,

[26] Specifically, farmers are told that the financial incentives is conditional on the farmer spending at least half a day with the laborer, at our training location. Because of the timing of the event, this generally required the farmers to return for a second day to finish training the laborer.

[27] Event staff do monitor the training to ensure that i) each pair remains separate from others ii) that farmers who are eligible to receive the financial incentive and wanted to claim it do engage with their trainee. Farmers at the beginning of the event are told they will not receive payment for just standing around, for instance.

[28] Our original sample for the Spillover Experiment consisted of 92 villages of which 12 were assigned to a T2-Labor Insurance Treatment. This treatment is discussed in Appendix 1.7.

[29] These provinces are Murumvya and Gitega.

[30] This last criteria ruled out a number of villages close to towns.

iii) spillover-employers and iv) spillover non-employers.

Sampling of individuals from villages. We recruit three sets of individuals from villages, employers who are provided financial incentives to train workers, non-training employers and trainees (a visualization of how employers are (see Figure 1.1).

Recruitment of trainer-employers. 1AF Field Officers (FOs) recruit trainer-employers, who consist of clients of 1AF in the village. At the time of this screening, the 1AF Field Officer was not aware of the treatment status of the village they were working in. Within each village, the FO was trained on a recruiting protocol for 1AF clients to serve as trainer-employers. Specifically 1AF clients were screened on, i) whether they knew row-planting and microdosage and ii) whether they regularly hired casual labor in the labor market. Each FO was instructed to bring 20-30 of these employers to the training event.

Recruitment of laborers. Eligible trainer-employers were then asked to bring a laborer to the training event. The laborer had to meet several criteria. First, the individual had to be someone who regularly supplied daily labor during the planting season. Second, the laborer had to be an individual that the trainer-employer would themselves be interested in hiring. Third, the laborer had to not know row planting and fertilizer microdosage. Fourth, the laborer could not be a member of the trainer-employer's household.

Eligibility of trainer-employers and laborers. At the training event, we conducted a second screening of trainer-employers and laborers based on the criteria above and pairs where the trainee was found ineligible were screened out of the project. A list of trainers and trainees was provided to the team prior to the event and this list was randomized. We then sampled approximately the first 18 of the eligible trainer-employers at the training event and first 12 of the eligible laborers to take surveys.

Recruitment of spillover households. In the spillover experiment we also randomly sampled three other sets of households in each village (see Panel b of Figure 1.1 for a diagram of groups sampled from the village). We randomly sample around 18 households who regularly hire labor during the planting season (spillover-employers), who were uninvolved in our training event. In addition, we also sample around 8 households per village who do not hire labor during the planting season: four that regularly supply labor, but are uninvolved in our training event, and four that are unengaged in the labor market. These proportions do not reflect the proportion of each type in the village: instead we oversample employers to have power to detect spillover effects on hiring and adoption, while sampling non-spillover-employers gives us some ability to examine mechanisms. These households were sampled using a random walk methodology.

Randomization. Randomization is at the village (local labor market) level. A simple randomization was conducted because time constraints made it impossible to collect substantial data about the village or sample prior to the first event, details on balance of the sample is provided in Section 1.5.3.

1.5 Data and Empirical Strategy

1.5.1 Data Collection

We measure outcomes for the spillover experiment over the course of the agricultural season, through enumerators' direct observation of the training event, self-reported survey data and audits of farmers fields. We measure hiring, employment and technology adoption through surveys during the planting season, at harvest, and during the planting season one year following the intervention.[31]

At the training event, enumerators measure the amount of time farmers spend with their trainee on a plot of land for each of the two days of the training event. In addition, to measure whether training translates into better skills in these techniques, we ask laborers to perform an incentivized practice test of the seeding technologies after the training event.

During the planting season, we conduct surveys with both employers and workers to measure labor market and farming activities. To construct measures of hiring, we ask each farmer whether they hired workers, and then ask for each worker hired i) the days that the worker worked, ii) the tasks completed in these days and the days spent completing the row planting or fertilizer microdosage during those days and, iii) payments made. In addition, questions about days worked and tasks completed were also asked for each family member that worked during the planting season. To measure changes in employment, we ask each farmer the number of employers that they worked for during the planting season. We then ask for each employer, i) the number of days they worked for this employer, ii) the payment received and iii) the tasks completed. Finally, individuals were also asked whether they did any other work during Season B, and total earnings from such work.

During the same survey, we ask farmers whether they adopted row-planting and fertilizer microdosage on their fields, and use field audits in order to validate responses. Prior to beginning of the survey, enumerators demonstrate to participants what correct spacing of fields consists of using tape measures and verbal descriptions, and then tell them that, at the end of the survey, one of their fields will be audited randomly to test whether their

[31] See Figure 1.1 for an overview of when data is collected.

description of the field lines up with how it is planted. Enumerators then elicit for each field, i) whether it is planted using row planting or broadcasting, and whether microdosage was adopted, and ii) for which proportion of the field. The core outcome is then the number of fields in which the majority of the field is planted using these techniques. To validate the survey responses, survey staff visit at least one field per farmer. Plots were selected randomly for each visit, and in each visit the survey team measures i) whether the field was seeded using row planting or traditional seeding practices and ii) conditional on using row planting the distances between rows and pockets at three randomly located points on the field. We design this audit to incentivize truthful reporting of how farmers plant their fields during the plot roster described above, as well as construct a non-self reported measure of adoption.

Finally, after harvest we conduct an additional survey measuring harvest outcomes. To do this for each crop we ask each farmer the quantities harvested and its price. We construct crop revenues by multiplying quantities of crops harvested by the price of the crop at the nearest market.[32] To measure profits, we subtract from this measure all other non-labor input costs, and subtract these and labor input costs.

1.5.2 Timeline

The experiment follows the Burundian Agricultural Season B, which runs from February to July —with most of the planting activities concentrated early in the season (see Section 1.2.1 and Appendix Figure A.1.2 for details). We conduct training events during December 2021 and January 2022. We measure hiring of daily laborers, adoption (self reported and field visits) and agricultural employment in a first visit between March and May 2022. We survey farmers on harvest outcomes and family labor during a second visit between July and September 2022. We conduct a follow up survey between June and August 2023 that measures planting outcomes one year on, as well as measuring the harvest of crops that were not ready for harvest in the initial harvest survey.

1.5.3 Summary Statistics

Table 1.1 presents descriptive statistics and a test for balance for the sample. Columns 1-3 show the means, standard deviations and a test of equality for treatment and control trainer-employers, while columns 4-6 and 7-9 present the same summary statistics for spillover-employers and trainee workers. The sample appears balanced with 3 of 60 tests exhibiting

[32]Because crops come in different varieties with different prices, we multiply the quantity harvested by the farmer's estimate of the price of that variety at the nearest market. We also show robustness to using the median price of respondents in the same area.

p<0.1, and balance is achieved on many important outcome variables, including wages, labor market and farm earnings. Spillover-employers in the treatment group do however adopt row planting techniques on slightly more fields at baseline, hence we control for this in regressions.

Laborers look different from employers along multiple dimensions. As has been found in other contexts, households of laborers are considerably poorer than households not supplying labor, having less land (15.7 versus 40-50 ares) and savings.[33] They are on average 34 years old and are almost equally likely to be a man or woman, which might reflect the fact that men or more likely to outmigrate for work during the planting lean season (Vinck, 2008). By construction, they are less knowledgeable of the agricultural technology, being less likely to have planted in lines in the past season, or in the past 5 years, and show low knowledge scores on a quiz about the techniques. Finally, almost all supplied labor the past season, of which they supplied on average of 11 days.

Employers do not select into training randomly, which leads to some important differences among the trainer-employer and spillover-employer samples. One difference arises from the screening criteria - trainer-employers exhibit high scores on the knowledge quiz of the technology and have all used the technology previously. By contrast, only 80% of spillover-employers previously used the technology, and they exhibit generally lower scores on the knowledge quiz. While both trainer-employers and spillover-employers are substantially larger farmers than laborers, spillover-employers are on average larger than trainer-employers, having 50 ares of land and 6.5 fields as opposed to trainer-employers 40 ares of land and 6 fields, with this likely reflecting the higher opportunity cost of time for such employers to attend the training event. Finally, while both groups are equally likely to hire, spillover-employers generally hire more days of labor than trainer-employers: hiring 29 days of labor as opposed to 20 days.

1.5.4 Empirical Strategy

We use the following empirical specification to measure Intent-to-treat (ITT) effects in the Spillover Experiment:

$$y_i = \beta T_{v(i)} + \gamma X_i + \varepsilon_i \tag{1.3}$$

where y_i is the outcome of individual i, $T_v(i)$ is a dummy for the treatment status of village v (a function of the individual whose outcome is being measured), and X_i is a vector of baseline controls for individual i. Standard errors in all regressions are clustered at the village level. In this specification, β measures the average treatment effect of being an individual in a treatment village on outcome y_i.

[33]Fink et al. (2020) find workers more likely to provide ganyu (labor during the planting season) if they are liquidity constrained or among the lowest asset quintile.

We run unweighted regressions for the specification above for each group: 1) trainer-employers, 2) trainees and 3) spillover employers. We also use inverse probability weighting to estimate the total impact of the intervention across the three groups.

When running heterogeneity analyses, we use the following specification:

$$y_i = \beta_1 T_{v(i)} + \beta_2 Het_i + \beta_3 T_{v(i)} \times Het_i + \gamma X_i + \varepsilon_i \quad (1.4)$$

where Het_i is a measure of heterogeneity.

1.6 Spillover Experiment: Measuring who Captures Returns

In this section we present the results of the spillover experiment, which speaks to Prediction 1 (that training generates spillovers to other actors uninvolved in training themselves). We first show that financial incentives in treatment villages substantially increase trainer-employers' willingness to train. We then show how this changes the employment and adoption of row-planting of trainees, before turning to how this changes the hiring and adoption of row-planting of trainer-employers and spillover-employers.

1.6.1 Willingness to Train and the Stock of Skilled Labor (First-Stage)

In response to the offer of financial incentives, trainer-employers in treatment villages become more likely to train their paired trainee in T1-Financial Incentives villages, as compared to control villages.[34] Almost 80% of trainer-employers in the T1-Financial Incentives villages train the laborer they brought to the training event, as compared to less than 1% of trainer-employers in the control group (see Figure 1.3).[35] Spending time training may not translate into a meaningful change in the skills of workers if farmers engage in training in limited ways, are incapable of teaching or laborers are unwilling or unable to learn. To test whether training translates into meaningful changes in the skills of trainees, we ask trainees to perform

[34] We measure whether the employee is trained by whether the trainer-employer spends at least half a day training the trainee at our event

[35] This does not imply that no training occurred in these villages, rather that few farmers passed the threshold of training required from farmers in treatment villages to qualify for incentives. Moreover, this does not rule out that in control villages farmers may have trained these laborers at times/places not observed by enumerators.

a timed, incentivized row-planting task, which measured laborers' ability to complete well spaced row planting under time pressure, the skill that employers in these villages primarily hire.[36] As Figure A.1.3 shows, trainees in T1-Financial incentives villages perform better in this task post training: we reject the null hypothesis that the treatment and control score is the same at the 1% significance level. Taken together, these results are suggest that training results in an immediate increase in the skills of some laborers in row planting and microdosage.

1.6.2 Changes to Trainee employment and farming decisions

Prediction 1 states that training generates returns for those not bearing the cost of training. This requires that workers use the skills they are trained in, and may capture some of the returns of training through higher wages. Table 1.2 provides evidence consistent with this hypothesis, by documenting the treatment effect of being in a T1-Financial incentives village on trainees' employment, earnings and wages. Trainees use the skills they are trained in: column 1 shows that trainees in treated villages are employed for 3.43 days more days doing row-planting or microdosage than trainees in control villages (p-value<0.01). This is a sizable change, with trainees substituting toward row planting work for almost half of the days that they work during the planting period (which totals around ~ 8 days, Column 3). Trainees primarily substitute away from other work tasks rather than increase the number of days worked, as total days of employment for trainees in treated villages increases by 0.8 overall (an increase of 13%), an increase is not significant at conventional levels. There is no impact of training on the *extensive* margin of employment during the planting season: around 84% of the trainees do any agricultural work during the planting period in both treatment and control villages (Column 5, p=0.98).

Prediction 1 states that returns to training are captured by those not bearing the cost, and these returns might include increased wages for workers trained. We find evidence consistent with this hypothesis: among the sample of trainees supplying *any* agricultural labor–, wages increase by 8.2% in treatment villages (p=0.02). This wage increase, coupled with a slight change in employment, leads however to a total earnings increase by almost 20% (p=0.02). Finally, column 7 documents an additional channel though which trainees might capture the returns from training: adoption of row-planting on their own fields. Specifically, we find that trainees in treatment villages plant 1.3 more fields using the correct row planting techniques as compared to only 0.23 among trainees in control villages (p<0.01).

[36]Laborers were scored on the number of rows and pockets planted at correct distances within a short period of time.

These results provide evidence consistent with trainers failing to capture the returns to training, due to workers' wages rising, and potentially through them using row-planting on their own fields.

The effectiveness of the training appears to be due to changing the beliefs of workers, providing them with technical information about how to row plant, as well as allowing them to practice and receive feedback on row-planting, lowering the cost of effort of doing so.[37] Survey evidence from a sample of laborers in other villages suggests that many laborers do not perceive that there are returns to the adoption of row-planting on their own fields. The substantial increase in the likelihood that trainees adopted row planting on their own fields (Column 7 of Table 1.2 suggests trainer-employers during training may have changed trainees' beliefs about the technology and about its profitability on their own farms. Additional evidence, however, suggests that the training transmitted meaningful changes in knowledge and skills, rather than simply changing the beliefs of trainees. We find that trainees in treatment villages have better technical awareness of details of the row planting process, as measured by a knowledge quiz that we administered during subsequent surveys, with this greater awareness persisting for almost a year after training, suggesting that training also transmitted technical information to trainees. Finally, the fact that there is a change in the performance of trainees in the incentivized planting task after training, suggests that during the training their ability to do row-planting under time pressure improved. Anecdotally, it seems that trainees found most helpful the process of practicing row-planting while receiving feedback from trainer-employers.

1.6.3 Who Hires Newly Skilled Labor? Measuring Hiring Spillovers Following Training

Trainer-employers' Hiring and Technology Adoption. Prediction 1 states that trainer-employers may not capture returns from training because other employers hire newly trained workers. We find that trainer-employers in treatment villages may not hire as much of their trainee's labor as they desire following training. Column 1 of Table 1.3 measures the likelihood that a trainer-employer attempted to hire the trainee that they invited to the training event.[38] While 57% of employers in the control group attempt to hire the trainee that they invited to train, 72% of employers in the treatment group do the same. This increase likely reflects the fact that unskilled labor is easier to find and more interchangeable than skilled labor. Column 2, however, shows that trainer-employers in treatment villages are *less* likely to

[37] Prior literature has found mixed results on the effectiveness of agricultural training interventions. See for instance, Kondylis et al. (2017); Udry (2010); Aker and Jack (2021).

[38] This data was collected in a follow up survey with a random subsample of trainer-employers.

successfully hire their trainee, suggesting that after being trained, it becomes harder to hire these workers. Trainer-employers in treatment villages are 16 percentage points more likely to state they were unable to hire their paired trainee ($p<0.01$). Columns 1 and 2 combined suggest that trainer-employers are unable to hire trainees 41% of the time conditional on trying to hire them, as compared to 55% of the time in treatment villages.

Trainer-employers in treatment villages do hire more labor to do row-planting, however. Trainer-employers increase by 46% the number of days that they hire to do skilled tasks (an increase of 0.81 days on a base of 1.76 days, $p<0.01$, Column 5). Partially this increase reflects more days of hiring the trainee that they themselves trained (0.52 days increase (Column 3)) while partially it reflects a spillover involving hiring trainees of other trainer-employers at the training event (Column 4, which shows that there is a 0.84 day increase in hiring any trainee at the training event to do the skilled labor task). This increase suggests that trainer-employers do capture some of the benefits of their own training, while also benefiting from the spillover created by living in a treatment village. Table A.2.2 shows that this hiring in the wage labor market translates into a net increase in total labor (family and hired) for row-planting, with trainer-employers in treatment villages using 10% more days of labor (family and hired) for row-planting on their fields ($p=0.09$). Increased hiring facilitates more adoption of row-planting: trainer-employers in treatment villages plant 18.6% more fields using improved row spacing practices than in control villages (an increase of 0.46 fields from a base of 2.47 fields, Column 6, $p<0.01$).

Spillover-Employers' Hiring and Technology Adoption Prediction 1 states that trainer-employers may not capture returns from training because some trained workers work for other employers, and potentially generating returns for these employers. Table 1.4, which shows treatment effects for spillover-employers on hiring workers to do row planting, and adopting row-planting on their fields, provides evidence consistent with employers capturing returns from training. Spillover-employers in treatment villages hire 55% more days of labor to do row-planting (an increase of 1.18 days, $p<0.01$ (Column 2)), with all of this increase in hiring driven by the hiring of trainees who were invited to the training event (Column 1). Table A.2.2 shows that this hiring in the wage labor market does not substitute for family labor, with spillover-employers using 11% more days of labor (family and hired) for row-planting on their fields ($p=0.09$).

The magnitude of the treatment effect on hiring for spillover employers is comparable or larger than that which they find for trainer-employers. This is possibly driven by the non-random selection of employers to train. Appendix Table A.2.4 provides some evidence that treatment effects are driven particularly by larger farmers, who are more likely to be in the spillover-employer pool. In this table, we run treatment effect heterogeneity regressions for the cross section of employers by land size and the number of days they previously hired, both proxies for the size of the farmer, and both characteristics that are imbalanced across trainer and spillover-employers. We find that the treatment effects are larger for both hiring and adoption among larger employers, and some evidence that they are also larger for employers who hire more days of labor. Therefore, it seems plausible that the magnitude of the treatment effects in the spillover group is driven partially the fact that this group comprises larger farmers.

Spillovers to other employers might reflect regular labor market churn, or may reflect that trainees become *more* likely to work for other employers once trained (for instance, because of "poaching" by other employers). Figure A.1.5 provides suggestive evidence that employers may find it harder to retain workers after being trained. The blue bar in the figure shows that trainees in the current season work on average only 30% of the days for an employer they worked for in the previous season. The red bar in the figure shows the treatment effect on the number of days of labor provided by trainees doing row planting (about 3 days, Column 1 of Table 1.2). The green bar multiplies these two quantities to predict the number of days trainer-employers would be expected to hire workers to do row-planting given this rate of turnover, which slightly more than one day. However, the observed treatment effect on trainer-employers hiring of their own trainee to do row planting is much lower, at around half a day (yellow bar). This suggests that trainer-employers inability to hire their own trainee after training might not reflect regular labor market churn, and could be due to additional sorting or poaching of workers in response to being trained.[39]

We find that the increase in hired labor facilitates the adoption of more row-planting on farmers fields. Spillover employers in treatment villages plant 0.45 more fields using an increase of 23.7% (Column 3 of Table 1.4, p<0.01).

1.6.4 Spillovers to Farm Profitability from Training

The prior sections demonstrate that hiring, wages and adoption of row planting change in treatment villages. In this section, we measure the returns to training, specifically changes to farm profitability and labor market earnings, for trainer-employers, trainees and spillover-employers. To measure farm profits, we complete crop rosters of harvest sizes for all crops planted during the season, and multiply these quantities by local prices for these crops. We then subtract from this the cost of all labor and non-labor farm inputs during the season.[40]

[39]One caveat related to this result is that we did not ask laborers the number of days that they worked for an employer of the previous year in the first season of the project– therefore the blue bar is constructed from data from year 2 of the project. Therefore an alternative interpretation of this figure is that rate of turnover among trainees in the control group was much higher in year 1.

[40]Non-labor farm inputs include spending on fertilizer, seed, compost, pesticides, land rental payments and other input costs. Given thin land markets and the likely noisy measures of land acreage measured at

Given the challenges associated with valuing family labor, we compute farm profits with two bounds: i) a bound that values any family labor at a wage of 0 and ii) a bound that values family labor either at the individual's own wage, if observed, or at the value average wage observed in the village. We show treatment effects on these outcomes for each subgroup, as well as regressions weighted by the inverse sampling probability for each group, to obtain the average treatment effect for those in the sample.[41]

Consistent with Prediction 1, we find that training generates returns for workers and non-training employers. Table 1.5 shows treatment effects on farm revenues and profitability for trainer-employers, trainees and spillover-employers. Farm revenues increase by 8% on average (Panel A, p=0.08), including 10.6% for trainees (Panel B, p=0.06), 7% for trainer-employers (Panel C, p=0.09) and 8% for Spillover-Employers (Panel D, p=.11). These increases in farm revenues make sense given that the adoption of row-planting should increase yields if applied correctly. This increase in revenues leads to increased profits for all three groups, as we do not find meaningful changes in input costs for these groups, which leads profits in all three groups to rise (see Appendix Figures A.1.4a-A.1.4c for CDFs showing raw effects). Assuming a shadow value for family labor of zero, we find that farm profits increase by 10.8% on average for the sample (Panel A, Column 2, p=0.05). This average effect is made up of a 14.2% increase in profits for trainees (Panel B, Column 2, p=0.04), 9.2% increase in profits for trainer-employers (Panel C, Column 2, p=0.09) and 9.6% increase in profits for spillover-employers (Panel D, Column 2, p=0.10). This result is only sensitive to the assumed shadow family value of labor for trainees: while the profit magnitudes are largely the same when assuming that family labor is valued at the market wage rate for trainer-employers and spillover-employers, estimates of farm profitability for trainees becomes statistically insignificant for trainees when family labor is valued at the market wage, reflecting the increased time that this group spends on their own farm after being trained. We also find that trainees labor market earnings increase (Column 4), while there is no treatment effect on trainer-employers and spillover-employers labor market earnings, which is unsurprising that they do not regularly work for others in this period.

Given the noise in measuring farm profitability, we also use a second measure of whether farmers perceive positive returns to hiring and adoption, by measuring persistence in hiring and adoption decisions, a revealed-preference measure perceived returns. We see persistent behavioral changes across all three samples, although there is some evidence of weaker treatment effects a year after the training. Figures 1.4-1.6 display the core treatment effects for each trainer-employers, spillover-employers and trainees for two agricultural seasons: the

baseline, we do not measure the implicit rental value of land used as an input cost

[41]Note that this is not the same as village level effects since these regressions leave out individuals in the village who are uninvolved in the labor market, as well as other laborers. To the extent that there are negative or null effects on these individuals, ATE for the village as a whole will be lower.

season immediately following training, and the season 12 months later. Trainees in Financial Incentive villages remain continue supplying more days of labor utilizing the technologies and and applying the trained techniques on their own fields, although the magnitudes of these treatment effects are lower than in the first agricultural season. Similarly, Figure 1.5 shows that trainer-employers and spillover-employers adopt the techniques on 0.23 and 0.28 more fields than their control counterparts (p=0.01 and p=0.02) and hire 0.98 and 1.28 more days of labor to conduct the trained techniques, respectively. These persistent changes to behavior suggest that the training generates surplus, at least for some members of the population.

1.6.5 Training Underinvestment and the Incidence of Returns

Prediction 2 in Section 1.3 states that if skills are general, the market may provide too little training in such skills. To test whether there is underprovision, we compare the aggregated benefits from our training to the cost of provision and measure whether the ratio is larger than one. The ideal measure of the benefit of training would be the the aggregated willingness to pay of individuals in treated villages for the training to happen, with prior papers estimating this quantity using consumption changes (Bandiera et al., 2017). Since we do not have a measure of consumption, we instead measure the benefits of training as the change in earnings for trainer-employers, spillover-employers and trainees as a result of treatment (see Section 1.6.4), which maybe a reasonable proxy for consumption given that we find no changes in output prices. Two further assumptions are required for this proxy to be reasonable: first it should be that the increase in earnings is not driven by a large change in days worked (which would lead to an additional disutility of labor) and that the training does not yield another direct utility benefit or cost for some subpopulations (Hendren and Sprung-Keyser, 2020). We discuss the sensitivity of the estimate to these assumptions later in the section.

We compute the benefits of training as the treatment effects on total income (farm profits and labor market earnings) for each subgroup (trainer-employers, trainees, spillover-employers), assuming a discount rate of 10%. In line with the two year treatment effects on adoption, we assume that these earnings benefits persist for two seasons but depreciate at a rate of 50% (approximately in line with the magnitude of depreciation we observe for adoption over two years). We then aggregate these benefits according to the proportion of each group in the village.[42]

We compute the total costs of training by summing the cost of incentives for participants, the cost of training materials and the opportunity costs of time for participants in the training. This includes costs of equipment, such as the cost of land and equipment rental

[42]This corresponds to weights of approximately 25% for trainees, 25% for trainer-employers and 50% weights for spillover employers.

(16% of training costs) and financial incentives paid to trainers to train the trainees, and to trainers and trainees to attend (22% of total training costs). We also include an opportunity cost of time for the trainee (conservatively assuming 1 day of time) as well as the opportunity cost of time for the trainer to find the trainee (again, conservatively assuming that this takes 1 day) as well as train the trainee.[43] We value the time of the trainee at the average wage in the control group, and the value of time of the trainer at the average wage in the treatment group. Together these account for 14% of the training costs. Finally, we include the cost of staff time that was required to advertise the event, provide invitations, prepare and monitor the training (48% of total training costs). We assume that there are no fiscal externalities as a result of training.

Table 1.6 Panel E shows the Benefit-Cost ratio, aggregating the returns across these populations. In the first column, as is common in many training programs, it is assumed that the returns to training accrue only to the trainee - that is only the earnings accrued by trainees are their own fields and while working for others, are counted as benefits. In Column 2, added to these benefits are the additional farm earnings that the training generates for the trainer-employer as well. Finally, Column 3 adds to this total the additional farm earnings accrued by Spillover Employers in the same village.

We find large returns to training investments, consistent with training being underprovided by the market. Incorporating the returns to training for the trainee only delivers a benefit cost ratio of 0.5, and we can't reject the null that returns to training are less than or equal to one[44] (p=0.99). Incorporating the returns for the trainer as well increases the Benefit-Cost ratio to 1.2 (p=0.33). Including the returns to spillover-employers, however, increases the Benefit-Cost ratio to 2.4, suggesting that a dollar of training investment returns 2.4 dollars of returns (p=0.12). This large increase in the returns between column 2 and 3 is driven by the fact that trainer-employers and spillover-employers accrue similar returns to training, but there are a larger number of spillover-employers in the village population. The returns to training remain large even if we assume there are no returns to trainees, and if we assume the earnings benefits only last one year.

This benefit-cost ratio may not correctly measure the welfare impact of the training program for two reasons. First, it might overstate the welfare gains from training if training generates a labor supply response. If this were the case, then the willingness to pay for the human capital generated by the training would equal the change in wage generated from the training multiplied by the individual's days worked pre-training, i.e. the welfare metric nets out the portion of additional earnings arising from changes in days worked (Kline

[43] We include both the financial incentive paid to the trainer and the opportunity cost of time of the trainer to be as conservative as possible, as the the former may not reflect the latter if the employer perceives private returns to training.

[44] A benefit-cost ratio larger than one is a reasonable proxy for welfare improvements.

and Walters, 2016). However we do not observe statistically significant changes in days worked in response to training (Table 1.2 and Appendix Table A.2.2). Second, this table might also overstate the gains from training if there are losses from training from others in the community (for instance, other skilled laborers who lose earnings, or others who face additional competition for inputs). Given our sampling strategy, we cannot say much about these potential losses but it seems unlikely they would be sizable enough to erase the gains from training.

Taken together, these results suggest that there is underinvestment in training general skills, and moreover that the returns to training are large for employers, hard for any particular employer to capture. In the next section we test whether this spillover to other employers limits training investments.

1.7 Labor Guarantee Experiment: Design and Results

1.7.1 Motivating Evidence that separation limits training

Section 1.6 documents that when farmers train this generates positive total returns, but limited returns that are captured by the training employer.

We now test whether farmers would train if they could guarantee capturing a larger proportion of the returns from training, by ensuring that the worker returned following training (Prediction 3). This hypothesis is motivated by several stylized facts in our setting that suggest that it is hard to write contracts guaranteeing that a worker will return after being trained. First, employers state that laborers exhibit limited commitment in even very short term labor market contracting. Panel A of Figure 1.7 shows that almost half of employers in the spillover experiment control group report that at least one worker who they contacted workers to work for them during the planting season reneged on the agreement and did not show up to work as agreed. This points to quite severe constraints in enforcing even short term contracts during the planting period in which there is high demand for on farm labor, given that the modal employer contacts workers to work for them only several days prior to needing them to work. Enforcement of such contracts relies on the ability of employers to punish workers for defecting on contracts, and this in the enviroment is complicated by employers inability to observe the reason that workers do not arrive for work, and social norms that limit the ability of employers to sanction workers for reneging on contractual agreements for reasons outside of their control. While the most workers tell employers that they did not arrive because they were sick or had an accident (excuses that social norms prohibit employers from punishing workers for reneging on agreements), even though most employers believe (but cannot prove) that the worker took another job or worked on their own farm (see Panel B of Figure 1.7).

Finally, the training problem is further complicated by it requiring a long-term contract. Figure 1.2b shows that 70% of employers state that training cannot be done during the planting season itself due to the employer having no time, meaning that training must occur before the planting season. This further complicates the training contract, as uncertainty regarding the timing of rainfall, demand for inputs and timing of planting make it hard for employers to state precise contract terms (such as when the employee should return).

Given this suggestive evidence, we hypothesize that farmers do not train because the reliance on short-term contracts means that they cannot guarantee they will appropriate sufficient surplus from training to recoup the cost. We therefore use the Labor guarantee experiment to test Prediction 3 of Section 1.3, that employers train if the likelihood that workers separate after training reduces.

1.7.2 Design

To test if farmers become willing to train when they can guarantee that workers will work for them after being trained, we provide incentives for workers to work for particular employers during the planting season. We ask employers to identify a worker they would be interested in hiring who is unskilled in row-planting, and then offer the employer/worker pair one of the two following incentives, randomized at the individual level:

Control - The farmer and their laborer are told that the laborer will receive an unconditional financial transfer at planting time equal to approximately 65% of the daily wage for two days.
Labor Guarantee Treatment - The farmer and their laborer are told that the laborer will receive the same financial incentive as the laborers in the control group, *conditional* on returning to work for the farmer for two days during the planting season for two days of the farmer's choice.

This offer consists of several other features. First, we tell employers and laborers that our enumerators will be present in the village during the planting period, and will only make payments to workers if they observe the worker in-person working on the farmer's field. Second, employers and workers are told that our enumerators will also facilitate the worker coming to the employer, by checking with the employer when the worker is needed, and relaying this to the worker. Finally, we ask workers if they prefer this conditional payment in the planting season, or a smaller payment today, to understand if they are serious about this, and use our staff's relationship with 1AF FO's to make the offer credible.

Interpretation: The labor guarantee experiment increases the returns to training by making it more likely that workers return to work for employers after being trained. If farmers in this treatment become willing to train after receiving the contract, this suggests that they perceive positive returns to training, conditional on being able to appropriate the returns.

1.7.3 Sampling and Implementation

We sample an additional 6 villages that had not been chosen for the spillover experiment from the same list utilized in the spillover experiment. We then recruit trainer-employers and trainees following the same protocol: Appendix Table A.2.1 shows Balance for trainer-employers and trainees in the Labor guarantee experiment.

Within each village, we invite farmers and laborers to come to a central location to be surveyed. At this location, we screen participants according to the criteria in Section 1.4.3. We first privately explain the financial incentive to the worker that they would receive for working for the employer as part of the Labor guarantee experiment, and ask them if they prefer to receive this incentive or a small financial incentive today. We then survey trainer-employers and trainees (laborers) and explain to each their treatment conditions. Those in the Labor Guarantee Treatment are also told that our team will facilitate the matches during the planting season (communicating with workers when they are needed) and verify all work, so that payments to the workers will be conditional on our team verifying them working for the employer. We then tell respondents that a training event will be held the next week for several days, and that the event will consist of the provision of land, and equipment, that could be used to teach trainees planting techniques. We also tell respondents that we will not provide any training ourselves, and therefore that if trainers and trainees want to do training they would need to come as a pair. Enumerators also tell respondent that each pair would be separated from others (i.e. there would be no group training). Finally, we explain to all respondents that there are no other benefits of attending the training event and that not attending the training event would not influence the participant's participation in the study. Randomization in this experiment is conducted at the individual level.

1.7.4 Data Collection and Timeline

The Labor guarantee experiment follows the Burundian Agricultural Season A. We conduct baseline surveys, and described contracts to farmers in July 2023. We offer the training events in July and August 2023, at which time enumerators measure the amount of time farmers spend with their trainee on a plot land for each of the two days of the training even. We implement the labor guarantee in September, at which time enumerators visits farmers/households fields to observe laborers working. Finally we measure employment, hiring and farming outcomes for the Agricultural Season in October 2023.

1.7.5 Results

We find that employers in the Labor guarantee treatment become more willing to train workers knowing that the worker will return to work for them during the planting season (test of Prediction 3). Panel A of Figure 1.8 shows that farmers in the Labor guarantee treatment are 57 percentage points more likely to attend the training event for more than three hours, the minimum recommended amount of time required to train the worker (p¡0.01) . This suggests that in response to a higher likelihood of capturing training returns, employers perceive a positive return to training.

Given the large effect, we explore whether employers concerned by their ability to capture the gains from training respond to treatment. At baseline, we ask employers whether they contact workers to work for them and they do not show up to work, or whether they are concerned that if they train a laborer then the laborer will be unlikely to show up to work for them. 55% of employers agree with one of these two statements, and we classify these employers as having concerns about ”appropriation”. Panel B of Figure 1.8 shows that treated farmers who are concerned by appropriation are far more likely to train in response to treatment (an increase of 76 percentage points) as compared to farmers unconcerned by appropriation (an increase of 30 percentage points), implying that those primarily concerned by workers returning largely drive the effect.

Training appears to be effective and a meaningful decision: 90 percent of employers who train workers subsequently hiring them. Taken together, this evidence suggests that, at least for a subset of farmers, training would be more likely to occur if the labor market was structured so that employers perceived they could capture more returns from training.

1.8 Alternative Explanations

In the preceding sections, we interpret our findings as showing that inducing employers in some villages to train generates large returns captured by those in the labor market not incurring the cost of training, that this leads to overall welfare gains and that employer’s training investments are limited by a reliance on short term contracts, making it insufficiently likely that workers return to employers after being trained. In this section we consider alternate interpretations of this story.

1.8.1 Spillover Experiment

We interpret the magnitude and incidence of returns that we estimate in the spillover experiment to approximate the returns to training that farmers would accrue outside of experiment.[45] In this section we discuss alternate channels that our effects that might run through that would distort these estimates.

Learning/Social learning. Was there social learning of row planting and microdosage that enabled others in the village to increase adoption of these techniques without hiring skilled labor, magnifying the returns and spillovers from training beyond what would typically be estimated? This might have occurred because the training event signalled the importance of these skills to the trainer-employers, trainees, or others in the village, who then shared or sought this information from others. While we cannot rule out the existence of knowledge spillovers, several pieces of evidence point against this being the primary mechanism for our treatment effects. First, Panel B of Appendix Table A.2.3 shows that there was no change adoption or hiring for spillover non-employers, suggesting that such spillovers only appear to be impactful for those engaged in the labor market. Second Appendix Table A.2.5 shows that while spillover-employers who had not previously row planted drove some of the adoption treatment effects for this group as a whole, we see similar magnitude of effects also for spillover-employers *who previously used the techniques*, and where knowledge spillovers would be expected to have more muted impacts. Finally Panel A of Appendix Table A.2.3 shows heterogeneity in extensive margin adoption decisions for spillover-employers by several proxies for whether they previously used the techniques. This test is useful as one might expect if there were knowledge spillovers for this set of users to begin adopting without additional labor input. By contrast, we see instead that these farmers only adopt these techniques if they also hire labor (columns 4 versus 5 and 6 versus 7) suggesting that the hiring enabled the adoption. Finally, it should be noted that our event was probably not that unique in the village since 1AF at least annually has large group based trainings with its members which would also signal (potentially) the value of these skills.

Signalling. When designing the spillover experiment, to ensure a first stage we decided that it was crucial for us to be able to monitor the training to ensure there was no gaming of the training for the financial incentive. Monitoring the training was not feasible at the individual farms of each farmer, and therefore we used the training event as a way to ensure that we could monitor training at scale. The fact that a large training event was held in the village might have made it easier for trainees in the treatment to credibly signal to others that they had acquired the skills they were trained in. This fact might accentuate the magnitude of

[45]With the exception being that our experiment in some villages captures GE effects.

the spillovers that we find relative to a situation in which a farmer individually trains a worker on their own land. This seems unlikely, however, given that employers state that one of the reasons that they do not train is because the worker will use the skills elsewhere, suggesting that employers believe that workers have the ability to signal the skills to others once trained (see Figure 1.2).[46]

Revelation of Trainee Type. While not a feature of our experiment that would distort our measures of returns, an alternative channel for our findings could be that the training event served not to transfer human capital from trainer to trainee, but instead gave trainees a chance to reveal their type to employers. In this story, the training event still generates externalities, but through a mechanism of learning similar to Pallais (2014) or Terviö (2009). Three pieces of evidence suggest this is not the primary explanation for these findings. First, this requires this information to be shared with the spillover group by trainer-employers, which seems unlikely given that this would make it harder for the trainer-employers to hire back these workers. Second, something must be transmitted to laborers, given that laborers in treatment villages perform better in the incentivized planting task, and change their planting decisions on their own farm. Finally, this concern might be more prominent among trainees who did not previously work for their employer, as presumably this would create more scope for employer learning about worker type. Appendix Tables A.2.6 and A.2.7 show treatment effect heterogeneity for the core labor market outcomes by whether the employer and employee previously worked together, and we do not see meaningful evidence that the treatment effects are driven by trainees who did not previously work with their employer.

1.8.2 Labor Guarantee Experiment

We interpret the training response to training in the labor guarantee experiment as reflecting higher perceived returns to training conditional on receiving the contract. In this section we discuss alternate interpretations:

[46]Workers typically signal skills by demonstrating how to row plant on their own fields, to employers at the time that employers contact workers during the search process. Moreover, in subsequent seasons, farmers report learning from workers' fields if they are able to use the techniques. Finally, one fact that is particular to our context is that workers can use the skills on their own farms after being trained, and this on-farm work might require more time than working on the farm with traditional techniques. This increased on farm labor maybe an additional channel through which employers limit training investments that doesn't necessitate signaling skills.

Demand Effects. Willingness to train in this treatment may simply reflect experimenter demand. To alleviate this concern, as discussed in more detail in the preceding paragraphs, we did extensive audits of whether workers worked *and whether farmers themselves paid these workers* and found that they did so. Specifically, 90% of farmers that we classified as having trained the worker hired them as part of the contract, suggesting that the training was a meaningful decision.

Trainee Liquidity Constraints. Since the Labor guarantee experiment involves the a payment (conditional or unconditional) to the worker, an alternate interpretation of the results is that as a result of the treatment, workers paid trainer-employers to train them. This seems unlikely given that, a) the money was paid to the worker later in the planting season, b) money was paid unconditionally to control laborers, meaning that they could replicate the contract. This would mean that somehow this contract was only feasible for workers who received a contract to return to the farmer.

While this seems unlikely, we measure this directly by asking whether side payments were made from the worker to the trainer, and do not find any evidence for this. Specifically, at the time of contract implementation we asked workers if they made direct payments to the trainer to train them and zero said that they did.

Changes to Wage Bargaining as a Result of the Subsidy. It is possible that the subsidy paid to employers changed bargaining dynamics - for instance employers may have appropriated a share of the subsidy paid to employees via pass-through. Given that the subsidy was paid unconditional on training, it is unclear why this would lead to training, if the pass through of the subsidy is just a level shift in the profits of the trainer-employer. To induce training, it would need to be that the pass through of the subsidy to the employer was increasing in the level of training of the worker. It is unclear why one would expect this to occur, unless the subsidy was being passed through to the employer by the worker as an implicit form of payment.

Do we see large changes to the wage paid by the employer as a result of the subsidy? During the planting season, the average wage paid by an employer to a laborer is $\sim$ 2900 FBU. We observe slightly lower payments made to workers during the contract: the average payment made to a worker is 2625 FBU. This suggests the possibility of a subsidy pass-through rate of 10%. However, it is unlikely that this pass through is an implicit form of payment by trainees in return for training. The magnitude of the pass-through appears to be small, and would suggest that in return for training employers are willing to forgo a minimum of a half a day of a wage today to obtain 20% of the wage several months in the future.

1.9 Conclusion

In this paper, we show three facts consistent with a model where there is underinvestment in general skills training. First, we show that when farmers train workers, this generates returns for other employers, and that there may be underinvestment in training general skills. To show this, we conduct a village-level RCT that provides monetary incentives to farmers to train workers. We show that these incentives are effective in this context at generating skill transfer: workers trained in this sample work for other employers using these techniques and adopt the techniques on their own fields. We show that the returns to this training however, accrue not only to employers who train workers: while training employers employ more trained labor, other employers in the same labor market also hire this labor. Profits increase for both training employers and spillover employers. These training investments generate an increase in total social surplus but only one quarter of this surplus is captured by the employers who train workers themselves. Of the surplus not captured by employers who train, two-thirds is captured by other employers, and one-third by other workers.

We then show that a shift of the training surplus to training farmers induces farmers to train. To show this, we design a contract that increases the perceived probability that workers will return to work for farmers after being trained. We then show that this induces training and skill transfer: farmers are fifty percentage points more likely to train workers after receiving this contract.

This paper provides a complementary policy lever that can be used to facilitate the transmission of this information. Many models of social learning implicitly model the flow of information as happening passively, as farmers observe neighbors' fields and experiments with new technologies (Griliches, 1957) and has spawned a literature that has considered the optimal network members to diffuse information to. However, many costly policies such as agricultural extension, have had somewhat muted impacts in diffusing new technologies (Udry, 2010; Krishnan and Patnam, 2014). This paper, in the spirit of BenYishay and Mobarak (2019) and Chandrasekhar et al. (2022) assumes that information flow is not frictionless, and the transmission of information require effort on the part of those that hold it. Similar to their paper, we find that farmers may have insufficient incentives to transmit knowledge, and that policies that facilitate the diffusion of this information may have large multiplier effects.

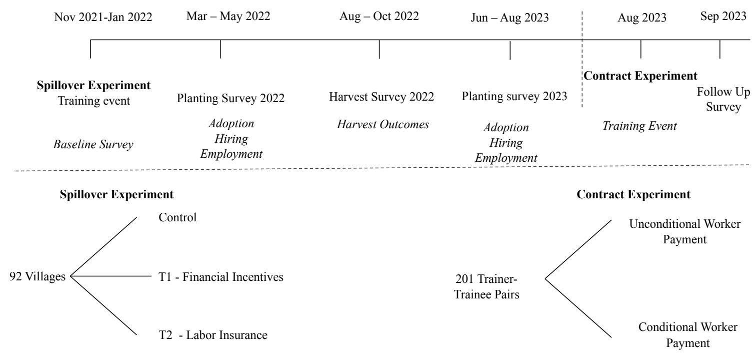
This paper also contributes to the discussion of technology adoption in LMICs. Facilitating the uptake of better agricultural technologies is a key policy objective in low income countries, and particularly in Sub-Saharan Africa, where yields have remained relatively low despite technologies existing that could conceivably increase farm returns (Bank, 2007). In this paper, we suggest that one mechanism that might contribute to underadoption are labor market frictions that lead farmers to underinvest in training workers in these new technologies. We provide evidence that the failure of employers to appropriate the returns to training may lead to meaningful underinvestment in new technologies.

Finally, this paper provides some evidence for "poaching externalities", or returns to training captured by other employers who do not incur the cost of training themselves. As discussed in a sizable theoretical literature, the existence of these returns could make standard contracting solutions to general skills investment decisions less likely to lead to first-best training investments (Acemoglu, 1997; Stevens, 1994). This fact, coupled with mounting evidence suggesting that those who seek information may face large costs in acquiring information, could justify more aggressive interventions in training markets.

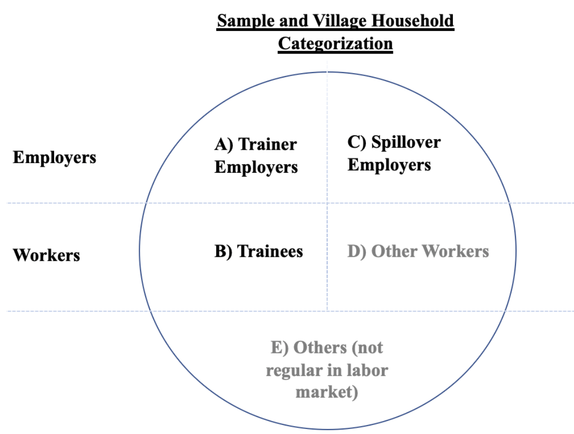
Many important questions remain, including which policy prescription, changes to contract structure, subsidies to training investments or otherwise, is optimal in such situations. Moreover, additional evidence is needed on how experience with new technologies interacts with incentives to diffuse information, and the effectiveness of farmer led training.

1.10 Tables and Figures

Figure 1.1: Experiment timeline, design, and sampling frame



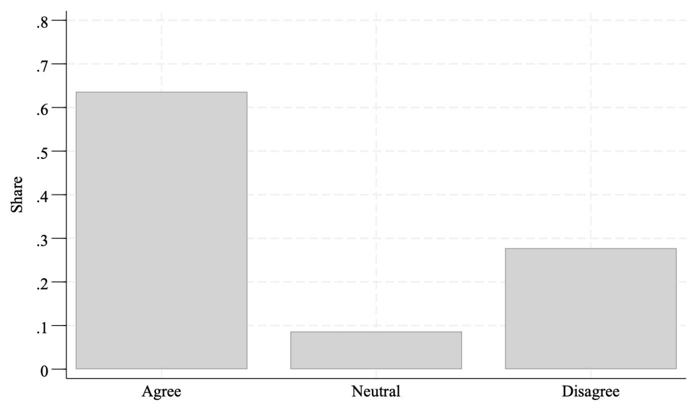
(a) Timeline of field operations and data collection, and sampling for each experiment



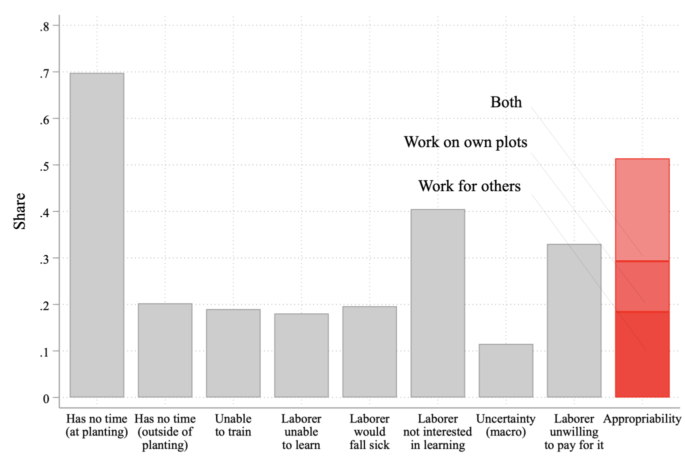
(b) Categorization of village households

Notes: This figure shows the experimental design and timeline of the intervention and surveys, as well as the primary outcomes measured in each survey. The second image shows different types of households in villages, and which households in particular we sampled for our project.

Figure 1.2: Motivating evidence: agricultural employers' beliefs about their ability to train, and reasons for not training



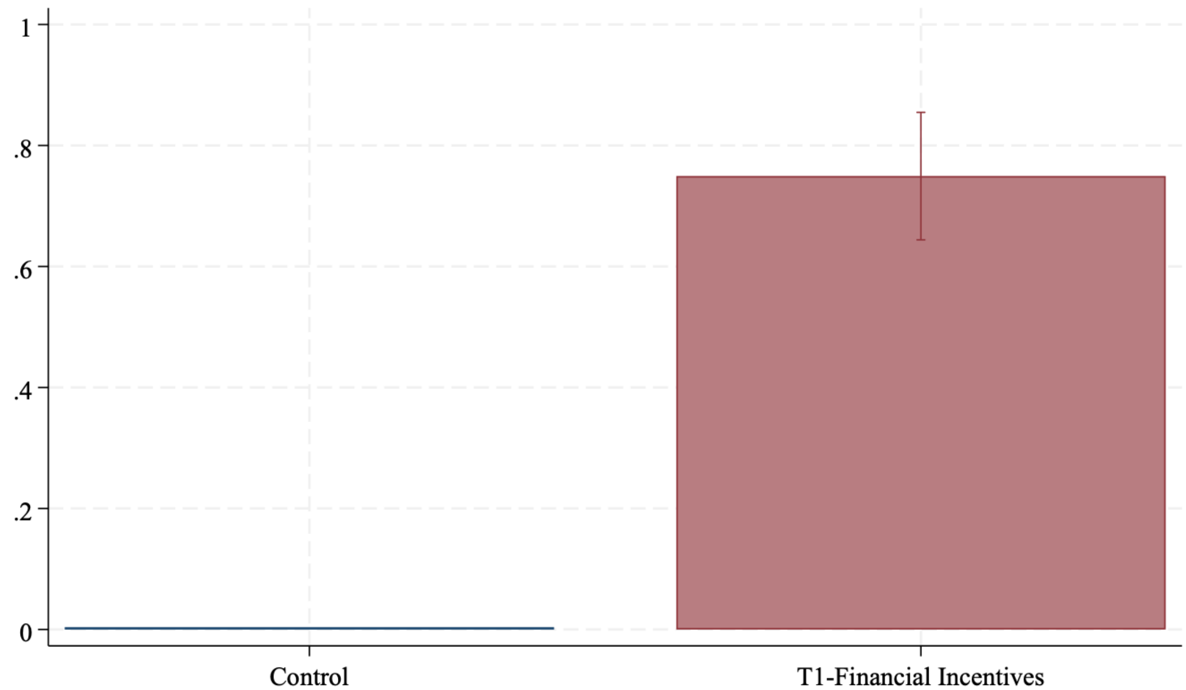
(a) Share of Employers agreeing that they are capable of training



(b) Proportion of Employers stating reason for not training

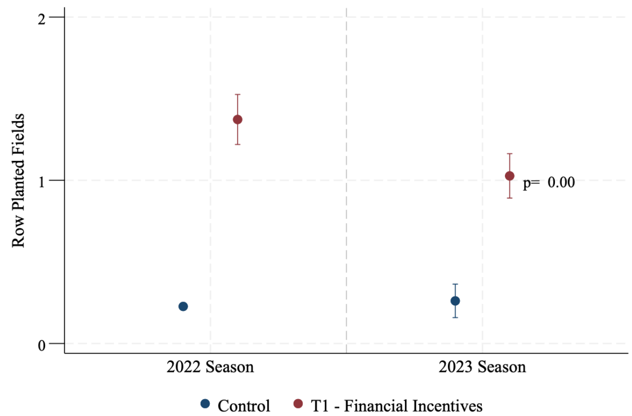
Notes: The first figure shows the share of agricultural employers who agree or disagree that they are capable of training workers. The second figure shows farmers' stated reasons for not having trained workers previously. The "Appropriability" bar in the second figure indicates the share of farmers stating that if they train the worker would go and work for others, would spend more time on their own fields and so not return, or both. In the second figure, farmers could state multiple reasons. The sample comprises a random sample of farmers regularly employing laborers, from villages outside the main experiment.

Figure 1.3: Spillover Experiment: proportion of trainer-employers training the paired worker in the Control and T1-Financial Incentives villages

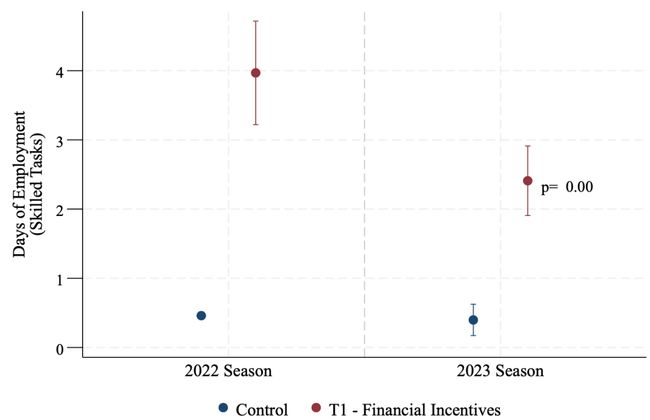


Notes: This figure shows willingness of trainer-employers to train their trainee in the different treatment arms of the Spillover Experiment. We plot the share of trainer-employers who trained according to our definition of spending at least 180 minutes supervised with the laborer. In the Control group, trainer-employers received an unconditional financial incentive. In T1-Financial Incentive villages farmers received a financial incentive conditional on this definition of training. Standard errors are clustered at the village level are displayed.

Figure 1.4: Spillover Experiment: trainees two season treatment effects on the adoption of row planting on own fields, and days worked in tasks involving trained techniques



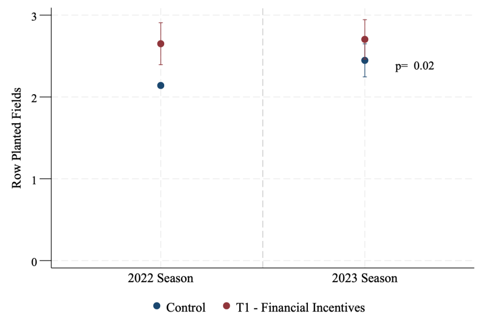
(a) Adoption of trained techniques (fields)



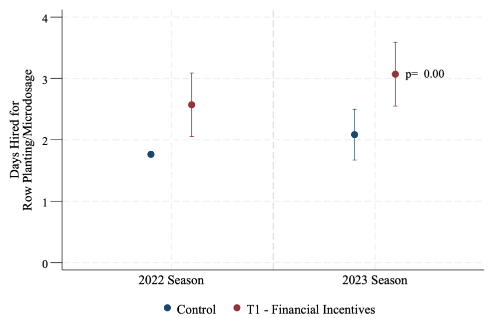
(b) Days of employment in trained techniques

Notes: This figure shows the number of fields on which row planting was adopted, and the number of days employed for tasks involving the trained techniques for trainees in the T1- Financial Incentives and Control Villages. Data is shown for two seasons: the 2022 season and the 2023 season. P-values for a test of equality of control and treatment coefficients in the second season are displayed.

Figure 1.5: Spillover Experiment: Trainer-employer two season treatment effects on adoption of trained techniques and skilled labor hiring



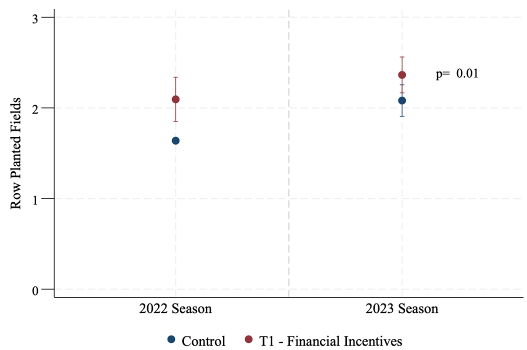
(a) Adoption of row planting (fields)



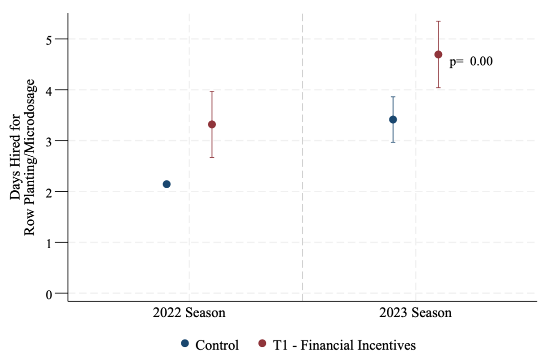
(b) Days of labor hired to do row planting/microdosage

Notes: This figure shows the number of fields on which row planting was adopted, and the number of days that labor was employed to do row planting/fertilizer microdosage, for trainer-employers in the T1-Financial Incentives and Control Villages. Data is shown for two seasons: the 2022 season and the 2023 season. P-values for a test of equality of control and treatment coefficients in the second season are displayed.

Figure 1.6: Spillover Experiment: Spillover-employer two season treatment effects on adoption of trained techniques and skilled labor hiring



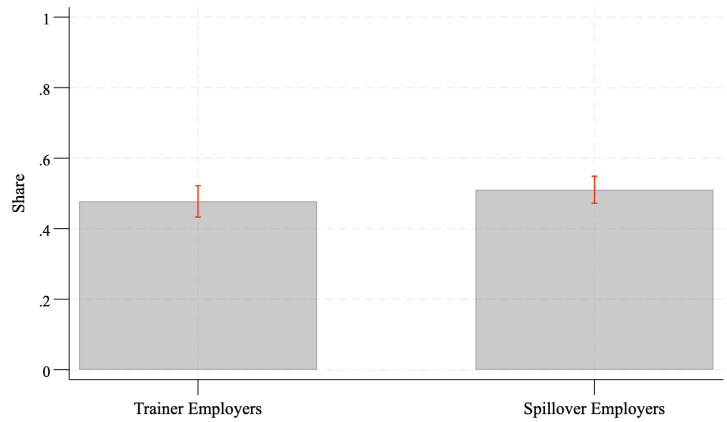
(a) Adoption of row planting (fields)



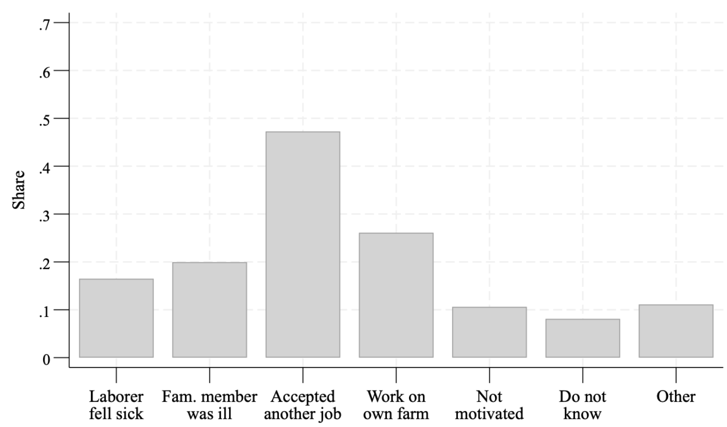
(b) Days of labor hired to do row planting/microdosage

Notes: This figure shows the number of fields on which row planting was adopted, and the number of days that labor was employed to do row planting/fertilizer microdosage, for spillover-employers in the T1-Financial Incentives and Control Villages. Data is shown for two seasons: the 2022 season and the 2023 season. P-values for a test of equality of control and treatment coefficients in the second season are displayed.

Figure 1.7: Motivating evidence: share of agricultural employers reporting contracting failures, and reported reasons for why a contracted laborer did not show up to work



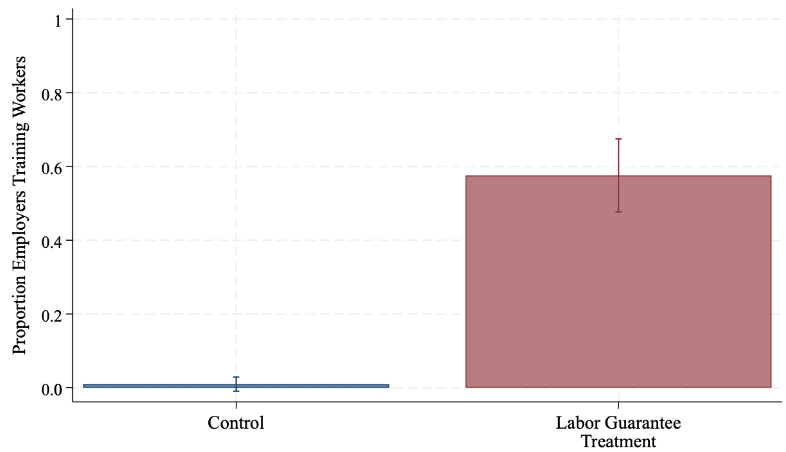
(a) Share of employers reporting that a worker they contracted to work did not show up when expected.



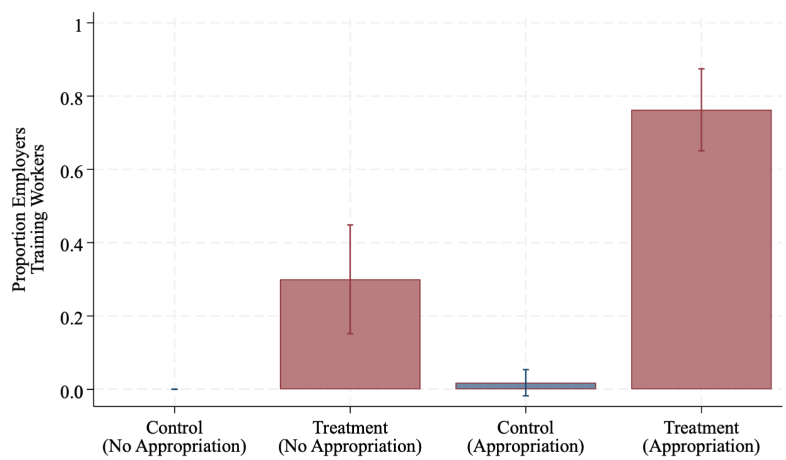
(b) Reported reasons for why the worker did not show up for an agreed-upon job.

Notes: The first figure shows the proportion of control group employers (trainer-employers and spillover-employers) who stated they contacted a worker to work for them, but she did not show up. This data is taken from the first planting season following the intervention. The second question asks the reason that employers believe the worker did not show up. This question was asked during the harvest survey to all the respondents who reported trying to hire at least one laborer during the previous Season B (regardless of whether they managed to hire one or not), but had at least one laborer who agreed to work for them yet did not show up to work.

Figure 1.8: Labor Guarantee Experiment: Impact of the contract on the trainer-employers' willingness to train and trainees' skill upgrade



(a) Share of trainer-employers attending the training event for at least half a day



(b) Proportion of trainer-employers training by concerns for appropriability

Notes: The top figure shows the proportion of farmers in the Labor guarantee experiment who attend the training event that we held for at least 3 hours. The bottom figure shows the proportion of employers who train in treatment by stated baseline concern for appropriability (whether trainee would return after being trained). Robust standard errors are shown.

Table 1.1: Spillover Experiment - Balance table

Variable	Trainer-Employers Control Mean (1)	Trainer-Employers Treatment Mean (2)	Trainer-Employers P-value Δ (3)	Spillover-Employers Control Mean (4)	Spillover-Employers Treatment Mean (5)	Spillover-Employers P-value Δ (6)	Trainees Control Mean (7)	Trainees Treatment Mean (8)	Trainees P-value Δ (9)
Demographics									
Age	43.7	43.6	0.89	47.0	47.6	0.58	34.0	34.8	0.36
	(11.1)	(11.5)		(13.9)	(14.1)		(11.0)	(11.3)	
No Education (binary)	0.3	0.3	0.34	0.3	0.4	0.23	0.3	0.4	0.39
	(0.5)	(0.5)		(0.5)	(0.5)		(0.5)	(0.5)	
Primary (binary)	0.6	0.6	0.30	0.6	0.5	0.40	0.6	0.5	0.68
	(0.5)	(0.5)		(0.5)	(0.5)		(0.5)	(0.5)	
Household Size	5.0	4.9	0.23	4.5	4.5	0.97	3.8	3.9	0.48
	(1.9)	(2.0)		(2.1)	(2.2)		(1.9)	(1.9)	
Unmarried	0.0	0.0	0.68	0.0	0.0	0.68	0.2	0.2	0.32
	(0.1)	(0.1)		(0.2)	(0.2)		(0.4)	(0.4)	
Male	0.6	0.6	0.44	0.5	0.5	0.10	0.4	0.4	0.87
	(0.5)	(0.5)		(0.5)	(0.5)		(0.5)	(0.5)	
Savings (log)	9.9	10.0	0.31	10.2	10.3	0.74	7.6	7.8	0.31
	(2.3)	(2.1)		(2.3)	(2.2)		(3.1)	(3.1)	
Baseline farm and knowledge									
Land (plots)	6.6	7.1	0.09	7.4	7.6	0.57	3.5	3.7	0.31
	(3.3)	(3.7)		(3.9)	(3.9)		(1.6)	(1.9)	
Land (ares)	43.0	45.2	0.46	51.9	55.0	0.47	15.7	17.0	0.53
	(38.8)	(37.5)		(46.8)	(51.2)		(21.9)	(19.5)	
Staples value (Past B)	142045.8	133083.8	0.31	148776.7	145203.1	0.69	34549.1	38635.1	0.26
	(94817.2)	(89028.2)		(100787.3)	(98655.2)		(28110.4)	(44164.8)	
Fields (Past B)	6.1	6.3	0.33	6.6	6.6	0.94	3.5	3.7	0.38
	(2.7)	(2.8)		(3.4)	(3.3)		(1.7)	(2.0)	
Row Planted Fields (Past B)	3.6	3.4	0.24	3.0	2.6	0.03	0.2	0.2	0.76
	(2.0)	(1.8)		(2.6)	(2.0)		(0.8)	(0.8)	
Planted in lines (Past season)	1.0	1.0	0.34	0.8	0.8	0.31	0.1	0.1	0.90
	(0.1)	(0.1)		(0.4)	(0.4)		(0.3)	(0.3)	
Previously used improved technology	1.0	1.0	.	0.8	0.8	0.28	0.1	0.1	0.91
	(0.0)	(0.0)		(0.4)	(0.4)		(0.2)	(0.2)	
Practice usage (past 5 years)	3.0	2.9	0.79	2.2	2.0	0.26			
	(1.8)	(1.8)		(2.0)	(1.9)				
Z Score Knowledge	0.7	0.8	0.18	0.3	0.3	0.81	-1.1	-1.2	0.13
	(0.6)	(0.6)		(0.9)	(0.9)		(0.4)	(0.4)	
Labor Market									
Past supply	0.2	0.2	0.38	0.2	0.2	0.27	1.0	0.9	0.27
	(0.4)	(0.4)		(0.3)	(0.3)		(0.2)	(0.2)	
Mandays supplied	2.0	1.7	0.45	0.6	0.7	0.80	11.7	11.1	0.56
	(4.3)	(4.3)		(2.5)	(2.8)		(9.0)	(10.1)	
Wage Past Season							2454.2	2355.1	0.16
							(544.7)	(483.5)	
Labor Market Earnings							28934.3	26484.5	0.37
							(23969.2)	(26333.3)	
Past demand	0.9	0.9	0.60	1.0	1.0	0.77	0.2	0.2	0.07
	(0.3)	(0.3)		(0.2)	(0.2)		(0.4)	(0.4)	
Mandays demand	20.0	20.6	0.71	29.0	33.0	0.12			
	(22.6)	(18.7)		(27.5)	(30.0)				
Observations	590	638		718	789		432	451	

Notes: This table shows the balance between all groups of farmers in the Spillover experiment. Variables were collected at baseline and correspond to the entire sample. In columns (1) (4) and (7), we show the mean and standard deviation for the variable of interest among the Control group. In columns (2) (5) and (8), we show the same for the Treatment group. Finally, in columns (3) (6) and (9) we show the p-value for a T-test of equality of means. The p-values of the difference in the means test come from SEs clustered at the village level.

Table 1.2: Spillover Experiment: Trainee employment, earnings and technology adoption

	Trained Techniques		Agricultural Work				Own Fields
	Total Days	Total Earnings	Total Days	Total Earnings	Worked	Wage	Modern Fields
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T1 - Financial Incentives	3.45	8555.15	0.86	3612.24	0.00	191.20	1.33
Train	(0.37)	(945.19)	(0.53)	(1441.92)	(0.03)	(77.31)	(0.09)
	[0.00]	[0.00]	[0.11]	[0.01]	[0.96]	[0.02]	[0.00]
Control mean	0.46	983.61	7.33	16852.41	0.84	2303.99	0.23
Obs.	848	848	848	848	848	710	848

Notes: This table shows the effect of the T1-Financial Incentive treatment on the employment and earnings of trainees –laborers selected by trainer-employers and invited to attend the training event. The dependent variable in Column 1 is the total number of days the laborer worked during the agricultural season doing work that involved the techniques taught during the training (row-planting and fertilizer microdosage). The dependent in Column 2 are the earnings from work that involved techniques taught in the training. Columns 3 and 4 are days worked and earnings from any kind of agricultural work during the agricultural season. Column 5 is an indicator variable for whether the laborer did any agricultural work during the agricultural season. Column 6 is the average wage earned by the trainee. Column 7 is the number of fields that the trainee planted using row planting. Standard errors are clustered at the village level.

Table 1.3: Spillover Experiment: Hiring & technology adoption among trainer-employers

	Tried Hire Paired Trainee	Unable Hire Paired Trainee	Paired Trainee Skilled Task	Trainees Skilled Task	Hired Skilled Task	Adoption Total
T1 - Financial Incentives	0.15	0.16	0.52	0.84	0.81	0.46
Train	(0.04)	(0.04)	(0.15)	(0.20)	(0.26)	(0.13)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Dependent Variable Unit	Binary	Binary	Days	Days	Days	Fields
Control mean	0.56	0.23	0.43	0.67	1.76	2.47
Obs.	555	555	1216	1216	1216	1216

Notes: This table shows the effect of the T1 - Financial Incentives Treatment on hiring of laborers for different agricultural tasks during the planting season, as well as the adoption of row planting on farmer's fields. The sample is comprised of trainer-employers (farmers who were invited to train laborers at the training event). Columns 1 and 2 are comprised of a randomly selected subsample of these individuals who we revisited after the initial survey. The dependent variable in column 1 is a binary variable equal to one if the farmer reported trying to hire the laborer they identified to bring to the training event during the planting season. The dependent variable in column 2 is a binary variable equal to one if the farmer reported trying to hire the laborer they identified to bring to the training event, and being unsuccessful in doing so. The dependent variable in column 3 is the number of days an employer hired the trainee who they invited to the training event to do skilled tasks on their fields (row planting and fertilizer microdosage). The dependent variable in column 4 is the number of days an employer hired any laborer who was invited to the training events to do skilled tasks on their fields (row planting and fertilizer microdosage). The dependent variable in column 6 is the number of fields in the farmer's household that were planted using row planting. Standard errors are in parentheses and p-values are in brackets. Standard errors are clustered at the village level.

Table 1.4: Spillover Experiment: Hiring & technology adoption among spillover-employers

	Trainees Skilled Task	Hired Skilled Task	Adoption Total
T1 - Financial Incentives Train	1.25	1.29	0.45
	(0.19)	(0.32)	(0.13)
	[0.00]	[0.00]	[0.00]
Control mean	0.23	2.14	1.90
Dependent Variable Unit	Days	Days	Fields
Obs.	1466	1466	1466

Notes: This table shows the effect of the T1 - Financial Incentives Treatment on hiring of laborers for different agricultural work during the planting season, as well as the adoption of row planting on farmer's fieds. The sample is comprised of spillover farmers uninvolved in the village training even. The dependent variable in column 1 is the number of days an employer hired a laborer who was a trainee to do the trained techniques (row-planting and fertilizer microdosage) on their fields. The dependent variable in column 2 is the number of days an employer hired any laborer to do the trained techniques (row-planting and fertilizer microdosage) on their fields. The dependent variable in column 3 is the number of fields the employer adopted row planting on. Standard errors are in parentheses and p-values are in brackets. Standard errors are clustered at the village level.

Table 1.5: Spillover Experiment: Incidence of training returns (farm profitability and total earnings)

	Farm Revenues	Farm Profits		Labor Market Earnings
		Wage = Own/Average	Wage = 0	
Panel A: Pooled	(1)	(2)	(3)	(4)
T1 - Financial Incentives	27,996	25,497	26,243	-598
Train	(15,023)	(12,949)	(12,737)	(783)
	[0.07]	[0.05]	[0.04]	[0.45]
Control mean	334,230	234,869	129,284	6,502
Obs.	3491	3491	3491	4060
Panel B: Trainees	(1)	(2)	(3)	(4)
T1 - Financial Incentives	16,551	16,408	12,084	3,382
Train	(8,558)	(7,786)	(7,634)	(1,409)
	[0.06]	[0.04]	[0.12]	[0.02]
Control mean	155,385	115,491	33,610	16,775
Obs.	839	839	839	839
Panel C: Trainer (Employers)	(1)	(2)	(3)	(4)
T1 - Financial Incentives	24,869	24,175	23,719	-732
Train	(14,690)	(13,507)	(13,281)	(635)
	[0.09]	[0.08]	[0.08]	[0.25]
Control mean	368,729	260,090	140,052	3,441
Obs.	1204	1204	1204	1204
Panel D: Spillover (Employers)	(1)	(2)	(3)	(4)
T1 - Financial Incentives	31,135	27,349	31,212	483
Train	(19,035)	(16,375)	(16,462)	(323)
	[0.11]	[0.10]	[0.06]	[0.14]
Control mean	409,011	282,978	175,626	859
Obs.	1448	1448	1448	1448

Notes: This table shows the effect of the T1 - Financial Incentives Treatment on the profits of farmers in treatment villages. Panel A shows regression results for the pooled sample (Trainers, Trainees, Spillover). Panel B shows regressions results for the sample of trainees only. Panel C shows regression results for the sample of trainer (employers). Panel D shows the regression results for the sample of spillover (employers). The dependent variable in column 1 is farm revenues. The dependent variable in column 2 is farm profitability, as total crop revenues less labor and non-labor input costs, valuing family labor at a wage of 0. The dependent variable in column 3 is the same as column 2 except that it assumes that any family labor is valued at the wage of either the individual themselves, or the average wage in the village. The dependent variable in column 4 is labor market earnings. Standard errors are in parentheses and p-values are in brackets. Robust standard errors are clustered at the village level.

Table 1.6: Spillover Experiment: Cost-benefit ratio of training

Panel A: Assumptions			
Seasons of Returns	2		
Depreciation of Returns	0.5		
Panel B: Parameters			
Discount Rate	0.1		
Panel C: Benefits			
Profits and Labor Market Earnings (Pooled)			
Proportion of Sample:			
	Trainee	Trainer	Spillover
	25%	25%	50%
Panel D: Cost			
Land/Equipment rental, Training Incentives, Staff time			
Opportunity cost of time (Trainee and Trainer Search)			
No Fiscal Externalities			
Panel E: Benefit/Cost			
	Trainee Only	Trainer/Trainee	Trainee/Trainer/Spillover
2 season	0.5	1.2	2.4
p ($H_0 : B/C \leq 1$)	0.99	0.33	0.12
1 season	0.3	0.8	1.7
p ($H_0 : B/C \leq 1$)	1.00	0.74	0.21

Notes: This table shows estimates of the benefit cost ratio associated with the training. Panel A shows the assumptions associated with the length of time the benefits last. Panel B presents another parameter. Panels C and D describe the Benefit and cost estimates. Panel E shows estimated benefit cost ratios and p-values for the test of the null that the ratio is less than or equal to 1. The first column in Panel E includes only the earnings benefits for trainees (increase in own field profitability and labor market earnings), the second column adds to this also the increase in farm profitability for trainer-employers, and the third column includes also the increase in farm profitability for spillover-employers. Panel F adjusts the one season Benefit-Cost ratio for trainees netting out the portion of earnings that is due to an increase in days worked.

Chapter 2

Kinship Pressure and Firm-Worker Matching

2.1 Introduction

Few businesses in low- and middle-income countries hire employees outside of their family network (Jayachandran, 2021). This fact is often attributed to contracting frictions and missing information markets, as hiring relatives might alleviate asymmetric information or reduce the scope for moral hazard (Chandrasekhar et al., 2020).

This paper proposes that some employers face pressure to offer financial assistance to extended family members in the form of employment, a hypothesis with potentially different productivity implications. This theory is consistent with a literature in anthropology (Akyeampong et al., 2014) as well as a long economics literature showing that individuals in low- and middle-income countries face redistributive pressure from their kinship network to share income (Dupas and Robinson, 2013; Jakiela and Ozier, 2016; Goldberg, 2017; Squires, 2018; Di Falco et al., 2018; Fiala, 2018; Carranza et al., 2022; Di Falco et al., 2019; Riley, 2022).

In this paper, I test whether pressure impacts employers' hiring decisions, quantify how costly it is for employers to tell their family they will not hire them, and finally estimate the returns to employers of hiring related as opposed to unrelated employees. Since family pressure is not randomly assigned, I use a set of field experiments to manipulate the amount of pressure employers face when they hire workers.[1] I vary pressure by offering some firms

[1] A common approach in the literature is for researchers to vary whether an individual takes an action publicly or privately and then measure how an individual's action changes in response to others observing it (Bursztyn and Jensen, 2015, 2017; Bursztyn et al., 2017; Breza et al., 2019; Carranza et al., 2022). This

plausible deniability in their hiring decisions; specifically, some firms receive a poster that suggests the firm may not have been eligible for the subsidy if they hired a relative.[2] I then measure whether firms that have plausible deniability to not hire from their family change who they decide to hire.

I run three separate experiments to test this hypothesis. The first experiment tests whether pressure impacts hiring under naturalistic hiring conditions. The second and third experiments have more involved protocols to extract more information about how employers make hiring decisions under pressure and measure the productivity impacts of hiring related and unrelated employees.

While the experiments differ, all three broadly follow the same design. Employers receive a subsidy to hire an employee so that I can observe hiring decisions.[3] Employers are then randomized to one of two conditions. In a control group, employers are told that they will receive a poster stating that they received a subsidy and selected the person hired. In the plausible deniability group, employers are told that they will receive a poster stating that they received a subsidy, but some employers were unable to use the subsidy for their preferred candidate. The poster in the second group makes it easier for employers to tell their family that they would not have received the subsidy if they hired a relative. To ensure that the treatment does not involve deception, employers in both treatment conditions are sometimes offered a take-it-or-leave-it subsidy for an employee who is not their preferred candidate.[4] I then measure the likelihood that the employer chooses to hire an employee who is related to them with the subsidy across treatment conditions.

In my first experiment, the "urban field experiment", I begin by testing whether social pressure alters hiring decisions for full-time employees among a sample of 253 small urban enterprises in Zambia. Firm owners are told that they might receive a 3-month subsidy to hire an employee and should identify a candidate for the job.[5] To create variation in plausible deniability, in the control group, employers are given a poster stating (among other things) that they were involved in selecting the employee. In the plausible deniability treatment, business owners receive a poster stating that some employers receive subsidies for employees they do not select, allowing employers to tell their family they would not receive the subsidy if they hired a relative.[6] I measure how the introduction of this poster changes who employers

design is infeasible in my setting since who is hired is almost always going to be observed by others.

[2]For this to be effective, it must be that the hiring norm is one where a business owner is punished if they *choose* to hire someone unrelated for a job that is desired. I show evidence for this being the case in my context.

[3]The process that I follow to offer subsidies generally follows the protocol of De Mel et al. (2019) who also offered subsidies to businesses to hire an employee.

[4]The control group's poster is also adjusted in such cases so as not to involve deception.

[5]Urban employers typically hire employees for medium-term durations (median = 6 months).

[6]To ensure that the treatment does not involve deception, employers are sometimes offered the subsidy

choose to hire.[7]

In the primary finding from this experiment, I show that plausible deniability decreases the likelihood of hiring a family member: employers who receive the plausible deniability poster reduce the likelihood of hiring from their family by around 20%. There is also suggestive evidence that employers in the plausible deniability treatment are able to select more productive employees. The employees selected by employers in the plausible deniability treatment are expected to earn higher wages, and are more likely to have some secondary education.

To measure the magnitude and incidence of these pressure costs, I turn to my second field experiment, the "urban mechanism experiment", conducted with 619 urban employers. This experiment broadly follows the same design as the urban field experiment, with two key deviations. First, before introducing variation in plausible deniability, employers are asked to identify a related and unrelated employee for the subsidy. Second, after treatment assignment, I offer employers varying subsidies for related and unrelated employees and use the Becker-Degroot-Marschak method to estimate how much additional subsidy must be offered to employers to induce them to hire an unrelated employee. Comparing this across treatment arms allows me to quantify the pressure cost associated with hiring unrelated employees.

In this second experiment, I replicate the result of the urban field experiment that access to the plausible deniability poster leads employers to choose their family members as employees less often: when employers are offered the same subsidy for each employee, they are around 25% less likely to choose to hire their relative. Consistent with the hypothesized mechanism, treatment effects are concentrated among employers who report more social pressure from their family at baseline: these employers reduce the hiring of related employees by more when offered plausible deniability, while those who do not report family pressure do not change the identity of who they hire across treatments. In addition, I find suggestive evidence that this pressure may lead to less productive employer-employee matches. Employers who perceive that their related employee would exert less effort for them than other employers reduce their hiring of related employees by around 30 percentage points.

I then estimate how costly it is for employers to tell their families they did not hire them for a job by measuring the difference in the amount of subsidy I need to offer employers to hire a non-relative in the control and plausible deniability conditions. Employers with plausible deniability reduce the amount of subsidy they must be offered to select a non-relative by around 4% of the median firm's monthly profits. This suggests that firms are willing to forgo

for an employee of their choosing and other times for another jobseeker. The poster received by the control group is adjusted in cases where the control group does not select the employee.

[7]In this experiment, subsidies are then implemented probabilistically, ensuring the experiment is feasible but choices are incentive-compatible.

hiring unrelated employees that they expect to increase their profits substantially to avoid telling their family they did not hire a family member.

I then consider a set of alternate interpretations for these results. Using variations in the basic experiment design, I show that 1) the signaling motive that I identify in my experiments for hiring related employees is perceived as a cost, not a benefit, for the majority of employers and 2) the results are not driven by the selection of employees who would not be the marginal employee outside the experiment. Taken together, these results provide some of the first evidence for a social pressure mechanism contributing to the prevalence of family businesses in low-income countries.

Why does this pressure persist when alternate transfer mechanisms, for instance, hiring the most productive employee and then making larger cash transfer payments to the family, seemingly offer benefits for all parties? And how does this mechanism co-exist with our common understanding that relationships alleviate contracting frictions? While a full discussion of these questions is beyond the scope of this paper, I provide two pieces of evidence suggesting that this norm may be sustained by disagreements between employers and other family members' beliefs about related employees productivity. While around half of firm owners expect that related employees work less hard for them than they would for unrelated employers, *almost all* employers expect their family to believe that their related employee would work harder for them than an unrelated employer. Moreover, the treatment effects in the Urban Mechanism Experiment are concentrated in jobs where the employer states that it is "easy to make excuses for poor performance", making it harder to verify the effort of the employee. Taken together, these findings suggest that hiring pressure might be concentrated in cases where there are disagreements in the family about an employee's performance and where it is difficult to update beliefs about this.

This paper extends a literature in development related to the role of redistributive pressure in distorting investment, savings, and labor supply decisions. A long experimental literature shows individuals take costly actions to hide income, while heterogeneity in field experiments suggests the effect of new savings technologies is particularly important for those who face redistributive pressure (Dupas and Robinson, 2013; Jakiela and Ozier, 2016; Goldberg, 2017; Squires, 2018; Di Falco et al., 2018; Fiala, 2018; Di Falco et al., 2019; Riley, 2022). However, there is limited field evidence for the effects of kinship pressure on margins beyond savings and investment, with the exception of a recent paper that documents impacts on labor supply (Carranza et al., 2022). I contribute to this literature by highlighting a new and economically important margin through which such pressure might manifest: hiring choices, and enacting a field experiment to quantify this pressure, as well as providing an explanation for how such pressure might persist.

This paper also speaks to a development literature on the functioning of labor markets in LMICs and the existence and microfoundations of frictions in these markets (Breza et al., 2019; Kaur, 2019; Breza et al., 2021). It also builds on two recent contributions to the literature that highlight social and redistributive purposes of work (Hussam et al., 2022; Macchi and Stalder, 2023). This paper extends these papers by proposing an alternate mechanism for redistribution through employment and documenting the productivity losses associated with work-related transfers.

This paper also builds on a sizable body of behavioral literature exploring the influence of social pressure and signaling motivations on economic behavior in a variety of domains (Bursztyn and Jensen, 2015; Bursztyn et al., 2017; DellaVigna et al., 2012; Karing, 2018). It also contributes to a related literature documenting how wriggle room and uncertain mappings from actions to outcomes can lead to sizable changes in behavior (Dana et al., 2007). This paper contributes to this literature by applying this insight to an important domain and providing an estimate of the productivity losses associated with this social pressure.

Finally, this paper also advances a sizable behavioral literature documenting that social ties across workers may distort workplace effort and team selection (Bandiera et al., 2005, 2009, 2013). I contribute to this literature by extending the focus from ties among workers to ties across employers and employees, and therefore introduce a possible distortion arising from social ties to a long literature that has considered the reasons for network-based hiring (Beaman and Magruder, 2012; Beaman, 2016; Pallais and Sands, 2016; Heath, 2018; Chandrasekhar et al., 2020). By blending these two literatures, I offer a complemetary explanation for some hiring from social ties - that this hiring reflects social norms - a fact that has been argued by Bertrand and Schoar (2006).

The remainder of the paper proceeds as follows. Section 2.2 outlines the context of the experiments. Section 2.3 provides the design of the two experiments. Section 2.4 provides summary statistics and the empirical strategy. Section 2.5 presents the results of the Urban Field experiment. Section 2.6 presents the results of the Urban Mechanism experiment. Section 2.7 presents the results of the Rural Mechanism experiment. Section 2.8 discusses alternate explanations for the empirical patterns. Section 2.9 discusses what leads to the persistence of pressure to hire. Section 2.10 concludes.

2.2 Setting

Urban Samples

The experiments with urban employers focus on owners of microenterprises and small and medium enterprises located in Chipata and Katete - two urban centers - in Eastern Province, Zambia. In February 2023, we conducted a listing of all markets located in these two cities.

Labor Markets and Hiring in Context

Attachment between workers and firms are relatively short in this context. Among firms with employees, the mean length of tenure of a current employee at a firm is 30 months, with the 25th percentile and median 6 and 24 months respectively. The average employee in the sample is paid around 600 ZMW per month (approximately 30 USD).

Cost does not appear to be a major factor in the decision to hire family versus non-family: the average wage paid for relative and non-relative employees is the same (though around 10% of relative employees are unpaid while the proportion of non-relative unpaid employees is lower).

Consistent with the prior literature, relatives do appear to be more likely to be hired for particular kinds of jobs. Consistent with Bloom et al. (2013), they are 12 percentage points more likely to be hired for a job in management, and 30 percentage points more likely to be hired for jobs involving money, and become increasingly likely to be hired as employers rank the job as requiring more trust. However, there relatives are also hired around 30% of the time for jobs that do not involve management, money and lower levels of trust, suggesting that other channels might also explain the reasons such individuals are hired.

The most common relatives hired are not immediate family: 22% are nephews and 20% are cousins. This is followed by more immediate relations: 17% are children of the owner, and 13 and 8 percent the brother and sister respectively. There do not appear to be significant differences between the type of jobs held by relatives, and the distance of the relative to the employer.

Rural Sample

The rural experiment focuses on farmers in villages around Chipata districts, who hire workers to shell their maize after harvest.

Labor Markets and Hiring in Context

The shelling of maize is a common task in villages that farmers hire workers to do, conditional on having a large harvest. The manual shelling of maize involves several steps: i) removing husks, if the maize has not already been dehusked, ii) shelling the maize, which typically involves beating the maize husks to knock the maize off of the cob, and iii) the cleaning of maize, which involves sifting maize shells out from dust, rocks and other debris that may have gathered among the shells.

The manual shelling of the maize can be completed in one of two ways. The maize can be put into a bag and beaten. This speeds up the process of knocking shells off of the cob, but increases the likelihood that shells maybe damaged - estimates of postharvest losses attributed to such shelling procedures range from 1-6%, and increase the likelihood of broken graints, which can command lower prices and are more susceptible to attacks by pests and insect attacks (Hodges and Maritime, 2012b,a). Alternatively, the maize can be shelled slowly by hand. This process is less likely to damage the maize, but requires effort and is time consuming (Nsubuga et al., 2021). Because of these differences in how the task can be completed, measuring both quantity and quality of the output is important.

Labor markets for this task are governed by short term contracts: the median contract to hire an individual for such work is just 3 days. Hiring relatives is common, among those employers who stated that they hired an individual to shell their maize in the prior year, 76% stated that they hired at least one relative to do so. Contracts are often agreed by an employer searching for a worker 1-2 days before needing them, and the work is usually completed at the household of the employer. The modal contract in the setting is a half day contract, with the worker starting in the morning and finishing around lunch time. Workers are paid either for a half day contract or by the volume of maize shelled - although there is no general prevailing wage in the village and the terms are generally bargained between employer and employee.

2.3 Experiment

Treatment Overview - Family Pressure and Hiring Choices

The key hypothesis of this paper is that employer's face social pressure when choosing who to hire as an employee, and that this pressure influences their hiring decision. To test this, I create an intervention that offers some employers more plausible deniability for not hiring a particular person. I then measure how employers who know they can make excuses for not hiring someone changes their hiring decisions.

To achieve this, I leverage the specifics of the norm in this setting. Employers report that they would expect to receive sanctions from their family if the family learnt that they chose to hire someone else for the job, when someone in the family wanted the job. However, they report that they are much less likely to face sanctions from the family if they hire an unrelated employee because they have no other choice (for example, because no one in the family wants the job). I use this fact to offer some employers the ability to tell others that they had "no choice" but to hire someone. I then measure how employers knowing they have this excuse changes who they choose to hire.

In all three experiments I vary what information is signalled to others about why a new employee was hired in the business. In the control condition, when employers choose to hire an employee, they know that others will learn that they selected to hire that employee. In the treatment condition, by contrast, employers instead are offered the ability to signal to others that the person who was hired may not have been the employee they selected. To ensure that the treatment does not involve deception, employers in both conditions are sometimes offered the subsidy for their selected employee, and sometimes not.[8]

Treatment Implementation - Urban Field Experiment

The experiment broadly follows the protocol used by De Mel et al. (2019) to offer employers a subsidy to hire an additional employee into their business for 3 months.[9] Employers are told that they will receive a subsidy for either a local jobseeker, or for a person of their choice. They are then told that during the period of the subsidy, they will be asked to display one of the following posters on their business:

- **Control Poster:** The poster states that the employer received a subsidy for their business, and that they were involved in selecting the employee.[10]
- **Plausible Deniability Poster:** The poster states the same information as in the control poster, but in describing the involvement of the employer in selecting the employee states that in many cases business owners receive subsidies for employees chosen by the program, rather than their own first choice.[11]

Enumerators explain to businesses in the treatment group how the poster could be used to justify not hiring a particular person.[12] Enumerators then leave and return several days later to ask who the employer identified to hire for the position. After this, employers probabilistically receive a subsidy either for their choice of who to hire, or a local jobseeker.

Treatment Implementation - Urban Mechanism Experiment

The protocol is the same as above except for the following deviations. During the baseline survey, a listing of the employer's relatives is taken. At the end of the baseline survey,

[8]In cases where the control group receives a subsidy for someone they did not select, their signal is adjusted to state that they were offered a subsidy for the selected employee by our project.

[9]To ensure there are no "coercive" hires, subsidies must not be used for household members or children under 18. A listing of household members it taken before the subsidy is explained, and the ID of the employee is checked before starting work for implemented subsidies.

[10]In cases where employers received a randomly assigned subsidy, the poster was adjusted to state this.

[11]This statement is not deception as in many cases the subsidy was ultimately provided for a randomly assigned employee, rather than the choice of the employer.

[12]No mention of family is made during this explanation.

employers are told that they might receive a subsidy to hire either a related employee, or an unrelated employee. They are asked to identify the two possible candidates before a second visit. To ensure that the related employee is related and the unrelated employee is not, the employer is told that if the subsidy is implemented, the identity of the hire will be checked against the initial list of relatives at baseline that was completed.

At the second visit, employers who have identified a related and unrelated employee are told that they will be allocated the subsidy either for the relative, the non-relative, or will be offered a subsidy in which they can choose between the two. Before choosing who to hire in each condition, employers are told that they will be assigned to one of two conditions, as before.

- **Control Poster:** The poster states that the employer received a subsidy for their business, and that they had were involved in selecting the hired employee.[13]
- **Plausible Deniability Poster:** The poster states the same information as in the control poster, except when describing the involvement of the employer in the hiring decision states that in many cases business owners receive subsidies for employees chosen by the program, rather than their own first choice.[14]

Enumerators explain to businesses in the treatment group how the poster could be used to justify hiring someone, and then telling others that the selected hire was not their choice.[15] Choices are then elicited at different subsidy levels for the related and unrelated employees, to obtain a compensating differential to hire the related employee. Employers are then probabilistically assigned to either the relative subsidy, non-relative subsidy, or the person they chose (at a randomly determined subsidy level).

Treatment Implementation - Rural Mechanism Experiment

The implementation is the same as in the previous experiment with the following deviations. At the second visit, rural employers are asked to bring the identified related and unrelated employees to the survey. It is then explained to everyone - the employer and the employees - that the employer might randomly receive the subsidy for one of the two employees, or might be able to choose between the two.

The employer is then taken to a private area, and asked to choose who they prefer to hire.

[13]In cases where employers received a randomly assigned subsidy, the poster was adjusted to state this.

[14]This statement is not deception as in many cases the subsidy was ultimately provided for a randomly assigned employee, rather than the choice of the employer.

[15]No mention of family is made during this explanation.

- **Control Choice Deniability Choice:** Before making their choice, the employer is told that if they are assigned to the choice group, the employees will be told who is hired, and that the reason the person is being hired is because the employer preferred them.

- **Plausible Deniability Choice:** Before making their choice, the employer is reminded that the employees will not learn if the person is being hired because of the employer's choice, or because of random assignment to the subsidy.

Choices are then elicited at different subsidy levels for the related and unrelated employees, to obtain a compensating differential to hire the related employee. Employers are then assigned to either the relative subsidy, non-relative subsidy, or the person they chose (at a randomly determined subsidy level)

Treatment Interpretation

In all three experiments, I make it easier for some employers to tell others later that they had no choice but to hire the person they received the subsidy for. I then measure how this plausible deniability, which is designed to reduce social pressure from others, changes their decision of who to hire.

The experiments offer tradeoffs in how I achieve this. In the Urban Field experiment, I achieve this in a light touch way, with no mention of family or relatives. This let's me observe choices the most naturalistic hiring conditions. However, I am unable to quantify the magnitude of any costs associated with this distortion. To achieve this, in the Urban Field experiment I ask employers to identify two candidates, and elicit preferences between the two, to obtain a compensating differential. This provides more information, at the expense of a (potentially) artificial hiring process that groups the world into two categories. Finally, the Rural Mechanism experiment enables me to implement the subsidy for all employers, and therefore measure the returns to hiring. However, the specifics of this setting mean that the variation I create in plausible deniability, is potentially artificial and heightened as compared to natural hiring conditions, since conceivably in real life employers are able to make hiring decisions without having them immediately blasted to their kinship network.

Screening and Selection

Urban Field experiment We conducted a census of all markets in Katete. During the screening survey, businesses were screened on i) whether the person managing the business was the owner, or was able to make hiring decisions for the business independently, ii) whether they were willing to hire an additional employee for their business if they received a subsidy of approximately 50% of the wage, iii) the individual had relatives of working age nearby who they believed would be willing to work for them, iv) if they were willing to hire a local jobseeker who we had identified could be suitable for jobs in this location.

Urban Mechanism experiment The protocol is the same as above except that 1) because we were less worried about informational spillovers, we used a smaller distance between businesses and 2) we did not screen on willingness to hire a local jobseeker since this was not required in this experiment.

Rural Mechanism Experiment Villages were selected based on their proximity to Chipata town. A random walk was then conducted by enumerators in villages, to select households. After visiting each household, employers were screened on i) whether they had any maize to shell, ii) whether they were willing and planned to hire someone to shell their maize later in the year and iii) whether they had adult relatives over 18 years of age in the same village who they would be able to hire to do the shelling.

Subsidy Details and Implementation

Urban Field and Mechanism Experiments The subsidy that the business owner receives can be used to hire any employee outside of the owners immediate family.[16] Businesses are told that they can set the terms of the hire (wage, working hours) as they like except that the employee must work full time, to enable enumerator audits. Business owners are told during the experiment that the subsidy they will receive is designed to cover around 60% of the wage of hiring a new employee, for three months.[17] Businesses are also told that they will be randomly audited, and payment of the subsidy will be conditional on successful audits. Finally, businesses are told that the identity of employees will be checked using identification cards on the first day of the subsidy.

On the date that the subsidy begins, our enumerator team visits the business that receives the subsidy and verifies the identity of the employee who is hired by the business using their NRC card. After the employees identity is verified, employers are told that the subsidy will be paid out at the end of the month, for 3 months, conditional on successful audits of the business. Our enumerator team then makes unannounced visits to the business on average 15 times over the course of the subsidy and verifies that the employee is working. Finally, after

[16] This restriction is made to stop coercive hires.

[17] This subsidy level was determined to be enough to induce hiring, and is similar to the subsidy amount used in similar studies (Alfonsi et al., 2020; De Mel et al., 2019).

$\sim$ 6 weeks, each employee is interviewed privately and asked what the employer is paying them in the preceding months. Employers in the urban context who have the subsidies implemented are visited on average twice each week and the presence of the employee is verified - any business failing multiple audits is disqualified from receiving further payments.

Following the experiment, the subsidy is implemented probabilistically in both urban samples.

Rural Mechanism Experiment The subsidy that the employer receives can be used to hire any employee outside of the owners immediate family.[18] Businesses are told that they can set the terms of the hire (wage, working hours) as they like except that the employee must work full time, to enable enumerator audits. At the time of being offered the subsidy, rural employers are told that they will receive a subsidy that will cover 60% of the average half-daily wage. The subsidy On both days that the subsidy is implemented our enumerator team visits the employer and employee. They attend early in the day to ensure that no shelling has already been done, and observe whether the employee has arrived. They then stay in the village during the course of the day. At the end of the half day, a subsidy is paid to the employer conditional on the employee working the full half day.

Following the experiment, the subsidy is implemented with $\sim$ 100% probability.

Randomization

In each experiment there are two randomizations. A first randomization determines if the employer is assigned to the Control or Plausible Deniability conditions. A second randomization determines who the employer receives the subsidy for. Finally, in the Urban Field and Mechanism experiments, a third randomization determines if the subsidy is implemented.

Randomization in the Urban Mechanism experiment was stratified by market, gender and whether the business had any non-casual employees to improve statistical power. Simple individual level randomisation was conducted in the Rural Mechanism and Urban Field experiment.

Outcome Measures

Urban Field Experiment The core outcome measure in this experiment is whether the employer chooses to hire a related employee. To understand the determinants of hiring choices, we also ask for information regarding why the employer chose to hire a particular person.

[18] This restriction is made to stop coercive hires.

Urban Mechanism Experiment In addition to the above outcome measure, we also measure the hiring preference of the employer at different subsidy levels, to obtain the compensating differential required to hire the relative employee.

Rural Mechanism Experiment In addition to the above outcome measures, we also measure hiring outcomes in this sample. Specifically, at the end of each day we measure the total output of maize shelled, the quality of this maize, and the wage paid to the employee.

2.4 Data and Empirical Strategy

Data sources

We rely on several sources of data for analysis. Our core measure of hiring choices is elicited by enumerators during the experiment. Whether an employer takes up a subsidy after the fact is validated by unannounced enumerator visits to employer's businesses, on average twice per week.

To measure the returns to hiring in the rural sample, enumerators also observed the output of shelling each day of the contract. Enumerators arrived at the villages and monitored that employees were present and working - subsidies were only paid out conditional on this. At the end of the workday, enumerators measured the total quantity of maize shelled, and also rated the quality, using the scale of Bold et al. To validate the quality measure, enumerators also purchased some maize from a subsample of employers, and sold it at nearby markets.

Empirical Strategy

The primary empirical specification used to measure the effect of social pressure on hiring choices is:

$$y_i = \alpha_i + \beta \text{Plausible Deniability}_i + \text{Strata}_i + \epsilon_i$$

Where y_i is the primary outcome of interest - either the choice of hiring a related employee or the compensating differential required to hire a related employee, and $Strata_i$ are stratification variables in the Urban Mechanism experiment. The primary coefficient of interest is β - the effect of being offered more plausible deniability on these outcomes.

In the Rural Mechanism experiment, we also use the following specification to measure the returns to hiring related and unrelated employees:

$$y_i = \alpha_i + \beta \text{Relative Subsidy}_i + \text{Subsidy Day}_i + \epsilon_i$$

Where y_i is the primary outcome of interest related to performance in the shelling task, either output produced, quality of output, or wage paid. $SubsidyDay_i$ is a fixed effect for whether it is the first or second day of the subsidy. The primary coefficient of interest is β which measures the effect of being assigned a subsidy for a related employee rather than an unrelated employee.[19]

2.5 Urban Field Experiment - Family Pressure Changes Hiring Choices

How does alleviation of family pressure change hiring choices? I turn first to the decision of employers in the Urban Field Experiment. As seen in Figure 2.2 Panel A, increased plausible deniability reduces the likelihood of hiring someone you are related by around 20%.

Is this finding consistent with the explanations given by firm owners for their hiring? At the end of each survey, enumerators ask business owners to explain why they hired the individual they hired. I then extract the most common words that are used to describe these choices, and correlate them with whether the employee is related to the employer. Panel A of Figure B.1.1 shows the relative likelihood of using a particular word to describe hiring a related (rather than unrelated) employee. The largest difference, by far, is the use of the word "help" to describe hiring of relatives.[20] Consistent with the prior literature, related employees are also more likely to be described as trustworthy, although this impact is relatively small and statistically insignificant. By contrast, unrelated employees are around ten percentage points more likely to be described as "reliable" and "hard"(working). Panel B of the same figure shows how the justifications of who was hired differ with treatment. The primary finding is that with Plausible Deniability, business owners are less likely to justify a hire as being hired to "help" and more likely to justify a hire as hiring because the person is reliable.[21]

[19]This speficiation drops employers who were assigned their choice of who to hire, and therefore measures the effect of random assignment to hiring a related versus unrelated employee.

[20]For example, one business owner stated "He has been asking me for a job for a while I believe this was an opportunity to help".

[21]This suggests that the decision to help the related employee is not driven by pure altruism, as this would not change with the signalling of the decision.

2.6 Urban Mechanism Experiment - Magnitude and Incidence of Hiring Cost

The Urban Field experiment establishes in a light touch manner that increased plausible deniability for employers changes their hiring decisions away from members of their family. To understand the size of these pressure costs, and who faces them in particular, I turn to the urban mechanism experiment. I first begin by replicating the results of the Urban Field Expeirment. I find similar outcomes in this setting. Consistent with the proposed mechanism that these changes in hiring preferences reflect responses to social pressure, Panel A of Figure 2.3, show heterogeneity in the treatment effects by whether these populations report receiving sanctions from their kinship network if they do not share with them. There are much stronger effects for employers who report not facing sanctions from not sharing with their family. Employers who face sanctions decrease their hiring of related employees substantially, while those who do not have sanctions see limited change.

These preferences to not hire some family members suggest that there maybe sizable productivity consequences for some businesses from hiring family. To better understand if this is the case, I turn to the Rural Mechanism Experiment that allows me to quantify the returns to hiring the marginal related and unrelated candidate.

2.7 Rural Mechanism Experiment

I replicate the results from the previous sections. In this sample and design, plausible deniability again changes hiring choices: Employers in the Control Choice condition choose their relative to work for them 78% of the time, whereas in the plausible deniability choice condition they choose to hire their relative around 25% less often. This effect is not that different from the Urban Field experiment, suggesting that the change to intensive signalling meaningfully changes treatment effects.

I find similar heterogeneity by the likelihood of social sanctions and of relative moral hazard. Rural employers who perceive that their family will sanction them if they do not give them what the ask for have much stronger treatment effects. I also construct a measure of moral hazard for this sample based on how much they expect their relative to shell for them as an employer, as opposed to others.[22] Again I find stronger treatment effects for those

[22]In the rural sample, employers are asked to forecast which of the employees - the related employee or non-related employee, would be able to shell more maize in a baseline shelling task designed to elicit productivity. The employers are then asked which employee would likely shell more maize for them. Employers facing a negative match specific effect are those whom forecast better performance of the relative on the baseline

who expect their relative to exert less effort for them for them. This heterogeneity again suggests that employers particularly susceptible to related employee frictions in contracting maybe particularly susceptible to pressure.

2.8 Alternative Explanations

The previous sections suggest that the prevalence of hiring relatives as employees may partially reflect redistributive norms that are perceived as a cost by firms, with potentially important productivity implications for the businesses. The next section discusses alternate explanations for the results.

Noisy Decisions

This project uses choices in the hiring experiment to infer employer's beliefs about productivity differences in employees as well as the signalling costs of the employer's choice. Doing so relies on businesses making considered choices, despite their choices only being implemented probabilistically. However, business owners might have learnt that their was a chance they would not receive a subsidy and not made their decisions carefully, or have assumed that they could receive a subsidy without hiring an individual. While it is not clear why noisy choices, or choices that seek to maximize revenues would lead to the treatment effects observed in this study, we do several checks to understand how prevalent such issues might be.

We see limited evidence that the hiring decisions in the experiment were not considered. Treatment effects for the subsample receiving the subsidy look similar to the overall sample. We also see choices translate into hiring for the vast majority of this subsample: all but one employer hire the employee randomly assigned or chosen by them at the end of the experiment, and all but two complete the subsidy period in the urban experiment. . These employers also receive this subsidy conditional only on successful audits of their business. Businesses were audited on average 15 times over the course of the subsidy, unannounced times and in 98% of cases the employee was found working at the business.

An alternative possibility is that employers in the urban experiment may have manipulated the subsidies in order to keep a portion of the cash themselves, or to reduce the amount of the wage paid by themselves so much so as to incur almost no cost from hiring the individual[23]. While the subsidies were formulated to cover 70% of the wage on average,

task as compared to the own employer task.

[23]This behavior was impossible in the rural experiment as the subsidy was paid directly to the employer and the payment to the worker was observed at the end of each day

similar to the levels provided in De Mel et al. (2019); Alfonsi et al. (2020), it is possible that the employer may have changed their behavior after the fact. Again, while it is not clear why this would generate the observed patterns of results, it would make the experiment potentially a lower stakes decision.

To check for this possibility, our surveyors conducted private interviews with hired employees around 2 months after the subsidy was initiated to ask them how much they had been paid by the employer in the previous months. We see limited evidence for the manipulation of wages: in 83% of cases workers reported being paid what the employer had told us they would pay the worker, in 10% of cases they reported being paid less and 5% of cases they reported being paid more.

Taken together, this evidence suggest that businesses made considered and high-stakes decisions in the experiment, that had considerable consequences for spending of the business owner themselves.

Employee Selection

The treatment effects in the experiment may reflect the effect of redistributive pressure on hiring decisions in the experiment, they may not carry forward to real hiring decisions if the candidates selected for the experiment do not reflect the marginal relative and non-relative candidates for actual jobs. In particular, the treatment effects found in the experiment may overstate the importance of redistributive pressure on hiring decision if the relative selected for the experiment were worse than the marginal relative candidate (because, for example, the employer thought that the experiment could be a cheap way to hire this employee, or because this employee heard about the experiment and demanded a job which would have not occured in the absence of the experiment), or if the non-relative selected were better than the marginal non-relative candidate (for example, if employers searched harder for this candidate because they knew there was a likelihood that they were going to be randomly assigned a subsidy for this individual, or because they were asked to search for a non-relative and therefore incurred a search cost that they would not normally have incurred). More subtly, the the marginal non-relative might be expected to be better than outside of the experiment because non-relatives maybe expected to perform better within the experiment (for example, because they believe non-relatives wont be likely to steal while partaking in the experiment).

We do several supplemental data collection exercises to test for these possibilities. First, at the end of the experiment, we ask all employers to rank the relative they selected for the experiment relative to the other relatives that they could have selected for the job, and ask them to explain why they chose that individual. Two-thirds of employers stated that this worker was the best available among their relatives, and that this was why the individual was selected, and 24% said the relative was the second best available. Only in around 10% of cases, did the employer state that the relative was lower ranked, and hired for some other reason (for instance, because the relative that they did want to hire had taken up another job and was not available).

Second, we conduct follow up surveys with around 90% of businesses that do not receive the subsidy around 3 months following the experiment, and ask them if they have hired any employees since the experiment ended, and if so who they hired. We find that within this period 21% of firms hired any employee to work into their business - among these 15.7% hired at least one relative permanent employee and 6.1% hired at least one non-relative permanent employee. In both cases we find a high likelihood of hiring the individual brought to our experiment: around 54.6% of relatives and 54.0% of non-relatives hired for a permanent position are the individual invited to the experiment - suggesting that in many cases this was an individual seriously considered by the employer for hiring even outside of the experiment.

Finally, a necessary design feature of the experiment, following De Mel et al. (2019) was to not allow businesses to hire employers to hire members of their household for jobs, in order to limit the possibility that employers might coerce members of their families to work for them unwillingly. However, this makes it possible that all of the treatment effects maybe driven by employers for whom the marginal candidate outside the experiment would have been a member of their household. We find this is not the case: during the survey we ask employers if they would have identified a different relative candidate if they were allowed to hire a member of their household and 25% of employers say yes. However, heterogeneity in treatment effects by this variables suggest that this does not account for the treatment effects: Appendix figure B.1.3 shows that among the 75% of employers who stated they did not have a desired replacement in their household there is still a sizable treatment effect of the poster.

An additional possibility is that the non-relative employee selected in the experiment would not be the marginal non-relative candidate - for instance, because the search costs of finding this non-relative were particularly high and search effort was made only because the job would be subsidised.

Several pieces of evidence suggest this is not the case. First - in the urban experiment we see high levels of hiring the non-relative in the experiment, particularly for casual (short term jobs after the experiment) - of all non-relatives hired for such jobs, the non-relative selected for the experiment is hired 57% of the time.[24] Second, in general in the urban experiment it seems like the non-relative selected for the experiment is a person that the

[24]Since we do not have a pure control, however, this could be driven by firms overcoming search costs as part of the experiment

firm owner knew well, and was unlikely to have searched very hard for. The average non-relative employee selected was known by the business owner for 6 years, in 85% of cases is someone that the firm owner said they see weekly or multiple times per week, in 30% of cases was someone who had previously worked for the employer. Therefore, it seems unlikely that the employer chose an individual that would be difficult or implausible for them to find outside of the experiment. Finally, in the rural experiment, we specifically measure the search costs associated with finding both the relative and non-relative employee. We find the difference in search time for the non-relative to be 3 minutes larger than that for relatives - making it implausible that this constitutes a meaningful barrier to hiring non-relatives in this context.

Finally, one could argue that the interaction of the experiment and the identity of the employee makes hiring non-relatives more desirable than it would be naturally. For example, employers might believe that non-relatives may be less likely to steal from them knowing they are part of the experiment, or they believe that the fact that the non-relative knows that the business is receiving the subsidy will make the non-relative more likely to work hard/not quit for another job. While these concerns are plausible, they might equally apply to related hires. Moreover, the fact that unrelated hires are hired outside of the experiment (see Appendix Figures B.1.6 and B.1.5) suggests that these employees are not undesirable outside of the experiment.

Welfare and Signalling Benefits

This paper suggests that redistributive pressure imposes a cost on businesses by distorting their hiring decisions, and that my experiment measures a signalling cost associated with employers choosing to hire an unrelated employee. An alternative view of the experiment is that employers derive a signalling benefit from choosing their relative (for instance, by allowing the employers to signal to others how desirable the employee is). In these alternate stories, my experiment by introducing noise in the mapping from choices to outcomes, might reduce signalling benefits that employers value.

To test for this, I allowed respondents in the rural experiment, prior to the making their choice, to decide if they preferred their choice to be made with or without plausible deniability, with the choice implemented for a small portion of the sample to make the choice incentive compatible. I find evidence that individuals prefer their choices to be made privately and perceive signalling their choices as a cost. Almost 90% of respondents prefer for their choice to be made privately.

Magnitude of Signalling Cost

One concern about the experimental design is that the level of social pressure in the control conditions maybe larger than normal. For example, knowledge of the experiment may lead family members to seek out the employer and ask questions of them that they would not normally. Moreover, particularly in the rural experiment, the control condition may appear unnatural as employers may not expect to have their hiring choices immediately amplified to the network, or have the unrelated employee be passed over so saliently. While these concerns do not undermine the test for the existence of pressure, they matter for interpreting the magnitudes, which maybe overstated because of these design features.

Three reasons suggest that this may not be a large cause for concern. First, the design of the Control Poster condition in the urban experiment removes many potentially "unnatural" features of the hiring decision arising in the rural experiment, and treatment effects are still large in this case.

Second, motivating evidence regarding the norm suggests that family is likely to learn that an individual is hired, and that ex-post excuses (eg. "You weren't available at the time") are unlikely to be accepted by the family. Therefore, while the experiment might heighten the signalling of the opportunity to others, this is unlikely to be a major driver of the pressure felt.

Finally, the literature on social pressure has generally found stronger effects when individuals are approached by others with a request (Lazear et al., 2012; DellaVigna et al., 2012). While our experiment abstracts away from requests for jobs these are common in our context, around 40% of employers report hiring a worker because of the worker approaching the employer requesting a job. If social pressure is particularly concentrated among such requests, our experiment may underestimate the prevalence of pressure to hire in the population.

Mechanism for Signalling Cost

This paper interprets the change in hiring patterns arising from observability of hiring choices as reflecting a cost anticipated by employers arising from the fact that they deviate from social expectations that they choose to hire their relative. However, alternative interpretations of the signalling cost are possible. For example, individuals might believe that if other employers observe their decision to not hire their relative, they will revise downwards their beliefs about the employees type, making it less likely that that individual is employed elsewhere in the future.[25] If this signalling to others mechanism is so strong, then the effect maybe driven by

[25]Employers might care about this because of altruism, or because they expect to be able to draw on income from this individual in the future, and therefore would like them to have more income.

altruism, rather than social pressure.

While we cannot entirely rule out alternative signalling stories, several pieces of evidence suggest this is not the primary reason for the behavioral response to choice signalling. First, this heightened signalling to others is most pronounced in the Rural Mechanism Experiment - but I find similar effects in all three experiments. Second, the fact that 90% of employers choose not to signal hiring when they have the choice suggests that there is limited upside to the employee of having it signalled to others that they were chosen by the employer - but if there is a range of abilities of related employees one might expect there is a positive signalling benefit for some related employees. Finally qualitative evidence suggests that the primary mechanisms that employers mention for choosing to hire their relative have to do with sanctions and disagreements with their family, rather than humiliation for the potential employee. Taken together, this suggests that these results reflect employers abiding by social expectations, rather than a deep concern of signalling negative news about their relative.

2.9 What generates pressure to hire?

Why should relatives side with employees who take undesirable actions toward their relative employers? Appendix Figure B.1.2 offers some suggestive evidence that this maybe because employers and their family have different beliefs about the performance of the relative as an employee. The figure shows employers beliefs, as well as their beliefs about the beliefs of members of their family, about whether the selected relative employee would work better or worse for them than for another employer who they were not related to. As noted previously there is substantial heterogeneity in the beliefs of business owners in the experiment regarding the performance of their relative in their business. However, in general employers perceive that there is likely to be disagreement between them and their relatives on this point: among business owners who believe that their relative will not work as well for them as they would for someone they are not related to, almost 75% believe that their relatives believe the employee would work harder for them than someone they are not related to. One also finds that a substantial proportion of the treatment effects from the experiment are concentrated in this region of disagreement: although the sample where there is agreement is small, there is no statistically significant impact of the treatment on the choice of who to hire when the employer thinks the relative will work less well for him and believes his relatives believe this as well.

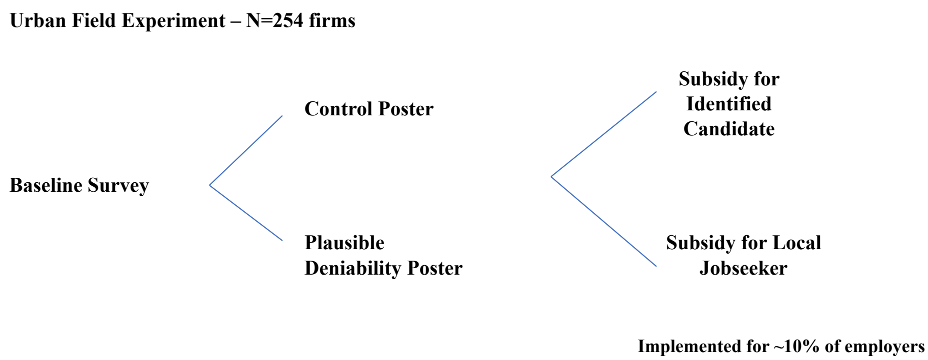
2.10 Conclusion

This paper tests whether some employers in LMICs hire family members as employees in due to redistributive pressures. In three experiments, I document that when employers know they can make excuses for not hiring someone, they are less likely to select a related employee for a job in their business. I find that employer's perceive this redistributive pressure as costly. Finally, by randomising subsidies for related and unrelated employees in one task, I find that there are productivity consequences from hiring related employees, who are less productive for firm owners.

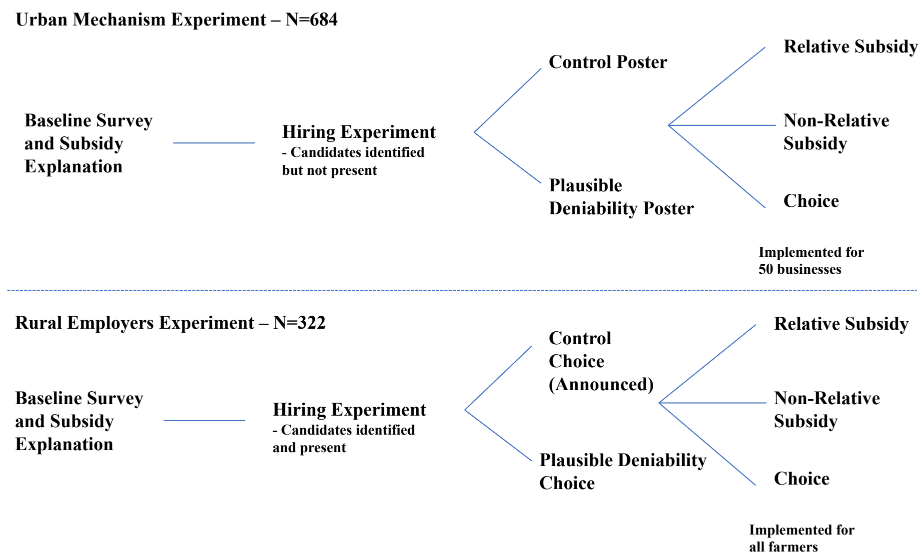
These findings have important implications for thinking about the functioning of labor markets in LMICs. While labor market frictions have been posited to be a constraint to firm growth, active labor market policies have generally been found to have muted effects (McKenzie, 2017). Similar to Macchi and Stalder (2023), this paper suggests that one explanation for this might be that work may have a redistributive role. Unlike Macchi and Stalder (2023), this paper suggests that partly the decision to give jobs might results from family negotiations and pressure, rather than employer's internalized preferences.

Many important questions remain. One is how formal policies with a redistributive role, such as social insurance schemes, might interact with such informal redistributive mechanisms. Another important question is whether alternate transfer mechanisms exist for business owners to alleviate hiring pressure, and what they entail.

2.11 Figures



(a) Experiment Design – Urban Field Experiment

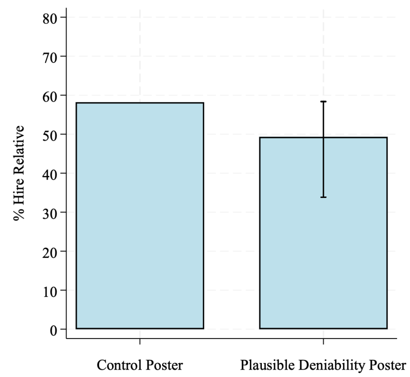


(b) Experiment Design – Urban and Rural Mechanism Experiments

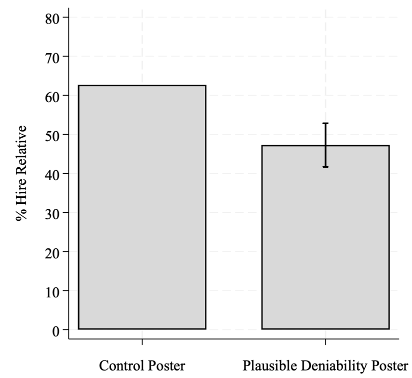
Figure 2.1: Overview of Field Experiment Designs

Notes: These figures show the field experiment designs. Panel A shows the experiment design of the Urban Field Experiment. Panel B shows the experiment design of the Urban Mechanism and Rural Mechanism Experiments.

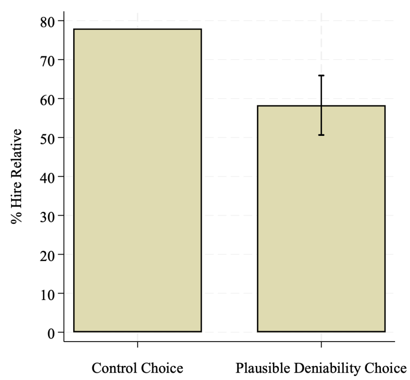
Figure 2.2: All Field Experiments: Treatment effects on hiring a related employee



(a) Urban Field Experiment

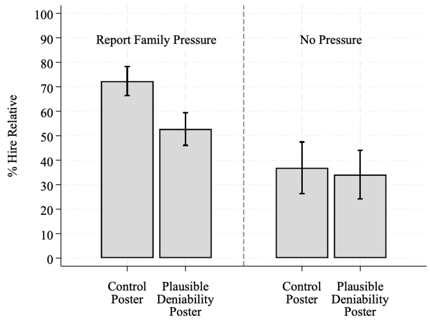


(b) Urban Mechanism Experiment

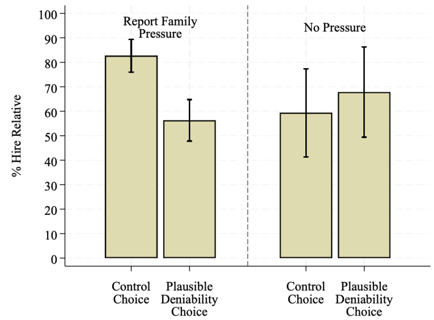


(c) Rural Mechanism Experiment

Notes: These figures show the proportion of employers who choose to hire a related employee in each experiment, across treatment conditions. Panel A shows the proportion of employers selecting a relative to hire with the subsidy in the Control conditions, as well as the proportion of employers selecting to hire a relative in the plausible deniability condition, using the coefficient of a regression of likelihood of hiring a relative on an indicator for Plausible Deniability, location and industry fixed effects. Panel B shows the proportion of employers selecting a relative to hire when offered the same subsidy for a related and unrelated employee in the Control and Plausible Deniability Poster conditions. Panel C shows the proportion of employers selecting a relative to hire when offered the same subsidy for a related and unrelated employee in the Control and Plausible Deniability Choice conditions. Robust standard errors are shown.



(a) Urban Mechanism Experiment



(b) Rural Experiment

Figure 2.3: Treatment Effect Heterogeneity in Urban and Rural Mechanism Experiments – By Stated Family Pressure (Threat of Sanctions)

Notes: These figures show the proportion of employers who choose to hire a related employee in each experiment, across treatment conditions, conditional on reporting redistributive pressure from their family. In the rural mechanism experiment, a respondent is coded as facing pressure if they agree or strongly agree that "If my relative learned that I chose my non-relative and not him for the job, I would face negative consequences from others". In the Urban Mechanism Experiment, employers who are described as facing redistributive pressure are those who agree or strongly agree that "If your relative expects something from you, and you don't help them, there can be strong consequences for you". Robust standard errors are shown.

Chapter 3

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

This chapter is coauthored with Ned Augenblick, Kelsey Jack, Supreet Kaur and Felix Masiye.

3.1 Introduction

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

We focus on a classic economic problem: deciding how much to spend today and how much to save for tomorrow's expenses. Solving this problem requires that individuals consider myriad future expenses, from large and certain ones (such as a car payment), to those that are small, irregular, and uncertain (such as gas, maintenance, and repair). Even though many of these expenses can be retrieved with thought or external prompting, our core hypothesis is that some of these expenses are not recalled when making decisions. In contrast, we posit that individuals have fewer issues when retrieving information about their inflows, which generally arise from a few large and consistent sources.[1] Using a simple model, we illustrate

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

how this asymmetry causes an individual to overestimate future savings and overspend today, such that she must cut back on future consumption when the neglected expenses (i.e., those she had previously failed to retrieve) become due. Consequently, an intervention that boosts the retrieval of more future expenses causes her to cut back on spending today, which allows her to spend more in the future.

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

To motivate the potential relevance of retrieval failures, Figure 3.1 documents large systematic bias in beliefs: farmers overestimate future savings and underestimate future expenses. Early in the agricultural year, we ask farmers to make incentivized forecasts about their future savings (i.e., maize stock). 78% of individuals think they will have higher savings 3 months later (at the start of the hungry season) than they actually do, with the average participant overestimating future savings by 81%. In addition, we ask individuals to predict their "worst case scenario": how much savings they will have left in the future if "everything that could possibly go wrong does go wrong." Strikingly, more than 60% of farmers end up with less savings than their worst-case savings forecast. They are similarly overoptimistic about savings levels 5 months in the future (the middle of the hungry season).

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

likely be larger and more volatile.

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

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

these expenses are largely comprised of items such as school fees, farm inputs, and medical costs—items that are unlikely to be temptation goods.

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

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

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

We randomize 197 farmers to receive either the full or placebo board. Relative to individuals' prior at baseline, the placebo expense board leads to no detectable changes in beliefs about non-food expenses. This suggests that simply going through the motions of planning

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

alone does not change beliefs. Rather, consistent with our hypothesis, the treatment expense board has large impacts: treated farmers expect to spend 42% more on non-food expenses than the control (p <0.001). Our model predicts that such changes in beliefs will make individuals feel "poorer," raising the shadow price of money and therefore lowering spending today. Consistent with this prediction, in a real-stakes opportunity to buy a discretionary consumption good (e.g., new clothing), treated individuals' willingness to pay for the good falls by 37% relative to the control group (p <0.001).

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

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

The immediate impacts of the longer field experiment match those from the mechanism experiment. First, in the plan developed through the retrieval exercise, treated farmers indicate that they expect to spend 62% more on non-food expenditures relative to their

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

baseline forecast ($p < 0.001$). Second, treatment participants are, on average, willing to pay 34% less for discretionary consumption goods ($p < 0.001$).[6]

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

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

Together, these results provide consistent support for our model of retrieval failures. In Section 3.7, we discuss alternative explanations such as soft commitment, intra-household bargaining and experimenter demand, and show that, while some might explain isolated findings, they cannot simultaneously account for the full set of results without relying on a form of retrieval failures. In addition, in Section 3.8, we complement our core findings by documenting impediments to learning from one's past or learning from others—helping explain why biased beliefs may persist despite experience.

Finally, to examine the external validity of our mechanism, we run a similar intervention in the United States, and discuss the results in Section 3.9. Specifically, in a survey of around 700 low income participants, we collect prior estimates of upcoming monthly income and expenses. Next, we lead subjects through a categorization-based retrieval exercise similar to

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

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

the one we implemented in Zambia, separately for income and expenses. We then elicit their posterior estimates. Subjects revise estimates of both income and expenses upward, but by a considerably larger margin for expenses. This is consistent with the idea that expense items are more susceptible to retrieval failures than are income items. These results suggest that retrieval failures may generalize across populations with varied economic circumstances.

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

Relatedly, a burgeoning body of theoretical work in economics models the implications of imperfect information retrieval (Mullainathan, 2002; Gabaix, 2019; Bordalo et al., 2020, 2023; Malmendier and Wachter, 2023). These models are inspired by psychology research on memory (Anderson and Milson, 1989; Kahana, 2012; Wimmer and Shohamy, 2012; Kahana and Wagner, 2023) and prediction and retrieval biases (Kahneman and Tversky, 1973; Tversky and Kahneman, 1974, 1983; Lichtenstein et al., 1978). However, direct field evidence on imperfect information retrieval, and particularly on the relevance of memory, has been limited. A notable exception is work that shows that sending individuals text message reminders to undertake a specific normatively desirable action (i.e., take a pill or save) can immediately increase compliance with that action (Pop-Eleches et al., 2011; Karlan et al., 2016). In contrast, our design does not direct people toward a specific action, but rather lowers the cost of retrieving the various pieces of information that are inside their minds.[8] We find that this drastically changes individuals' understanding of their overall maximization problem (i.e., through a substantive change in beliefs), with subsequent changes in behavior, including total household spending over multiple months. This offers complementary and naturalistic evidence that retrieval failures can cause large distortions in economic behavior. Moreover, our design enables us to directly test specific predictions of memory models: irregular and stochastic items are more subject to retrieval failures, and cuing finer and more homogeneous categories improves recall (Bordalo et al., 2023).

Our study relates closely to a large literature in psychology on the "planning fallacy": the

[8]As we discuss below, participants think through their expenses as a whole, and we do not provide participants any guidance on what kind of expense item should increase.

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

Relatedly, a line of work finds that making a concrete and detailed plan to undertake a specific task—such as voting or getting vaccinated—increases immediate task completion (Nickerson and Rogers, 2010; Milkman et al., 2011, 2013; Abel et al., 2019), though recent studies have argued this approach is less effective at changing repeated behaviors (Carrera et al., 2018). This literature discusses multiple potential mechanisms, from self-control to reference dependence (Beshears et al., 2016). In our mechanism experiment, simply articulating a plan via the placebo board has no apparent impact; it is only when the intervention induces information retrieval (i.e., via the full expense board) do we see effects. Potentially consistent with our findings, this body of work emphasizes the need for plans to be detailed and concrete. To the extent that detailed planning induces retrieval—for example, forcing individuals to recall other time commitments or obstacles that may otherwise be neglected—retrieval failures may be one (not mutually exclusive) mechanism for the empirical findings in this normative literature.

Finally, our paper contributes to the literature on the presence and causes of consumption smoothing failures. The existing literature has examined several micro-foundations for potential consumption smoothing failures, including missing markets (e.g., Burke et al., 2019; Fink et al., 2020), present bias (e.g., Shapiro, 2005; Ganong and Noel, 2019; Gerard

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

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

and Naritomi, 2021), and social pressure to share income with others (e.g., Dillon et al., 2021; Carranza et al., 2021; Jakiela and Ozier, 2016).[11] We augment this literature by offering evidence for an additional (and not mutually exclusive) channel: retrieval failures. We document that ameliorating retrieval failures can have sizable consequences for smoothing behavior, indicating first-order relevance as a mechanism.

3.2 Study Setting

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

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

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

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

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

3.3 Model

3.3.1 Model: Introduction

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

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

We present relatively intuitive results in less formal terms. Appendix Section C.3.1 contains proofs and a more formal discussion, while Appendix Section C.3.2 illustrates our predictions using a simple numerical example.

3.3.2 Model: Consumption Smoothing with Retrieval Failures

An individual is endowed with income Y, which she spends over three periods. In each period t, the individual must choose food consumption c_t, which produces utility $u(c_t)$. In addition to food, there are N other possible expenses, which stochastically arise at time t with probability $\pi_{it} \in [0, 1]$. That is, some expenses (such as household supplies) arise in every period, some (such as school fees) arise in only one period, and some (such as emergency

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

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

medical payments) arise stochastically. If expense i arises at time t, the individual chooses an amount e_{it} to spend on the expense and receives utility $v_i(e_{it})$. We assume that the functions $u(\cdot)$ and $v_i(\cdot)$ are increasing and concave, and that $u'(c_t) \to \infty$ as $c_t \to 0$ and $v_i'(e_{it}) \to \infty$ as $e_{it} \to 0$. To isolate the impact of our mechanism, we assume no time discounting, no borrowing, and use a simple-three period model, although modifying these elements does not change the main results.[14]

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

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

In other words, the individual solves the budget problem as if some potential expenses will not arise in the future. She remains unaware of these expenses until they arise at period t, at which point $\hat{\pi}_{it} = 1$. In each period, the individual observes which expenses arise for that period, decides on current spending on those expenses and consumption, and creates a state-contingent plan for future consumption and spending that satisfies her (perceived) budget. Then, she enters the next period, observes which expenses arise for that period, and repeats the process given her remaining wealth. The formal maximization problem is written in Appendix Section C.3.1, and requires additional definitions regarding uncertainty and realizations.[15]

For the most part, we do not explicitly model *which* expenses are subject to retrieval failures because it does not matter for our core results. However, in our setting, a particularly important example of a large, salient, predictable, and regular expense is food: maize (in the form of nshima patties) is consumed as part of every meal. To capture that food is likely to be recalled in our setting, we explicitly separate consumption c_t in the model and assume

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

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

that it is not subject to retrieval failures. This generates an additional ex-ante prediction: the individual should particularly under-estimate *non-food* expenditures. We leverage this more specific prediction in constructing our empirical tests (see discussion in Section 3.4).

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

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

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

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

Figure 3.1 and Appendix Figure C.1.1 provide empirical support for Initial Predictions 1 and 2. Of course, these patterns are also consistent with a variety of other explanations, such as naive quasi-hyperbolic discounting. Consequently, we use a targeted intervention to create a clean test for the presence of retrieval failures.

3.3.3 Model: Impact of Increasing Retrieval

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

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

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

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

Prediction 4. *[Confirming hypothesized impact of the intervention on beliefs] An individual who receives the retrieval intervention will predict higher expected spending on previously-neglected expenses.*

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

Prediction 5. *[Changes in perception of budget] The intervention will increase the individual's immediate perceived shadow price of money (the amount of utility gained from a marginal increase in wealth from her plan) and therefore lower her willingness to pay for discretionary goods.*

The third prediction concerns the long-term impact of these changes. Following the second prediction, an individual who appreciates more future expenses will spend and consume

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

less in order to save for these now-retrieved expenses. At some point in the future, these savings will allow her to spend more on consumption and expenses.

Prediction 6. *[Changes in savings, consumption, expenditures trajectory] The intervention will cause an individual to have a flatter average spending profile (lower initial spending and higher later spending). The individual will have weakly higher average savings throughout the cycle.*

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

3.3.4 Retrieval Failures and Quasi-Hyperbolic Discounting

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

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

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

3.4 Intervention and Study Sample

3.4.1 Intervention

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

As discussed in Section 3.3, to construct our intervention, we draw on a well-documented insight in the psychology literature: it is easier to recall items when they are grouped associatively in categories (Kahana, 2012; Bordalo et al., 2023). For example, an individual is more likely to remember that she buys laundry detergent three times a year if she is asked to recall "household supplies" rather than asked to recall "expenditures" as a whole. Consistent with this, the planning fallacy literature robustly documents that thinking in disaggregated categories tends to increase forecasts—a phenomenon referred to as the "segmentation effect" (Kahneman and Tversky, 1977; Buehler et al., 2010).

We leverage these insights in our intervention. We design an "expense board" that prompts individuals to think through their expenses category-by-category (see Figure 3.2). Using preparatory fieldwork, we identify seven major categories of spending: food (maize allocated to food expenses in each month of the year) and six non-food expense categories (school fees, household supplies, farm inputs, transfers to others, health shocks / other emergencies, and a residual "other" category). In selecting these seven categories, our goal is to design cues that are specific enough to assist with associative recall, but broad and general enough that every household could be expected to have positive expenses within each category. This helps avoid concerns that the categories convey information or normative guidance.[18] In addition, having a relatively small number of categories helps prevent fatigue and keeps the exercise tractable.

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

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

in beliefs: thinking through the categories in the expense board will have a disproportionate effect on the retrieval of *non-food expenses* (Intervention Prediction 1).

We undertake an exercise with individuals in the treatment group using this physical expense board (Figure 3.2, Panel A). To promote cognitive engagement, we provide individuals with thumbtacks that equal the number of bags of maize they currently have in savings. We then ask individuals to allocate these thumbtacks across categories on the expense board, in order to depict their spending plan for the coming year.

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

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

3.4.2 Sample and Summary Statistics

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

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

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

The 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 (a concern we discuss in more detail in Section 3.7.3). All study activities are conducted by enumerators at the participant's home.[20]

The demographic statistics are broadly the same for the mechanism and field experiments, as shown in Appendix Table C.2.1. Participants are on average around 44 years old, and most grew up on farms—indicating many years of experience in solving the annual "eat-the-pie" problem. Around 80% of our sample is male (female participants are often unmarried heads of household). Participants have on average around 15 bags of maize remaining from their harvest, which is around 50% of the total value of their maize harvest.[21]

3.5 Mechanism Experiment

3.5.1 Design

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

Specifically, treatment participants complete the full expense board with 6 non-food categories and 12 monthly food categories as discussed in the previous section and shown in Panel A of Figure 3.2. Control participants instead use a "placebo" board with only one non-food expense category and one food category, shown in Panel B of Figure 3.2. This design ensures that both treatment and control participants undertake an exercise with the same mechanics—articulating a spending plan where income equals expenditures—with the difference only in the degree of segmentation, i.e., the extent to which the categories will be useful for cuing retrieval.

Under the null hypothesis that individuals can fully retrieve information from their memory, there should be no difference in the spending plan reflected on the two types of boards. In

the hungry season (criteria (iii)) (see Fink et al., 2020).

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

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

contrast, if individuals face retrieval failures, then thinking through expenses with finer categories will promote associative recall, leading to increased retrieval—particularly for items that are more subject to retrieval failures. As discussed above, we predict this will lead to a disproportionate treatment effect on forecasted *non-food expenses* compared to forecasted food expenses.

3.5.2 Implementation

Figure 3.3 displays the timeline of activities for the mechanism experiment. First, we elicit all participants' prior of how much they will spend on non-food expenditures over the coming year.[22] Second, treatment and control farmers undertake the retrieval exercise: allocating their available savings (i.e., maize) using the full expense board or placebo board, depending on their treatment status. Third, we examine short term treatment effects using a willingness to pay exercise. Finally, we have the control group list their recalled expenses both before and after undertaking a second budget exercise with the full treatment board. This offers a within-person comparison of the different expenses that are retrieved using the placebo board and the full expense board. We use this to characterize the features of the items that are most subject to retrieval failures in our setting. Further details are provided in Appendix C.4.

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 3.4.2). Appendix Table C.2.1 shows that the treatment and control groups are balanced on most baseline covariates.

3.5.3 Results

Intervention Prediction 1: Increase in Perceived Expenses

Intervention Prediction 1 from our model is that individuals in the treatment group will forecast higher expected spending on previously neglected expenses due to improved retrieval. We test this prediction by comparing the number of bags allocated to non-food expenses

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

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 3.4. At baseline, the prior is the same on average across the treatment and control groups. Relative to the placebo board, the full treatment board substantially increases expected non-food expenses: the treatment group's allocation to non-food expenses is 38% higher (p <0.001) (see Appendix Table C.2.3). Furthermore, the distribution of the share of bags allocated to non-food expenses in the treatment group effectively stochastically dominates that of their prior and of the control (Appendix Figure C.1.3, Panel A).

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

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

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

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

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

Consistent with our theoretical prediction, the distribution in the control stochastically dominates that in the treatment (Figure 3.5, Panel A). On average, the willingness to pay in the treatment group is 36% lower than in the control (p <0.001) (Appendix Table C.2.4). This result is robust to including controls for baseline characteristics, item fixed effects, or a Tobit specification. These findings suggest that the intervention changed people's perceptions of their future expenses and made them feel "poorer," at least in the short term.

Types of Expenses Associated with Retrieval Failures

In the final step of the mechanism experiment, we aim to shed light on what types of expenses are most subject to retrieval failures. To achieve this goal, we add a set of steps for the *control group only*. First, control participants are asked to list all individual items they considered when constructing their allocation to the "non-food" category on the placebo expense board. For each item listed, the surveyor asks the amount of the expense, the time when the expense is expected to arise, the expected frequency of the expense, and the degree of certainty of the expense. Then, the control group undergoes the full retrieval exercise (i.e., with all 6 categories of non-food expenses). After completing the full expense board, the same follow up questions about items in the non-food categories are asked again.

Completing the full retrieval exercise causes the control group to increase their expected non-food expenses by 37.5% (Appendix Table C.2.3, column 2), which is very close to the between-subject treatment effects above. By examining which items are added to expenses when going from the placebo board to the full board, we can characterize what kinds of items were initially neglected. Table 3.1 provides the results of regressing whether an expense was forgotten on different characteristics of that expense. Expenses that are small, infrequent, irregularly-timed, and uncertain are more prone to retrieval failures. These results are broadly consistent with what one might expect under a cognition-based mechanism for retrieval failures, and particularly under imperfect memory (Bordalo et al., 2023).

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

3.6 Field Experiment

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

3.6.1 Experimental Design

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

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

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

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

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

[27]Specifically, all participants across both groups is shown the labels, and each expenditure category (the non-food expenses and consumption for each month of the year) is then explained to the participant. They

enables us to examine the impact of the expense board alone over a substantial time horizon, before labels are introduced.

3.6.2 Implementation

The field experiment activities were conducted between late August 2019 and Sept 2020.[28] Figure 3.6 provides an overview of the field experiment timeline and activities. At the start of implementation, most households had just completed shelling their maize. The intervention is embedded in the baseline survey (Visit 1). All participants take a brief baseline survey, which includes information about baseline savings (e.g., maize) and other demographic variables. As in the mechanism experiment, they are asked for their "prior" forecast of non-food expenses. The treatment group completes the retrieval exercise using the full expense board shown in Panel A of Figure 3.2. 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 piece of clothing, solar panel, or radio—which they would like a chance to purchase later; we then elicit the willingness to pay at the end of the Visit 1 survey to test Intervention Prediction 2. For the treatment group, the retrieval intervention takes around 45 minutes, and the entire survey (i.e., all Visit 1 activities) takes about 90 minutes.

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

Our first follow up is 2 months after the baseline (Visit 2), when we collect data on savings (maize and cash), expenditures, and labor supply. We use the outcomes collected in this visit to test for the impacts of the retrieval exercise on consumption smoothing behavior before labels are provided. For participants that chose to take-up the labels at the end of the

are then told that if they want, they can attach these labels to their maize bags, to mark what they thought they would spend on each category. Consequently, if there are no retrieval failures but the labels provide soft commitment to sophisticates, then both treatment and control groups should be able to benefit from them equally.

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

baseline survey, the enumerator offers to attach labels to their maize bags; the participant chooses which labels to attach to their remaining maize bags.[29] We complete two additional visits between December and March (Visits 3 and 4), and again measure savings (maize and cash), expenditures, and labor supply.[30] Because planting begins in December, we also collect basic data on farm investment during these two visits. Data collection was paused in March 2020 due to the COVID-19 pandemic.

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

The field experiment was run with 837 farmers. Participants were drawn from 113 villages, sampling up to 14 households per village (using the sample selection criteria described in Section 3.4.2). We randomize at the individual (rather than village) level in order to improve statistical power. However, this design choice generates some scope for spillovers between participants—for example, control participants may learn about the intervention from treatment households, or may pressure them to share their extra savings during the hungry season. Note that such spillovers would only dampen our measured treatment effects, and only those 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 C.2.1). Appendix C.4 describes the protocols for the field experiment in detail.

3.6.3 Intervention Predictions 1 and 2: Immediate Effects

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

First, as in the mechanism experiment, the intervention increases participants' forecasted expenses. In the mechanism experiment, we are able to compare forecasts in the treatment and control group. Here, since the control group does not complete a retrieval exercise, we instead rely on a comparison within the treatment group between the stated prior non-food expenses and the sum of the non-food categories in the retrieval exercise. As shown in

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

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

Panel B of Figure 3.4, the intervention increases the allocation to non-food expenses by 60% ($p < 0.001$), consistent with Intervention Prediction 1. Furthermore, as in the mechanism experiment, the distribution of forecasted non-food expenses after the intervention stochastically dominates that of the baseline forecasts (i.e., the distribution of priors) (Appendix Figure C.1.3, Panel B).

Second, the intervention decreases the willingness to pay for a discretionary good. Panel B of Figure 3.5 qualitatively matches the results from the mechanism experiment. The distribution of the willingness to pay of control individuals effectively stochastically dominates that of the treatment, with an average change of 34% ($p < 0.001$), consistent with Intervention Prediction 2. This result is robust to including controls for baseline characteristics, item fixed effects, or a Tobit specification (Appendix Table C.2.4).

3.6.4 Intervention Prediction 3: Long-Term Effects on Saving

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

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

$$y_{it} = \sum_{j=1}^{3} \beta_j \mathbb{1}\left(\text{Treatment}_i \times \text{Visit}_{j(t)}\right) + \sigma_t + X_i'\theta + \varepsilon_{it} \tag{3.1}$$

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

The key coefficients of interest are the β_js. We estimate period-specific treatment effects as our preferred measure of impact for two reasons. First, we do not expect the treatment effects to be constant over time. For instance, since the group that we study face an "eat-the-pie" problem, savings should decline throughout the year and differentially between

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

treatment and control, whose levels should converge as they approach the next harvest. Second, outcomes may be non-monotonic. For instance, if the treatment group realizes their available budget is smaller than they thought, they will reduce immediate spending, resulting in more savings available to support higher spending in future periods. In that case, an average treatment effect over the year would mask heterogeneity in the response over time.

3.6.4.0.2 Results Intervention Prediction 3 states that the treatment will lead to weakly higher savings, due to lower initial spending (followed by higher spending at later dates). Figure 3.7 provides evidence consistent with this prediction. The y-axis of the figure measures total spending (on food and non-food expenditures), using a normalized version of the difference between participant savings in each visit. This effect is displayed starting in October, the beginning of Visit 2, and ending in early March, the end of Visit 4.[32] The figure documents that treatment households immediately decrease spending after the intervention, leading to increased savings. As a result, they are able to spend more in later parts of the year, with a crossing point around the end of November, after which the treatment group uses 5-10 kilograms more of maize per week, through the duration of the project. Overall, this results in a flatter spending profile across the year for treated households.

We examine the effect on the evolution of savings stocks directly in Table 3.2. Our primary specification measures savings as the sum of the amount of unprocessed maize in storage and the value of cash savings (converted to maize equivalents). Treated households held 100 kilograms more than those in the control group at the first follow up (Visit 2), an increase of around 15% relative to the control group (column 1, $p = 0.026$). This first follow up was on average 43 days after our baseline visit, and occurred *prior* to attaching labels to participants' maize bags. Consequently, this 15% treatment effect on savings in the first 43 days reflects the impact of the expense board alone, without labels.

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

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

[33]This pattern of savings suggests that treated participants were also able to delay some maize sales

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

We consider two potential concerns regarding the interpretation of our savings results. First, our savings measures require a number of assumptions to aggregate cash and grain savings. Results are robust to alternative assumptions about how to convert cash into kilograms of maize (Appendix Table C.2.6, columns 1 and 2). Second, the ideal savings measure would reflect all assets held by households that could conceivably be used as savings, since rural agricultural households save in multiple forms (Fafchamps et al., 1998). Our measure of savings, by contrast, reflects cash and unprocessed maize only. While these are the two primary vehicles for saving in this context, we may be missing some substitution of savings across asset classes. To help alleviate this concern, we perform several tests. We first incorporate processed maize into our measure of savings, to ensure that our savings outcome does not just capture substitution of the treatment group from processed to unprocessed maize (Table C.2.6, column 3). Next, we examine total expenditures by treatment on household assets, including livestock, that could conceivably function as savings (Appendix Table C.2.7).[34] These robustness checks show that, if anything, we are undercounting the savings effect: treatment participants (insignificantly) increase their net holdings of livestock by selling fewer large livestock during the hungry season.

3.6.5 Long-Term Effects on Other Outcomes

3.6.5.0.1 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 (3.1). First, we measure food consumption during each follow up visit. Following Fink et al. (2020), we record the number of meals per day consumed by adult members of the household over the past two days.

relative to the control. We observe a statistically significant delay of 11 days in the sale of the first maize bag in the treatment group, and a positive but statistically insignificant increase in the price per kilogram at the time of sale. Thus, lower early expenditures may lead to overall higher income from later maize sales, though these effects are too small to be measured in our sample.

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

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

Finally, we measure spending on other inputs (fertilizer, herbicides and pesticides) and agricultural output (total harvest quantity and value) during Visit 5, with questions covering the full agricultural cycle. Since these outcomes are measured only in a single survey, we estimate treatment effects following equation (3.1), but with one observation per participant. Overall, we report results at the frequency at which they were collected: if we collected a given outcome in multiple survey rounds, we report effects separately by round; if we only collected it once, we report one aggregate effect over the recall period.

3.6.5.0.2 Results Consistent with the patterns of expenditure smoothing, Table 3.3 suggests that treated participants slightly reduced consumption of the staple food at the first follow up survey visit, although this reduction is not statistically significant at conventional levels (column 1). The treatment group has slightly higher consumption of staple meals at the beginning of the hungry season (Visit 3, $p = 0.079$), although this difference disappears toward the end of the hungry season in Visit 4 (consistent with the savings results).[36]

Under-saving has implications not only for consumption—and therefore household welfare—but also for productive investments and future income. In our setting, planting of crops occurs after savings have begun to decline, which may affect both labor and non-labor investments. Table 3.3 indicates that the increased savings induced by the treatment reduced the likelihood of selling household labor (ganyu): during the hungry season, treated households engaged in 32% fewer days of wage labor on others' farms , relative to the control group mean (column 2) ($p = 0.043$), and we find no corresponding increase earlier in the year. Note that this helps explain some of the muted impacts on hungry season consumption: the increased wage earnings among the control group (column 3) are used to purchase food ($p = 0.008$).

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

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

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

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

3.7 Alternative Explanations

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

3.7.1 Reminders and Increased Salience of the Need to Save

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

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

Determining the precise impact of the labels is more difficult, since they were differentially adopted by the treatment group. The labels were designed to help participants memorialize the results of the retrieval exercise. However, the labels might act as a simple nudge to consider spending decisions more carefully. There are two reasons we believe that the latter explanation is unlikely to play a large role. First, past literature demonstrates that the impact of attentional reminders is typically small (Gabaix, 2019) and short-lived (Carrera et al., 2018). And, very similar to our setting, Burke et al. (2019) cross-randomize a maize bag labeling intervention in Kenya and find little effect on behavior. Second, in the field experiment, a follow up survey documents that treated farmers perceive the impact of the labels as largely tied to the intervention: less than 2% report that the labels would be helpful in the absence of the retrieval exercise (Appendix Figure C.1.4, Panel A) and participants value the labels with the intervention at nearly 10 times their value for labels alone (Appendix Figure C.1.4, Panel B).[37]

3.7.2 Present Bias and Soft Commitment

As we discuss in Section 3.3.4, models of present bias have a maintained assumption—common to most economic models—that individuals have unconstrained access to information that they "know." Therefore, these models would predict that an intervention that manipulates only retrieval will have no impact on perceptions of expenses or immediate behavior, contrary to our findings.

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

While the field experiment was designed to target retrieval, the treatment group was differentially exposed to other components, such as spending more time with an interviewer, discussing a consumption plan, and potentially receiving labels. One might argue that some of these additional components could be used as a "soft" commitment device by partially-

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

sophisticated present-biased participants to incentivize their future selves to take actions that better align with current preferences.

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

3.7.3 Intrahousehold Coordination

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

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

3.7.4 Demand Effects

Some of our results may be susceptible to experimenter demand effects. For instance, while our retrieval intervention was constructed carefully to avoid making normative statements or suggesting allocations, treated participants may have perceived some demand to reduce immediate spending, leading to dampened valuations in the willingness to pay exercise. It seems difficult, however, to square this mechanism with other results. For instance, it is unclear why treated participants would perceive experimenter demand to spend more on non-consumption items (Appendix Figure C.1.3). If anything, the perceived demand in this context would probably be to save more for future food consumption, which would suggest that beliefs would shift in the opposite direction, toward allocating a *smaller* proportion of their maize stock toward future non-food expenses. More specifically, to the extent that finer categorization itself is a signal of experimenter preferences, the full treatment board has 12 food categories (one for each month) and 6 non-food categories. Consequently, in the mechanism experiment, when participants go from the placebo board to the full board, there is a larger increase in food categories; it is therefore unclear why the full expense board would signal demand for lower food consumption.

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

Finally, it seems implausible that a demand effect alone would generate substantial (objectively observed) savings increases over a 3-6 month period, as we show in Table 3.2.

3.8 Extensions: The Persistence of Biased Beliefs

Figure 3.1 documents remarkably skewed beliefs among highly experienced agents. Our study focuses on one particular explanation that can generate such bias in beliefs: retrieval failures. The main focus of our study is to test for the presence of retrieval failures, and their resultant impacts on consumption smoothing. In this section, we go beyond this core focus, and present suggestive evidence on additional mechanisms for why biased beliefs may persist despite experience.

To motivate this line of enquiry, note that, even under the presence of retrieval failures, the patterns in Figure 3.1 still present a puzzle: even if individuals do not remember all their future expenditures, they could realize that they always run out of maize earlier than expected, and debias their beliefs about future savings this way. In line with this, the planning fallacy literature highlights three ways in which individuals could debias their beliefs about the future. The first, analytical forecasting, is thinking through all components of the problem (i.e., all future expenses) and computing the correct forecast. Our paper demonstrates that, due to retrieval failures, people do not do this perfectly.

Second, alternatively, individuals could use recall based forecasting: thinking about their own past history (i.e., when maize has run out in past years) to form a more accurate assessment of the future (i.e. at what point maize is likely to run out this year). We investigate whether individuals learn from past experience in two ways. We first look in the cross-section at whether individuals with more experience—proxied by age—hold more accurate beliefs. 85% of individuals in the lowest age quartile are overoptimistic, compared to 73% in the top age quartile, where the mean age in each quartile group is 28 and 62, respectively (Appendix Figure C.1.5).[38] Therefore, while there is some suggestive evidence for updating over decades of experience, experience appears insufficient to eliminate the overoptimistic bias in forecasts.

Why aren't people learning from their own experience? We gather additional evidence by returning to our field experiment sample in September 2020, one year after our initial intervention. At this time, we ask participants to recollect their savings forecasts and realized savings in the preceding year. We find evidence for systematically biased memory, where individuals recall the past as being better than it was (Appendix Figure C.1.6). More than 70% of participants recall having more maize savings than they actually did at Visit 3. This difference is meaningful: the average recall bias can explain 80% of the average forecast error, meaning that if participants actually had the number of bags they recalled, their forecast error at baseline would be only 20% of the size observed.[39] Note that while participants exhibit rosy memory, they do not completely disregard the hungry season: on average, they recall having 52% less maize in Visit 4 (hungry season) than in Visit 3 (early hungry season), while the actual decline in maize was 63.6%. This bias in memory is consistent with psychology literature on the planning fallacy (Roy et al., 2005; Griffin and Buehler, 2005), and may slow learning but is too small to fully explain the beliefs we document.

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

[38]Appendix Figure C.1.5 shows the cumulative distribution function of forecast error—(incentivized) forecasts of future savings minus realized savings—for the bottom and top age quartiles.

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

[40]To be precise: let the participants' actual forecast about future savings at baseline be $\hat{S}_3$, i.e., it is their forecast at baseline of how many bags of maize they would have in storage at Visit 3. We represent realized

findings together, we can compare a participant's recollection of their forecast error, to their actual forecast error. Appendix Figure C.1.7 shows participants' actual forecast error, as measured by baseline forecasts of future savings less actual savings, and recalled forecast error, as measured by recall of baseline forecast less recalled of savings. This measure shows that participants appear somewhat naive about their overoptimism: while more than 80% of participants exhibit positive forecast errors, fewer than 40% recall that their forecasts were overoptimistic. Together, these pieces of evidence shed light on why experience alone is not enough to debias over-optimism.

The third potential approach to debiasing beliefs is reference-based forecasting: using the experiences of others similar to oneself to form more accurate beliefs. During the same follow-up final survey visit, we ask questions about own forecasted savings and, later in the survey, about the forecasted savings of other households like their own. We find participants place themselves at the extremes of the distribution: more than 50% of participants forecast that they will have more maize than all ten similar households for whom they provided forecasts (Appendix Figure C.1.8). While this exercise was not incentivized, it is consistent with other experiments that find evidence of asymmetric naivete (Fedyk, 2018), and with literature on the planning fallacy that shows that people underestimate their own task completion time but not that of others (Buehler et al., 1994).

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

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

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

---

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 $\bar{\hat{S}}_3 - \bar{S}_3$.

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

Before showing results from a segmentation exercise, we first report respondents' perceptions of how income and expenses match their forecasts in Figure 3.8. When presented with the scenario, "There are months where *more* expenses come up than I had initially expected," only 13% say this happens "rarely or never". In contrast, 59% rarely or never have *less* expenses come up than expected. In other words, when it comes to expenses, there is directional bias in the realization relative to expectations: individuals are generally more likely to be unpleasantly "surprised" (Panel A). In contrast, we see no such substantive asymmetry for income (Panel B). Finally, similar to the pattern for expenses, individuals state they often end up with less savings than they anticipated (Panel C). These patterns roughly match what one would expect under retrieval failures.

We then examine whether engaging in a segmentation exercise—thinking through categories—affects beliefs about income and expenses. First, we collect priors: we ask participants to forecast total expenses for the coming month. We also collect forecasted monthly income. Second, participants undergo a category-based elicitation for expenses and for income. Finally, we ask participants make revised forecasts for both income and expenses.[41] Example screenshots from the survey are shown in Appendix Figure C.1.9.

Figure 3.9 shows the resultant impact on beliefs from the segmentation exercise: the CDFs of the posterior, relative to the prior, for expenses and incomes. Consistent with Intervention Prediction 1 of the model, we observe that 58% of the sample revises expenses upward, with a mean change in forecasted expenses of 13%. While the mean update in forecasted income is also positive, it is substantively smaller: the mean change is 4.6%, and the modal change is 0%, with 59% of respondents having no change in their income forecast after the segmentation exercise. The p-value of a test of whether the mean change in income equals the mean change in expenses is ¡0.01).

These results offer suggestive evidence that (i) overoptimism in forecasted expenses and savings, but not in income, is common among low income Americans, and (ii) fine categories aid retrieval of expenses, and—to a lesser degree—income. Notably, when asked to explain the divergence between their prior and posterior estimates, 70% of respondents say that

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

forgetting expenses was an "important" or "very important" reason for the discrepancy. Tracking longer term outcomes in this population would be more challenging than in our field setting in Zambia, given the diversity of income flows and additional complexity of expenditures. However, these findings suggest that similar retrieval interventions may yield consumption smoothing benefits across a range of settings.[42]

3.10 Discussion and Conclusion

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

We see these failures as a consequence of the fundamental limits of the human retrieval system. Although we can imagine a role for willful neglect of the budget problem, our results are somewhat inconsistent with standard models of optimally-chosen motivated beliefs (e.g., Bénabou and Tirole, 2016).[43] Even if individuals are intentionally shifting their beliefs, our results show that the structure of the memory system shapes the form of manipulation. For example, it is unclear why someone who desires to willfully neglect expenses would be impacted by looking at finer categorizations, unless those categorizations somehow force undesired retrieval. Similarly, it is unclear why unconstrained willful manipulation would lead to misperception of small, irregular and uncertain expenses rather than of larger expenses or income. This suggests that any form of willful neglect would involve intentionally avoiding thinking deeply about the problem (Bolte and Raymond, 2022), which leads to specific retrieval failures arising from an imperfect retrieval system. Moreover, if individuals were

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

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

holding motivated beliefs, then we might expect under-estimation of expenses to be accompanied by over-estimation of income—contrary to our findings in the US sample. Finally, even if individuals are willfully refusing to think about their expenses, our results challenge the notion that the they are optimally trading off the benefits of distorted beliefs with the costs of distorted behavior: although precise welfare statements are naturally challenging, the utility costs of budget distortions are substantial in our context and our evidence suggests that the benefits of more comprehensive retrieval are large.

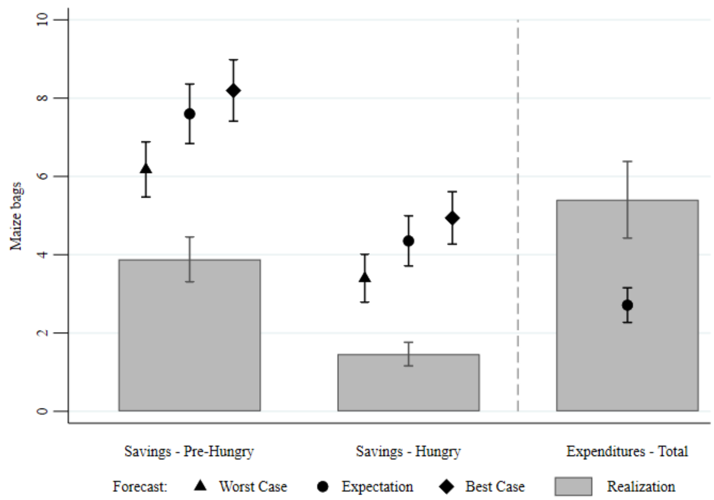
If individuals are intentionally avoiding thinking about their expenses, they may have preferred to avoid the retrieval intervention. We investigate this in follow-up data collection when we return one year after the initial intervention. We ask treatment farmers their willingness to pay to receive the intervention again. Over 90% are willing to pay for the intervention, and we find no evidence of a desire to avoid it in other qualitative questions. This suggests that participants find the intervention welfare improving, but also have difficulty replicating or completing it on their own (perhaps due to a range of other issues, including present bias).

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

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

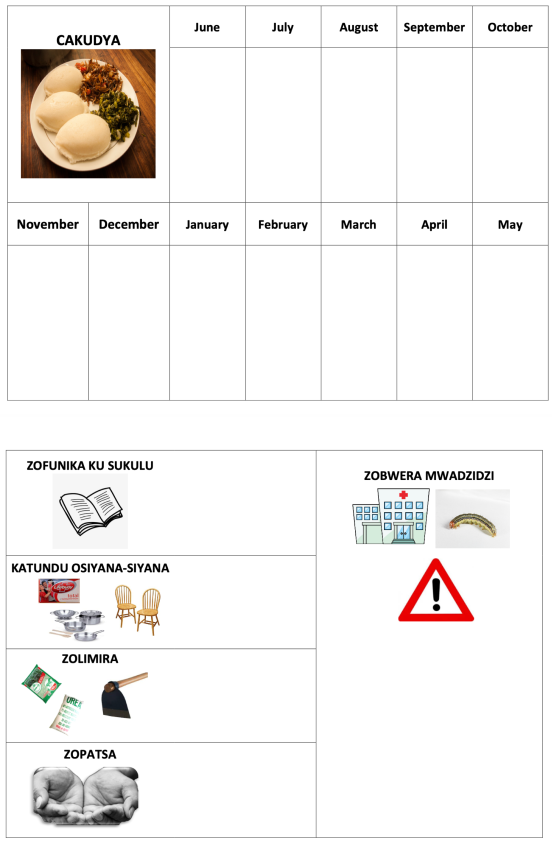
3.11 Tables and Figures

Figure 3.1: Overoptimism in savings and expenditure forecasts (field experiment)



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

Figure 3.2: Expense board and categories

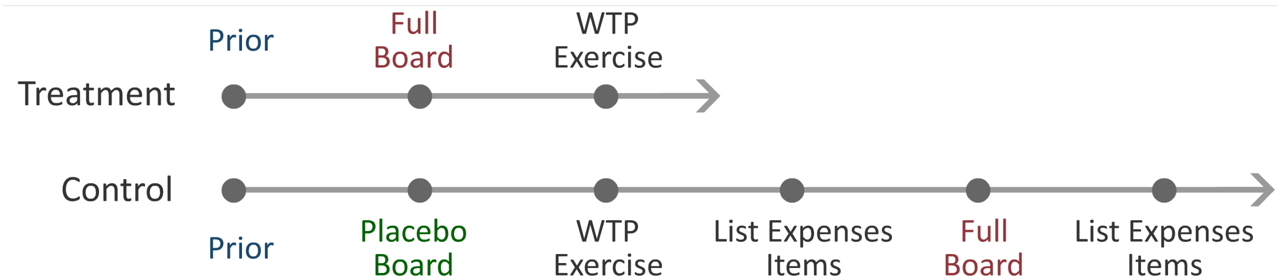


(a) Panel A: Full expense board

(b) Panel B: Two category expense board

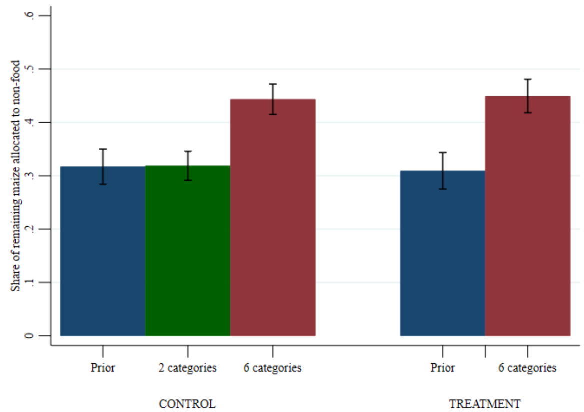
Notes: Panel A: The full expense board used by the treatment group to allocate their savings to expense categories. As part of the treatment, participants receive a set of pins, with each pin representing one bag of maize from their savings. Participants assign pins to food consumption (allocated by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

Figure 3.3: Timeline (mechanism experiment)

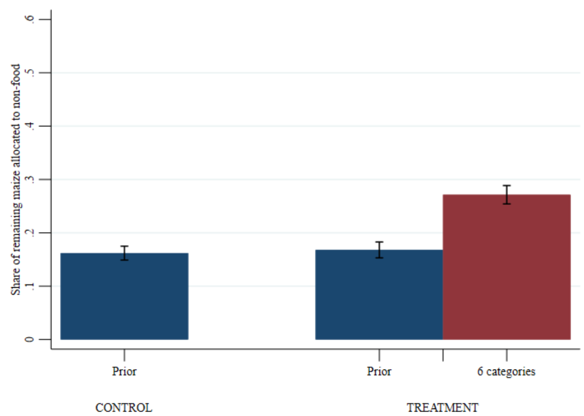


Notes: Mechanism experiment timeline. All participants are asked for their prior of non-food expenses. They are then randomized into treatment and control groups. The treatment group completes the expense board in Panel A of Figure 3.2 and the control group completes the board in Panel B. Both groups then complete the willingness-to-pay exercise. Finally, the control group goes through an additional set of activities to identify the previously-neglected expense items.

Figure 3.4: Share of maize bags allocated to non-food expenses



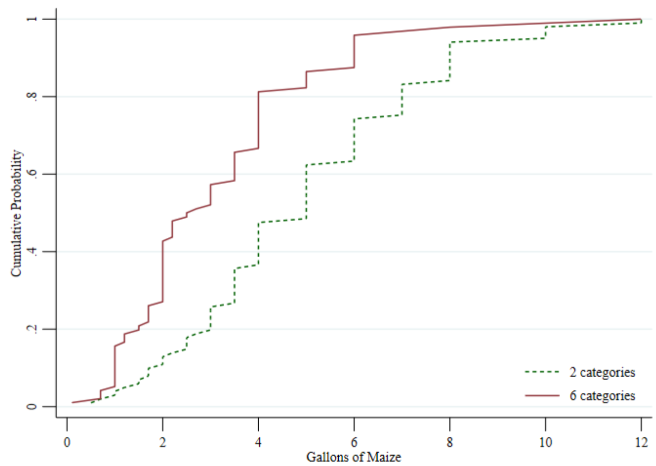
(a) Panel A: Mechanism experiment



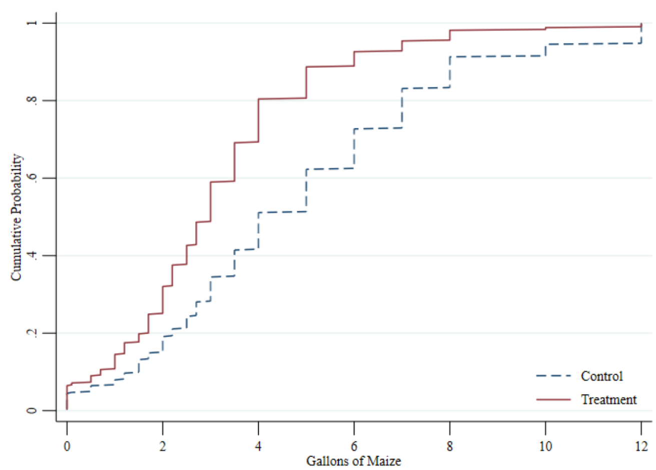
(b) Panel B: Field experiment

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

Figure 3.5: Willingness to exchange maize for discretionary consumption



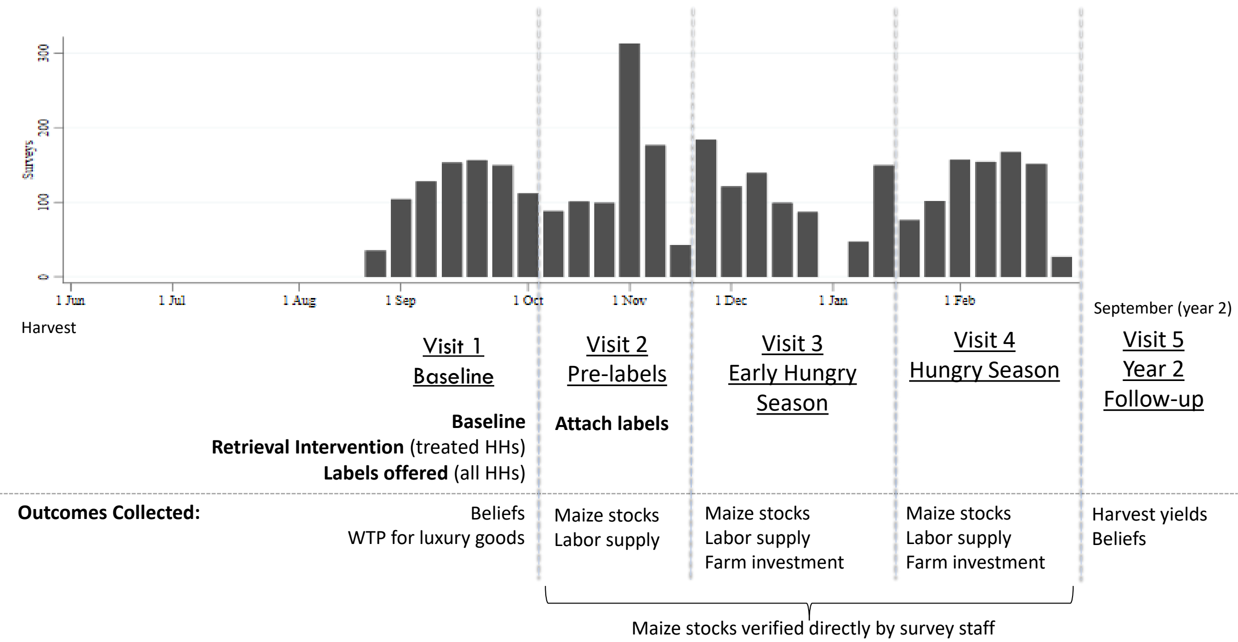
(a) Panel A: Mechanism experiment



(b) Panel B: Field experiment

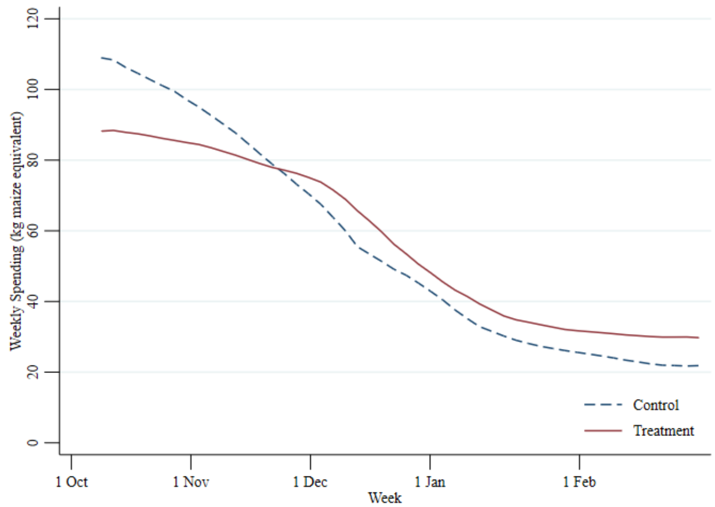
Notes: CDF of the willingness to pay (WTP) for a discretionary consumption items. WTP is elicited using the Becker-DeGroot-Marschak method after the retrieval exercise. In Panel A the green dashed line shows the WTP in the control group (two-category exercise), and the maroon solid line shows the WTP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

Figure 3.6: Timeline and data collection (field experiment)



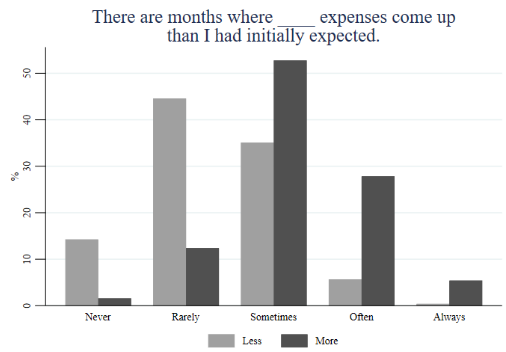
Notes: Field experiment timeline, including the intervention and data collection. Vertical bars represent the number of participants interviewed each week over the course of the study, by data collection rounds. The intervention was administered at the same time as the baseline survey (Visit 1). Labels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

Figure 3.7: Implied consumption path based on observed savings (field experiment)

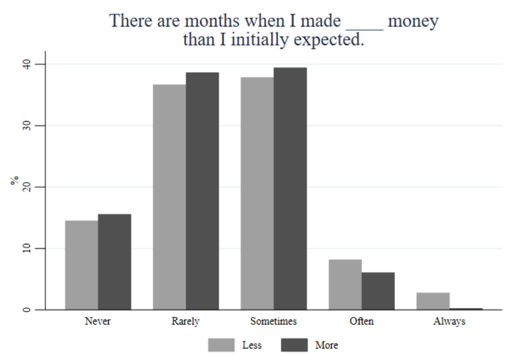


Notes: Smoothed consumption paths for treatment (maroon solid line) and control (blue dotted line) participants in the field experiment. The dependent variable is constructed as the difference in savings (measured as kilograms of maize plus the maize value of cash savings), divided by the number of weeks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

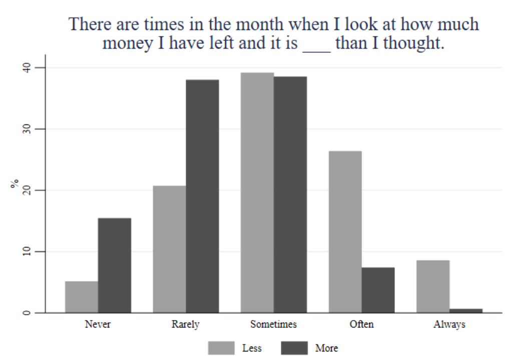
Figure 3.8: Asymmetry in income and expenses (U.S. sample)



(a) Panel A: Expenses



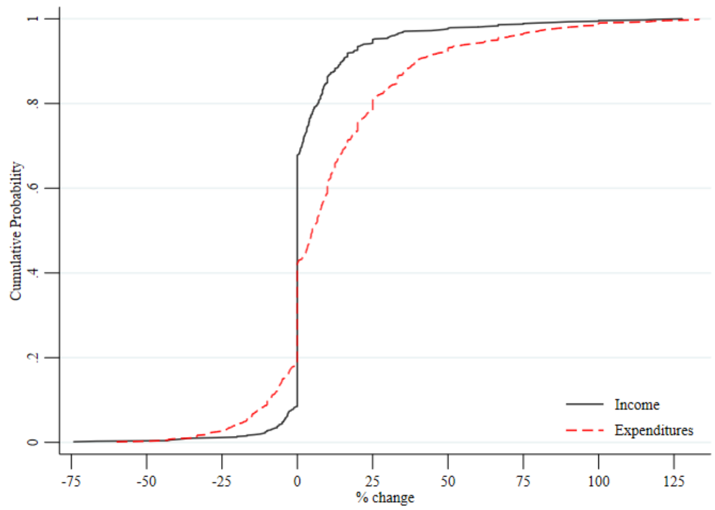
(b) Panel B: Income



(c) Panel C: Savings

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

Figure 3.9: Percentage change before/after retrieval exercise (U.S. sample)



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

Table 3.1: Characteristics of forgotten expenses (mechanism experiment)

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

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

Table 3.2: Impact of the retrieval exercise on savings (field experiment)

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

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

Table 3.3: Impact of the retrieval exercise on consumption and investment (field experiment)

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

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

Chapter 4

Dissertation Conclusion

This dissertation focuses on three different market failures in low income countries, and measures their contribution to low firm and worker productivity, alone or through their interaction with psychological biases.

Chapter 1 focused on the market for training. Through a field experiment it showed that a lack of long term labor contracts might preclude the formation of this market. In the second field experiment, it was shown that missing markets for formal insurance might lead firms to higher related, rather than unrelated, employees, as a means of informal redistribution. Finally, the third chapter showed that in a market with recourse for borrowing, a psychological bias in which individuals might fail to recall all future expenditure may lead to under-saving and over consumption.

Bibliography

1AF, O. A. (2016). Comprehensive impact report. *A Decade of Measurement and Impact. URL: https://oneacrefund.org/documents/15/Comprehensive_Impact_Report_One_Acre_Fu nd. pdf [02/2022].*

Abel, M., Burger, R., Carranza, E., and Piraino, P. (2019). Bridging the intention-behavior gap? the effect of plan-making prompts on job search and employment. *American Economic Journal: Applied Economics*, 11(2):284–301.

Acemoglu, D. (1997). Training and innovation in an imperfect labour market. *The Review of Economic Studies*, 64(3):445–464.

Acemoglu, D. and Pischke, J.-S. (1998). Why Do Firms Train? Theory and Evidence. *The Quarterly Journal of Economics*, 113(1):79–119.

Acemoglu, D. and Pischke, J.-S. (1999a). Beyond becker: Training in imperfect labour markets. *The economic journal*, 109(453):112–142.

Acemoglu, D. and Pischke, J.-S. (1999b). The structure of wages and investment in general training. *Journal of political economy*, 107(3):539–572.

Adhvaryu, A., Bassi, V., Nyshadham, A., Tamayo, J., and Torres, N. (2023). Organizational responses to product cycles. *Available at SSRN 4403515.*

Adhvaryu, A., Gauthier, J.-F., Nyshadham, A., and Tamayo, J. A. (2021). Absenteeism, productivity, and relational contracts inside the firm. Technical report, National Bureau of Economic Research.

Aggarwal, S. (2018). Do rural roads create pathways out of poverty? evidence from india. *Journal of Development Economics*, 133:375–395.

Aguiar, M. and Hurst, E. (2005). Consumption versus expenditure. *Journal of Political Economy*, 113(5):919–948.

Aker, J. C. and Jack, K. (2021). Harvesting the rain: The adoption of environmental technologies in the sahel. Technical report, National Bureau of Economic Research.

Akyeampong, E., Bates, R. H., Nunn, N., and Robinson, J. (2014). *Africa's development in historical perspective*. Cambridge University Press.

Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., and Vitali, A. (2020). Tackling youth unemployment: Evidence from a labor market experiment in uganda. *Econometrica*, 88(6):2369–2414.

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

Atkin, D., Chaudhry, A., Chaudry, S., Khandelwal, A. K., and Verhoogen, E. (2017). Organizational barriers to technology adoption: Evidence from soccer-ball producers in pakistan. *The Quarterly Journal of Economics*, 132(3):1101–1164.

Autor, D. H. (2001). Why do temporary help firms provide free general skills training? *The Quarterly Journal of Economics*, 116(4):1409–1448.

Bandiera, O., Barankay, I., and Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, 120(3):917–962.

Bandiera, O., Barankay, I., and Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4):1047–1094.

Bandiera, O., Barankay, I., and Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, 11(5):1079–1114.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2017). Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, 132(2):811–870.

Bank, W. (2007). *World development report 2008: Agriculture for development*. The World Bank.

Barrett, C. B., Islam, A., Mohammad Malek, A., Pakrashi, D., and Ruthbah, U. (2022). Experimental evidence on adoption and impact of the system of rice intensification. *American Journal of Agricultural Economics*, 104(1):4–32.

Basu, K. and Wong, M. (2015). Evaluating seasonal food storage and credit programs in east indonesia. *Journal of Development Economics*, 115:200–216.

Beaman, L. (2016). Social networks and the labor market.

Beaman, L., BenYishay, A., Magruder, J., and Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *American Economic Review*, 111(6):1918–1943.

Beaman, L. and Magruder, J. (2012). Who gets the job referral? evidence from a social networks experiment. *American Economic Review*, 102(7):3574–3593.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Beekman, G. and Bulte, E. H. (2012). Social norms, tenure security and soil conservation: evidence from burundi. *Agricultural systems*, 108:50–63.

Behaghel, L., Gignoux, J., and Macours, K. (2020). Social learning in agriculture: does smallholder heterogeneity impede technology diffusion in sub-saharan africa?

Behrman, J. R. (1999). Labor markets in developing countries. *Handbook of labor economics*, 3:2859–2939.

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

BenYishay, A. and Mobarak, A. M. (2019). Social learning and incentives for experimentation and communication. *The Review of Economic Studies*, 86(3):976–1009.

Berman, J. Z., Tran, A. T., Lynch Jr, J. G., and Zauberman, G. (2016). Expense neglect in forecasting personal finances. *Journal of Marketing Research*, 53(4):535–550.

Bertrand, M. and Schoar, A. (2006). The role of family in family firms. *Journal of economic perspectives*, 20(2):73–96.

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

Blattman, C. and Dercon, S. (2018). The impacts of industrial and entrepreneurial work on income and health: Experimental evidence from ethiopia. *American Economic Journal: Applied Economics*, 10(3):1–38.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? evidence from india. *The Quarterly journal of economics*, 128(1):1–51.

Bolte, L. and Raymond, C. (2022). Emotional inattention. Technical report, Working paper.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., and Shleifer, A. (2023). Memory and probability. *Quarterly Journal of Economics*, 138(1):265–311.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2020). Memory, attention, and choice. *Quarterly Journal of Economics*, 135(3):1399–1442.

Breza, E., Kaur, S., and Krishnaswamy, N. (2019). Coordination without organization: Collective labor supply in decentralized spot markets. Technical report, National Bureau of Economic Research.

Breza, E., Kaur, S., and Shamdasani, Y. (2021). Labor rationing. *American Economic Review*, 111(10):3184–3224.

Bridle, L., Magruder, J., McIntosh, C., and Suri, T. (2020). Experimental insights on the constraints to agricultural technology adoption.

Brown, G., Hardy, M., Mbiti, I. M., Mccasland, J. L., and Salcher, I. (2022). Can financial incentives to firms improve apprentice training? experimental evidence from ghana. *World bank policy research working paper*, (8851).

Brunnermeier, M. K., Papakonstantinou, F., and Parker, J. A. (2008). An economic model of the planning fallacy. Technical report, National Bureau of Economic Research.

Bryan, G., Chowdhury, S., and Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, 82(5):1671–1748.

Buehler, R., Griffin, D., and Peetz, J. (2010). The planning fallacy: Cognitive, motivational, and social origins. In *Advances in experimental social psychology*, volume 43, pages 1–62. Elsevier.

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

Burke, M., Bergquist, L. F., and Miguel, E. (2019). Sell low and buy high: arbitrage and local price effects in kenyan markets. *Quarterly Journal of Economics*, 134(2):785–842.

Bursztyn, L., Fujiwara, T., and Pallais, A. (2017). 'acting wife': Marriage market incentives and labor market investments. *American Economic Review*, 107(11):3288–3319.

Bursztyn, L. and Jensen, R. (2015). How does peer pressure affect educational investments? *The quarterly journal of economics*, 130(3):1329–1367.

Bursztyn, L. and Jensen, R. (2017). Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9:131–153.

Cafer, A. M. and Rikoon, J. S. (2018). Adoption of new technologies by smallholder farmers: the contributions of extension, research institutes, cooperatives, and access to cash for improving tef production in ethiopia. *Agriculture and human values*, 35:685–699.

Caicedo, S., Espinosa, M., and Seibold, A. (2022). Unwilling to train?—firm responses to the colombian apprenticeship regulation. *Econometrica*, 90(2):507–550.

Card, D., Ibarrarán, P., Regalia, F., Rosas-Shady, D., and Soares, Y. (2011). The labor market impacts of youth training in the dominican republic. *Journal of Labor Economics*, 29(2):267–300.

Card, D., Kluve, J., and Weber, A. (2018). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3):894–931.

Carranza, E., Donald, A., Grosset, F., and Kaur, S. (2021). The social tax: Redistributive pressure and labor supply.

Carranza, E., Donald, A., Grosset, F., and Kaur, S. (2022). The social tax: Redistributive pressure and labor supply. Technical report, National Bureau of Economic Research.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (2018). The limits of simple implementation intentions: Evidence from a field experiment on making plans to exercise. *Journal of Health Economics*, 62:95–104.

Chandrasekhar, A. G., Duflo, E., Kremer, M., Pugliese, J. F., Robinson, J., and Schilbach, F. (2022). Blue spoons: Sparking communication about appropriate technology use. Technical report, National Bureau of Economic Research.

Chandrasekhar, A. G., Morten, M., and Peter, A. (2020). Network-based hiring: Local benefits; global costs. Technical report, National Bureau of Economic Research.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

Chetty, R. and Saez, E. (2013). Teaching the tax code: Earnings responses to an experiment with eitc recipients. *American Economic Journal: Applied Economics*, 5(1):1–31.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American economic review*, 100(1):35–69.

Crossley, T. F. and Winter, J. K. (2014). Asking households about expenditures: what have we learned? In *Improving the measurement of consumer expenditures*, pages 23–50. University of Chicago Press.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33:67–80.

De Janvry, A., Sadoulet, E., and Suri, T. (2017). Field experiments in developing country agriculture. In *Handbook of economic field experiments*, volume 2, pages 427–466. Elsevier.

De Mel, S., McKenzie, D., and Woodruff, C. (2019). Labor drops: Experimental evidence on the return to additional labor in microenterprises. *American Economic Journal: Applied Economics*, 11(1):202–235.

Deaton, A., Grosh, M., et al. (1998). Consumption. *Designing Household Survey Questionnaires for Developing Countries: lessons from ten years of LSMS experience. Washington, USA. The World Bank*, pages 1–78.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The quarterly journal of economics*, 127(1):1–56.

Dercon, S. and Gollin, D. (2014). Agriculture in african development: theories and strategies. *Annu. Rev. Resour. Econ.*, 6(1):471–492.

Dercon, S. and Krishnan, P. (2000). Vulnerability, seasonality and poverty in ethiopia. *Journal of Development Studies*, 36(6):25–53.

Deutschmann, J. W., Duru, M., Siegal, K., and Tjernstrom, E. (2022). Relaxing multiple agricultural productivity constraints at scale. *Available at SSRN 4479905*.

Di Falco, S., Feri, F., Pin, P., and Vollenweider, X. (2018). Ties that bind: Network redistributive pressure and economic decisions in village economies. *Journal of Development Economics*, 131:123–131.

Di Falco, S., Lokina, R., Martinsson, P., and Pin, P. (2019). Altruism and the pressure to share: Lab evidence from tanzania. *Plos one*, 14(5):e0212747.

Dillon, B., De Weerdt, J., and O'Donoghue, T. (2021). Paying more for less: why don't households in tanzania take advantage of bulk discounts? *World Bank Economic Review*, 35(1):148–179.

Dodd, D. E. and Jolliffe, I. T. (2001). Early detection of the start of the wet season in semiarid tropical climates of western africa. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 21(10):1251–1262.

Donkor, E., Owusu-Sekyere, E., Owusu, V., and Jordaan, H. (2016). Impact of row-planting adoption on productivity of rice farming in northern ghana. *Review of Agricultural and Applied Economics (RAAE)*, 19(395-2016-24360):19–28.

Donovan, K., Lu, W. J., and Schoellman, T. (2023). Labor market dynamics and development. *The Quarterly Journal of Economics*, 138(4):2287–2325.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from kenya. *American economic review*, 101(6):2350–90.

Dupas, P. and Robinson, J. (2013). Savings constraints and microenterprise development: Evidence from a field experiment in kenya. *American Economic Journal: Applied Economics*, 5(1):163–192.

Dusabumuremyi, P., Niyibigira, C., and Mashingaidze, A. (2014). Narrow row planting increases yield and suppresses weeds in common bean (phaseolus vulgaris l.) in a semi-arid agro-ecology of nyagatare, rwanda. *Crop Protection*, 64:13–18.

Emerick, K., De Janvry, A., Sadoulet, E., and Dar, M. H. (2016). Technological innovations, downside risk, and the modernization of agriculture. *American Economic Review*, 106(6):1537–1561.

Fafchamps, M., Udry, C., and Czukas, K. (1998). Drought and saving in west africa: are livestock a buffer stock? *Journal of Development Economics*, 55(2):273–305.

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

Fentie, A. and Beyene, A. D. (2019). Climate-smart agricultural practices and welfare of rural smallholders in ethiopia: Does planting method matter? *Land use policy*, 85:387–396.

Fiala, N. (2018). Returns to microcredit, cash grants and training for male and female microentrepreneurs in uganda. *World Development*, 105:189–200.

Fink, G., Jack, B. K., and Masiye, F. (2020). Seasonal liquidity, rural labor markets, and agricultural production. *American Economic Review*, 110(11):3351–92.

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

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of political Economy*, 103(6):1176–1209.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, volume 2, pages 261–343. Elsevier.

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

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

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

Gollin, D., Lagakos, D., and Waugh, M. E. (2014). The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993.

Griffin, D. and Buehler, R. (2005). Biases and fallacies, memories and predictions: Comment on roy, christenfeld, and mckenzie (2005).

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, pages 501–522.

Guiteras, R. P. and Jack, B. K. (2018). Productivity in piece-rate labor markets: Evidence from rural malawi. *Journal of Development Economics*, 131:42–61.

Haaland, I., Roth, C., and Wohlfart, J. (2023). Designing information provision experiments. *Journal of Economic Literature*, 61(1):3–40.

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

Heath, R. (2018). Why do firms hire using referrals? evidence from bangladeshi garment factories. *Journal of Political Economy*, 126(4):1691–1746.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Hodges, R. and Maritime, C. (2012a). Postharvest weight losses of cereal grains in sub-saharan africa. *Natural Resources Institute, University of Greenwich*, 24.

Hodges, R. J. and Maritime, C. (2012b). Postharvest quality losses of cereal grains in sub-saharan africa. *African Postharvest Losses Information System*, 22.

Howard, J., Crawford, E., Kelly, V., Demeke, M., and Jeje, J. J. (2003). Promoting high-input maize technologies in africa: the sasakawa-global 2000 experience in ethiopia and mozambique. *Food Policy*, 28(4):335–348.

Hussam, R., Kelley, E. M., Lane, G., and Zahra, F. (2022). The psychosocial value of employment: Evidence from a refugee camp. *American Economic Review*, 112(11):3694–3724.

Islam, M. and Beg, S. (2021). Rule-of-thumb instructions to improve fertilizer management: Experimental evidence from bangladesh. *Economic Development and Cultural Change*, 70(1):237–281.

Jack, B. K. (2013). Market inefficiencies and the adoption of agricultural technologies in developing countries.

Jakiela, P. and Ozier, O. (2016). Does africa need a rotten kin theorem? experimental evidence from village economies. *The Review of Economic Studies*, 83(1):231–268.

Jayachandran, S. (2021). Microentrepreneurship in developing countries. *Handbook of labor, human resources and population economics*, pages 1–31.

Jeong, D. (2021). Creating (digital) labor markets in rural tanzania. *Available at SSRN 4043833*.

Jones, M., Kondylis, F., Loeser, J., and Magruder, J. (2022). Factor market failures and the adoption of irrigation in rwanda. Technical report, National Bureau of Economic Research.

Kahana, M. J. (2012). *Foundations of human memory*. Oxford University Press.

Kahana, M. J. and Wagner, A. D. (2023). *Oxford Handbook of Human Memory*. Oxford University Press.

Kahneman, D., Lovallo, D., and Sibony, O. (2011). Before you make that big decision. *Harvard Business Review*, 89(6):50–60.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4):237.

Kahneman, D. and Tversky, A. (1977). Intuitive prediction: Biases and corrective procedures. Technical report, Decisions and Designs Inc Mclean Va.

Karing, A. (2018). Social signaling and childhood immunization: A field experiment in sierra leone. *University of California, Berkeley*, 2.

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

Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural Decisions after Relaxing Credit and Risk Constraints. *The Quarterly Journal of Economics*, 129(2):597–652.

Kaur, S. (2019). Nominal wage rigidity in village labor markets. *American Economic Review*, 109(10):3585–3616.

Kline, P. and Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Kling, J. R., Mullainathan, S., Shafir, E., Vermeulen, L. C., and Wrobel, M. V. (2012). Comparison friction: Experimental evidence from medicare drug plans. *Quarterly Journal of Economics*, 127(1):199–235.

Kondylis, F., Mueller, V., and Zhu, J. (2017). Seeing is believing? evidence from an extension network experiment. *Journal of Development Economics*, 125:1–20.

Krishnan, P. and Patnam, M. (2014). Neighbors and extension agents in ethiopia: Who matters more for technology adoption? *American Journal of Agricultural Economics*, 96(1):308–327.

Kruger, L. (2016). The timing of agricultural production in hazard-prone areas to prevent losses at peak-risk periods: A case of malawi, madagascar and mozambique. *Jàmbá: Journal of Disaster Risk Studies*, 8(2):1–9.

Kuhn, M. A. (2021). Electronic benefit transfer and food expenditure cycles. *Journal of Policy Analysis and Management*, 40(3):744–773.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018). Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849.

Lazear, E. P., Malmendier, U., and Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1):136–163.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., and Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6):551.

Loewenstein, M. A. and Spletzer, J. R. (1999). General and specific training: Evidence and implications. *Journal of Human Resources*, pages 710–733.

Ma, X., Nakab, A., and Vidart, D. (2024). Human capital investment and development: The role of on-the-job training. *Journal of Political Economy Macroeconomics*, 2(1):000–000.

Macchi, E. and Stalder, J. (2023). Work rather than just cash: Informal redistribution among employers and workers in kampala, uganda.

Macchiavello, R. and Morjaria, A. (2023). Relational contracts: Recent empirical advancements and open questions. Technical report, National Bureau of Economic Research.

Malmendier, U. and Wachter, J. A. (2023). *Oxford Handbook of Human Memory*, chapter Memory of past experiences and economic decisions. Oxford Univeristy Press.

Mansingh J, P. and Deressa Bayissa, D. (2018). Awareness of improved practices of teff by smallholder farmers in chaliyadistrict, west shoa zone, ethiopia. *IMPACT: International Journal of Research in Applied, Natural and Social Sciences*, 6(2):91–98.

Mason, A., Ludvig, E., and Madan, C. (2023). *Oxford Handbook of Human Memory*, chapter Conditioning and associative learning. Oxford Univeristy Press.

McKenzie, D. (2017). How effective are active labor market policies in developing countries? a critical review of recent evidence. *The World Bank Research Observer*, 32(2):127–154.

Merfeld, J. D. and Morduch, J. (2023). Poverty at higher frequency. *Working Paper*.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, 108(26):10415–10420.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2013). Planning prompts as a means of increasing preventive screening rates. *Preventive Medicine*, 56(1):92–93.

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

Nickerson, D. W. and Rogers, T. (2010). Do you have a voting plan? implementation intentions, voter turnout, and organic plan making. *Psychological Science*, 21(2):194–199.

Niragira, S., D'Haese, M., D'Haese, L., Ndimubandi, J., Desiere, S., and Buysse, J. (2015). Food for survival: Diagnosing crop patterns to secure lower threshold food security levels in farm households of burundi. *Food and nutrition bulletin*, 36(2):196–210.

Nsubuga, D., Kabenge, I., Zziwa, A., Kiggundu, N., Wanyama, J., and Banadda, N. (2021). Improving maize shelling operation using motorized mobile shellers: A step towards reducing postharvest losses in low developing countries. In *Maize Genetic Resources-Breeding Strategies and Recent Advances*. IntechOpen.

Palacios-Lopez, A., Christiaensen, L., and Kilic, T. (2017). How much of the labor in african agriculture is provided by women? *Food policy*, 67:52–63.

Pallais, A. (2014). Inefficient hiring in entry-level labor markets. *American Economic Review*, 104(11):3565–3599.

Pallais, A. and Sands, E. G. (2016). Why the referential treatment? evidence from field experiments on referrals. *Journal of Political Economy*, 124(6):1793–1828.

Paxson, C. H. (1992). Using weather variability to estimate the response of savings to transitory income in thailand. *American Economic Review*, pages 15–33.

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

Peetz, J., Buehler, R., Koehler, D. J., and Moher, E. (2015). Bigger not better: Unpacking future expenses inflates spending predictions. *Basic and Applied Social Psychology*, 37(1):19–30.

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

Pigou, A. C. (1912). *Wealth and welfare.* Macmillan and Company, limited.

Pop-Eleches, C., Thirumurthy, H., Habyarimana, J. P., Zivin, J. G., Goldstein, M. P., De Walque, D., Mackeen, L., Haberer, J., Kimaiyo, S., Sidle, J., et al. (2011). 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(6):825.

Riley, E. (2022). Resisting social pressure in the household using mobile money: Experimental evidence on microenterprise investment in uganda.

Ritchie, H. and Roser, M. (2013). Crop yields. *Our World in Data.* https://ourworldindata.org/crop-yields.

Rosenzweig, M. R. (1988). Labor markets in low-income countries. *Handbook of development economics*, 1:713–762.

Roy, M. M., Christenfeld, N. J., and McKenzie, C. R. (2005). Underestimating the duration of future events: Memory incorrectly used or memory bias? *Psychological Bulletin*, 131(5):738.

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

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

Squires, M. (2018). Kinship taxation as an impediment to growth: experimental evidence from kenyan microenterprises. Technical report, Working Paper. London: DfID.

Stevens, M. (1994). A theoretical model of on-the-job training with imperfect competition. *Oxford economic papers*, 46(4):537–562.

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

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

Terviö, M. (2009). Superstars and mediocrities: Market failure in the discovery of talent. *The Review of Economic Studies*, 76(2):829–850.

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

Tversky, A. and Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4):293.

Udry, C. (2010). The economics of agriculture in africa: Notes toward a research program. *African Journal of Agricultural and Resource Economics*, 5(311-2016-5540):284–299.

Vandercasteelen, J., Dereje, M., Minten, B., and Taffesse, A. S. (2020). From agricultural experiment station to farm: The impact of the promotion of a new technology on farmers' yields in ethiopia. *Economic Development and Cultural Change*, 68(3):965–1007.

Vandercasteelen, J., Dereje, M., Minten, B., Taffesse, A. S., et al. (2016). Synopsis: Row planting teff in ethiopia: Impact on farm-level profitability and labor allocation. Technical report, International Food Policy Research Institute (IFPRI).

Vandercasteelen, J., Regassa, M. D., Minten, B., and Taffesse, A. S. (2014). Perceptions, impacts and rewards of row planting of teff. *Available at SSRN 2530422*.

Verwimp, P. and Muñoz-Mora, J. C. (2018). Returning home after civil war: food security and nutrition among burundian households. *The Journal of Development Studies*, 54(6):1019–1040.

Vinck, P. (2008). Comprehensive food security & vulnerability analysis burundi. *WFP*.

Wachter, J. A. and Kahana, M. J. (2019). A retrieved-context theory of financial decisions. Technical report, National Bureau of Economic Research.

Wachter, J. A. and Kahana, M. J. (2023). Associative learning and representativeness. *Working Paper*.

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

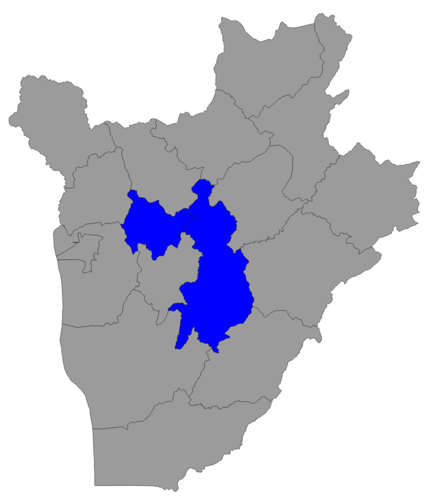
Zeweld, W., Van Huylenbroeck, G., Tesfay, G., and Speelman, S. (2017). Smallholder farmers' behavioural intentions towards sustainable agricultural practices. *Journal of environmental management*, 187:71–81.

Appendix A

Under-training by Employers in Informal Labor Markets: Evidence from Burundi

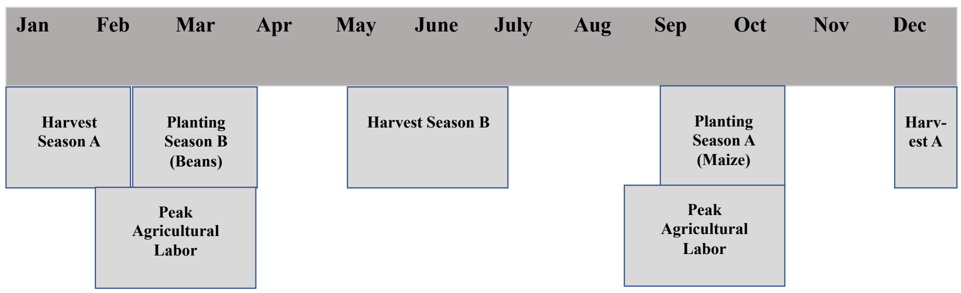
A.1 Appendix Figures

Figure A.1.1: Map of Study Area



Notes: This figure shows a map of the areas (highlighted in blue) that the study was located in, in Burundi.

Figure A.1.2: Burundian Agricultural Calendar

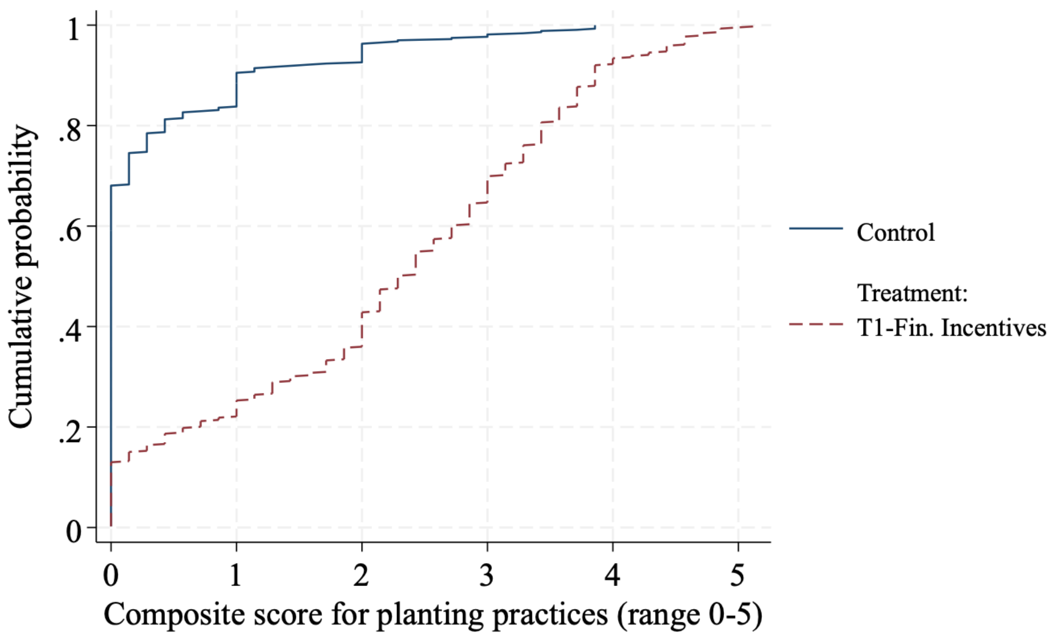


Planting Season – Time Period Per Task (Season B)

Land Preparation
Planting/Fertilizer Application – 2 weeks
Application of Tuteurs
Weeding
Harvesting

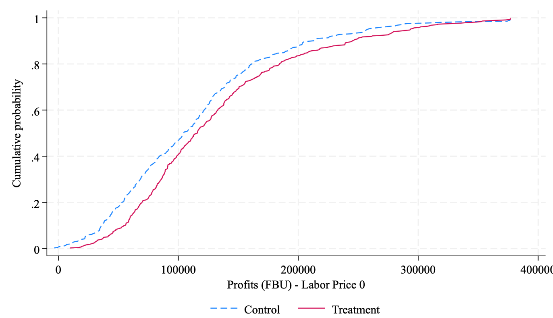
Notes: This figure shows details of the agricultural calendar and labor requirements in Burundi. The figure is based partially on a similar figure in Vinck (2008)

Figure A.1.3: Improvement of laborer's skills in incentivized task

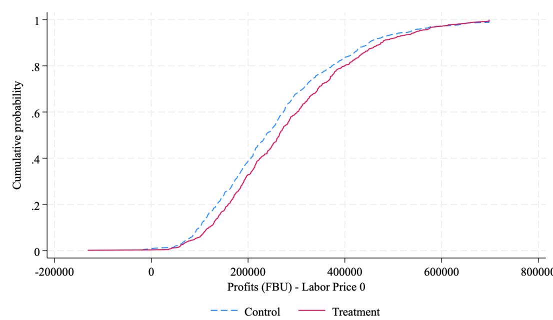


Notes: This figure shows the cumulative distribution function of scores (from 0 to 5) obtained by laborers (trainee) in an incentivized planting practice activities after training happened. The blue solid line shows the CDF of scores in the control group, the red dashed line shows the score in the treatment group. Laborers were shown a small plot of land of equal size, and given 4 minutes to plant in "modern" way. At the end, the enumerator would measure the distance between pockets and rows, which we translated in a score ranging from 0 to 5. This task tested two crucial aspects for employability of laborers: (i) accuracy of spacing between pockets, (ii) speed. Laborers were told that, conditional on performing above a certain threshold in the task, they would be entered in a lottery to win a prize of the value of one day of unskilled labor.

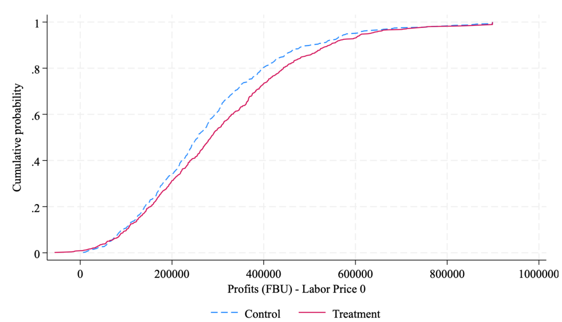
Figure A.1.4: Spillover Experiment: profit distribution



(a) Trainees



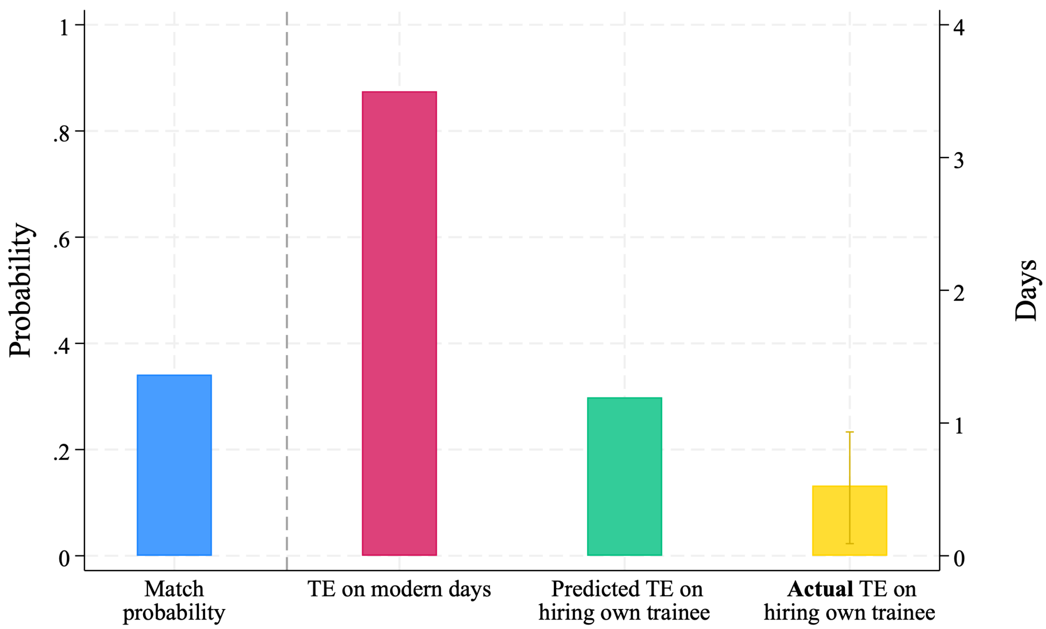
(b) Trainer-employers



(c) Spillover-employers

Notes: This figure shows the CDF of Farm profits for trainees (Panel A.1.4a) trainer-employers (Panel A.1.4b) and spillover employers (Panel A.1.4c) in the Spillover Experiment, in both T1-Financial Incentives and Control Villages. Profits are winsored at the 99th percentile. The price of family labor in these figures is assumed to be 0.

Figure A.1.5: Predicted and measured training Spillovers



Notes: This figure shows the spillovers predicted by the training intervention, and those measured. The blue bar is the days weighted churn rate as measured for control group trainees - the interpretation of this number (0.3) is that if a trainee worked for an employer in the prior year, then they spend 30% of their working days in the current year working for them. This is measured for the trainees in the second season of the experiment. The red bar is the number of additional days that trainees in the treatment group work in the labor market using the techniques that they were trained to do. The green bar is the number of days of modern practices that one would therefore predict using the control group churn rate that employers who train would be able to hire their own train for to do these techniques: it is the number in the blue bar multiplied by the number in the red bar. The yellow bar is the number of days that trainer-employers actually hired their trainee to use these techniques

A.2 Appendix Tables

Table A.2.1: Balance table - Labor Guarantee Experiment

Variable	Trainees Control Mean (1)	Trainees Treatment Mean (2)	Trainees P-value Δ (3)	Trainer-Employers Control Mean (4)	Trainer-Employers Treatment Mean (5)	Trainer-Employers P-value Δ (6)
Demographics						
Age	34.2	34.8	0.71	45.1	46.7	0.40
	(12.6)	(11.0)		(11.7)	(14.2)	
Sub-Primary Education	0.7	0.7	0.77	0.8	0.9	0.26
		(0.4)	(0.4)	(0.4)	(0.3)	
Household Size	4.9	5.1	0.43	6.0	5.3	0.02
	(2.2)	(2.0)		(2.1)	(2.0)	
Unmarried	0.2	0.2	0.43	0.0	0.0	0.43
	(0.4)	(0.4)		(0.2)	(0.1)	
Male	0.2	0.3	0.09	0.5	0.3	0.04
	(0.4)	(0.5)		(0.5)	(0.5)	
Savings (FBU)	7.9	8.5	0.31	10.8	10.9	0.72
	(4.2)	(3.9)		(2.1)	(1.9)	
Baseline farm and knowledge						
Land (plots)	4.7	5.2	0.16	9.0	8.1	0.13
	(2.2)	(3.0)		(4.7)	(4.5)	
Land (ares)	14.9	17.9	0.08	39.8	41.8	0.59
	(12.6)	(12.1)		(22.2)	(30.4)	
Beans value (Past B)	101935.3	130094.4	0.04	322141.2	294563.6	0.22
	(77877.7)	(114055.5)		(173752.9)	(141726.2)	
Fields (Past B)	4.7	5.1	0.16	7.5	7.3	0.55
	(1.9)	(2.8)		(3.2)	(2.9)	
Planted in lines (Past season)	0.1	0.1	(0.94)	1.0	1.0	
	(0.3)	(0.3)		(0.0)	(0.0)	
Row planted all fields (number of times past 5 years)				1.1	1.2	0.83
				(1.8)	(1.9)	
Labor Market						
Past supply	1.0	1.0		0.1	0.1	0.96
	(0.0)	(0.0)		(0.2)	(0.2)	
Mandays supplied	14.0	12.7	0.24	0.3	0.3	0.99
	(7.7)	(7.3)		(1.1)	(1.1)	
Past demand	0.1	0.2	0.12	1.0	1.0	
	(0.4)	(0.4)		(0.0)	(0.0)	
Mandays demand				23.5	19.7	0.24
				(21.8)	(23.5)	
Observations	102	99		102	99	

Notes: This table shows the balance between trainer-employers and trainees in the Labor guarantee experiment. Variables were collected at baseline and corresponds to the entire sample. In columns (1) and (4), we show the mean and standard deviation for the variable of interest among the Control group. In columns (2) and (5), we show the same fro the Treatment group. Finally, in columns (3) and (6) we show the p-value for a T-test of equality of means. The p-values of the difference in means test come from SEs clustered at the village level.

Table A.2.2: Hired and Family Labor

Panel A: Trainer-Employers	Days Family Skilled Labor (1)	Days Family and Hired Skilled Labor (2)
T1 -Financial Incentives	0.73	1.57
Train	(0.82)	(0.91)
	[0.37]	[0.09]
Obs.	1205	1205
Control mean	13.17	15.08
Panel B: Spillover-Employers		
T1 -Financial Incentives	0.29	1.45
Train	(0.75)	(0.83)
	[0.70]	[0.09]
Obs.	1453	1453
Control mean	10.42	12.71

Notes: This table shows the effect of the Financial incentive treatment on total household labor applied to the trained techniques, as well as the sum of household and hired labor for the training techniques. Panel A shows regression results for the sample of Trainer-employers. Panel B shows regression results for the sample of spillover-employers. The dependent variable in column 1 is total days of family labor used to do the trained techniques (row planting and fertilizer microdosage). The dependent variable in column 2 is total days of hired and family labor to do the trained techniques. Standard errors are clustered at the village level.

Table A.2.3: Social Learning

Panel A: Spillover Employers	Knowledge Score (1)	P(Adoption and Hired) (2)	P(Adoption and Not Hired) (3)	P(Adoption and Hired) (4)	P(Adoption and Not Hired) (5)	P(Adoption and Hired) (6)	P(Adoption and Not Hired) (7)
T1-Financial Incentives	0.43	0.19	-0.10	0.13	0.05	0.05	0.00
Train	(0.12)	(0.03)	(0.03)	(0.05)	(0.04)	(0.06)	(0.05)
	[0.00]	[0.00]	[0.00]	[0.02]	[0.28]	[0.01]	[1.00]
T1 X [Heterogeneity]				0.08	-0.18	0.04	-0.12
	()	()	()	(0.06)	(0.05)	(0.06)	(0.06)
	[]	[]	[]	[0.18]	[0.00]	[0.56]	[0.04]
Obs.	1466	1466	1466	1466	1466	1466	1466
Control mean	0.52	0.36	0.40	0.36	0.40	0.36	0.40
Heterogeneity Var				Previously Used	Previously Used	Used last season	Used last season
Panel B: Spillover Non-Employer	Any Plots Modern (1)	Plots Modern (2)	Hired Modern (3)				
T1-Financial Incentives	0.04	0.10	0.10				
Train	(0.05)	(0.08)	(0.08)				
	[0.45]	[0.20]	[0.20]				
Obs.	584	584	584				
Control mean	0.42	0.50	0.11				

Notes: This table shows the effect of the Financial incentive treatment on various measures of hiring, adoption and knowledge for spillover farmers. Panel A is restricted to farmers who are spillover employers. Panel B is restricted to farmers who are in the spillover sample but were identified as non-employers at baseline. The outcome in panel a, column 1 is a composite spacing knowledge score. The outcome in panel a, column 2, 4 and 6 is the probability that a farmer adopted and hired labor for the skilled task. The outcome in Panel a, column 3, 5 and 7 is the probability that a farmer adopted and without hiring any labor for the skilled task. Regressions in Panel A are run either directly on the treatment, or as an interaction with a measure of heterogeneity: whether the farmer previously used the modern seeding practices, and whether the farmer planted in lines the prior season. The outcome in Panel B, column 1 is whether the farmer planted any fields with the modern seeding techniques. The outcome in Panel B, column 2 is whether the number of fields planted with the modern seeding techniques. The outcome in Panel B, column 3 is the number of days of labor hired to do the skilled task. Robust standard errors are clustered at the village level.

Table A.2.4: Employer Heterogeneity in Treatment Effects

Employers (All)	Days Hired Skilled Task (1)	Fields Row Planting (2)	Days Hired Skilled Task (3)	Fields Row Planting (4)
T1 - Financial Incentives	0.10	0.13	0.59	0.37
Train	(0.52)	(0.21)	(0.32)	(0.13)
	[0.85]	[0.53]	[0.07]	[0.01]
T1 - Financial Incentives	0.16	0.04	0.04	0.00
Train x [Covariate]	(0.08)	(0.03)	(0.02)	(0.00)
	[0.06]	[0.10]	[0.04]	[0.37]
Heterogeneity Measure	Fields	Fields	Days Hired	Days Hired
Obs.	2682	2682	2682	2682
Control mean	2.0	1.9	2.0	1.9

Notes: The table shows treatment effect heterogeneity for the sample of employers. The outcome variables in columns 1 and 3 are the number of days the individual hired to a worker to do tasks involving the trained techniques. Outcome variables in columns 2 and 4 are the number of fields planted using the trained techniques. The heterogeneity measure in columns 1 and 2 are the number of fields the farmer planted the prior agricultural season. The heterogeneity measure in columns 3 and 4 are the number of days the employer hired the prior agricultural season. Regressions are weighted using inverse probability weights to make the results representative of employers at the village level. Robust standard errors are clustered at the village level.

Table A.2.5: Adoption by Spillover Employers - Heterogeneity by knowledge/past usage

	Fields Row Planting (1)	Fields Row Planting (2)
T1 - Financial Incentives	0.45	0.40
Train	(0.14)	(0.17)
	[0.00]	[0.02]
T1 X [Heterogeneity]	0.05	0.10
	(0.17)	(0.20)
	[0.77]	[0.63]
Obs.	1466	1466
Control mean	1.64	1.64
test Treat+TreatxHet=0	0.00	0.00
Heterogeneity Var	Previously Used	Used last season

Notes: This table shows intensive margin treatment effects on the number of fields planted using the improved agricultural technology for the sample of spillover employers. Regressions show heterogeneous effects by the variable mentioned at the bottom of the table. The heterogeneity variable in column 1 is whether the employer previously had used the improved planting technique. The heterogeneity variable in column 2 is whether the employer previously had used the improved planting technique the prior season. Robust standard errors are clustered at the village level.

Table A.2.6: Treatment Effect Heterogeneity by whether hired previously - Trainees

Trainees	Days Modern (1)	Wage (2)	Employers (3)
T1 - Financial Incentives	3.31	111.25	0.17
Train	(0.50)	(102.76)	(0.13)
	[0.00]	[0.28]	[0.189]
T1 X Hired Previously	0.15	125.97	0.05
	(0.68)	(106.86)	(0.18)
	[0.83]	[0.24]	[0.795]
Obs.	730	614	730
Control mean	0.46	2,304	1.9
p-val T+TxHired=0	0.00	0.02	0.10

Notes: This table shows the effect of the Financial incentive treatment on the labor market outcomes of trainees allowing for heterogeneous effects for whether the trainee previously worked for the trainer. The outcome variable in column 1 is the number of days that the trainee worked doing the skilled labor task. The outcome in column 2 is the wage. The outcome in column 3 is the number of employers that the trainee worked for. Regressions are run on an indicator variable for treatment status, an indicator variable for having hired the trainee previously, and the interaction term of the two, as well as controls. Robust standard errors are clustered at the village level.

Table A.2.7: Treatment effect heterogeneity by previous hiring – Trainer-employers

Trainees	Hired Own Trainee (1)	Hired Trained Technique (2)	Unable to hire trainee (3)
T1 - Financial Incentives	0.55	0.16	0.16
Train	(0.15)	(0.05)	(0.06)
	[0.00]	[0.00]	[0.01]
T1 X Hired Previously	-0.05	0.03	-0.03
	(0.23)	(0.06)	(0.08)
	[0.83]	[0.58]	[0.74]
Obs.	1216	1216	484
p-val T+TxHired=0	0.02	0.00	0.01
Control mean	0.43	0.39	0.24

Notes: This table shows the effect of the Financial incentive treatment on the hiring practices of farmer-trainers conditional on whether they had previously hired the trainee they identified. The outcome variable in column 1 is an indicator variable equal to 1 if the individual reported hiring the trainee they identified to bring to the training event. The outcome variable in column 2 is an indicator variable equal to 1 if the individual reported hiring any individual to perform the skilled labor task. The outcome variable in column 3 is an indicator variable equal to 1 if the individual reported being unable to hire their trainee after attempting to hire them. Regressions are run on an indicator variable for treatment status, an indicator variable for having hired the trainee previously, and the interaction term of the two, as well as controls. Robust standard errors are clustered at the village level.

Table A.2.8: Information spread in household

	Trainer Employer		Trainee Laborers		Spillover Employer	
	Indicator for individual mentioned providing skilled labor on own farm					
	Self	Another Family Member	Self	Another Family Member	Self	Another Family Member
	(1)	(2)	(3)	(4)	(5)	(6)
T1 - Financial Incentives	0.02	0.05	0.66	0.27	0.02	0.07
Train	(0.01)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
	[0.14]	[0.11]	[0.00]	[0.00]	[0.59]	[0.02]
Control mean	0.95	0.65	0.14	0.14	0.78	0.473
Obs.	1205	1205	842	842	1453	1453

Notes: This table shows treatment effects on the likelihood that an individual, or family member, works in tasks on their own farm related to the trained techniques. Columns 1 and 2 relate to the sample of trainer-employers. Columns 3 and 4 relate to the sample of trainees. Columns 5 and 6 relate to the sample of spillover-employers. The dependent variable in odd numbered columns is the likelihood that the respondent engaged in work on their own farm involving the trained techniques. The dependent variable in even numbered columns is the likelihood that the another individual in the same household engaged in work on their own farm involving the trained techniques. Robust standard errors are clustered at the village level.

Table A.2.9: Changes to Farm Labor in Response to Treatment

	Farm Labor	
	Total Household Labor	Respondent Farm Labor
Panel A: Trainees	(1)	(2)
T1 - Financial Incentives	1.19	-0.03
Train	(1.62)	(0.77)
	[0.47]	[0.97]
Control mean	33	20
Obs.	842	842
Panel B: Trainer (Employers)		
T1 - Financial Incentives	1.47	-0.42
Train	(2.05)	(0.69)
	[0.48]	[0.54]
Control mean	52	26
Obs.	1205	1205
Panel C: Spillover (Employers)		
T1 - Financial Incentives	-0.91	-1.06
Train	(1.62)	(0.87)
	[0.65]	[0.23]
Obs.	1453	1453
Control mean	47	25

Notes: This table shows treatment effects on family farm labor provided during the planting period. The outcome in column 1 is total household supply of labor. The outcome in column 2 is the respondent's own supply of farm labor. Panel A corresponds to the sample of trainees. Panel B corresponds to the samples of trainer-employers. Panel C corresponds to the sample of spillover-employers. Standard Errors are clustered at the village level.

A.3 Additional Context

This section provides some additional details not provided in the main context section of the paper.

A.3.1 Agriculture in Burundi

Burundi is among the poorest countries in the world and the third most densely populated country in Africa.[1] Due to small average farmable land and low agricultural productivity, the country faces a persistent risk of food insecurity (Verwimp and Muñoz-Mora, 2018). Burundi has three agricultural seasons per year. We conduct the experiment over the course of Season "B". This season generates most of the country's agricultural production, and lasts from February to July. The staple crop grown during this season is beans, which we use as the focus of the training intervention. As described in Section A.3.2, farmers plant beans either according to traditional practices, or using improved agricultural techniques.

A.3.2 Row Planting and Fertilizer Microdosage - External Validity and Training Importance

While this project focuses on a failure to adopt modern planting techinques in Burundi, this is fact is not limited to this context. Other studies have found that only ∼55% of teff farmers in Ethiopia and ∼10% of rice farmers in Ghana adopted (Fentie and Beyene, 2019; Donkor et al., 2016), despite several efforts to promote the technology. These studies have also confirmed the increased yields of row planting for these crops in agronomic trials and have cited the increased labor intensiveness of row planting as an impediment to adoption - in Ethiopia 100% of surveyed farmers who cited constraints to adoption mentioned labor intensity as a major impediment to the adoption of row planting (Fentie and Beyene, 2019). Research in other contexts also suggests that while the technique is seemingly simple, farmers have stated that viewing fields or having an oral description of the technique is not sufficient for them to adopt it - rather they require practical demonstrations and training in order to learn these techniques (Cafer and Rikoon, 2018), and has found relatively large impacts of training in the techniques, with almost no impact of extension without training (Zeweld

[1] Agriculture is the dominant sector of the Burundian economy, representing approximately 50% of the GNP and 80% of its exports (Beekman and Bulte, 2012). According to the World Bank (2020), approximately 86% of the Burundian population is rural, composed mostly of subsistence farmers. The average Burundian household consumes around 72% of what it produces, and the rest is either marketed or exchanged through social networks Niragira et al. (2015).

et al., 2017). In a similar context, Deutschmann et al. (2022) provide evidence for the importance of training, finding evidence that part of the impact of the O1AF is relaxing informational constraints. Deutschmann et al. (2022) also find some evidence for potential knowledge depreciation of these techniques among Kenyan farmers, again pointing to the potential importance of training for skill accumulation and retention.

A.3.3 Agricultural Labor Markets

Farm labor is supplied more by women than by men - with 90% of women involved in agricultural activities in rural households but only 65.5% of men(Vinck, 2008). Partially this is due to the fact that men migrate and provide seasonal labor(Vinck, 2008). This tendency of women to be involved in farm labor is particularly true for staple rather than cash crops. This tendency of both men and women to provide substantial amount of farm labor, both on and off own farm is consistent with other contexts (Fink et al., 2020; Guiteras and Jack, 2018; Palacios-Lopez et al., 2017). Finally, consistent with other contexts, laborers in such villages tend to be among the poorest in the village: another study found agricultural laborers to have access to limited land, and to be among the poorest in terms of assets in the village (Fink et al., 2020; Vinck, 2008).

The tasks of row planting and fertilizer microdosage are concentrated in a narrow window of around 2 weeks after the onset of the rains. Consistent with many farmers utilizing rainfed agriculture in Sub-Saharan Africa, Burundian farmers perceive higher returns to planting soon after the onset of rains, particularly for Season B given the already shorter growing season (Dodd and Jolliffe, 2001). Moreover, farmers are also encouraged to plant early by the government, which generally announces specific windows of time by which farmers must complete their planting. 1AF data documents a strong positive correlation between planting earlier and planting using row-spacing rather than broadcasting, consistent with the possibility that farmers change planting techniques as they run out of time.

There is one additional labor market that exists which is the market for exchange labor - however this constitutes a minor portion of overall labor input as employers in this setting obtain only 2-3 days of labor from this, so it constitutes a less important source of labor input overall.

A.3.4 Training Markets

What makes learning row planting a skill that requires training rather than information that can be described verbally, or learnt by observation from people's fields? Farmers describe three features that require training. First, they mention that row planting rather than broadcasting requires several pieces of equipment (sticks to measure distance, ropes to measure line to ensure planting is straight) and that understanding how to utilize this equipment is complicated from observation. Second, they note that while distances and techniques can be communicated verbally, often individuals find it hard to replicate without learning by doing. Finally, the repetition and practice of the technique is important to develop using the technique into a skill that can be sold in the labor market.

A.4 Additional Treatments

In the Spillover experiment, 12 villages are also assigned to the following treatment:

T2 - Labor Insurance - In addition to what was provided in control in the spillover experiment, farmers in this condition were told that if they trained their laborer for two days for 180 minutes, we would ensure that this laborer, or another similarly skilled laborer, would return to work for them for 2 days at the prevailing unskilled wage in the village.

Implementation To implement this treatment without coercion we do the following. Before any treatment status of a village was revealed, we asked the 1AF Field Officer and local authorities to tell us: 1) what the approximate daily wage was of an individual who worked for another individual planting beans in the traditional way and 2) what the daily wage was of an individual who worked for another individual planting beans using the modern planting practices. On the day of the event, we explain to trainer-employers that if they train their trainee in the modern planting practices, our team would guarantee that, at planting time, a laborer skilled in these techniques would return to work for the farmer, and that the employer would be required to pay that laborer the wage that workers tend to work for when planting in the traditional manner.

To incentivize the worker to return after being trained by the trainer-employer, it was explained to workers that, if they returned to work for this employer during the planting season, then they would receive an additional top-up to their wage paid by the enumerator team. This in total ensured that the worked would receive a slightly higher wage than the skilled wage paid in the village conditional on returning to this employer.

We construct the design of this treatment in order to ensure that employers capture more of the returns of training conditional on training (by ensuring that the worker did not separate, and that the wage did not rise) while also not coercing the worker into working, by also offering them a wage that was higher than the wage paid for skilled labor in the village, and in cases where this was refused or the worker was not available for other reasons, providing another similarly skilled worker to the trainer-employer. To ensure that this offer did not lead to collusion/fake jobs, trainer-employers and trainees were told that all work would need to be scheduled in advance with the enumerator team, and payment to the worker at the end of the day occurred only if it was clear they had worked the entire day, and after the employer had paid their portion of the wage to the worker.

Results/Interpretation We interpret the results as suggesting that farmers become willing to train when they know that they will capture returns from training (because the wage does not rise and the laborer returns). However, there are several alternate interpretations of this treatment.

The first is that farmers are just willing to pay to forgo search costs in the planting season. In this interpretation, farmers do not care if the laborer they hire is skilled or unskilled, but are willing to pay a cost today (including a cost of effort involved in training) in order to overcome future search costs. This seems unlikely to explain the results since most employers hire the employee they train for the trained techniques. Moreover, the search costs associated with hiring an unskilled laborer are reported by trainer-employers in general as being relatively low. Therefore, it does not seem likely that this explanation explains the results.

The second possibility is that farmers respond to the training incentive purely because of the wage subsidy. This is unlikely to explain the results - the wage that the employer was calibrated to pay was designed to match the average unskilled wage in the village, and so there was not supposed to be any benefit of training if the employer did not perceive a positive return to training. Moreover, even if the employer did perceive that some subsidy to the worker might be passed through, this subsidy was not large, and is unlikely to explain the magnitude of the training response we observe.

The third possibility is that farmers perceive the returns to training as zero/negative, but value skilled labor, and believe that if they trained they could say that the worker did not worker show up and ask us to replace the worker with another skilled laborer. The fact that we see zero instances of this in practice suggests that this is not a primary motivation, but it is hard to rule out entirely. In order to rule out this possibility, the Labor guarantee experiment did not include a replacement worker in the event the originally contracted worker did not show up.

A.5 Data Collection Details

Adoption During the planting survey, we ask all respondents to draw a map of all their fields, and then complete a plot roster. For each field we ask the crop (or crops) planted, and whether that field was planted using row planting and fertilizer microdosage techniques. Conditional on planting in rows, farmers were also asked details about the distance between rows and pockets, and the consistency of this distancing across the plot (i.e. on what proportion of the field they used these techniques). These questions were incentivized by a portion of the respondents survey compensation being offered conditional on the responses to these questions being found to be true during an audit of the farmers' fields.

At least one field audit was conducted per respondent in which enumerators visited farmers fields, and observed the planting method.

Hiring We measure hiring outcomes at the worker-task level. During the planting survey, we ask all employers to state the number of workers who were hired. For each worker we then ask the name. We match the name against the list of trainees to determine whether they were an individual invited to the training event, or not. We then ask for each laborer 1) the number of days worked, which tasks were done, the total amount paid, if the work involved the trained tasks, the number of days that were spent specifically on those tasks. In some surveys, we did not ask this last question - and therefore use an approximation based off of the task combination to estimate the number of days that were conducted on the trained task.[2]

Employment in Agricultural Labor Markets We measure agricultural employment outcomes at the worker-task level. During the planting survey, we ask all in the sample to state the number of employers who they worked for during the agricultural season. We then ask for each employer the number of days worked, which tasks were done, the total amount paid, if the work involved the trained tasks, the number of days that were spent specifically on those tasks.

Farm outcomes At harvest time, we ask farmers the quantities of all crops harvested. Prices at the nearest market are elicited, and aggregated at the village level to create an average village level price. The amount spent on non-labor inputs is elicited for each input (eg. fertilizer, pesticide, hired land). Due to thin land markets, we are unable to elicit an

[2]Results are not changed if instead of measuring the number of days conducted on row planting and fertilizer microdosage, we instead use as an outcome measure the number of days on tasks that included row planting and fertilizer microdosage.

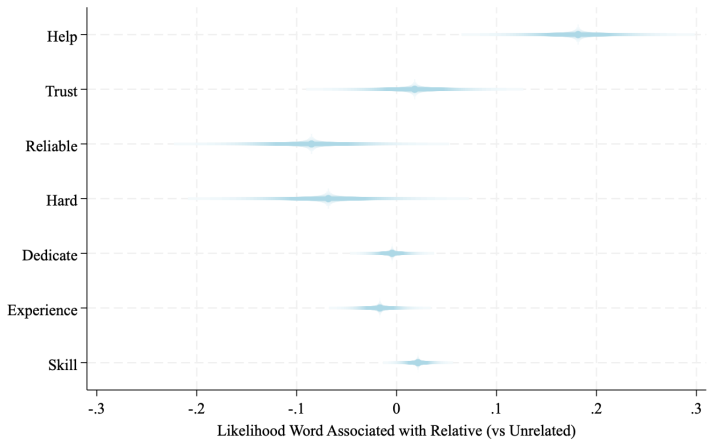
average rental price of land, and therefore in all profit calculations we do not subtract off the opportunity cost of own land utilized.

Appendix B

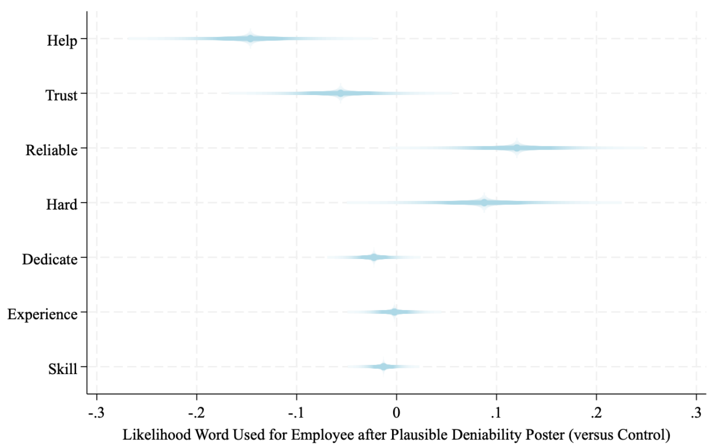
Kinship Pressure and Firm-Worker Matching

B.1 Appendix Figures

Figure B.1.1: Urban Field Experiment – Relative likelihood of using certain words for justifying the hiring choice



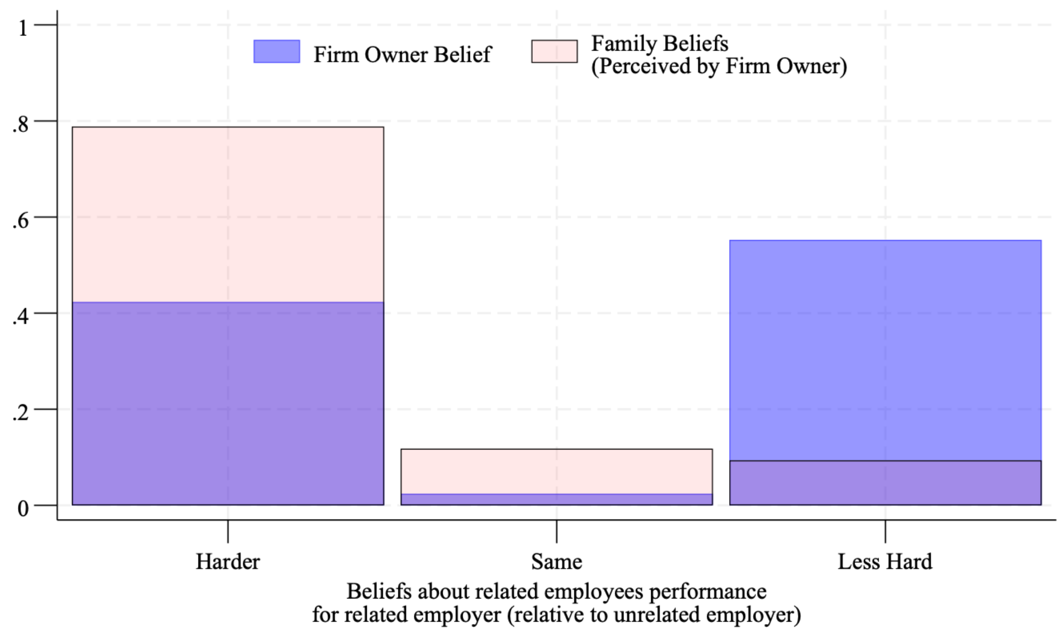
(a) By whether the employee is related



(b) By Plausible Deniability Treatment

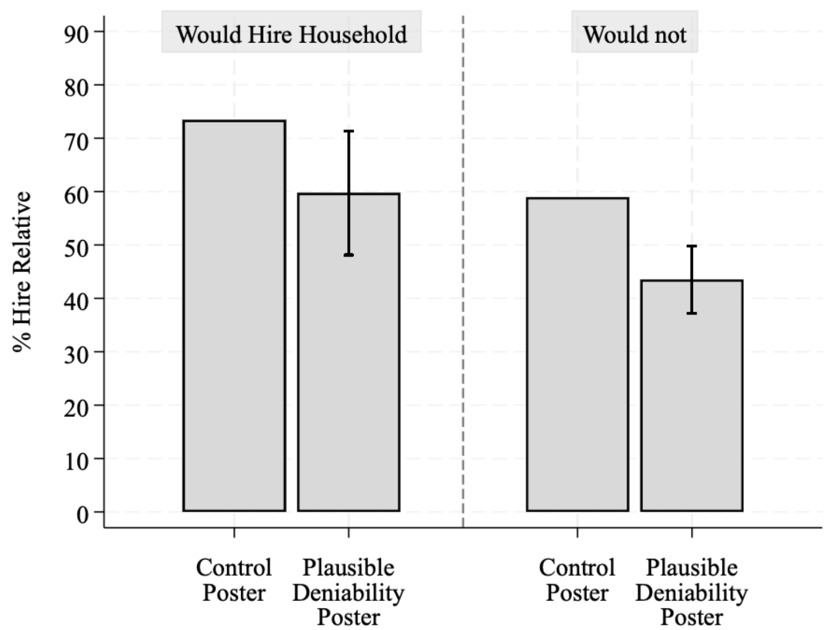
Notes: These figures show the likelihood of using particular words to rationalize hiring choices in the Urban Field Experiment. The coefficients in Panel A are regression coefficients from regressing the word in question on a dummy variable for whether the employee selected for the job was related to the employer, or not. Therefore it measures the increased likelihood that a particular word is used to justify hiring a relative rather than a non-relative. Panel B shows regression coefficients from regressing the word in question on a dummy variable for whether the employer was assigned to the Plausible Deniability poster, or not. Therefore it measures the increased likelihood that a particular word is used to justify hiring an employee in the plausible deniability condition. All regressions also control for location and industry fixed effects. Robust standard errors are shown.

Figure B.1.2: Urban Mechanism Experiment - Disagreement between Employer and Family Regarding Work Effort of Employee



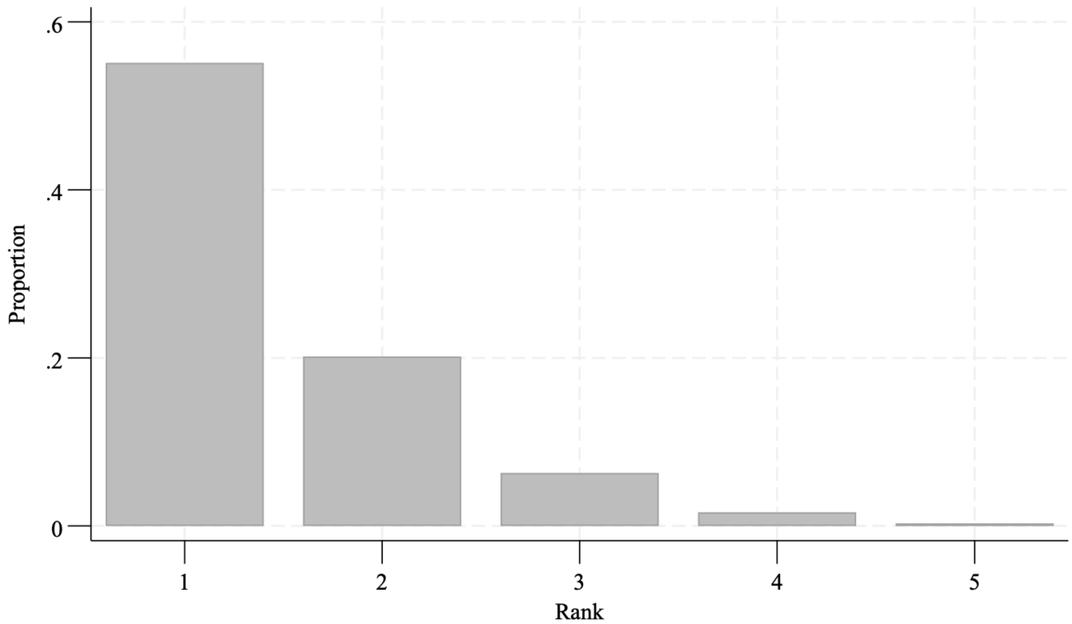
Notes: This figure shows the belief of the the employer, and the second order beliefs of the employer about their family's beliefs, of the related employee's work effort. Each bar displays the proportion who perceive that the related employee will work harder, the same, or less hard for their related employer than other unrelated employers. The blue bars show the beliefs of the firm owner. The red bars show the beliefs of the firm owner of their family's beliefs.

Figure B.1.3: Urban Mechanism Experiment - Treatment Effect heterogeneity conditional on employer stating that the marginal relative hire is a member of their household



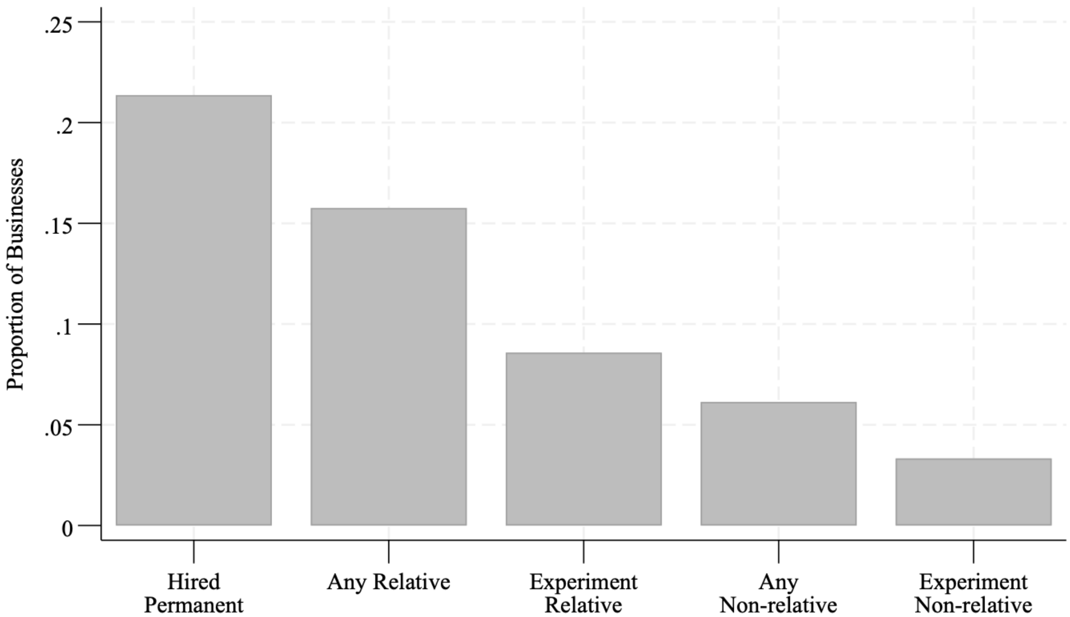
Notes: This figure shows the likelihood of a relative being selected in the Urban Mechanism experiment, conditional on the employer stating at baseline that the maringal relative hire was, or was not a member of their household.

Figure B.1.4: Urban Mechanism Experiment - Employer Ranking of Selected Relative among Relative Options



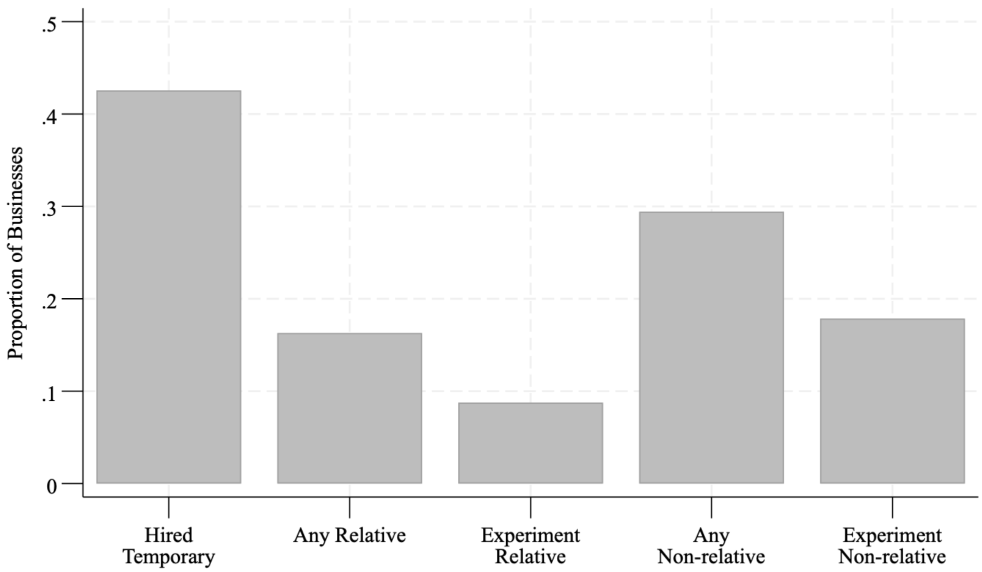
Notes: These figures show how employers ranked the relative selected for the experiment among the relatives they listed at baseline as a prospective employee. 1 indicates best available. Data was collected for employers not assigned the subsidy following the experiment.

Figure B.1.5: Urban Mechanism Experiment - Identity of employees hired for employers not given subsidy following experiment - Permanent jobs



Notes: These figures show which permanent employees, if any, employers in the experiment hired after the experiment. Each bar is an indicator variable equal to one if the employer hired that category of employee. Data was collected for employers not assigned the subsidy following the experiment.

Figure B.1.6: Urban Mechanism Experiment - Identity of employees hired for employers not given subsidy following experiment - Casual jobs



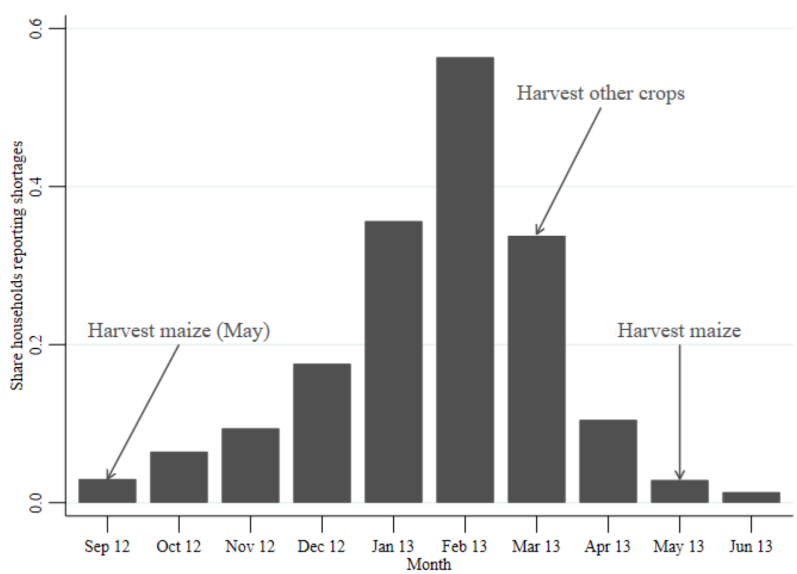
Notes: These figures show which casual employees, if any, employers in the experiment hired after the experiment. Each bar is an indicator variable equal to one if the employer hired that category of employee. Data was collected for employers not assigned the subsidy following the experiment.

Appendix C

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

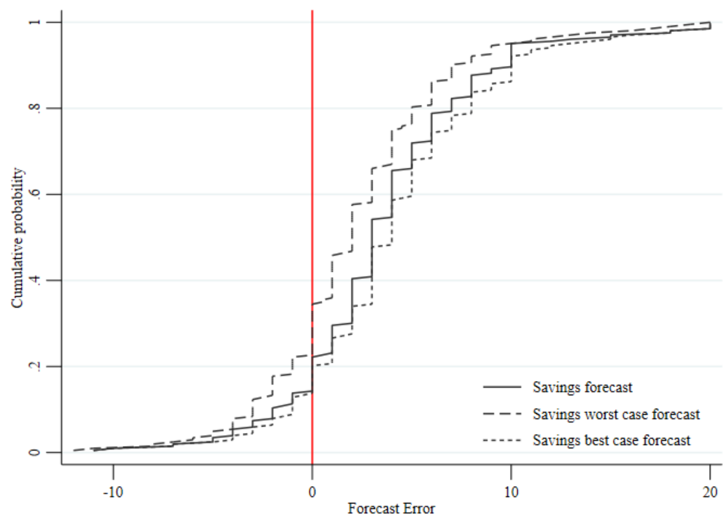
C.1 Appendix Figures

Figure C.1.1: Reported food shortages by month among Zambian farmers

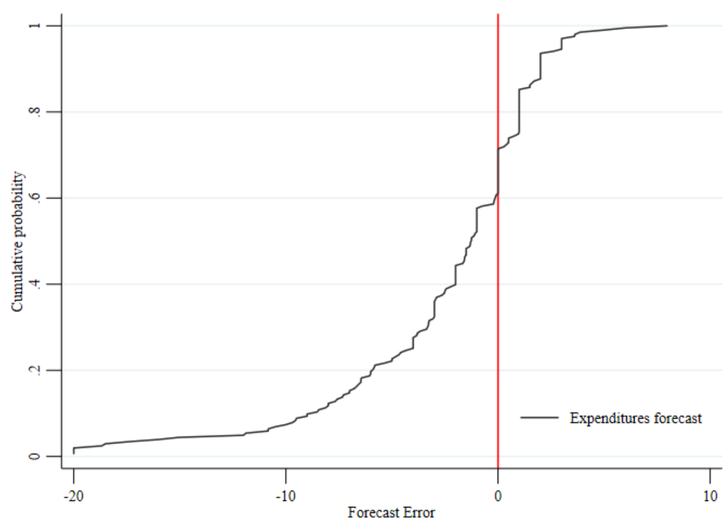


Notes: Proportion of farmers in Eastern Zambia reporting food shortages by month. The data come from Fink et al. (2020). The sample consists of farming households located in Chipata district, Zambia.

Figure C.1.2: Overoptimism in savings and expenditure forecasts (field experiment)



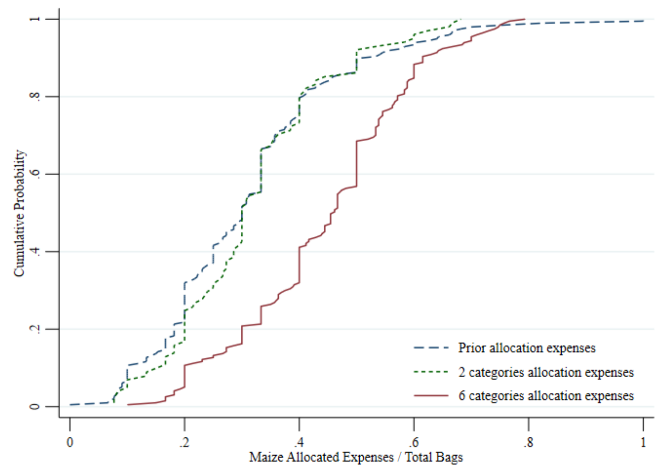
(a) Panel A: Savings forecast



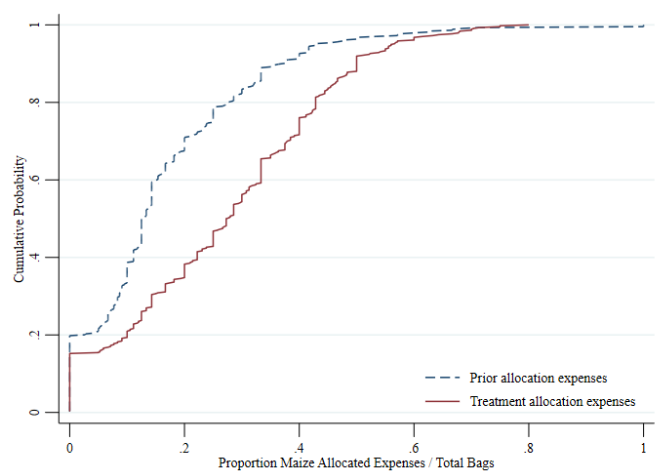
(b) Panel B: Expenditure forecast

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

Figure C.1.3: Proportion of current maize wealth allocated to non-food expenses



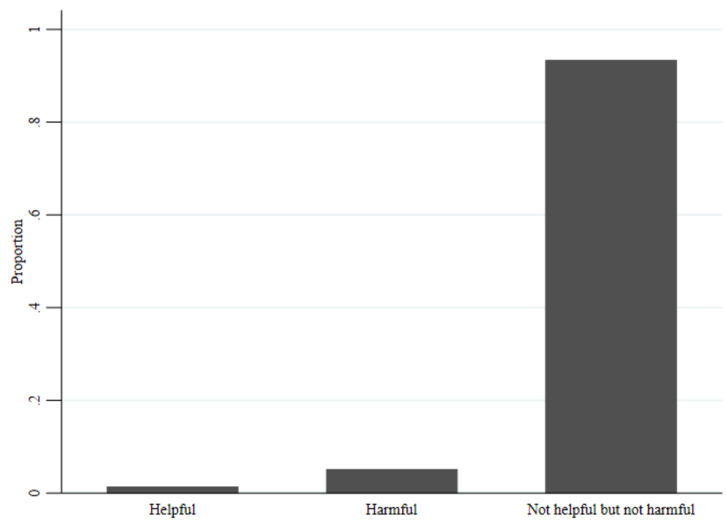
(a) Panel A: Mechanism experiment



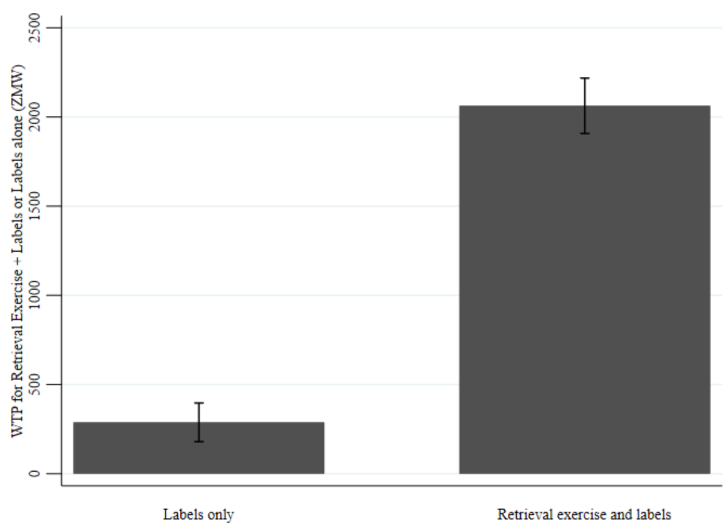
(b) Panel B: Field experiment

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

Figure C.1.4: Evidence of value of labels (field experiment)



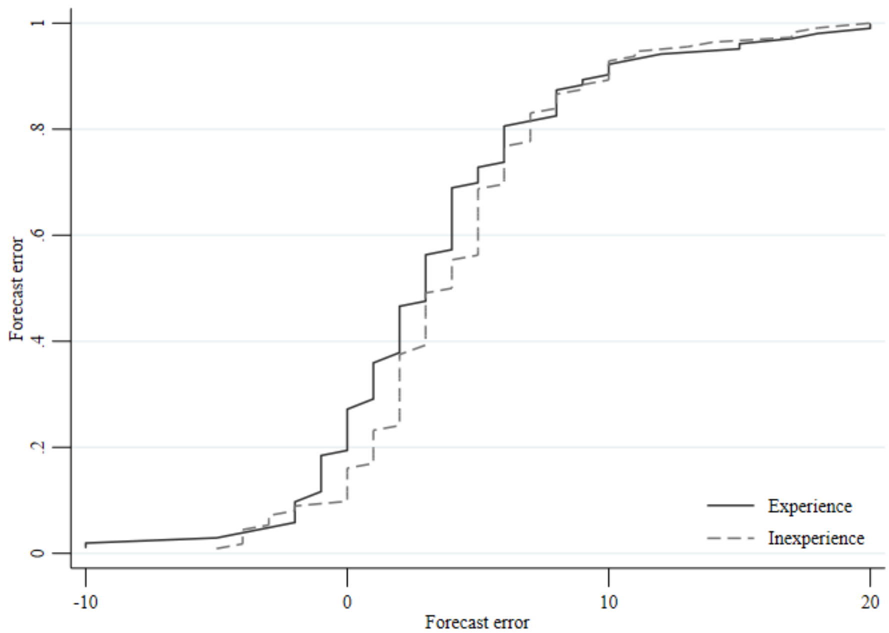
(a) Panel A: Hypothetical effect of labels without retrieval exercise



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

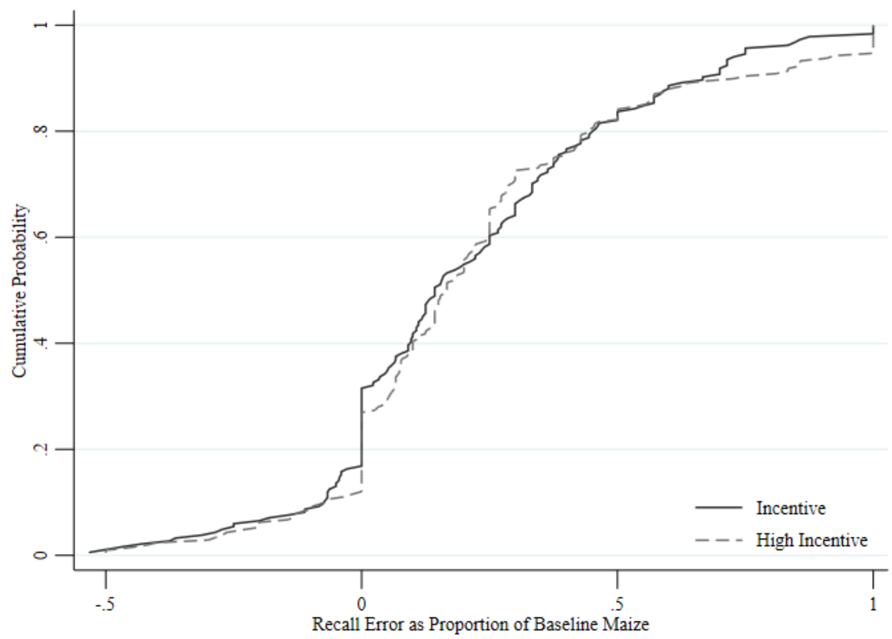
Notes: Panel A shows the proportion of participants who indicated that being given labels alone would have been 1) helpful, 2) harmful or 3) neither helpful nor harmful if labels were not accompanied by the retrieval exercise. The questions was asked during the second year of the project. The sample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

Figure C.1.5: Forecast versus realized maize savings, by experience (field experiment)



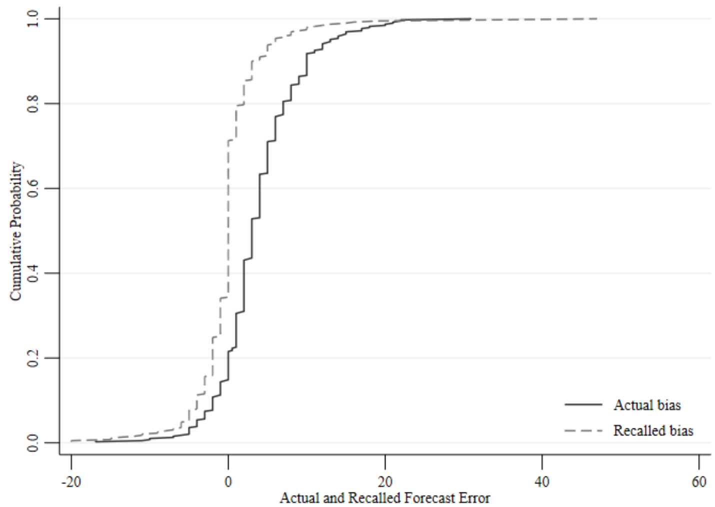
Notes: Forecasts versus realizations of savings. Experience is proxied by age: individuals in the lowest age quartile (mean age 28 - dashed line) are categorized as inexperienced, those in the top age quartile (mean age 62 - solid line) as experienced. The sample is restricted to control participants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

Figure C.1.6: Memory error – Recalled maize savings versus actual savings (field experiment)



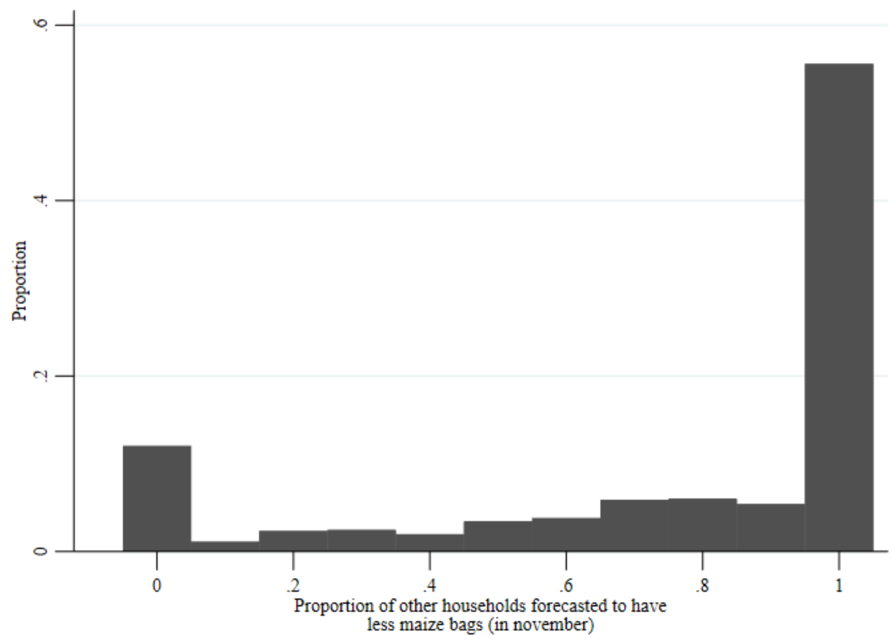
Notes: Recollection of savings in maize bags at the time of a past survey visit minus the number of maize bags measured during that past visit. The solid line shows the low incentive condition for accurate recall; the high incentive condition (dashed line) doubles the payoff for accuracy.

Figure C.1.7: Sophistication of past overoptimism (field experiment)



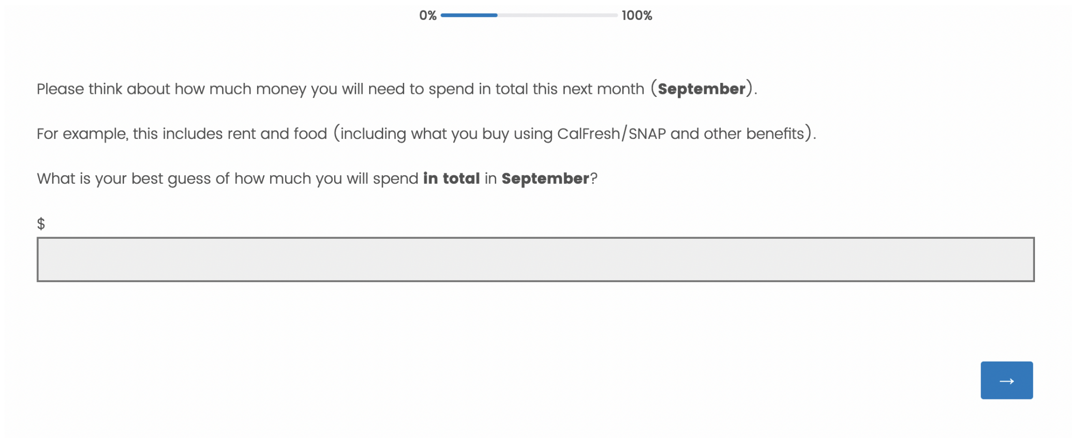
Notes: Actual forecast error and recalled forecast error. Actual forecast error is measured as forecasted savings minus realized savings (dark grey solid line). Recalled forecast error is the household's recalled forecast minus recalled savings (light grey dashed line). The sample is restricted to control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

Figure C.1.8: Own forecast, relative to forecasts for other similar households (field experiment)



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

Figure C.1.9: Prolific Survey Screenshots


0% 100%

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

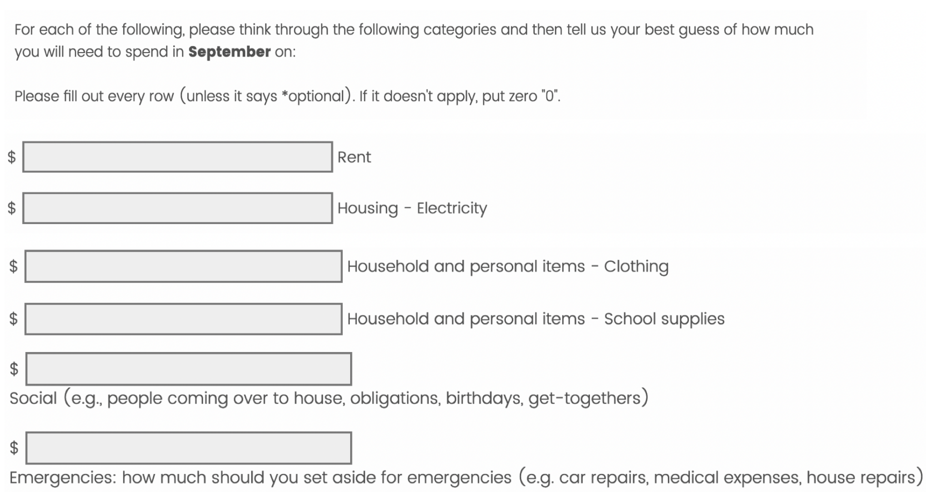
For example, this includes rent and food (including what you buy using CalFresh/SNAP and other benefits).

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

$

→

(a) Panel A: Aggregated Expenses


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

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

$ Rent

$ Housing - Electricity

$ Household and personal items - Clothing

$ Household and personal items - School supplies

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

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

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

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

C.2 Appendix Tables

Table C.2.1: Balance

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

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

Table C.2.2: Attrition (field experiment)

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

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

Table C.2.3: Allocation of bags to non-food expenses according to the exercise (mechanism experiment)

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

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

Table C.2.4: Willingness to exchange maize for consumption goods

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

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

Table C.2.5: Allocation of bags to non-food expenditures according to the category of expense and the timing of the question (control group, mechanism experiment)

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

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

Table C.2.6: Savings results - Robustness checks (field experiment)

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

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

Table C.2.7: Impact of retrieval exercise on assets (field experiment)

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

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

Table C.2.8: Impact of incentives and retrieval exercise on recall of past savings (field experiment)

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

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

C.3 Appendix B: Model Proofs and Simple Example

C.3.1 Proofs of Full Model

Definitions

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

Maximization Problem

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

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

$$\text{s.t.} \ \sum_{t=1}^{3} c_t(\Theta) + \sum_{t=1}^{3} \sum_{i=1}^{N} e_{it}(\Theta) = Y \text{ for all } \Theta \in \boldsymbol{\Theta}$$

A mistaken person who fails to retrieve some expenses solves the same problem, but uses

$\hat{\pi}_{it}$ to construct $\hat{Pr}(\Theta)$:

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

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

Proof: Solution

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

$$u'(c_3) = v_i'(e_{i,3}) \equiv MU_3(\Theta, s_3) \text{ for realized expenses in period 3} \tag{C.1}$$

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

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

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

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

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

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

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

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

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

$$u'(c_1) = v_i'(e_{i,1}) \equiv MU_1(\Theta) = E_{\Theta|\mathcal{I}_1}[MU_2(\Theta, s_2)] = E_{\Theta|\mathcal{I}_1}[MU_3(\Theta, s_3)] \tag{C.3}$$

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

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

Finally, the impact of the intervention moves the person toward the rational benchmark. As discussed above, all results about the relative comparisons of a mistaken person to a

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

[2] The direction on period-3 non-food expenses is ambiguous. An unanticipated period-2 expense can cause savings to drop relative to period-1 expectations to the point that both consumption and non-food expenses drop in period 3 relative to period-1 expectations.

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

Formally-Stated Predictions

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

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

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

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

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

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

C.3.2 Simple Model

In this section, we work through a simplified example of our model with no stochasticity and limited expenses to help with intuition. The example is intentionally stark. Harvest income is 1 ($Y = 1$) and expenses all share the same utility function as consumption ($v_i(\cdot) = u(\cdot)$). We assume that a person has two expenses in addition to food (school fees and herbicide) which are only paid in period 2. That is, $\pi_{1,2} = \pi_{2,2} = 1$ and $\pi_{1,1} = \pi_{2,1} = \pi_{1,3} = \pi_{2,3} = 0$. Therefore, a person with perfect retrieval is maximizing five components $\sum_i^5 u(e_i)$ under the constraint that $\sum_i^5 e_i = 1$.

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

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

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

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

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

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

C.4 Protocols

C.4.1 Field experiment

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

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

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

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

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

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

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

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

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

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

Finally, they were told that regardless of their choice, we would return to conduct more surveys with them in a month's time.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C.4.2 Mechanism experiment

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

The sample consists of 197 households, half of them randomly assigned to a treatment group and the other half to a control group. Participants were drawn from 28 villages, with up to 14 participants per village (using the sample selection criteria described in Section 3.4.2). For the mechanism experiment, due to logistical constraints, most villages were located near a town. Consequently, some of our screening criteria were relaxed: not all households reported food shortages during the hungry season, and households were more likely to have some small alternative sources of income.

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

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

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

- Treatment group (6 category budget board): The exercise was exactly the same as described in section C.4.1: the respondent was first provided with a sheet of paper representing five categories of non-food expenses (school, farming inputs, household goods, transfers and emergencies) and then with another sheet of paper representing food consumption per month. The respondent was asked to allocate the thumb tacks to each of the categories according to its plan on future expenses.

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

After the willingness to pay exercise, the respondent was asked to list all the items he/she had in mind for the non-food category or categories when choosing how many bags of maize to allocate.[3] For each item volunteered by the respondent, the surveyor asked the amount of the expense,[4] the time when the expense is expected to be realized,[5] and the expected frequency of the expense.[6] and the degree of certainty of the expense.[7] This was done both for the control and the treated groups.

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

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

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

[5]The options were: “this week”, “later this month”, each following months and “don’t know”.

[6]The options were: ”once”, ”multiple” and ”uncertain”.

[7]The options were: “certain” and “uncertain”.