

Learning to See the World’s Opportunities: Memory, Mental Experiencing, and the Economic Lives of the Vulnerable*

Nava Ashraf[†] Gharad Bryan[‡] Alexia Delfino[§] Emily A. Holmes[¶]
Leonardo Iacovone^{||} Christian Johannes Meyer^{**} Ashley Pople^{††}

March 29, 2026

Abstract

The capacity to imagine possibilities beyond one’s experience is fundamental to economic life. This ability draws on memory and becomes distorted by experiencing trauma. To what extent does this distortion influence how new experiences are interpreted, and can it be repaired? In this paper, we answer these questions in two RCTs where we introduce and evaluate Guided Mental Experiencing (GME), a scalable intervention that trains individuals to construct vivid, causally structured simulations of future success and is designed to repair the cognitive impacts of trauma. In a sample of refugees in Ethiopia, consistent with the repair hypothesis, GME alone improves simulation quality, labor supply, income, and food security. In a sample of vulnerable would-be entrepreneurs in Colombia, we test both hypotheses. Traditional business training harms trauma-exposed participants—consistent with the idea that history shapes how people respond to new interventions. Adding GME reverses these effects—consistent with cognitive repair. We contribute evidence that trauma-induced constraints are malleable, introduce new tools to measure simulation, and provide large-scale evidence on the behavioral and economic effects of GME.

*We thank Pedro Bordalo, Stefan Dercon, Xavier Gabaix, Karla Hoff, Alex Imas, Matthew Gentzkow, Supreet Kaur, Elizabeth Kensinger, Asim Khwaja, Eliana La Ferrara, David Laibson, Kate Orkin, Andrei Shleifer, Chris Woodruff, and seminar participants at BEAM, CEPR Development Workshop, Columbia, Ethiopian Policy Studies Institute, EUI, Harvard/MIT Development Seminar, LUISS, Maastricht, Marseille, Oxford, Padova, RWI Berlin, and Toulouse for very helpful comments. We gratefully acknowledge funding for the Colombia RCT by IPA Peace and Recovery Program, J-PAL Post-Primary Education, J-PAL Jobs and Opportunity Initiative, the Rockefeller Foundation, and the Institute of Global Affairs at the LSE for the Colombia RCT. We gratefully acknowledge funding for the Ethiopia RCT by the “Building the Evidence on Protracted Forced Displacement” partnership of United Kingdom’s Foreign, Commonwealth and Development Office (FCDO), the United Nations High Commissioner for Refugees (UNHCR), and the World Bank Group (WBG) as well as the Prospects partnership by the Ministry of Foreign Affairs of the Kingdom of the Netherlands. In Colombia, we thank Innovations for Poverty Action Colombia for their support in data collection and partnership management, and we are grateful to Maria Paula Gómez Franco, Jose David Angel Castaño, Camilo Andres Hernandez Estrada, Mariana Medina, and Manuel Pardo for their input in the curriculum design, training and supervision of the implementation team. In Ethiopia, we thank Yehualashet Gebremedhn Astera, Victoria Strokova, and Benyam Worku for guidance and support throughout the project and RRS management and staff for their collaboration. Moustafa El-Kashlan, Miguel Fajardo-Steinhäuser, Mathis Greussing, Zijun Lin, George MacDonald, Yacine Ouahioune, Cármen Goncalves Tome Da Silva Pegas, Misha Seong and Ada Gianassi provided excellent research assistance. The Colombia RCT was registered with the American Economic Association RCT Registry as AEARCTR-0004695 (<https://www.socialscienceregistry.org/trials/4695>). It received ethics approval from the Innovation for Poverty Action Institutional Review Board USA (IPA IRB Protocol No. 14638) and the London School of Economics Research Ethics Committee (Ref No. 00769c). The Ethiopia RCT was registered with the American Economic Association RCT Registry as AEARCTR-0008934 (<https://www.socialscienceregistry.org/trials/8934>). It received ethics approval from the University of Oxford (Ref No. ECONCIA21-22-23) and the London School of Economics (Ref No. 51684).

[†]Department of Economics, London School of Economics: N.Ashraf1@lse.ac.uk

[‡]Department of Economics, London School of Economics: g.t.bryan@lse.ac.uk

[§]Bocconi University: alexia.delfino@unibocconi.it

[¶]Uppsala University: emily.holmes@uu.se

^{||}The World Bank: liacovone@worldbank.org

1

^{**}Oxford Martin School and Department of Economics, University of Oxford:
christian.meyer@oxfordmartin.ox.ac.uk

^{††}The World Bank and University of Oxford: ashley.pople@economics.ox.ac.uk

1 Introduction

Economists have long recognized that investing in the future takes more than resources—it requires the ability to clearly imagine what lies ahead. As Böhm-Bawerk wrote in 1889:

Provision for the future makes no inconsiderable demands on our intellectual strength...
The present always gets its rights. It forces itself upon us through our senses.... But
the future we must anticipate and picture.¹

Decades of research in neuroscience and psychology explain how this anticipation is done – we picture the future by reconfiguring our memories of the past, a process called episodic future simulation (Schacter et al., 2008; Addis, 2020). Indeed, a growing body of economic research illustrates how our experiences of the past shape our decisions today and beliefs about the future (Malmendier and Nagel, 2011; Bordalo et al., 2020).

When past experiences are particularly negative, this capacity to picture the future becomes impaired. For most of the world, memories of the past are shaped by adversity. More than 70% of the global population has experienced trauma, and by 2030, over 60% of the world’s poor will live in fragile states marked by conflict and violence (Kessler et al., 2017). Because we imagine the future through the lens of the past, this trauma impairs the ability to plan and move forward (McNally et al., 1995; Kleim et al., 2014). The tragedy is that in wake of trauma – just when people most need to envision a better future and take steps to achieve it –their ability to do so is compromised.

Adversity can thus distort forward-looking choice, which raises two fundamental questions. First, the repair question: can the damage of past trauma be undone? Second, the interpretation question: do the effects of past experience end with current choices, or do they shape how new information is processed—is program impact determined by content alone, or does it depend on the history of the participants? We address these questions in three steps. First, we draw on insights from psychology showing that mental simulation is a plastic and potentially repairable capacity. We use this to design a new intervention—Guided Mental Experiencing (GME)—that helps individuals construct vivid, causally structured simulations of goal-relevant futures (Holmes et al., 2007; Pile et al., 2021; Pham and Taylor, 1999; Oettingen, 2012). GME adapts principles from imagery-based therapy but is delivered in a non-clinical, scalable format by lay facilitators.² Second, we embed GME into

¹“Indeed, to have any effect in the future, we must form a double series of anticipations. We must be able to form a mental picture of what will be the state of our wants, needs, feelings, at any particular point of time. And we must be able to form another set of anticipations as to the fate of those measures which we take at the moment with a view to the future.” - The Positive Theory of Capital and Interest (1889)

²By mental experiencing we mean the process of vividly and multi-sensorially imagining experiences in our mind as if we were living them. This necessarily involves thinking in mental images and using all five senses, which also evokes much greater emotion (Pearson et al., 2008; ?). This process is also known in the psychology literature as mental simulation, mental imagery, visualization, mentalization or imagination. We use these terms as synonyms throughout the paper.

two large-scale randomized controlled trials to test the repair and interpretation questions. The first RCT, conducted with refugees in Addis Ababa, Ethiopia, evaluates whether improving future simulation can change economic behavior in a population with high trauma exposure. The second RCT, in Bogotá, Colombia, compares a standard entrepreneurship program to a GME-enhanced version among vulnerable individuals, to examine whether cognitive support alters how training is received and interpreted.

Third, we develop new tools to measure the quality of mental simulation—assessing both specificity and emotionality—in participants’ narratives about the past and future. These allow us to capture changes in the structure and affective depth of imagined futures and link them to economic outcomes. Our results provide affirmative answers to both of our questions. Cognitive constraints induced by trauma can be repaired cost-effectively. And trauma shapes not only current decision-making, but how individuals interpret and respond to new information—altering the effectiveness of development programs.

In Ethiopia, we test the repair hypothesis: can improved simulation change economic behavior? GME leads to significant improvements in participants’ ability to vividly and emotionally simulate future scenarios, with average cognitive gains of 0.10 standard deviations. These translate into substantial behavioral and economic improvements: labor supply increases by 5 percentage points (over a control mean of 21%), earnings rise by 0.14 standard deviations, and food insecurity falls by 0.22 standard deviations. Participants also report higher life satisfaction and reduced functional impairment. These results demonstrate that trauma-induced constraints are not immutable.

In Colombia, we test both the repair and interpretation hypotheses. Traditional business training, delivered without cognitive support, backfires: it reduces the quality of imagined business-related futures by 0.16 standard deviations, especially among individuals with high trauma exposure. This decline is accompanied by lower income (between -0.14 and -0.22 standard deviations) and lower likelihood of getting funding for the business (by 6 percentage points), suggesting that distorted simulation impairs how training is processed and acted upon. When GME is added, these effects are reversed: simulation quality improves by 0.18 standard deviations, income returns to baseline, the likelihood of receiving any business funding rises by 6 percentage points and business survival increases by 5 percentage points. The same intervention that increased employment in Ethiopia makes development programming trauma-robust in Colombia.

Our findings make four contributions. First, we show that the cognitive scars of trauma are malleable: the quality of mental simulation can be significantly improved, with meaningful effects on behavior and economic outcomes. Second, we demonstrate that history shapes not only present behavior but also how new experiences are interpreted—affecting whether programs help or harm. Third, we introduce a practical, scalable intervention that improves future simulation capacity. Fi-

nally, we contribute new tools for measuring the specificity and emotionality of mental simulation in field settings.

Our work builds on and complements several strands of literature. [Alan and Ertac \(2018\)](#) study a classroom-based program in Turkey that encourages primary school children to consider future consequences. [John and Orkin \(2022\)](#) evaluate a light-touch intervention in Kenya that prompts low-income women to imagine their future selves. Both aim to increase the salience of future rewards and have been shown to shift behavior in domains such as health and savings. Related work in psychology shows that individuals who vividly imagine and emotionally connect with their future selves are more likely to make forward-looking choices ([Hershfield et al., 2011](#)). We build on this work by both focusing on and measuring specificity and emotionality in mental simulation. We designed with trauma in mind, to repair the cognitive capacity to construct emotionally grounded, domain-specific simulations of the future. GME targets adults who have experienced trauma and shows that repairing this simulation capacity can protect against backfire in programs that demand future planning. Due to this focus, we implement and test field measurement of affect and specificity in the quality of mental simulation and its economic impacts.

We also contribute to the growing economics literature on psychological and cognitive interventions. Prior work has shown that light-touch adaptations of cognitive behavioral therapy (CBT) can improve mental health and economic outcomes in low-income settings ([Haushofer et al., 2020](#); [Barker et al., 2022](#); [Ridley et al., 2020](#)). GME differs in focus: CBT targets mood and beliefs, whereas GME targets the construction of future experiences. GME is thus a cognitive complement to mood-based therapy—focused not on how people feel in the present, but on what they are able to imagine about their future.

Finally, we speak to research on how memory and experience shape economic behavior. Recent work highlights the role of selective recall and memory-based simulation in risk perception, judgment, and belief formation ([Bordalo et al., 2020, 2024](#); [Schacter et al., 2007](#)). Other research shows that life experience shapes economic behavior through long-lasting imprint effects ([Malmendier and Nagel, 2011](#); [Malmendier and Wachter, 2021](#)). We extend this work by explicitly centering and measuring the role of affect in simulations of both memory and future scenarios, by showing that trauma-induced distortions in memory and simulation can be repaired, and, crucially, that repairing this capacity has meaningful economic effects. We also show that history not only alters decision making, it can significantly negatively alter how individuals respond to new opportunities and learning if they don't have the scaffolding of cognitive repair.

2 Trauma, GME and Economic Decision Making

In this section we sketch a model of how trauma affects decision-making. The framework highlights that trauma alters how people retrieve memories and imagine future events, and shows how these cognitive distortions affect the perceived value of forward-looking actions. Rather a new conceptualization, the framework is a summary of existing research which motivates the design of our Guided Mental Experiencing (GME) intervention to ‘repair’ the negative cognitive impacts of trauma.

We use the framework to generate testable implications of the impact of GME, and these form the backbone of our empirical analysis of in Ethiopia. We also discuss why programs may fail to be trauma-robust— leading to harm for vulnerable populations—and explain how GME can mitigate this risk. These ideas form the basis for our analysis in Colombia. While the exposition here is verbal, a formal model is provided in [Appendix A](#).

2.1 A Model of Trauma and Decision-Making

Trauma-related disorders are disorders of memory ([Brewin and Holmes, 2003](#); [Brewin, 2014, 1998](#); [Cohen and Kahana, 2022](#); [Rubin et al., 2008](#)). The hallmark memory distortion associated with trauma is intrusive memories of the traumatic event. These take the form of sensory (predominantly visual) memories which spring to mind unbidden ([Ehlers et al., 2004](#); [Iyadurai et al., 2019](#)). While everyone can experience involuntary memory recall, the distinguishing feature of traumatic reactions is that these recalls are aversive, unwanted and have traumatic content ([Berns et al., 2007](#); [Berntsen and Rubin, 2002](#)).

Following research in cognitive psychology, we assume that forward-looking decision-making relies on a process known as episodic simulation—the construction of imagined future events using fragments of past experiences ([Schacter et al., 2007](#); [Addis, 2020](#); [Peters and Büchel, 2010](#)). When an individual considers a forward-looking action—such as applying for a job or starting a business—she imagines possible outcomes using cues drawn from memory. If recalled experiences are detailed, emotionally grounded, and suggest the possibility of success, this simulation is likely to motivate action. If they are vague, negative, or emotionally aversive, it may discourage planning or risk-taking.

Trauma impairs the cognitive process of mental simulation through two distinct, but reinforcing, mechanisms: *distortion* and *avoidance*.

- **Distortion** occurs because traumatic memories are encoded with high emotional salience, making them more likely to be retrieved—even in unrelated contexts ([Brown et al., 2014](#); [Kahana, 2020](#); [Bordalo et al., 2024](#)). Critically, this distortion can spread.

In the period after trauma, individuals continue to simulate choices in daily life. If the emotional environment is saturated with fear, failure, or shame, even unrelated simulations may become negatively colored—biasing downward individuals’ simulations of utility outcomes of economic choices (Ehlers and Clark, 2000; Brewin et al., 2010). These simulations are then stored as new memory traces—encoding imagined failure or hopelessness even in domains where no real setback has occurred (Pearson et al., 2008; Schacter et al., 2008).

Experimental work supports this dynamic. When neutral images (e.g., the Eiffel Tower) are paired with emotional content (e.g., a crying child), individuals later rate the neutral image itself as emotionally charged—and even misattribute this emotion to related but unseen stimuli (Madan and Kensinger, 2021). This “emotional bleed” implies that affective distortions may generalize beyond the original context, covertly shaping how individuals evaluate a broader set of future possibilities (Kleim et al., 2014).

- **Avoidance**, Avoidance arises because the act of simulating itself becomes emotionally unpleasant. Simulating the future may trigger anxiety, hopelessness, or re-experiencing, prompting individuals to disengage from mental construction of the future (Foa and Kozak, 1986; Bomyea et al., 2016). The result is vague and incoherent beliefs rather than concrete simulations—what Schacter et al. (2007) describe as a “blank future.”

As a result, individuals may appear passive or unmotivated—not because they lack aspirations, but because they lack usable simulations to support action.

Panel A of Diagram 1 provides a schematic representation of the argument.

2.2 The Role of GME

Guided Mental Experiencing (GME) aims to counteract distortion and avoidance in a targeted domain by guiding individuals to generate vivid, goal-specific simulations of future success and the path there (including possible failures). This draws on psychological evidence that structured, high quality mental simulation can improve planning, enhance motivation, and reduce avoidance (Pham and Taylor, 1999; Oettingen, 2012; Holmes et al., 2009). High quality mental simulation involves a high degree of specificity and emotionality: essentially as close as possible to a true experience.

A large body of evidence in psychology underscores the mechanisms through which the quality of mental simulation is malleable and can be improved. A recent meta-analysis of 27 experiments shows that prompting individuals to vividly recall emotionally rich autobiographical events (“episodic specificity induction”) results in a significant improvement across all studies in the specificity of future simulation on imagination tasks (Lauro, 2024).³

³Recent research finds that women can be more sensitive and responsive to episodic specificity than men, especially

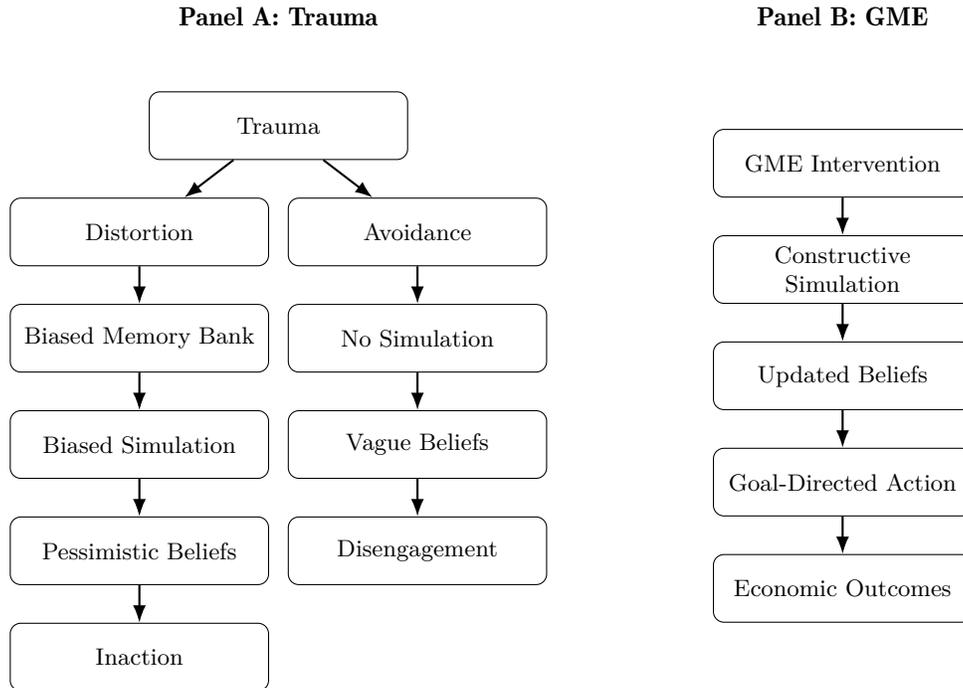


Diagram 1. Panel A shows how trauma impairs decision-making via distortion and avoidance. Panel B shows how GME restores the simulation-belief-action pathway.

The same meta-review finds that episodic simulation induced for specificity improves performance on forward-looking tasks requiring complex reasoning and sequencing. For example, participants primed with episodic recall solved the Tower of Hanoi puzzle more efficiently, suggesting stronger sub-goal construction and means–ends reasoning.

Designing a structured training to improve the quality of mental simulation beyond the laboratory and that can have economic impact, particularly for populations who have faced great trauma, requires identifying a potential population who not only has a history of trauma but also a shared economic goal, such as having a successful business as in Colombia or creating a better economic life for their family, as among the refugees in Ethiopia. The exercises that make up GME guide individuals through concrete, causally structured simulations—e.g., renting a stall, arranging goods, interacting with customers in the case of Colombia. These imagined experiences form new entries in the memory system, which can serve as positive anchors in subsequent simulations (Pile et al., 2021; Holmes et al., 2007).

with high retrieval support (Fuentes and Desrocher, 2013). We show heterogeneity by gender in the latter part of the paper.

The goal salience helps motivate participants to engage. GME does not attempt to rewrite the entire memory bank; it targets a specific goal domain. This not only enhances feasibility but may also increase demand for the intervention, particularly in low-income settings where mental health services are often underutilized due to economic priorities (Haushofer and Fehr, 2014).

In this sense, GME acts as a cognitive bridge between where participants are and where they want to go. It builds a set of new, emotionally appropriate memories – a bare-bones scaffold –that can be used to support forward planning without relying on distorted or aversive content. These simulations are typically confined to the specific goal domain in which they are practiced; GME does not construct a general-purpose cognitive skill, but rather a targeted set of usable experiences that can be drawn upon in decision-making. As these imagined futures are rehearsed, they are encoded into memory, becoming increasingly available for future simulation and action.⁴

This conceptual framework leads to four claims, which structure our empirical analysis in Ethiopia. We divide our claims into primary claims (those that are essential to the effectiveness of GME and its key channels) and secondary claims, those that are supportive of the mechanism that we hypothesize between simulation and outcomes.

Primary claims:

1. GME reduces distortion and avoidance in mental simulation in the goal domain;
2. GME improves economic outcomes and subjective well-being.

Secondary (mechanism) claims:

1. GME leads to more concrete, plausible, and forward looking beliefs and plans; and
2. GME increases motivation and supports taking active action (rather than avoiding or passive behavior).

2.3 Traditional Programs and Trauma Non-Robustness

Traditional anti-poverty programs—from business training to cash transfers—often require participants to imagine and plan for their future. These designs implicitly assume intact capacity for simulation and belief formation (Banerjee et al., 2015; Karlan et al., 2014). But as discussed, trauma may impair these processes.

For trauma-affected individuals, program-relevant simulations may retrieve distorted memory traces, leading to pessimism and inaction. Worse, engaging with such programs may reinforce negative

⁴We are grateful to Elizabeth Kensinger for her insight that GME may be providing cognitive scaffolding by engaging hippocampal - rather than amygdala - binding, enabling participants to overcome some of the event-segmentation challenges associated with trauma which is linked to overgeneral memory (Williams et al., 2022).

simulation pathways by adding new negatively valenced content to memory (Hayes et al., 2012). This creates a feedback loop: future-oriented programming elicits painful or incoherent imagery, leading to disengagement and diminished program impact (Bomyea et al., 2016).

We refer to this vulnerability as *trauma non-robustness*: the idea that standard programs may perform worst for those most harmed by past adversity.

GME is designed to mitigate this risk. When delivered alongside traditional programming, it supports participants in creating vivid, positively valenced simulations that seed the memory database with constructive content. This makes program content more accessible and emotionally manageable, and supports downstream behavior and outcomes (Holmes et al., 2009; Schacter et al., 2007). This leads to four testable claims for our Colombia analysis:

Primary claims:

1. Traditional training may increase distortion and avoidance in mental simulation in the goal domain. GME reduces or eliminates these impacts;
2. Traditional training may worsen economic outcomes and subjective well-being, GME reduces or eliminates these impacts.

Secondary (mechanism) claims:

1. Traditional training may result in more pessimistic or incoherent beliefs. GME reduces or eliminates these impacts;
2. Traditional training may undermine motivation and propensity to take action. GME reduces or eliminates these impacts.

3 Context, GME Intervention, and Experimental Design

3.1 Context and Populations

We evaluate GME in two high-adversity settings: Eritrean refugees in Addis Ababa, Ethiopia, and low-income aspiring entrepreneurs in Bogotá, Colombia.⁵

Economic constraints differ across settings. In Ethiopia, only 19 percent of refugees work despite 59 percent of comparable Addis Ababa residents being economically active (Table 1, columns 1 and 3). While refugees lacked work permits at the time of the study, informal employment was de facto not criminalized. At baseline thirty-nine percent were actively searching for employment, suggesting

⁵Appendix Sections D.1 (Ethiopia) and E.1 (Colombia) provide detailed discussions of the experimental context.

available opportunities exist but remain inaccessible. Conditional on employment, working refugees earn approximately half the wages of the general population (USD 98 versus USD 189).

In Colombia, participants live in a context where micro and small firms comprise over 87 percent of businesses. Sixty-seven percent earn below minimum wage, averaging USD 200 monthly. While 55 percent already operate businesses, government partners reported that traditional economic programs consistently failed among trauma-affected populations.

Government partners in both countries identified trauma as a likely barrier to the effectiveness of economic programs. These concerns are reflected in rates of trauma exposure and probable post-traumatic stress disorder (PTSD) in our experimental samples.

In Ethiopia, 98 percent of the sample report having directly experienced or witnessed at least one type of trauma in their lives, and 13 meet the threshold for probable PTSD – more than three times the rate in the general global population. Common trauma types include life-threatening illness or accidents (92 percent of the sample, compared to 34 percent in the global general population), exposure to war or armed conflict (76 percent of the sample), and physical assault (70 percent of the sample).

In Colombia, trauma exposure is similarly prevalent, with eighty-three percent of participants reporting having experienced or witnessed traumatic events. 22 percent screen positive for PTSD. Nearly half report war-related trauma, and 61 percent report having witnessed or survived an assault – again far exceeding global averages.

Qualitative interviews with participants in both contexts echoed the concern by policymakers: Participants expressed difficulty imagining concrete steps toward improving their livelihoods – a pattern consistent with our conceptual framework.

3.2 GME Intervention and Experimental Designs

Our GME intervention directly addresses the distortion and avoidance mechanisms identified in our model by guiding individuals to generate vivid, goal-specific simulations that form new entries in the memory system. Consistent with our framework, these repeated simulations create positive anchors that can be recalled without interference from trauma memories, thereby reducing the probability of drawing distorted memory traces when considering future economic actions. By building this cognitive scaffold in a targeted domain, GME restores the simulation-belief-action pathway.

We implemented this theoretical approach through structured sessions with three key elements: (1) repeated practice of concrete, causally structured simulations in the goal domain; (2) exercises designed to enhance specificity and emotionality—the key components of high-quality mental sim-

ulation; and (3) generation of positive anchors to serve as new memory entries accessible without trauma interference. In both countries, GME was adapted to the local context, delivered by trained non-specialist facilitators, and implemented through government partnerships. We next describe how these core principles were operationalized in each setting.

In Ethiopia, we partnered with the Ethiopian Refugee & Returnee Service (RRS) to recruit 1,652 Eritrean refugees in Addis Ababa through their centralized administrative database. Participants were randomly assigned to either GME (n=900) or control (n=900), stratified by gender and a cross-randomized priming treatment not discussed in this paper (Figure 1).

In Ethiopia, the GME intervention consisted of four one-hour sessions delivered by trained refugee facilitators from the same community. Sessions focused on imagining a potential life in Addis Ababa in the immediate term, building a safe mental space for economic decision-making, identifying specific actions to access economic opportunities, and developing concrete plans to improve economic outcomes. Each session featured two to three GME exercises guided by pre-recorded audio from clinical psychologists to ensure standardization. Sessions were held in a community center, with a referral system in place for participants in need of additional mental health support.⁶

In Colombia, the experiment evaluated whether integrating GME into standard economic support programs can enhance their effectiveness for vulnerable populations. We partnered with Bogotá’s District Department of Social Integration (SDIS) and District Department of Economic Development to recruit 1,967 vulnerable aspiring entrepreneurs through a multi-channel media campaign advertising an entrepreneurship program.⁷ We randomized participants into three arms: business training plus GME (n=956, 48 percent), traditional business training (n=558, 28 percent), and a no-intervention control (n=453, 23 percent), stratifying assignment by subdivision, sex, age, entrepreneurship status, and income level (Figure 2).

Both the traditional and the GME-enhanced training program consisted of ten three-hour group sessions covering identical entrepreneurship themes in the same order (e.g., product development, marketing, budgeting). The key distinction was that the GME-enhanced program included three to four mental experiencing exercises per session - always tied to the business content of the day - while the traditional treatment replaced these with group discussions, role plays, and written work of the same content and time length. For example, in one session, traditional participants completed a written task listing how their business matched their skills. In the GME version, participants were guided to imagine running their business—focusing on how it felt, what they

⁶This included a risk protocol and referral mechanism for individuals with severe symptoms. Of the 24 identified for additional evaluation, 17 agreed to seek care and were referred to a local provider.

⁷As part of their mandate, the SDIS were required to train numerous groups that face economic and social challenges. As a result, they brought to us a sample that includes victims of conflict and Venezuelan migrants, as well as low-income youth, LGBTQ groups, entrepreneurs with disabilities or carers thereof, the formerly homeless, and the elderly. Each of these groups is labeled “subdivision” by our field partner and in the experiment.

saw and heard, and what specific actions they were taking. By embedding GME exercises directly into business-relevant content, we ensured that the new memory entries created through simulation were immediately linked to the entrepreneurial domain. This design follows our model’s prediction that domain-specific positive anchors are most effective when they can be readily recalled in the context of relevant economic decisions. Trainers in both arms used standardized scripts and visual materials, and sessions were held in consistent venues across neighborhoods.

For detailed descriptions of recruitment procedures, intervention contents, and experimental protocols, see Supplementary Materials in Appendices [D.2](#) and [E.2](#).

4 Measurement and Empirical Strategy

We study the causal impact of GME using survey data collected at baseline and post-intervention in both Ethiopia and Colombia. In Colombia, baseline data were collected in two waves between July and December 2019 (corresponding to the two implementation waves of the program). Due to COVID-19 restrictions, follow-up surveys were conducted by phone in May and November 2020 – approximately seven and thirteen months after the intervention – and are pooled in our analysis. In Ethiopia, baseline surveys were conducted in-person between May and July 2022, and a follow-up took place three months later, approximately six weeks after the intervention. We present our Ethiopian results, where we direct test GME, first. Outcome measures are pre-specified in pre-analysis plans registered in the AEA RCT Registry.⁸ Supplementary Materials in Appendices [D.5](#) (Ethiopia) and [E.5](#) (Colombia) provide additional detail on measurement in each context.

We divide our causal claims into two parts. First, we have two primary outcomes of interest: mental simulation and economic performance. As economists our overall goal is to understand economic outcomes, and our main claim in this paper is that mental simulation can be altered by GME, and this leads to economic gains. We are also clear about how to measure mental simulation and economic outcomes, and what we expect these outcomes to capture.

Second, we have measures that lie on the causal path between simulation and economic outcomes. These are beliefs and goal-oriented actions. Where relevant, composite indices are constructed by standardizing and averaging across related items. Measures are tailored to each context and harmonized across settings where possible.

⁸Ethiopia RCT ID: AEARCTR-0008934 ([doi:10.1257/rct.8934-1.1](https://doi.org/10.1257/rct.8934-1.1)); Colombia RCT ID: AEARCTR-0004695 ([doi:10.1257/rct.4695-1.7](https://doi.org/10.1257/rct.4695-1.7)). The construction of our main outcomes, model specifications, and heterogeneity analyses follow the plans laid out in our pre-analysis. Reports that detail deviations from each PAP are available at bit.ly/lswopapreport_eth (Ethiopia) and bit.ly/lswopapreport_col (Colombia).

4.1 Measurement of Primary Outcomes

Simulation

Our GME program is aimed at improving the quality of simulation through increased specificity and emotionality of mental simulation.

In Colombia, the need to survey over the phone led us to measure the quality of mental simulation using two short scales, which are adapted versions of the Prospective Imagery Task (PIT) (Stöber, 2000) and the Spontaneous Use of Imagery Scale (SUIS) (Nelis et al., 2019). The adapted PIT presents participants with three positive and three negative scenarios, across both the general and business domains, and asks them to imagine a related future event (e.g., “the COVID-19 pandemic is over, and you save enough money to buy an asset you really want” or “the COVID-19 pandemic is over, and your business closes”). Respondents describe their simulation of these scenarios according to its level of vividness and emotionality on a five-point Likert scale. The adapted SUIS assesses the respondents’ propensity to make use of simulation in business and general settings. They are presented with a statement e.g., “When I need to go to a meeting, I picture the route in my mind before going” and asked to rate how much they agree with it on a five-point Likert scale. We aggregate measures of self-reported specificity and emotionality from the PIT and frequency of use from the SUIS to construct an overall index for the quality of mental simulation.

The Ethiopian context gives us the opportunity to improve on these measures by enabling longer, in-person surveys, making the link to memories as well as future scenarios, and recording responses which are then blindly coded. For that purpose, we utilize a combined and adapted version of the PIT (Stöber, 2000) and the Autobiographical Memory Test (AMT) (Griffith et al., 2009). Participants are asked to describe specific memories or future events they recall based on different scenario prompts. For both past and future recollections, they are presented with a positive, a negative and two neutral scenarios as prompts to recall a specific memory or imagine a future occurrence (e.g., “A friend of yours gets fired from a job” or “You receive a gift from a neighbor”).⁹ For each scenario, the quality of respondents’ mental experiences is scored according to its level of specificity and emotionality by two treatment-blind research assistants following a pre-defined coding scheme. Coders are also asked to assess the scenarios’ emotional valence (from very negative to very positive). To create the quality of mental simulation index, we average the office-coded scores on specificity, emotionality, and frequency of positive scenarios (among neutral prompts) across both past and future scenarios¹⁰. To corroborate the link between past and future simulations, Figure 3 shows that the quality of memories and future imagined scenarios are positively related with each

⁹Contrary to the standard AMT, we use scenarios and not words as prompts to better reflect economic contexts.

¹⁰We also ask respondents to describe their mental simulation according to its level of specificity and emotionality (on a Likert scale from 1 to 5). Self-reports and back-office coded measures are positively correlated in the Ethiopia sample (see Figure B2)

other. This supports our choice of including both past and future elicited scenarios in our main indices.

Appendix B shows that our mental simulation indices are correlated with economic behavior and vary by baseline trauma, indicating that they capture meaningful differences across individuals in line with our framework.

Economic Outcomes

We align economic measures with each intervention’s focus, and pre-specify the outcomes and analysis.

In Ethiopia, where GME targeted general livelihood improvement, we measure: employment status, weekly earnings (from all sources), consumption expenditure, and food security. In Colombia, where GME was embedded in entrepreneurship training, we measure: (i) an earnings index combining personal income and business sales, and (ii) business survival (operating or temporarily closed). Given Colombia’s data collection during COVID-19, we assess outcomes both pre- and post-lockdown.

Mental Health and Welfare

We measure trauma exposure and symptoms in our samples at baseline using context-appropriate scales, in order to use them for heterogeneity. Trauma exposure is assessed using a contextually-relevant trauma history checklist (Life Events Checklist, LEC-5) that captures whether respondents experienced or witnessed a traumatic event, and its type (e.g., assault, natural disaster, torture).

PTSD symptoms are measured using scales that assess subjective distress caused in the past month by the most traumatic event ever lived. In Ethiopia, we use the PCL-5 (Weathers et al., 2013), which is a PTSD checklist following the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). In Colombia, we use the 22-item Impact of Event Scale-Revised (IES-R, see Weiss and Marmar, 1997). For both scales, symptoms include “trouble concentrating”, “disturbing dreams” and “trying not to think about the event.” We refer to participants as “high trauma” when their scores are above practitioners’ thresholds for a probable PTSD diagnosis.¹¹

In terms of outcome measures related to mental health and wellbeing, we include in our endline survey in Ethiopia a measure of functional impairment using the World Health Organization Disability Assessment Schedule (WHODAS), reverse-coded so that higher scores indicate better functioning.

¹¹We use a cutoff score of 31 for PCL-5 (Bovin et al., 2016) and a cutoff score 33 for IES-R (Creamer et al., 2003). We used different scales because we followed the guidance of local trauma experts in each context, which prioritized the availability of a culturally adapted and translated version. The two scales are known to be moderately to highly correlated.

We also collect subjective well-being using the Cantril Self-Anchoring Ladder, a visual scale of life satisfaction.

In Colombia, we measure psychological distress with the Kessler-6 (K6), a short screener for mood and anxiety symptoms (Kessler et al., 2002). In one of our two follow-up surveys, we also measure life satisfaction through the Cantril Self-Anchoring Ladder.¹²

4.2 Measurement of Secondary Outcomes

These measures help to build the causal chain between an improvement in the quality of mental simulation and economic outcomes. Not all of them were pre-specified, and therefore are classed as “secondary” but help to illuminate the path to economic impact.

Beliefs

In Ethiopia, we collected measures related to refugees’ beliefs about their opportunities in the labor market and intentions related to their stay in the country. In particular, participants were asked to report their likelihood of remaining in Addis Ababa under different hypothetical probabilities of receiving a work permit. We also elicited respondents’ reservation wage (with and without the permit), expectations of receiving a work permit, of finding work (with and without the permit) and of the time to find a job.

In Colombia, our measures of beliefs are limited because we focused on collecting behavioral data due to time constraints in our phone surveys.¹³

Actions

In Ethiopia, we measure goal-directed actions by looking at respondents’ involvement in the labor market. Our main indicators are the total hours they worked in the past week and the kind of work they did (for example, wage jobs versus casual or unpaid work). Because there are many (informal) work opportunities in Addis Ababa, we see the number of hours worked as mostly a choice made by the individual, rather than something decided by an employer. We also assess efforts to find work by creating a job search index. This index reflects how much effort people put into looking for a job by combining information on hours spent searching, the number of calls made, the different channels used, and the money spent on these activities.

In Colombia, we measure entrepreneurial actions by looking at various behaviors tied to managing a business. Specifically, we ask whether participants secured new financing for their enterprise,

¹²Life satisfaction is the only main outcome that was not pre-registered in the Colombia PAP, as it was added only in our final follow-up survey. We include it here to maintain consistency with the Ethiopia experiment.

¹³Appendix Table C12 shows treatment effects on the few belief-related questions we did ask in Colombia.

invested in their business (for example, by buying new tools or creating marketing materials), introduced a new product or service, hired employees or worked in the business at all.

4.3 Empirical Strategy

Ethiopia

For all outcomes, we present intent-to-treat estimates for individual i in stratum s using the following specification:

$$y_{is} = \alpha + \beta \cdot GME_i + \delta_s + \epsilon_{is} \quad (1)$$

where GME_i is an indicator variable for being assigned to the GME treatment and δ_s are controls for stratification variables (gender and a priming treatment not discussed here). As randomization was at the individual level, we report Eicker–Huber–White standard errors. Unless noted otherwise, all outcome variables are standardized and expressed in standard deviations of the control group and indices are constructed by averaging across standardized sub-items, following our pre-analysis plan and [Kling et al. \(2007\)](#).

Colombia

For all our outcomes, we present intent-to-treat estimates for household i in stratum s and survey wave w using the following specification:

$$y_{isw} = \alpha + \beta \cdot T_i + \delta_s + \delta_w + \epsilon_{isw} \quad (2)$$

where δ_s are controls for stratification variables (subdivision interacted with the experimental wave, gender, age group, entrepreneurship status and income group) and δ_w are survey wave fixed effects. Standard errors are clustered at the household level.¹⁴ Unless noted otherwise, all outcome variables are standardized and expressed in standard deviations of the control group. Indices are constructed by summing across standardized sub-items and re-standardizing the overall index. As specified in the PAP, we limit the sample to individuals who accepted to participate in the training (86 percent of the sample).

Our main tables show three specifications comparing different treatment groups as denoted by T_i : i) traditional training versus no intervention, ii) GME treatment versus traditional training, and iii) GME treatment versus no intervention.¹⁵ The comparison with the no-intervention group provides

¹⁴We collapse the data to the household level to account for the small share of individuals who reported living in the same address as another participant and were thus randomized in the same treatment.

¹⁵As specified in our PAP, due to insufficient sample size and our implementation partner’s preferences, we do not have two subdivisions in the control group for one of the experimental waves. In order to use the full available

an estimate of the effect of receiving a traditional or GME-augmented entrepreneurship program. We interpret differences between the two training programs as the impact of using GME techniques relative to standard pedagogical techniques (e.g., think-pair-share, pros and cons lists).

Robustness and Multiple Hypothesis Testing

For all the outcomes, Appendix Tables C2, C3 and C4 show results from specifications augmented with a set of controls X_i selected using Post-Double Selection Lasso (Belloni et al., 2014).

Throughout all the main tables, we use false discovery rate corrections to adjust for multiple hypothesis testing (Benjamini et al., 2006). We apply these FDR corrections separately to the sets of simulation outcomes, economic and mental health outcomes, and secondary outcomes, reflecting our belief that findings in each of these domains are of independent interest.

5 Results

In Ethiopia, we find evidence of impact across the entire causal chain, from simulation to beliefs and income-generating behaviors. In Colombia, we test two core hypotheses: first, that traditional business training can reduce economic performance for trauma-affected individuals by impairing simulation; and second, that Guided Mental Experiencing (GME), delivered alongside training, can reverse these effects.

5.1 GME Impact on Mental Simulation Quality

Ethiopia

GME successfully improves mental simulation capacity among Ethiopian refugees. Participants assigned to GME show a 0.09 standard deviations increase in their Mental Simulation Index ($se = 0.030$), driven primarily by enhanced *emotionality* (+0.13 SD, $q = 0.01$) and more *positivity* (+0.11 SD, $q = 0.01$) (Panel A, Table 2). These improvements reflect what our theory predicts: in Ethiopia, GME appears to repair emotional distortion, enabling participants to imagine futures with appropriate affect and specificity.

The intervention proves especially powerful for those starting with the most impaired simulation capacity. Estimated quantile treatment effects indicate improvements across the distribution in the simulation index and emotionality sub-index. While specificity shows no impact on average, participants in the lowest percentiles – those whose simulations were vaguest at baseline – experience meaningful gains in detail and concreteness (Appendix Figure C1b). This distributional pattern aligns with GME’s design as cognitive repair.

sample, we prefer running separate regressions comparing the GME treatment to each of the other two arms.

The estimated treatment effects on mental simulation are robust to covariate adjustment with PDS Lasso (Appendix Table C2) and do not appear to be driven by social desirability bias (Appendix Table C13).

Colombia

Consistent with our theory, in Colombia we find evidence that, without GME, mental simulation in the goal domain becomes more distorted and avoidant.. Table 2 (Panel B) shows that the traditional training reduces the quality of simulation in the domain of entrepreneurship. Participants in this group score -0.16 SD ($p < 0.1$) on a business-domain simulation index, which combines self-reported specificity, emotionality, and the frequency of use of mental simulation. This is consistent with our theory: when training demands future-oriented reasoning in a domain where memory is distorted, simulation may become degraded.

GME reverses this effect. Compared to the traditional group, GME increases simulation by $+0.18$ SD ($p < 0.05$), driven by improvements in emotionality ($+0.17$ SD, $p < 0.05$) and specificity ($+0.13$ SD, $p < 0.1$). However, GME does not significantly outperform the control group in terms of simulation outcomes.¹⁶

Both the decline of simulation quality in the traditional training and restoration in the GME training are limited to the business domain. We observe no change in simulations outside that context (Table 2, Panel B, Columns (1) and (3)). This pattern is consistent with the domain-specific mechanism outlined in the theory section: imagined futures are easier to construct when memory structures in that domain are already accessible or recently activated. The difference with respect to the Ethiopia results also reflect design choices. In Colombia, GME exercises were integrated into one-third of the sessions, all tied to business themes. In Ethiopia, GME spanned the full program and addressed general economic life. The broader effects in Ethiopia may reflect both greater exposure and greater generalizability of the simulated domain.

¹⁶Business scenarios were only asked to individuals who had a business at the time of the surveys. Results comparing GME and traditional training are unchanged when we limit the sample to households with a business at baseline (51%; $\text{coeff}=0.181$, $p=0.058$). We find no differences in baseline business status across treatment groups, supporting monotonicity: in the first follow-up, 84.5% of GME and 81% of traditional training households still had a business ($p=0.29$); in the second, 75.5% and 73.7%, respectively ($p=0.66$). Table C15 shows that even when we focus only on those with a business at baseline, there is no improvement in simulation quality outside the business domain. This suggests that learning patterns in this sub-sample are similar to those in the overall sample.

5.2 GME Impact on Economic Outcomes and Mental Health

5.2.1 Ethiopia

Economic Outcomes

Table 3 (Panel A) shows significant economic gains among GME participants in Ethiopia, in line with our primary theoretical claims. The likelihood of working increases by 0.12 SD ($p < 0.05$). This corresponds to an increase of 5 percentage points on a control mean of 21.3% ($p < 0.05$, see Table C10). Gains are concentrated in wage employment rather than casual or unpaid work, suggesting a reallocation toward more stable and structured labor (Table C11).¹⁷ In line with this increase in economic activity, weekly earnings increase by 0.14 SD ($p < 0.05$). This corresponds to a non-standardized increase of 113 Ethiopian birr per week ($p = 0.026$), or a 53% increase over the control mean.¹⁸ In line with the improvement in labor market engagement, food security increases by 0.22 SD ($p = 0.001$) and the coefficient on the treatment dummy for total consumption expenditure is also positive, but not statistically significant.¹⁹

Mental Health and Welfare

GME improves both functional and subjective wellbeing. WHODAS scores (reverse-coded so higher indicates better functioning) increase by 0.16 SD ($p < 0.005$). Life satisfaction improves by 0.13 SD ($p < 0.05$). These results suggest that cognitive and affective wellbeing may be jointly responsive to simulated mental experience.

5.2.2 Colombia

Economic Outcomes

Traditional training harms economic performance (Table 3, Panel B). Participants in this group earn significantly less than control in both the pre-COVID period (-0.22 SD, $p < 0.05$) and during the COVID pandemic (-0.14 SD, $p < 0.05$). These effects are not explained by differences in attrition or baseline characteristics (see Table C3 for the inclusion of controls using PDS Lasso).

¹⁷In Table C11, we disaggregate treatment effects on labor market outcomes at the extensive and the intensive margin by the type of work. While our experiment is not powered to detect effects separately for work in wage employment, own business, casual employment, and unpaid apprenticeships, we take the significant increase in wage work as suggestive evidence that treated participants increase their work in better paid, more desirable activities.

¹⁸While this effect on earnings is large, we caution that this is an increase from a very low base (215 birr in control group, around 4 USD) and is driven by a small share of actively working individuals (see Table C10 for non-standardized labor market outcomes).

¹⁹The GME intervention might influence labor outcomes by changing norms around the acceptability of working. Eritrean refugees in Addis Ababa are largely not formally allowed to work at the time of the data collection, though informal work is common. Using a “rule orientation” scale (Fine et al., 2016) to measure attitudes toward rules, we see no variation in treatment effects, suggesting the program doesn’t just encourage those more willing to break rules to work (Table C14).

Instead, they align with our theoretical claim that asking participants to simulate detailed future actions—without first addressing trauma-related memory distortions—may lead to disengagement or negative reinforcement. In some cases, such planning exercises may even reactivate intrusive or aversive memories.

Compared to participants in the traditional group, participants in the GME-enhanced training arm perform better (Table 3, Panel B.2). GME reverses all the negative effects of the traditional training, raising our IHS index of pre-COVID earnings by +0.19 SD ($p < 0.05$) and COVID earnings by +0.14 SD ($p < 0.05$).

Because earnings are highly skewed, we use an inverse-hyperbolic sine (IHS) transformation for our earnings indices, standardized relative to the no-intervention group. Since the IHS blends intensive and extensive margins, we also look at quantile treatment effects to explore distributional impacts. Figure C2 shows that treatment effects are much noisier in the upper tail, consistent with the IHS placing less weight on higher values. It also reveals that the drop in earnings from traditional training is concentrated below the median: compared to the control group, traditional trainees have significantly lower earnings at the 5th and 10th percentiles, indicating more participants with no or very low income, and also show significant declines at intermediate percentiles (20–50) in both pre-COVID19 and COVID19 periods.

The gap in earnings between the two trainings is not driven by a differential likelihood of having a business in the few months right after the intervention. However, we find that participants assigned to the GME group are more likely to keep their business open during the COVID-19 pandemic than those in the traditional training (by +0.05 percentage points, $p < 0.10$, Column (4) of Table 3), perhaps pointing to higher resilience.

The last row of Table 3 shows GME does not significantly outperform the control group in any of these economic outcomes, even if the point estimates are generally positive.

Mental Health and Welfare

Neither traditional training nor GME produce significant changes in psychological wellbeing. We find no effects on our main measure of psychological distress, reverse-coded so higher indicates better health, or on life satisfaction. Unlike in Ethiopia, these results suggest that GME’s mechanism in Colombia operates through cognitive restoration rather than affective repair.

5.3 Characterizing the Causal Path: Beliefs and Goal-Oriented Actions

In this section, we test our secondary claims: that GME helps people form more concrete, realistic, and forward-looking plans and beliefs, which in turn increases their motivation to take action

instead of avoiding or staying passive. The data from Ethiopia allow us to fully trace this causal pathway. The measures from Colombia is more limited and can only give partial insights into these mechanisms.

5.3.1 Ethiopia

Beliefs

Participants assigned to the GME program are more likely to express a desire to remain in Addis under hypothetical work permit scenarios, with an average effect size of +0.18 standard deviations ($q = 0.004$).²⁰ This increased willingness to stay is accompanied by more optimistic beliefs about both the likelihood of finding a job and the difficulty of securing one (see Columns (2) to (4), Table 4). On average we see no change in the reservation wage (Column (5), Table 4), indicating that the types of jobs people aim for remain similar. Together, these results suggest that the intervention enhances participants' confidence in the value of planning and taking action within the local context.

Actions

GME leads to an increase in total labor supply. Not only does employment rise on the extensive margin, but weekly hours worked also increase by 0.15 standard deviations ($p = 0.030$). This effect is of about three hours per week, relative to a control group mean of eight hours per week. This result points to an increased motivation of participants to engage in the labor market actively, reflecting a greater willingness to allocate time and effort toward work opportunities.²¹

Contrary to expectations, the job search index declines by 0.10 SD ($q = 0.017$). This may reflect a reduction in search effort among those who found jobs or became more engaged in existing work. However, it could also indicate disengagement or a shift away from active opportunity-seeking, and is thus potentially inconsistent with the theoretical expectation that GME increases goal-directed planning. To test among these hypotheses we split the sample by baseline work status (Table C9). Those who were already in work see a statistically significant reduction in search effort, and we see no such effect in the sample that were not in work, suggesting the effect is driven by a reduction in on-the-job search. Consistent with this interpretation, we also find a qualitative decline in the reservation wage for participants who were working at baseline, implying that the perceived value of their current jobs has increased. While spurring action by those refugees who were inactive at baseline, the program in Ethiopia also seems to have increased participants' awareness of the value of their current activities, in line with more grounded and realistic beliefs about their livelihoods.

²⁰We asked about the intention to stay in Ethiopia in the next three years with no work permit, or varying the likelihood of receiving it (1%, 50%, 99%). The outcome in the table reports the average response across the scenarios.

²¹Qualitative evidence suggest that informal work opportunities are available to our sample of Eritrean refugees in Addis Ababa, mostly through community networks, with refugees engaging in multiple forms of casual and self-employment that allow flexible hours (see Appendix C.1 and Table C11).

Simulation to Outcomes

Figure 6 shows a direct connection between changes in mental simulation quality and economic outcomes. The analysis focuses on refugees who were not working at the start of the study (82% of the sample), assigned to the treatment group. The figure compares their mental simulation quality at baseline (x-axis) and at the endline (y-axis). Refugees who remain unemployed by the endline show little to no improvement in mental simulation quality. By contrast, every participant who found employment by the endline demonstrated noticeable improvements across the entire range of mental simulation quality. This evidence supports our mechanism claims — that improved mental simulation helps individuals form more concrete, plausible, and forward-looking plans and increases motivation to take active steps — ultimately translating into improved economic behavior.

Additional Intermediate Outcomes

Panel A of Table C12 reports additional outcomes related to beliefs and preferences. GME mainly improves future-oriented beliefs and patience, consistent with our theoretical claim that GME fosters more concrete, forward-looking mindsets that support motivation and planning for future goals. Specifically, we find that the intervention led to statistically significant increases in optimism (0.17 SD, $p < 0.01$), measured using the Revised Life Orientation Test, a 10-item psychometric scale to assess levels of optimism versus pessimism in life. We also find smaller effects on aspirations (0.11 SD, $p < 0.05$), and on self-reported levels of trust and patience (both of 0.11 SD, $p < 0.05$). We find no meaningful impacts on self-efficacy or risk aversion. This suggests that GME helped participants develop more hopeful expectations about the future and a greater willingness to wait for future gains or take risks in the social domain, even if the program did not broadly change all psychological traits.²²

5.3.2 Colombia

Business Actions

The traditional training weakens participants' business engagement (Table 4, Panel B). Compared to the control group, those in the traditional training arm are significantly less likely to have secured new funding for their business (by about 6 percentage points, $p < 0.10$) and also receive smaller amounts of funding on average (0.15 SD lower, $p < 0.10$). They are also somewhat less likely to spend any time working on their business or to have employees, although these differences are not statistically significant. These findings align with the idea that distorted mental simulations can lead people to disengage from their goals

²²A mediation analysis shows that the change in beliefs is the strongest mediator, explaining about 30% of the increase in employment, followed by optimism at 11%.

In contrast, GME reverses such disengagement in business activities. Compared to the traditional training, GME increases the likelihood of getting new funding and the average amount of funding received ($p < 0.05$), as well as time spent working in the business (by 5 percentage points, $p < 0.05$). GME participants are also more likely to introduce new products, invest in assets or marketing, and manage employees, although these differences are modest and not statistically significant with respect to the traditional group. Yet, when compared to receiving no intervention, we see that GME trainees are significantly more likely to introduce new products (by 6 percentage points, $p < 0.05$) and invest in marketing (by 6 percentage points, $p < 0.10$), pointing to net gains in entrepreneurial attitudes. Overall, these patterns suggests that GME helps counter the cognitive costs of trauma that might otherwise undermine engagement with traditional programming.

Data collected from about 50 percent of the original sample in 2021 (two years after the intervention) support this interpretation. As shown in Table C16, GME-trained business owners are more likely to take initiative in low-cost marketing activities that expose them to potential rejection or criticism, such as seeking feedback from customers or competitors. In contrast, participants in the traditional training tend to focus on more “internal” tasks—like record-keeping or expense planning—which, although framed as good business practices, may actually function as a way to avoid taking real, outward-facing actions that could lead to emotional discomfort or failure. This fits well with our theoretical framework: GME helps reduce avoidance and distorted thinking, encourages the development of concrete, future-oriented plans, and motivates individuals to take proactive steps—even when these involve emotional risks they might otherwise shy away from.

Additional Intermediate Outcomes

Table C12, Panel B, presents additional outcomes from Colombia, highlighting how the interventions may have influenced beliefs and actions beyond core business activities.

We find no effects on participants’ self-reported confidence to handle challenges, measured as an index which aggregates scores on psychological scales of self-efficacy and resilience.

We also find no significant effects on perceptions of financial safety nets — such as having enough savings or cash to manage short-term needs — nor on COVID-19 mitigation behaviors, indicating that the intervention did not substantially change participants’ immediate sense of financial security or their crisis-specific adaptations.

Columns (4) to (6) examine broader economic actions outside the business domain. GME participants are somewhat more likely to report having saved any money than their peers in control (0.048 SD, $p < 0.10$), but there is no significant impact on the amount saved. Likewise, there is no detectable effect on whether individuals engaged in wage work. This pattern suggests that while GME may have encouraged a slight increase in precautionary saving behavior, it did not

meaningfully shift broader labor market choices outside the business domain. The concentration of the GME effects in actions within the business domain suggests that when mental simulation is scaffolded effectively, people become more willing to take forward-looking actions specifically tied to their goals.

5.4 Exploring Variation by Baseline Trauma

We examine whether the effects of the intervention differ by baseline trauma, using sources of heterogeneity that were pre-specified in both contexts. Our theoretical framework suggests that GME may help reduce the negative impact of trauma on economic outcomes. Therefore, we expect to see larger treatment effects among individuals with higher initial levels of trauma symptoms. However, because trauma levels are already high across both of our samples, there may be limited variation in actual trauma burden. As a result, our ability to detect meaningful heterogeneity by trauma is likely to be constrained, and these analyses should be interpreted with caution. Given that women generally experience a higher trauma burden—and this pattern also holds in our data—we additionally explore whether treatment effects differ by gender.²³

5.4.1 Ethiopia

We define respondents who score above the PCL-5 cut-off for probable PTSD at baseline as having “high trauma.” Table C5 shows treatment effects in the low vs high trauma samples.²⁴ The coefficients are larger among highly traumatized participants for five of our seven main outcomes. While treatment effects in both groups are generally positive and significant, the statistical significance of the difference in difference varies. For example, in the control group, participants in the “high trauma” category report life satisfaction that is 0.44 standard deviations lower than others. The GME treatment almost closes this gap, increasing life satisfaction in this group by 0.35 SD. A similar pattern appears for the WHODAS score: high-trauma refugees report significantly poorer health in the control group, but under GME, their health is on par with that of low-trauma peers. The GME program also raises the likelihood of employment by 0.30 SD among highly traumatized participants, compared to 0.09 SD in the lower trauma group, although the difference between the two effects is not statistically significant at conventional levels ($p=0.19$).

Yet, when we look at an overall standardized index that combines all the main outcomes—covering mental simulation, economic activity, and mental health—we find that the treatment effects for the

²³Randomization was stratified by gender in both experiments, and we pre-registered gender heterogeneity for the Ethiopia experiment.

²⁴We run an interacted model: $y_{is} = \alpha + \beta_1 \cdot GME_i + \beta_2 \cdot HighTrauma_i + \beta_3 \cdot GME_i \cdot HighTrauma_i + \delta_s + \epsilon_{is}$. The coefficients reported in Table C5 are β_1 (treatment effects in the low trauma sample), $\beta_1 + \beta_3$ (treatment effects in the high trauma sample) and β_2 (effect of high trauma in control). At the bottom of the table, we include a test of $\beta_1 = \beta_1 + \beta_3$. We use the same approach for heterogeneity by gender and in the Colombia experiment.

high trauma group are significantly larger than for the low trauma group ($p=0.04$, Column 1 of Table C5). This suggests that the program is generally more effective for individuals with higher initial levels of trauma.

For female participants, the results are more mixed (Table C6). Women show the same increase in mental simulation as men, in a statistical sense. However, while women in the control group generally do worse than men on most measures of economic activity, the treatment effects on economic outcomes are also generally weaker for women: in three out of four outcomes, the effects are smaller, and we cannot reject the null of equal treatment effects between the two genders. The only outcomes where we find significantly stronger effects for women are in mental health and welfare. For example, the treatment improves the WHODAS score by 0.27 standard deviations for women, compared to just 0.05 for men ($p=0.03$ on the gender difference).

These heterogeneous effects are visualized in Panel (a) of Figure 4 for trauma heterogeneity and Panel (a) of Figure 5 for gender heterogeneity.

5.4.2 Colombia

We define respondents who score above the IES-R cut-off for probable PTSD at baseline as having “high trauma.” Table C7 shows treatment effects in the low vs high trauma samples. Just like in the Ethiopia setting, we find evidence that treatment effects are largest among participants with high trauma burden, although the statistical significance of these effects is mixed.

On both simulation and economic outcomes, the traditional training appears especially harmful for participants with high trauma levels. For example, the simulation index declines by 0.20 SD in the low trauma group ($p < 0.10$), and by an even larger 0.25 SD in the high trauma group ($p > 0.10$). Similarly, pre-COVID-19 earnings fall by 0.21 SD in the low trauma sample ($p < 0.10$) and by 0.44 SD among highly traumatized participants ($p < 0.05$), although we cannot reject the null of equal effects across the two groups.²⁵

Because high trauma participants seem to be most adversely affected by the traditional training, they also stand to gain the most from GME. In this group, GME increases business-domain simulation by 0.36 SD ($p < 0.05$) and pre-COVID earnings by 0.36 SD ($p < 0.10$), compared to smaller effects of 0.15 SD ($p < 0.10$) and 0.20 SD ($p < 0.10$), respectively, in the low trauma group.

To gain more statistical power, we also aggregate all main outcomes into an overall standardized index, reported in Column (1). This confirms the pattern: GME raises the index by 0.18 SD for high trauma participants, compared to 0.10 SD for their low trauma peers ($p = 0.34$ for the

²⁵We obtain more precise estimates when defining “high trauma” using both the IES-R threshold and an indicator for being a registered conflict victim or a recent Venezuelan migrant, which increases the share of participants classified as “high trauma” from 25% to 50% of the sample.

difference). Moreover, Panel C of Table C7 shows that GME has a positive and significant effect on high trauma participants relative to the non-intervention group on this index, highlighting a potential net benefit of GME for those with higher trauma levels.

Overall, these patterns support the hypothesis that traditional training is not trauma-robust, and that GME serves to protect the simulation system against collapse when cognitive capacity is compromised.

Table C8 shows heterogeneity by gender. Our results on both mental simulation and economic outcomes are driven by female participants. Treatment effects on female participants are stronger across all key variables of interest, except for having an operating business. At baseline, women reported a similar number of traumatic experiences as men but experienced a higher trauma burden according to the IES-R scale. However, gender-based heterogeneous effects persist even after controlling for trauma, suggesting that gender-specific factors beyond trauma burden—such as differences in sensitivity and responsiveness to episodic specificity (Fuentes and Desrocher, 2013)—are important in shaping training outcomes.

Overall, these heterogeneous treatment effects confirm the key role that traumatic experiences and gender play in shaping people’s returns from participation into economic programs.

Panel (b) of Figure 4 illustrates these differential effects by trauma, while Panel (b) of Figure 5 shows the corresponding patterns by gender.

6 Robustness and Validity

Measurement Validity and Survey Bias. To address concerns about social desirability bias in our mental simulation measures, we use blind enumerator-coded outcome measurement and cross-validate them with self-reported outcomes (in Ethiopia, see Figure B2). Results remain robust across both types of measurement (Table C1).

A remaining concern is that the types words that the participants use – on which treatment-blinded research assistants base their coding – can be influenced by social desirability bias. To address this, Panel A of Table C13 presents estimates on mental simulation indices of the treatment dummy interacted with participants’ score in the Marlowe-Crowne Social Desirability Scale (Crowne and Marlowe, 1960; Dhar et al., 2022) measured at endline. The coefficients on the interaction between the treatment and the social desirability score tend to be negative or close to zero, and we can never reject the null hypothesis of no effect, suggesting that people most concerned with social approval do not respond differently to our mental simulation questions (Panel A, Table Table C13). We also find no interaction between the Marlowe-Crowne Social Desirability Scale and answers to our main

outcomes (Panel B, Table [Table C13](#)).

Balance and Experimental Integrity. Treatment assignment was randomized and stratified by key demographic and geographic characteristics. In Ethiopia, we stratified assignment by gender and a priming treatment (not discussed here). Appendix [Table D1](#) shows that baseline characteristics are well balanced across treatment groups, with no significant differences on observable variables. A joint orthogonality test fails to reject equality of means across arms ($p = 0.60$), confirming the integrity of the randomization. In Colombia, we stratified assignment on subdivision, experimental wave, gender, age group, entrepreneurship status and income group. Appendix [Table E1](#) shows that treatment assignment is also balanced on observables, and a joint orthogonality test fails to reject equality of means across arms ($F(15,1818)=0.65, p=0.83$).

Program Take-up and Compliance. In Ethiopia, 75 percent of participants assigned to GME attended at least one session, with an average attendance of 2 out of 4 sessions. Take-up was higher among older and more trauma-exposed individuals (Appendix [Table D2](#)), consistent with the intervention’s focus on repairing simulation capacity among those most affected. In Colombia, attendance rates were 67% for GME and 63% for the traditional training. Of those attending at least one session, 48% in GME and 53% in traditional attended seven or more of the ten sessions and qualified for a certificate.²⁶ Take-up is broadly balanced across demographic characteristics ([Table E2](#)), but also in Colombia older people and those who suffered a larger number of traumatic experiences are more likely to participate in at least one session.

Attrition and Sample Composition. In Ethiopia, 84 percent of baseline respondents completed the endline survey, so overall attrition was low. The treatment group was 6 percentage points more likely to reply to the survey than the control. Nevertheless, Appendix [Table D3](#) shows that baseline variables do not systematically predict attrition. Moreover, the endline sample remains balanced across arms, with the exception of years spent in Addis and a marginal difference in total income (significant at the 5% and 10% level, respectively; Appendix [Table D4](#)). In Colombia, 77 percent of our sample participated in at least one of the two follow-up surveys, and we find no differences in response rate between the GME and the traditional training treatment arms in both follow-up surveys. While the control group was around 4 percentage points less likely to reply to any of the two follow up surveys than the treatment arms ($p<0.10$), we find negligible observable differences among follow-up respondents between the control and the treatment groups ([Table E4](#))

Comparison with Other Related Studies. Our results on labor market outcomes among refugees compare favorably to other psychological interventions or job assistance programs.

²⁶By way of comparison, attendance rates in the [McKenzie and Woodruff \(2014\)](#) review of business training programs in developing countries range from 39 to 92%, with a mean of 64%.

For example, [Angelucci and Bennett \(2024\)](#) and [Bhat et al. \(2022\)](#) find that pharmacotherapy and psychotherapy reduce depression but have no impact on economic outcomes. Similarly, [Haushofer et al. \(2020\)](#) report no effects on psychological or economic outcomes from a five-week psychotherapy program. In contrast, [Patel et al. \(2017\)](#) show that a brief behavioral activation program increases reported working time in the short term, comparable to our findings in Ethiopia. Within three months of a group-based cognitive behavioral therapy program in Ghana, [Barker et al. \(2022\)](#) find improvements in self-reported health similar to those we observe in Ethiopia. Finally, the Generalized Self-Efficacy intervention in [McKelway \(2025\)](#) raises women’s work engagement by 26%, closely matching the extensive margin effects among our refugee sample. Experiments that try to shift participants’ aspirations similarly find mixed results on economic outcomes. [McKenzie et al. \(2022\)](#) observe a reduction in borrowings and business investment, while [Rojas Valdes et al. \(2022\)](#) find small and statistically insignificant increases in microenterprise sales from a “hope-inducing intervention”. An exception is [Orkin et al. \(2023\)](#), who find that an aspirations workshop increased household’s annual labor supply by 5% and consumption expenditure by 4% of a placebo mean.²⁷

Recent evidence on job search assistance shows that support can meaningfully improve labor market outcomes, but also underscores the need to tailor programs to individual circumstances for greater effectiveness ([Carranza and McKenzie, 2024](#)). Our approach highlights the added value of incorporating trauma-awareness to boost impact in fragile contexts.

Cost Effectiveness The group-based intervention in Colombia cost on average USD 532 per targeted beneficiary, whereas the individual intervention in Ethiopia costs on average USD 229.²⁸ In Ethiopia, the GME intervention leads to a gain in unconditional monthly earnings of 52.74 percent over the control group – albeit from a very low base of about USD 16.60 per month (Appendix [Table C10](#)). While this estimate is based on a limited number of non-zero earnings observations and is sensitive to extreme values, it is suggestive of the intervention’s cost effectiveness. This compares favorably, e.g., to the vocational training programs reviewed by [McKenzie \(2017\)](#), which yield an average earnings gain of 17 percent at a typical cost of USD 500 to USD 1,700 per person trained.

7 Conclusion

This paper starts from two questions: Can the cognitive effects of trauma be repaired? And does trauma shape not only behavior, but also how new experiences are interpreted? Across two randomized trials, we find evidence in the affirmative. Guided Mental Experiencing (GME), a

²⁷A recent meta-analysis suggests a more positive interpretation of mental health treatments on labor market outcomes ([Lund et al., 2024](#)), with magnitudes in line with what we find.

²⁸Detailed cost estimates are available in Online Appendix Sections [D.6](#) (Ethiopia) and [E.6](#) (Colombia).

scalable simulation-based intervention, improves economic behavior in trauma-affected populations. In Ethiopia, GME alone increases labor supply, income, and food security. In Colombia, GME prevents the harms caused by traditional business training among trauma-exposed participants. These results suggest that the constraints imposed by trauma are neither fixed nor neutral: they can be modified, and they matter for how people engage with economic opportunity.

The findings also extend the literature on history and choice. Prior work has shown that past experience shapes preferences and beliefs. We show that it also shapes how individuals respond to new programs. In this sense, trauma functions not only as a shock to cognition, but as a modifier of policy effectiveness—a form of treatment effect heterogeneity rooted in personal history.

We do not interpret GME as a substitute for structural interventions. Rather, it may be a necessary complement: many development programs assume the capacity for goal-directed planning, but do not attempt to support it. GME provides scaffolding; it helps individuals engage with these programs more effectively by restoring the cognitive infrastructure required for future-oriented decision-making.

Our results also raise new questions. In Ethiopia, we observe improvements in mental health and life satisfaction, though these gains are concentrated in specific subgroups and not clearly linked to economic outcomes. In Colombia, we see no improvements in well-being. This suggests that our economic results are unlikely to be driven solely by changes in mood. Understanding whether GME affects behavior through affective channels beyond simulation, such as confidence or emotional regulation, remain open areas for further research.

In Colombia, GME mitigates the negative impacts of traditional business training but does not yield a net positive effect relative to no intervention. This likely reflects two opposing forces: traditional programs may trigger negative memories that hinder engagement, while GME teaches participants to create more constructive mental experiences. Because most programs serve both trauma-affected and non-affected individuals, such negative effects may be obscured in average impacts. Our findings suggest that GME can buffer these risks while still benefiting others. Yet, while we concentrate on the trauma-robustness aspect of GME, there is tantalizing evidence that GME may be especially effective for those with the most severe trauma burden, outperforming even the no-intervention group.

It is also important to note that our trial in Colombia concluded just before the COVID-19 pandemic, limiting the time participants had to apply new skills. Moreover, none of the course contents addressed how to navigate broader economic shocks. Thus, we interpret our net effects as a lower bound of what could have achieved under normal circumstances or with content more directly linked to business strategy in difficult times. Integrating GME within a successful program tailored to

current economic circumstances may lead to positive net effects, presenting an important area for future research.

The capacity to imagine the future is foundational for making important economic decisions, ranging from human capital development to savings and investment. This is especially true in uncertain and risky domains, such as entrepreneurship and job search. Through training in mental experiencing, individuals can learn to see the world's opportunities in the future, even after difficult experiences in the past.

References

- Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology* 11(2), 233–259.
- Alan, S. and S. Ertac (2018). Fostering Patience in the Classroom: Results from a Randomized Educational Intervention. *Journal of Political Economy* 126(5), 1865–1911.
- Angelucci, M. and D. Bennett (2024). The economic impact of depression treatment in india: Evidence from community-based provision of pharmacotherapy. *American economic review* 114(1), 169–198.
- Banerjee, A., D. Karlan, and J. Zinman (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics* 7(1), 1–21.
- Barker, N., G. Bryan, D. Karlan, A. Ofori-Atta, and C. Udry (2022). Cognitive Behavioral Therapy among Ghana’s Rural Poor Is Effective Regardless of Baseline Mental Distress. *American Economic Review: Insights* 4(4), 527–45.
- Belloni, A., V. Chernozhukov, and C. Hansen (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives* 28(2), 29–50.
- Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93(3), 491–507.
- Berns, G. S., D. Laibson, and G. Loewenstein (2007). Intertemporal choice—toward an integrative framework. *Trends in cognitive sciences* 11(11), 482–488.
- Berntsen, D. and D. C. Rubin (2002). Emotionally charged autobiographical memories across the life span: The recall of happy, sad, traumatic and involuntary memories. *Psychology and aging* 17(4), 636.
- Bhat, B., J. De Quidt, J. Haushofer, V. H. Patel, G. Rao, F. Schilbach, and P.-L. P. Vautrey (2022). The long-run effects of psychotherapy on depression, beliefs, and economic outcomes. Technical report, National Bureau of Economic Research.
- Bomyea, J., E. D. Johnson, and A. J. Lang (2016). Information processing in PTSD: Evidence for biased attentional, interpretation, and memory processes. *Psychopathology Review* 4(3), 218–243.
- Bordalo, P., G. Burro, K. Coffman, N. Gennaioli, and A. Shleifer (2024). Imagining the future: memory, simulation, and beliefs. *Review of Economic Studies* 92(3), 1532–1563.
- Bordalo, P., N. Gennaioli, and A. Shleifer (2020). Memory, Attention, and Choice. *The Quarterly Journal of Economics* 135(3), 1399–1442.
- Bovin, M. J., B. P. Marx, F. W. Weathers, M. W. Gallagher, P. Rodriguez, P. P. Schnurr, and T. M. Keane (2016). Psychometric properties of the PTSD checklist for diagnostic and statistical manual of mental disorders—fifth edition (PCL-5) in veterans. *Psychological assessment* 28(11), 1379.
- Brewin, C. R. (1998). Intrusive autobiographical memories in depression and post-traumatic stress disorder. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 12(4), 359–370.
- Brewin, C. R. (2014). Episodic memory, perceptual memory, and their interaction: foundations for a theory of posttraumatic stress disorder. *Psychological bulletin* 140(1), 69.
- Brewin, C. R., J. D. Gregory, M. Lipton, and N. Burgess (2010). Intrusive images in psychological disorders: Characteristics, neural mechanisms, and treatment implications. *Psychological Review* 117(1), 210–232.

- Brewin, C. R. and E. A. Holmes (2003). Psychological theories of posttraumatic stress disorder. *Clinical psychology review* 23(3), 339–376.
- Brown, A. D., D. R. Addis, T. A. Romano, C. R. Marmar, R. A. Bryant, W. Hirst, and D. L. Schacter (2014). Episodic and semantic components of autobiographical memories and imagined future events in post-traumatic stress disorder. *Memory* 22(6), 595–604.
- Carranza, E. and D. McKenzie (2024). Job training and job search assistance policies in developing countries. *Journal of Economic Perspectives* 38(1), 221–244.
- Chen, G., S. M. Gully, and D. Eden (2001). Validation of a new general self-efficacy scale. *Organizational research methods* 4(1), 62–83.
- Cohen, R. T. and M. J. Kahana (2022). A memory-based theory of emotional disorders. *Psychological review* 129(4), 742.
- Creamer, M., R. Bell, and S. Failla (2003). Psychometric properties of the Impact of Event Scale—Revised. *Behaviour Research and Therapy* 41(12), 1489 – 1496.
- Crowne, D. P. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology* 24(4), 349–354.
- Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping Adolescents’ Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review* 112(3), 899–927.
- Ehlers, A. and D. M. Clark (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy* 38(4), 319–345.
- Ehlers, A., A. Hackmann, and T. Michael (2004). Intrusive re-experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory* 12(4), 403–415.
- Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics* 133(4), 1645–1692.
- Fine, A., B. Van Rooij, Y. Feldman, S. Shalvi, E. Scheper, M. Leib, and E. Cauffman (2016). Rule orientation and behavior: Development and validation of a scale measuring individual acceptance of rule violation. *Psychology, Public Policy, and Law* 22(3), 314.
- Foa, E. B. and M. J. Kozak (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin* 99(1), 20–35.
- Fuentes, A. and M. Desrocher (2013). The effects of gender on the retrieval of episodic and semantic components of autobiographical memory. *Memory* 21(6), 619–632.
- Griffith, J. W., J. A. Sumner, E. Debeer, F. Raes, D. Hermans, S. Mineka, R. E. Zinbarg, and M. G. Craske (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory* 17(6), 609–623.
- Haushofer, J. and E. Fehr (2014). On the psychology of poverty. *science* 344(6186), 862–867.
- Haushofer, J., R. Mudida, and J. P. Shapiro (2020). The Comparative Impact of Cash Transfers and a Psychotherapy Program on Psychological and Economic Well-being. Working Paper 28106, National Bureau of Economic Research.
- Hayes, J. P., M. B. Vanelzakker, and L. M. Shin (2012). Emotion and cognition interactions in PTSD: A review of neurocognitive and neuroimaging studies. *Frontiers in Integrative Neuroscience* 6, 89.

- Hershfield, H. E., M. P. Daniel, D. M. Guy, A. L. Bryan, and J. Fox (2011). Future self-continuity: how conceptions of the future self transform intertemporal choice. *Annals of the New York Academy of Sciences* 1235(1), 30–43.
- Holmes, E. A., A. Arntz, and M. R. Smucker (2007). Imagery rescripting in cognitive behavior therapy: Images, treatment techniques, and outcomes. *Journal of Behavior Therapy and Experimental Psychiatry* 38(4), 297–305.
- Holmes, E. A., T. J. Lang, and D. M. Shah (2009). Developing interpretation bias modification as a “cognitive vaccine” for depressed mood: imagining positive events makes you feel better than thinking about them verbally. *Journal of abnormal psychology* 118(1), 76.
- Iyadurai, L., R. M. Visser, A. Lau-Zhu, K. Porcheret, A. Horsch, E. A. Holmes, and E. L. James (2019). Intrusive memories of trauma: A target for research bridging cognitive science and its clinical application. *Clinical psychology review* 69, 67–82.
- John, A. and K. Orkin (2022). Can simple psychological interventions increase preventive health investment? *Journal of the European Economic Association* 20(3), 1001–1047.
- Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology* 71, 107–138.
- Karlan, D., A. L. Ratan, and J. Zinman (2014). Savings by and for the poor: A research review and agenda. *Review of Income and Wealth* 60(1), 36–78.
- Kessler, R. C., S. Aguilar-Gaxiola, J. Alonso, C. Benjet, E. J. Bromet, G. Cardoso, L. Degenhardt, G. de Girolamo, R. V. Dinolova, F. Ferry, S. Florescu, O. Gureje, J. M. Haro, Y. Huang, E. G. Karam, N. Kawakami, S. Lee, J.-P. Lepine, D. Levinson, F. Navarro-Mateu, B.-E. Pennell, M. Piazza, J. Posada-Villa, K. M. Scott, D. J. Stein, M. Ten Have, Y. Torres, M. C. Viana, M. V. Petukhova, N. A. Sampson, A. M. Zaslavsky, and K. C. Koenen (2017). Trauma and PTSD in the WHO World Mental Health Surveys. *European Journal of Psychotraumatology* 8(sup5), 1353383.
- Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. Normand, E. E. Walters, and A. M. Zaslavsky (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological medicine* 32(6), 959–976.
- Kessler, R. C., P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry* 62(6), 593–602.
- Kleim, B., B. Graham, S. Fihosy, R. Stott, and A. Ehlers (2014). Reduced specificity in episodic future thinking in posttraumatic stress disorder. *Clinical Psychological Science* 2(2), 165–173.
- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica* 75(1), 83–119.
- Koenen, K. C., A. Ratanatharathorn, L. Ng, K. A. McLaughlin, E. J. Bromet, D. J. Stein, E. G. Karam, A. Meron Ruscio, C. Benjet, K. Scott, and et al. (2017). Posttraumatic stress disorder in the World Mental Health Surveys. *Psychological Medicine* 47(13), 2260–2274.
- Lauro, J. G. (2024). The effect of episodic specificity inductions on cognitive tasks involving episodic retrieval: A quantitative review. *Neuropsychologia*, 108884.
- Lund, C., K. Orkin, M. Witte, et al. (2024). The effects of mental health interventions on labor market outcomes in Low- and Middle-Income Countries. *NBER Working Paper No. 32423*.

- Madan, C. and E. Kensinger (2021). Exploring the generalisation of affect across related experiences: A study of affective bleed and memory precision.
- Malmendier, U. and S. Nagel (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking? *Quarterly Journal of Economics* 126(1), 373–416.
- Malmendier, U. and J. A. Wachter (2021). Memory of Past Experiences and Economic Decisions. In M. J. Kahana and A. D. Wagner (Eds.), *The Oxford Handbook of Human Memory*, Volume 2. Oxford University Press.
- McKelway, M. (2025). Women’s Self-Efficacy and Economic Outcomes: Experimental Evidence from India. Available at SSRN 5158816.
- 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.
- McKenzie, D., A. Mohpal, and D. Yang (2022). Aspirations and financial decisions: Experimental evidence from the Philippines. *Journal of Development Economics* 156, 102846.
- McKenzie, D. and C. Woodruff (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer* 29(1), 48–82.
- McNally, R. J., N. B. Lasko, M. L. Macklin, and R. K. Pitman (1995). Autobiographical memory disturbance in combat-related posttraumatic stress disorder. *Behaviour research and therapy* 33(6), 619–630.
- Moher, D., S. Hopewell, K. F. Schulz, V. Montori, P. C. Gøtzsche, P. J. Devereaux, D. Elbourne, M. Egger, and D. G. Altman (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 340.
- Nelis, S., E. A. Holmes, J. W. Griffith, and F. Raes (2019). Mental imagery during daily life: Psychometric evaluation of the Spontaneous Use of Imagery Scale (SUIS). *Psychologica Belgica* 54(1), 19.
- Oettingen, G. (2012). Future thought and behaviour change. *European Review of Social Psychology* 23(1), 1–63.
- Orkin, K., R. Garlick, M. Mahmud, R. Sedlmayr, J. Haushofer, and S. Dercon (2023). Aspiring to a better future: can a simple psychological intervention reduce poverty? Technical report, National Bureau of Economic Research.
- Patel, V., B. Weobong, H. A. Weiss, A. Anand, B. Bhat, B. Katti, S. Dimidjian, R. Araya, S. D. Hollon, M. King, et al. (2017). The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. *The Lancet* 389(10065), 176–185.
- Pearson, J., C. W. G. Clifford, and F. Tong (2008). The functional impact of mental imagery on conscious perception. *Current Biology* 18(13), 982–986.
- Peters, J. and C. Büchel (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-medioprefrontal interactions. *Neuron* 66(1), 138–148.
- Pham, L. B. and S. E. Taylor (1999). From thought to action: Effects of process-versus outcome-based mental simulations on performance. *Personality and Social Psychology Bulletin* 25(2), 250–260.
- Pile, V., G. Williamson, A. Saunders, E. A. Holmes, and J. Y. F. Lau (2021). Harnessing emotional mental imagery to reduce anxiety and depression in young people: an integrative review of progress and promise. *The Lancet Psychiatry* 8(9), 836–852.

- Ridley, M., G. Rao, F. Schilbach, and V. Patel (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science* 370(6522).
- Rojas Valdes, R. I., B. Wydick, and T. J. Lybbert (2022). Can hope elevate microfinance? Evidence from Oaxaca, Mexico. *Oxford Economic Papers* 74(1), 236–264.
- Rubin, D. C., D. Berntsen, and M. K. Bohni (2008). A memory-based model of posttraumatic stress disorder: evaluating basic assumptions underlying the PTSD diagnosis. *Psychological review* 115(4), 985.
- Schacter, D. L., D. R. Addis, and R. L. Buckner (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience* 8(9), 657–661.
- Schacter, D. L., D. R. Addis, and R. L. Buckner (2008). Episodic simulation of future events: concepts, data, and applications. *Annals of the New York Academy of Sciences* 1124(1), 39–60.
- Scheier, M. F., C. S. Carver, and M. W. Bridges (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of personality and social psychology* 67(6), 1063.
- Schein, J., C. Houle, A. Urganus, M. Cloutier, O. Patterson-Lomba, Y. Wang, S. King, W. Levinson, A. Guérin, P. Lefebvre, and L. L. Davis (2021). Prevalence of post-traumatic stress disorder in the United States: a systematic literature review. *Current Medical Research and Opinion* 37(12), 2151–2161.
- Schwarzer, R. and M. Jerusalem (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, and M. Johnson (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs*, pp. 35–37. Windsor, UK: NFER-NELSON.
- Sinclair, V. G. and K. A. Wallston (2004). The development and psychometric evaluation of the Brief Resilient Coping Scale. *Assessment* 11(1), 94–101.
- Smith, B. W., J. Dalen, K. Wiggins, E. Tooley, P. Christopher, and J. Bernard (2008). The brief resilience scale: assessing the ability to bounce back. *International journal of behavioral medicine* 15, 194–200.
- Stöber, J. (2000). Prospective cognitions in anxiety and depression: Replication and methodological extension. *Cognition & Emotion* 14(5), 725–729.
- Weathers, F. W. et al. (2013). The PTSD Checklist for DSM-5 (PCL-5). *National Center for PTSD*.
- Weiss, D. S. and C. R. Marmar (1997). The Impact of Event Scale-Revised. In J. P. Wilson and T. M. Keane (Eds.), *Assessing psychological trauma and PTSD*, pp. 399–411. The Guilford Press.
- Williams, S. E., J. H. Ford, and E. A. Kensinger (2022). The power of negative and positive episodic memories. *Cognitive, Affective, & Behavioral Neuroscience* 22(5), 869–903.

Figures

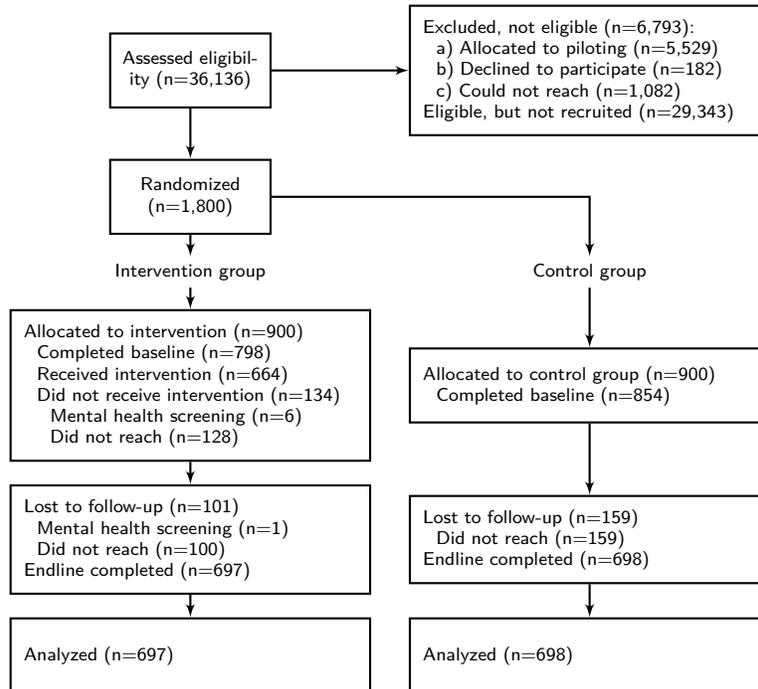


Figure 1. Ethiopia Trial Flow Diagram

Notes: This flowchart follows the Consolidated Standards of Reporting Trials (CONSORT) recommendations (Moher et al., 2010). Received intervention refers to participants who began the first session of the GME intervention. Eligibility was assessed on the universe of all registered refugees that satisfied the inclusion criteria (refugees aged 18 to 50 with at least grade 7 education). Mental health screening was conducted at every point of contact with respondents. Six participants were excluded from the treatment group at baseline and referred to further evaluation. One participant was excluded from the treatment group at the point of the first intervention and referred for further evaluation.

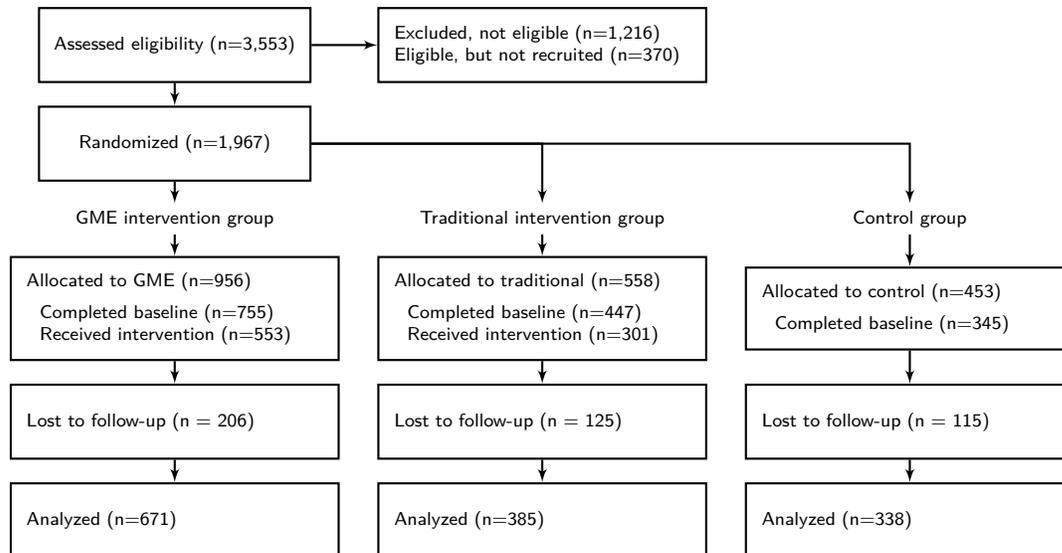


Figure 2. Colombia Trial Flow Diagram

Notes: This flowchart follows the Consolidated Standards of Reporting Trials (CONSORT) recommendations (Moher et al., 2010). To be eligible, participants needed to demonstrate entrepreneurship potential by reporting having a business or plans to launch a business in the following three months. We further defined entrepreneurship potential to encompass those who could describe their business or business idea in a few words and classify them by sector. In order to account for the need to self-fund transportation costs to the training centers, we limited eligibility to those applicants who reported either non-zero income or business sales in the past six months. Eligible applicants were also required to be literate, over the age of 18 years and to provide three points of contact. “Received intervention” reports the number of people who attended at least one session. “Lost to follow-up” reports the number of people who did not participate in both the midline and endline surveys. The sample for the analysis includes only people who accepted to participate in the training.

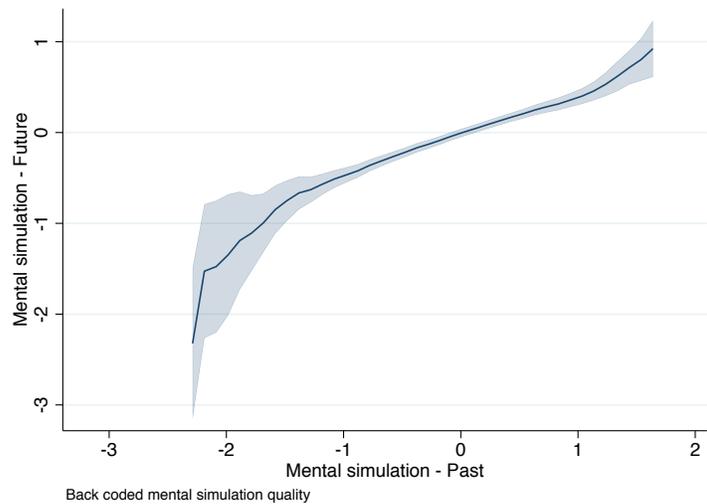


Figure 3. Ethiopia: Correlation between Quality of Past and Future Mental Simulation

Notes: The figure shows the correlation between the quality of mental simulation in past and future scenarios, from the respondents of the baseline survey in Ethiopia. The quality of mental simulation includes coded scores on specificity, emotionality and frequency of positive scenarios (computed among neutral scenarios).

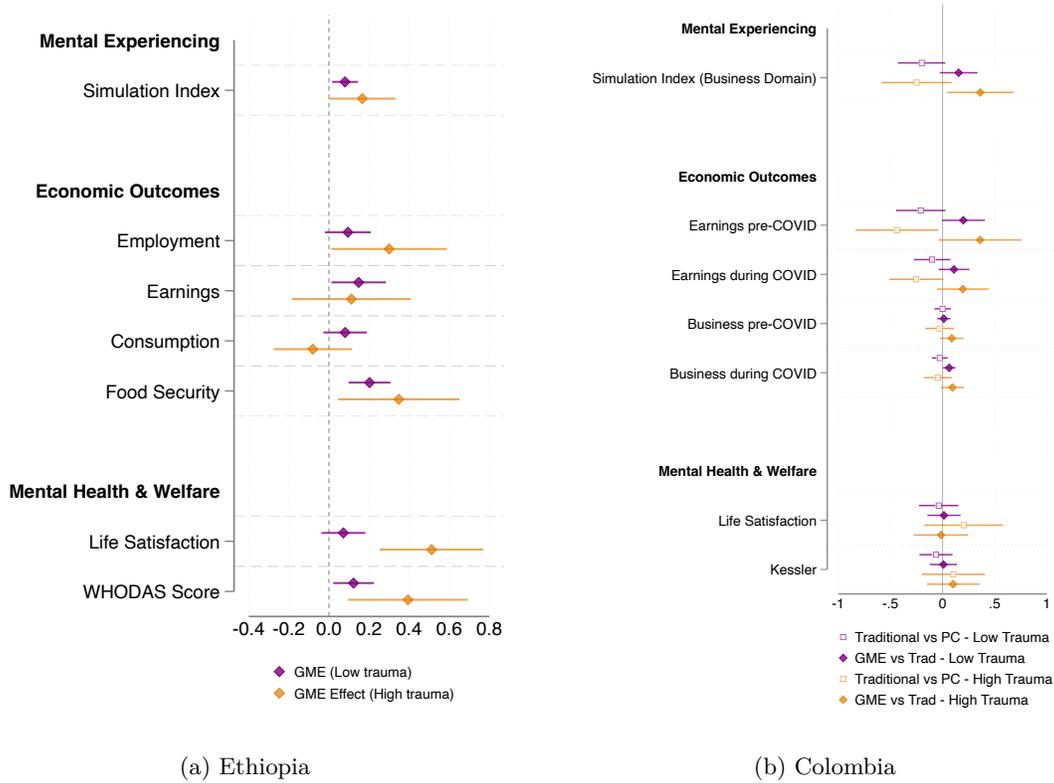
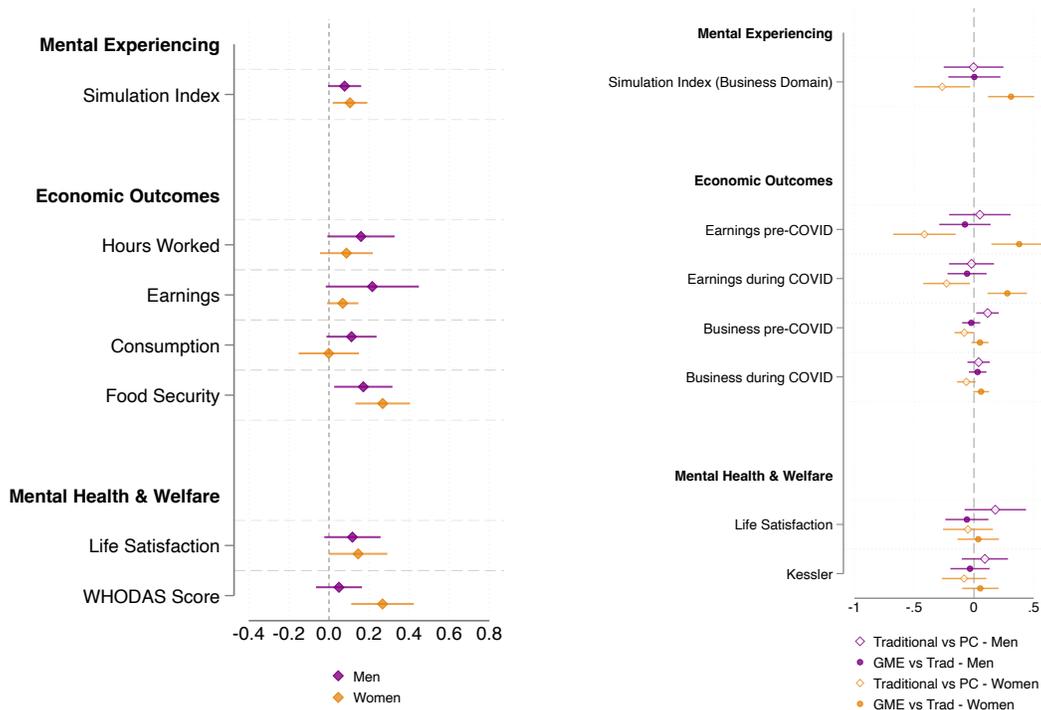


Figure 4. Heterogeneous Treatment Effects on Main Outcomes, by Trauma

Notes: The figure shows heterogeneous treatment effects on main outcomes by baseline trauma. For the Ethiopia results in panel (a), the figure plots coefficient estimates from specification (1) augmented with an interaction term for baseline trauma. “High Trauma” is defined as a baseline score of more than 31 on the PCL-5, a 20-item self-report measure that assesses 20 symptoms of PTSD as defined in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Outcomes and regression specifications are defined as in Panel A of Tables 2 and 3. The coefficients shown in the figure are the treatment effects in the low (in purple) and high (in orange) trauma samples. In Colombia (panel b), we categorize respondents as “High Trauma” if they score 33 or higher on the Impact of Event Score (IES-R) at baseline. Outcomes and regression specifications are defined as in Panel B of Tables 2 and 3. The figure in panel (b) reports four different coefficients from two regressions. One regression compares the traditional training with the control (empty squares), and the other the GME training against the traditional one (full diamonds). Each coefficient shows the treatment effect within the sample of people with high (in orange) or low (in purple) trauma. 90 percent confidence intervals are shown for all coefficient estimates. Both high trauma dummies correspond to our pre-registered definitions. Tables C5 and C7 report the corresponding regression results.



(a) Ethiopia

(b) Colombia

Figure 5. Heterogeneous Treatment Effects on Main Outcomes, by Gender

Notes: The figure shows heterogeneous treatment effects on main outcomes by gender. For the Ethiopia results in panel (a), the figure plots coefficient estimates from specification (1) augmented with an interaction term for gender. Outcomes and regression specifications are defined as in Panel A of Tables 2 and 3. The coefficients shown in the figure are the treatment effects in the male (in purple) and female (in orange) samples. In Colombia (panel b), outcomes and regression specifications are defined as in Panel B of Tables 2 and 3. The figure in panel (b) report four different coefficients from two regressions. One regression compares the traditional training with the control (empty diamonds), and the other the GME training against the traditional one (full circles). Each coefficient shows the treatment effect within the sample of women (in orange) or men (in purple). 90 percent confidence intervals shown for all coefficient estimates. Tables C6 and C8 report the corresponding regression results.

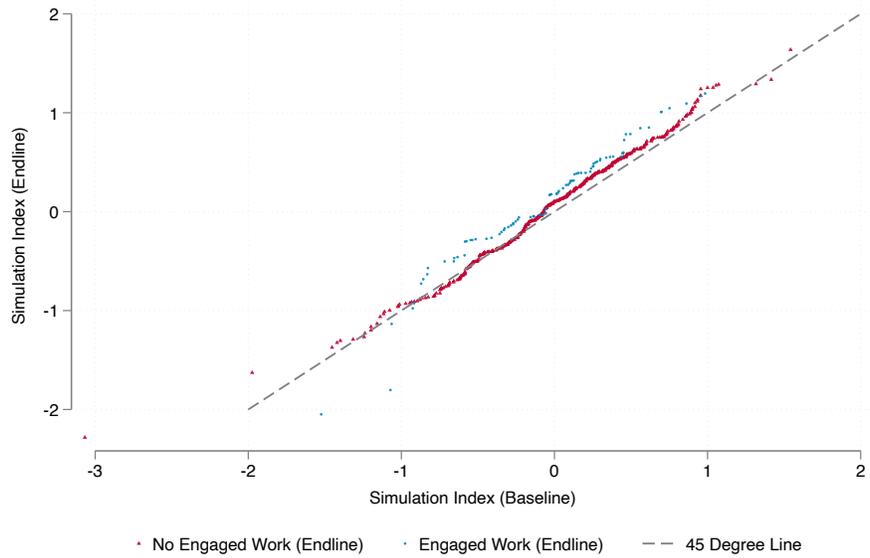


Figure 6. Ethiopia: Quantile-Quantile Plot: Overall Imagery Index and Work Engagement

Notes: The figure compares the Quantile-Quantile plots of the overall simulation index at baseline and endline between treatment group individuals who were not working at baseline but were working at endline, and treatment group individuals who were not working in both baseline and endline.

Tables

Table 1. Economic Vulnerability and Trauma in Our Samples

	Experimental Samples		Comparison Samples	
	(1)	(2)	(3)	(4)
	Ethiopia	Colombia	Ethiopia [†]	Colombia [‡]
<i>Panel A: Socioeconomic characteristics and economic activity</i>				
Age	28.6	32.0	31.5	33.5
Household size	1.8	3.5	2.10	3.1
Engaged in productive activity (fraction of sample) ^(a)	0.19	1.00	0.59	0.55
Wage if employed (US\$ per month) ^(b)	98.19	n/a	189.10	n/a
Earnings below minimum wage (fraction of sample)	n/a	0.67	n/a	0.51
	Ethiopia	Colombia	World General Population	
<i>Panel B: Post traumatic stress disorder^(c)</i>				
At risk of PTSD (fraction of sample above threshold)	0.13	0.22	0.04 [¶]	
<i>Panel C: Exposure to trauma (lived or witnessed)</i>				
Any trauma (fraction of sample)	0.98	0.83	0.61** to 0.70*	
<i>Number of traumas</i>				
Mean	5	2.8		
Median	4	2		
1 trauma (fraction of sample)	0.10	0.24	0.27**	
2 traumas (fraction of sample)	0.12	0.20	0.15**	
3 traumas (fraction of sample)	0.13	0.17	0.10**	
4 or more traumas (fraction of sample)	0.65	0.26	0.10**	
<i>Panel D: Exposure to specific types of trauma</i>				
Assault, incl. sexual, weapon (fraction of sample)	0.71	0.61	0.23*	
War, torture, death, displacement (fraction of sample)	0.77	0.48	0.13*	
Life-threatening illness and accidents (fraction of sample)	0.92	0.51	0.34*	
<i>Panel E: Worst trauma lived or witnessed</i>				
Less than 3 years ago (fraction of sample)	0.44	0.32		
Respondent's or someone's life was in danger (fraction of sample)	0.78	0.81		

Notes: (a) For Ethiopia, economic activity refers to any wage employment or self-employment in the last 7 days. In our Colombia sample, eligibility was limited to participants who had non-zero income or business sales in the past six months, so the entire sample is economically active by definition. For the Colombian general population, we use the national estimate of the (formal and informal) employment to population ratio for people that are at least 15 years old. (b) Conditional mean for respondents who reported a wage income, in US\$ per month, converted using period average exchange rates from the World Development Indicator database. (c) In Ethiopia, we use the PCL-5 post-traumatic stress disorder (PTSD) checklist, a 20-item self-report measure of PTSD symptoms. In Colombia, we use the Impact of Event Score Revised (IES-R) scale, a 22-item self-report measure of PTSD symptoms.

Sources: † Ethiopian National Labor Force and Migration Survey 2021. We restrict the sample using the same criteria as our experimental sample (residence in Addis Ababa, 18 to 50 years old, with at least junior high school education completed). Wage income from the 2021 survey is inflated to 2022 using Ethiopia's Consumer Price Inflation from the World Development Indicator database to match our experimental sample. ‡ Based on the 2021 data from the Departamento Administrativo Nacional de Estadística (DANE) Gran Encuesta Integrada de Hogares (GEIH) for all Colombia. ¶ Koenen et al. (2017)

* Kessler et al. (2017) ** Kessler et al. (2005) § Schein et al. (2021)

Table 2. GME Treatment Improves Mental Simulation Quality

Panel A: Ethiopia						
	(1)		(2)	(3)		(4)
	Simulation Index			Sub-indices		
			Specificity	Emotionality		Frequency of Positives
GME Treatment	0.091*** (0.030)		0.030 (0.044) [0.197]	0.132*** (0.048) [0.010]		0.113*** (0.039) [0.010]
Mean DV in Control	0.00		0.00	-0.01		0.01
N in Control	665		665	665		664
N in Treatment	667		667	667		665

Panel B: Colombia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Simulation Index, by Domain			Sub-indices (Business Scenarios)		
	All Scenarios	Business Scenarios	Non-business Scenarios	Specificity	Emotionality	Frequency of Use
<i>Panel B.1: Traditional vs Control</i>						
Traditional Training	-0.069 (0.075)	-0.157* (0.089)	-0.049 (0.074)	-0.129 (0.080) [0.249]	-0.122 (0.081) [0.249]	-0.073 (0.086) [0.249]
<i>Panel B.2: GME vs Traditional</i>						
GME Training	0.026 (0.063)	0.178** (0.074)	-0.006 (0.061)	0.133* (0.068) [0.054]	0.167** (0.065) [0.034]	0.090 (0.072) [0.083]
<i>Panel B.3: GME vs Control</i>						
GME Training	0.014 (0.066)	0.046 (0.076)	0.004 (0.064)	0.037 (0.070) [1.000]	0.044 (0.069) [1.000]	0.023 (0.073) [1.000]
Mean DV in Control	0.00	0.00	0.00	0.00	0.00	0.00
Mean DV in Traditional	0.00	-0.10	0.00	-0.10	-0.10	0.00
N in Control	550	390	550	390	390	390
N in Traditional	656	456	656	454	454	456
N in GME	1140	839	1140	835	835	838

Notes. This table shows ITT effects of the treatments on the quality of mental simulation. Panel A shows results for the Ethiopia sample, where indices are based on treatment-blind coding done in the back-office, based on participants' recordings. Panel B shows results for the Colombia sample, where indices are based on participants' self-reports. Specificity refers to the clarity of the images the respondent is able to generate. Emotionality refers to the strength of the emotions these images entail. Frequency of positives refers to the frequency of positive images appearing when respondents are presented with neutral scenarios. Frequency of use refers to the use of mental images in everyday life. Differences in the number of observations across columns in Panel B are due to the business scenarios being asked only to respondents with a business. Further details on the outcomes are reported in the main text (Section 4), Appendix D.5 for Ethiopia and E.5 for Colombia. All the regressions control for randomization strata and, for Colombia, survey wave fixed effects. Parentheses report robust standard errors for Ethiopia, and clustered standard errors (at the household level) for Colombia. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 3. GME Treatment Effects on Economic Outcomes, Mental Health and Welfare

Panel A: Ethiopia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Economic Outcomes			Mental Health & Welfare		
	Employment	Earnings	Consumption	Food Security	Life Satisfaction	WHODAS Score
GME Treatment	0.124** (0.054) [0.019]	0.143** (0.063) [0.019]	0.056 (0.050) [0.046]	0.219*** (0.051) [0.001]	0.131** (0.052) [0.015]	0.158*** (0.049) [0.004]
Mean DV in Control	-0.00	0.00	0.00	0.00	-0.00	-0.00
N in Control	698	698	698	698	698	698
N in Treatment	697	697	697	697	697	697
Panel B: Colombia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Economic Outcomes			Mental Health & Welfare		
	Earnings (Pre-COVID19)	Earnings (COVID19)	Operating Business (Pre-COVID19)	Operating Business (COVID19)	Life Satisfaction	Kessler
<i>Panel B.1: Traditional vs Control</i>						
Traditional Training	-0.222** (0.096) [0.140]	-0.144** (0.071) [0.140]	0.001 (0.032) [1.000]	-0.022 (0.031) [1.000]	0.040 (0.083) [1.000]	-0.012 (0.070) [1.000]
<i>Panel B.2: GME vs Traditional</i>						
GME Training	0.191** (0.083) [0.070]	0.141** (0.060) [0.070]	0.020 (0.027) [0.518]	0.048* (0.025) [0.084]	-0.003 (0.065) [0.931]	0.017 (0.057) [0.843]
<i>Panel B.3: GME vs Control</i>						
GME Training	-0.007 (0.066) [1.000]	0.040 (0.057) [1.000]	0.031 (0.029) [1.000]	0.033 (0.028) [1.000]	0.040 (0.077) [1.000]	0.011 (0.064) [1.000]
Mean DV in Control	0.00	0.00	0.70	0.70	0.00	0.00
Mean DV in Traditional	-0.20	-0.10	0.70	0.60	0.10	0.00
N in Control	323	539	333	552	248	545
N in Traditional	380	642	392	659	304	647
N in GME	665	1115	679	1145	521	1133

Notes. This table shows ITT effects of the treatments on economic outcomes (Columns 1 to 4 for Panel A and Panel B) and mental health and welfare (Columns 5 and 6 for Panel A and Panel B). For Ethiopia (Panel A), earnings are computed as weekly income from wage and self-employment. Employment is a standardized dummy equal to 1 if respondents had been engaged in any type of work in the past week. Consumption is an index of total expenditures, excluding those linked with outgoing and ongoing transfers such as remittances and other transfers. Food security is an index defined as in the Social Economic Survey of Ethiopia and includes reverse-coded questions about, e.g., not having enough food or the number of days with less preferred food. Life satisfaction is based on the Cantril Self-Anchoring Striving Scale, a self-reported wellbeing score. The WHODAS score refers to the World Health Organization Disability Assessment Schedule, which measures self-reported health levels (and it is reverse-coded so that higher levels indicate better health). For Colombia (Panel B), earnings (Columns 1 and 2) are defined as the inverse-hyperbolic sine of earnings (including business sales and income from other activities). Operating business (Columns 3 and 4) is an indicator for having an operating business. Kessler is the standardized score of the Kessler K6 Distress scale (reverse coded so that higher values indicate higher wellbeing). Life satisfaction is based on the Cantril Self-Anchoring Striving Scale, and was only asked in the second follow-up survey. Further details on the outcomes are reported in Appendix D.5 for Ethiopia and E.5 for Colombia. All the regressions control for randomization strata and, for Colombia, survey wave fixed effects. Parentheses report robust standard errors for Ethiopia, and clustered standard errors (at the household level) for Colombia. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 4. Characterizing the Causal Chain: Beliefs and Actions

Panel A: Ethiopia							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Beliefs				Actions		
	Intent to Stay	Probability of Work Permit	Probability of Finding Work	Time to Find Work	Reservation Wage	Job Search	Hours Worked
GME Treatment	0.178*** (0.053) [0.002]	0.192*** (0.054) [0.002]	0.204*** (0.054) [0.002]	-0.092* (0.051) [0.032]	0.030 (0.054) [0.090]	-0.097** (0.042) [0.018]	0.145** (0.056) [0.011]
Mean DV in Control	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00
N in Control	698	698	698	698	698	698	698
N in Treatment	697	697	697	697	697	697	697
Panel B: Colombia							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Business Actions						
	New Financing: Any	New Financing: Amount	Introduced New Product(s)	Investment: Any Business Asset	Investment: Marketing	Worked in Business	Employees: Any
<i>Panel B.1: Traditional vs Control</i>							
Traditional Training	-0.064* (0.037) [0.392]	-0.147* (0.076) [0.392]	0.031 (0.032) [0.686]	0.009 (0.032) [1.000]	-0.003 (0.038) [1.000]	-0.033 (0.033) [0.686]	-0.020 (0.031) [1.000]
<i>Panel B.2: GME vs Traditional</i>							
GME Training	0.061** (0.030) [0.117]	0.131** (0.062) [0.117]	0.017 (0.028) [0.409]	0.015 (0.027) [0.409]	0.044 (0.032) [0.199]	0.053** (0.026) [0.117]	0.025 (0.027) [0.378]
<i>Panel B.3: GME vs Control</i>							
GME Training	0.015 (0.034) [0.723]	0.024 (0.071) [0.723]	0.062** (0.029) [0.269]	0.039 (0.029) [0.424]	0.061* (0.035) [0.325]	0.020 (0.030) [0.714]	0.028 (0.028) [0.662]
Mean DV in Control	0.30	0.00	0.20	0.50	0.30	0.50	0.60
Mean DV in Traditional	0.30	-0.10	0.30	0.50	0.20	0.50	0.50
N in Control	327	327	379	555	251	549	498
N in Traditional	382	382	456	658	306	651	561
N in GME	669	669	807	1143	526	1133	1018

Notes. This table shows ITT effects of the treatments on beliefs (Panel A, Columns 1 to 5) and actions in both Ethiopia (Panel A, Columns 6 and 7) and Colombia (Panel B, Columns 1 to 7). For Ethiopia, intent to stay is the standardized average of the intended duration of stays in Ethiopia across various scenarios that vary the likelihood of obtaining a work permit. Probability of work permit is the expected probability of receiving a formal residence permit that allows for working and starting a business in the next three years. Probability of finding work is the expected probability of finding some work in the next three years conditional on not being given a formal work permit and staying in Ethiopia. Time to find work is the expected number of weeks needed to find casual work that pays about 1,300 birr per month. Reservation wage is respondents' reservation wage conditional on not being given a formal work permit. Job search is an index aggregating the number of hours spent, the number of calls made, the number of channels used, and the money spent in searching for any job. Hours worked is the total number of hours spent working in the past week. For Colombia, New financing: any is an indicator for receiving any new funding for the business. New financing: amount is the amount of funding received for the business. Introduced new product(s) is an indicator of whether the respondent introduced a new product or service. The two investment outcomes in Columns 4 and 5 are indicators for having invested in any business asset (e.g., machines, tools, software) or marketing, respectively. Worked in business is an indicator of whether the respondent spent any hours working in their business in the last week. Employees: any is an indicator for employing any worker in the business. Variables in Columns 1, 2 and 5 of Panel B were asked in only one follow-up survey. Further details on the outcomes are reported in Appendix D.5 for Ethiopia and E.5 for Colombia. All the regressions control for randomization strata and, for Colombia, survey wave fixed effects. Parentheses report robust standard errors for Ethiopia, and clustered standard errors (at the household level) for Colombia. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Online Appendix

Learning to See the World’s Opportunities: Memory, Mental Experiencing, and the Economic Lives of the Vulnerable

Nava Ashraf, Gharad Bryan, Alexia Delfino, Emily A. Holmes,
Leonardo Iacovone, Christian Johannes Meyer, Ashley Pople

Table of Contents

A	A Model of Trauma and Decision-Making	2
A.1	The Cognitive Impacts of Trauma	3
A.2	Economic Implications of Trauma	5
A.3	Overcoming Trauma: GME	6
B	Mental Simulation Measures: Correlation with Economic Outcomes and Validation	7
C	Additional Results	10

A A Model of Trauma and Decision-Making

A decision maker (DM) considers which actions $a \in A$ to take. To fix ideas, we can think of the set A as containing actions that could help the DM get a job. The benefit of an action is $u(a)$, the cost of an action is c , and she wishes to take all actions for which $u(a) > c$. Ex-ante she has no information to distinguish between actions in A and we assume a randomly chosen action has a negative expected utility $\frac{1}{\#A} \sum_{a \in A} u(a) < c$, so without more information she will not take any action, and will receive utility 0.²⁹

Given this, our DM values information. Denote $A' = \{a \in A : u(a) > c\}$ to be the set of actions that are worth taking. The expected value of information that reveals whether a randomly chosen action is in A' is

$$\hat{V}(A) = \frac{1}{\#A'} \left(\sum_{a \in A'} u(a) - c \right).$$

Our DM can generate information about actions by *recalling* memories, and then using the recalled memories to *simulate* the likely utility of taking the action. Memories are recalled from a set E , with each memory $e \in E$ encoding details of an event, such as the utility experienced ($u(e)$), environmental factors, and actions taken. The probability of recall depends on the similarity between memory and action. We follow [Bordalo et al. \(2024\)](#) and assume that the probability of recalling memory e when considering action a is given by

$$r(e, a; E) = \frac{s(e, a)}{\sum_{e' \in E} s(e', a)}, \quad (3)$$

where $s(e, a)$ is a symmetric function that measures the similarity between e and a . As highlighted by [Bordalo et al. \(2024\)](#) a key implication of this recall function is *interference*. Suppose the set E contains both relevant and irrelevant memories, where an irrelevant memory is one that does not provide useful information about action a . As the set of irrelevant memories increase the denominator of (3) increases, and so the probability that a relevant memory is recalled decreases.

Having drawn memory e , our decision maker then uses it to simulate the outcome of taking action a , with simulated expected utility given by

$$U^s(a, E) = \sum_{e \in E} r(e, a; E) u(e).$$

We denote this $U^s(a, E)$ (simulated utility) to distinguish it from true expected utility $U(a)$.

When our DM simulates she pre-experiences the outcome, i.e. experiences part of the utility from the drawn event $\kappa u(e)$, where $\kappa \in (0, 1)$ captures the idea that simulation is not as vivid as reality, but is an experience. Thus the expected utility that comes from pre-experiencing a is

$$U^p(a, E) = \kappa U^s(a, E).$$

This expected utility depends on action a because, although she may lack direct information about the utility of a itself, she knows which life events a resembles and can predict which memories will be recalled. For some actions the DM may expect a negative utility from the pre-experiencing.

The process of simulation for an action $a \in A$ generates information $V(A)$ that we assume is some fraction of the value of perfect information $\hat{V}(A)$, and also a per-experiencing utility $U^p(a, E)$. Our DM decides to

²⁹We think of this as concentrating on the set of risky but potentially worthwhile actions.

recall and simulate action a if

$$V(A) + U^p(a, E) > 0.$$

The set A can be divided into a subset that the DM will choose to simulate,

$$S(A, E) = \{a \in A : V(A) + U^p(a, E) > 0\} \quad (4)$$

and its complement $A/S(A, E)$ that the DM chooses not to simulate. We refer to $S(A, E)$ as the simulation set. After simulating an action in $S(A, E)$, she takes the action if $u(e) > c$ for the drawn memory and, because a randomly chosen action has a negative expected utility, she does not take any action that she has not simulated.³⁰ The simulation process then defines a set of actions that the DM will take, which depends on E

$$\mathcal{A}(A, E) = \{a \in S(A, E) : U^s(a, E) > c\}. \quad (5)$$

A.1 The Cognitive Impacts of Trauma

To understand the impact of trauma memories on cognition it is useful to divide the set E into three subsets: $E^R(A)$, relevant undistorted memories; E^T , trauma memories; and $\bar{E}^R(A)$, relevant but trauma distorted memories.

In the simplest case, $E^R(A)$ consists of memories of previous attempts to get a job, and records the true utility of those events. We assume that $E^R(A)$ is unbiased, in the sense that if only $E^R(A)$ is used for recall and simulation, then simulated expected utility is an unbiased estimator of true utility

$$\mathbb{E}U^s(a, E^R(A)) = U(a).$$

On average, people do not make mistakes if they rely on undistorted and relevant memories.

The set E^T is the set of trauma memories. These memories have two characteristics. First, they are unlikely to be directly relevant, e.g., trauma memories are rarely directly relevant to looking for a job. In the language of [Bordalo et al. \(2024\)](#) they are non-domain specific. Second, they are – by definition – *very* low u . The first characteristic means that trauma memories can potentially interfere with the recall of relevant memories. Denote e^R to be any relevant memory, then from (3)

$$r(e^R(a), a; E^R(A) \cup E^T) < r(e^R(a), a; E^R(A)).$$

When our DM goes looking for useful information about how she will fare finding a job, her mind may well recall a trauma memory which provides no useful information.

The second characteristic implies that when trauma memories do interfere, they create distorted thinking and unpleasant pre-experiencing. On average

$$U^s(a, E^R(A) \cup E^T) < \hat{U}^s(a, E^R(A)) \quad \& \quad U^p(a, E^R(A)) < U^p(a, E^R(A) \cup E^T).$$

Because the trauma memories are likely have very low similarity with an economic choice a these effects are likely to be small on average, but to be quite devastating when they do occur. If the recalled event was very low utility, pre-experienced utility $\kappa u(e)$ will be very low. We think of this as a flashback: in attempting to think about the future, the trauma survivor re-experiences a trauma memory, which leads to strong negative feelings.

If this were the only impact of trauma it might not have strong economic consequences. However, as discussed above, this is not where the impact of trauma stops, because the negative emotions from trauma

³⁰It would not change anything if we allowed the DM to draw multiple memories.

can spread through the memory system, creating a new set of distorted memories: $\tilde{E}^R(A)$. Every time a memory is recalled and an action simulated a new memory is created. This is a memory of the re-experiencing and is a composite of the old memory and the context at the time of recall, which includes the decision maker’s current emotions. We can capture this by assuming that the memory e' , of recalling the memory e , has utility

$$u(e') = \alpha u(e) + (1 - \alpha)\bar{u},$$

where \bar{u} is the current emotional state. This new memory is appended to the existing set so $E' = E \cup \{e'\}$. If emotional state is subject to noise that is uncorrelated with recall, then $\mathbb{E}(\bar{u}) = 0$, and this formulation will predict a slow forgetting of emotional value with the new memories having - on average - utility closer to zero.

This creation of new memories allows the impacts of trauma to spread across domains. A key aspect of trauma is that it negatively impacts emotional state for an extended time after the event. Let the level of utility during this period be $\tilde{u} \ll 0$. For any memory e recalled while \tilde{u} is the emotional state, a new negatively distorted memory \tilde{e} will be created with

$$u(\tilde{e}) = \alpha u(e) + (1 - \alpha)\tilde{u} \ll u(e).$$

This process continues over time: the initial trauma creates negative emotions, which update existing memories, which leads to the recall of more negative memories, which in turn negatively impacts emotion, which updates additional memories, etc. As discussed above, this leads the impact of a sufficiently negative trauma to spread throughout the memory system. It is as if the negativity of the trauma invades the past, and hence the pre-experiencing of the future. We model this as creating an additional set of memories $\tilde{E}^R(A)$ which are relevant to the action A , for example because they are all about looking for a job, but have distorted utility $u(\tilde{e}) \ll u(e)$.

The creation of these negatively distorted domain relevant memories then leads to two additional cognitive changes beyond flashbacks.

1. Avoidance: those with many traumatic experiences avoid simulating the future

$$S(A, E^R(A) \cup E^T \cup \tilde{E}^R(A)) \subset S(A, E^R(A)),$$

where $S(A, \cdot)$ is defined in (4) as the set of actions the DM chooses to simulate.

2. Distortion: those with many traumatic experiences have a negatively distorted view of themselves and the future

$$U^s(a, E^R(A) \cup E^T \cup \tilde{E}^R(A)) \ll U(a).$$

The combined impact of avoidance and distortion is that the size of the action set is reduced

$$\mathcal{A}(A, E^R(A) \cup E^T \cup \tilde{E}^R(A)) \subset \mathcal{A}(A, E^R(A)),$$

where $\tilde{\mathcal{A}}(A, \cdot)$ is defined in 5. Trauma leads our DM to appear unmotivated and inactive, because she underestimates how many actions are worthwhile – she cannot see the world’s opportunities.

Overall, this model, a very simple extension of [Bordalo et al. \(2024\)](#), gives the three key diagnostic criteria for PTSD: flashbacks, distortion and avoidance. A simple way to conceptualize the model is that a trauma survivor wants to think about the future, but her mind spontaneously turns to recalling either trauma memories, or trauma distorted memories, which interfere with her ability to recall relevant and accurate information. If these recollections are painful enough, she may stop thinking about the future (avoidance), and if she does continue to think about the future, her simulations will be overwhelmingly negative (distortion).

A.2 Economic Implications of Trauma

The model has strong economic implications. Most directly, trauma survivors will be unmotivated and inactive, even outside of the domain of the trauma, and will not be able to see the world's opportunities. Trauma has a direct negative impact on the economic welfare of trauma survivors.

Beyond this direct effect, our model of trauma also has implications for traditional anti-poverty programs. Many of these programs are designed to remove the poor's economic constraints, under the argument that the poor are experts in their own lives, and just need opportunities. Microfinance, cash transfers and land titling have all been justified in this way. Our model of trauma suggests that, in populations where large numbers of people have experienced trauma, these programs will not be as effective: while economic constraints exist, cognitive constraints also limit the ability to fully leverage the opportunities these programs provide.

A second type of anti-poverty program focuses on education rather than direct opportunity (e.g., teaching a man to fish). These programs guide individuals on actions to take and encourage them to consider the opportunities those actions might create. For example, a business training program might teach someone about a set of actions B , one of which might be the need to approach a bank for credit. In our model this amounts to a) giving the individual an experience in which the utility of going to the bank $u(b)$ is greater than c and b) encouraging the decision maker to pre-experience going to the bank.

To see the impact of these two forces, suppose that prior to the training program our DM had not simulated any of the business actions B , and had a set of relevant experiences $E^R(B)$ that were undistorted, but that she also had a large stock of negatively distorted experiences and trauma memories \tilde{E} . The program will have two impacts on her set of memories. First, is a *teaching effect*: the information from the program that $u(b) > c$ creates a new relevant memory e^1 that is added to the memory bank. Second, is a *distortion effect*: the DM's simulation of the outcome of going to the bank also creates an experience e^2 that enters the memory bank. This memory will, on average, be negatively distorted by the potential interference of the memories in \tilde{E} . These two effects go in opposite directions with the updated set of relevant memories being $E^R(B) \cup \{e^1\} \cup \{e^2\}$.

The overall impact of the teaching and distortion effects is an empirical question, but with a population that is highly traumatized it is possible that the program will backfire. Let B be the set of actions discussed in the program, then if the distortion effect dominates the teaching effect

1. Induced avoidance:

$$S(B, E^R(B) \cup \{e^1\} \cup \{e^2\}) \subset S(B, E^R(B)).$$

so the DM may stop simulating advocated actions.

2. Induced distortion:

$$U^s(b, E^R(B) \cup \{e^1\} \cup \{e^2\}) < U^s(b, E^R(B)) \quad \forall b \in B,$$

so that the DM considers advocated actions to be low expected utility if she does simulate.

Overall, these two effects will combine to *reduce* the likelihood that the DM take actions that are within the domain of the program's teaching

$$\mathcal{A}(B, E^R(B) \cup \{e^1\} \cup \{e^2\}) \subset \mathcal{A}(B, E^R(B))$$

Induced avoidance also has an interesting testable implication. Traumatized populations who participate in a training program, and for whom the distortion effect dominates the teaching effect should be less likely to simulate actions associated with the program. This can be seen as a test of the hypothesis that some experiences can interfere with the process of simulation.

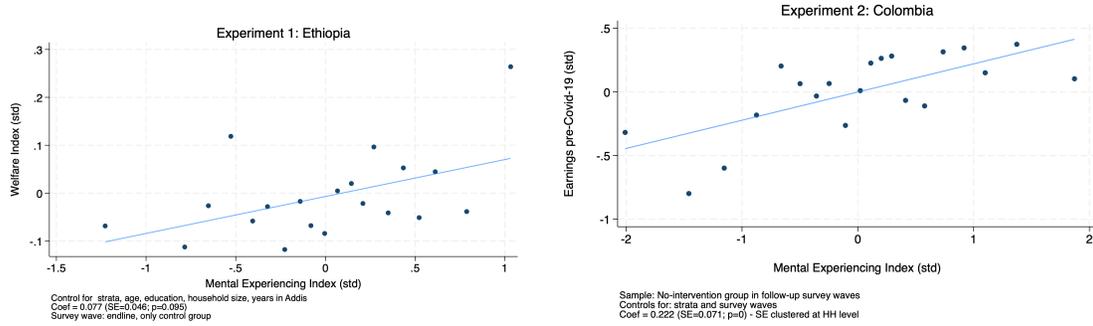
A.3 Overcoming Trauma: GME

The goal of GME is to make a memory pathway from a current state to a desired state that does not cross through the set of distorted memories \tilde{E} . This is done by guiding a DM to simulate an action b that is useful (for example going to the bank as above) but very carefully ensure that a realistic picture is created of the utility associated with taking that action. To model this, suppose that the desired action is b . GME involves carefully creating a set of memories or imagined events $E^R(b)$ such that $U^s(b, E^R(b)) = U(b)$. The implication being that after the program

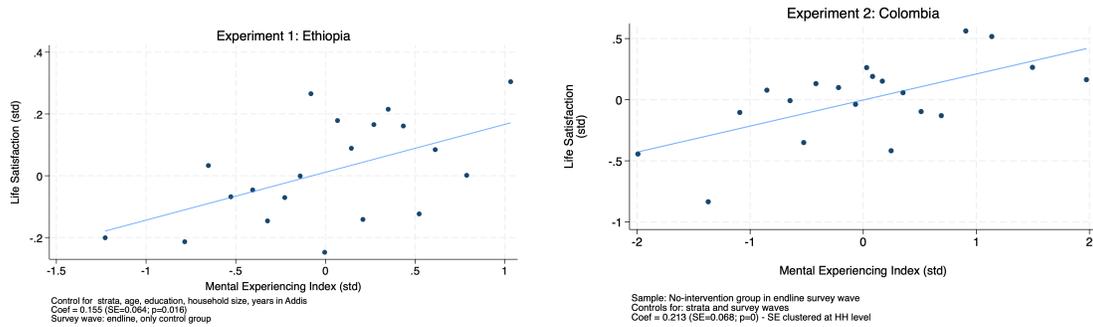
$$|U^s(b, \tilde{E} \cup E^R(b)) - U(b)| < |U^s(b, \tilde{E}) - U(b)|$$

so that the DM's simulations are more accurate.

B Mental Simulation Measures: Correlation with Economic Outcomes and Validation



(a) Mental Simulation and Economic Welfare



(b) Mental Simulation and Life Satisfaction

Figure B1. Mental Simulation Quality and Economic Outcomes

Notes: The figure shows binned scatter plots for the correlation between our mental simulation index and different measures of economic outcomes (labeled “mental experiencing” in the graphs). Both variables on the x-axis and y-axis are unconditional indexes in the post-intervention surveys, only for the control group. All the figures control for stratification variables. Within each panel, the figure on the LHS is for Ethiopia and the figure on the RHS is for Colombia. For Ethiopia, mental simulation measures are back-coded, and include specificity, emotionality and positivity. For Colombia, mental simulation measures are self-reported, including specificity, emotionality and frequency of use. In Panel (a), the welfare index in Ethiopia includes unconditional weekly wage, total expenditure, food security, frequency of savings. In Colombia, it is the earnings index for the pre-COVID19 period. In Panel (b), life satisfaction is the standardized response to the Cantril ladder question (on a scale from 1 to 10). For Colombia, this question was asked only in the second follow-up survey.

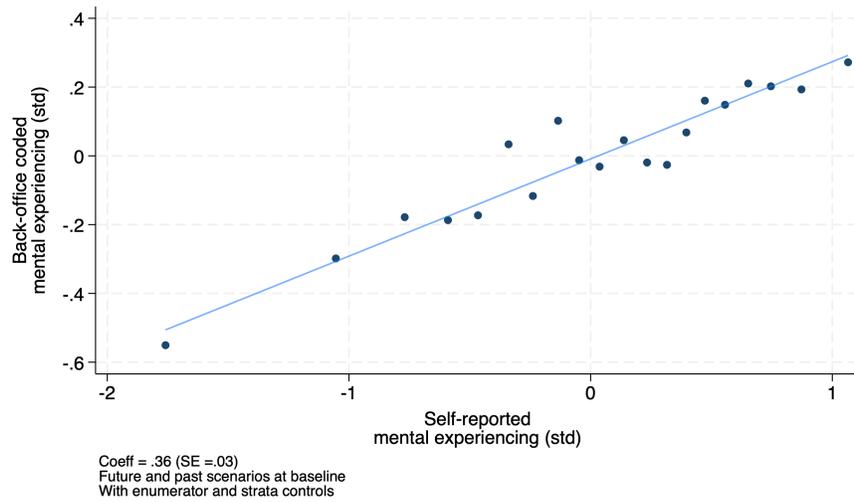


Figure B2. Mental Simulation Quality: Correlation between Back-Coded and Self-Reported Measures

Notes: The figure shows a binned scatter plot for the correlation between back-coded and self-reported measures of mental simulation quality from the Ethiopia experiment (labeled “mental experiencing” in the graphs). Both indexes average standardized indexes of specificity, emotionality and frequency of positive scenarios (out of the neutral ones proposed). For self-reported measures, the latter frequency was asked to the enumerator conducting the survey with the respondent. The figure controls for strata and enumerator fixed effects.

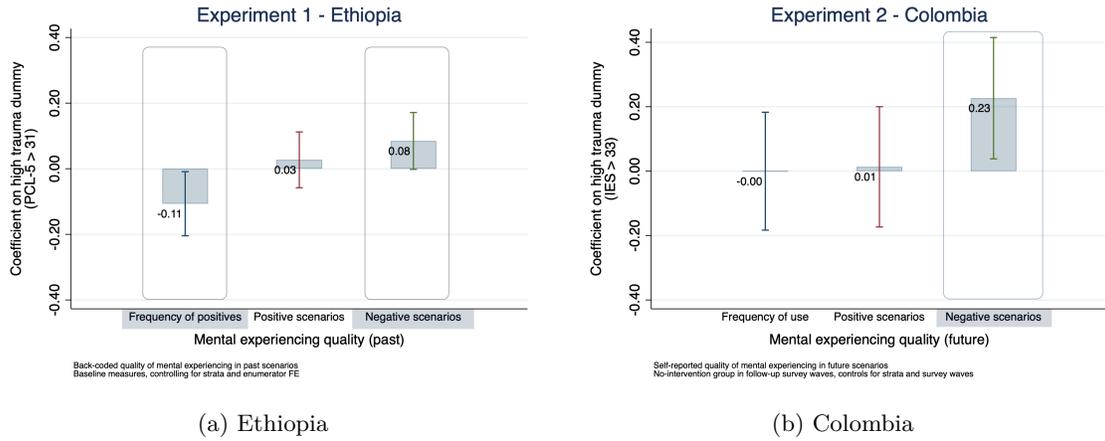
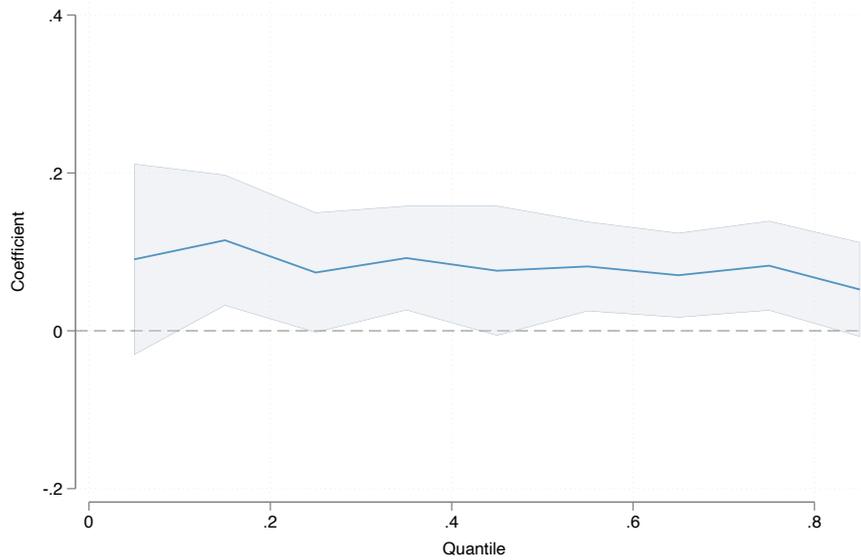


Figure B3. Mental Simulation Quality and Trauma

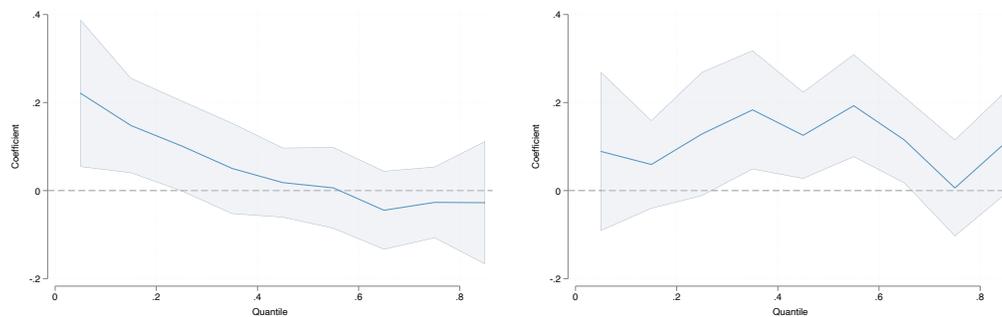
Notes: The figure reports coefficients on an indicator variable for “High trauma” symptoms for different measures of mental simulation quality (labeled “mental experiencing” in the graphs): frequency, quality of positive scenarios (including specificity and emotionality) and quality of negative scenarios (including specificity and emotionality). In panel (a), the first bar on the left indicates the frequency with which respondents come up with positive images when prompted with neutral scenarios. All the measures in panel (a) are back-coded by treatment-blind research assistants, are oriented to the past and come from the baseline survey for the full sample. Controls include strata and enumerator fixed effects. In panel (b), the first bar on the left indicates the frequency with which respondents use mental simulation in their everyday life. All the measures in panel (b) are self-reported, oriented to the future and come from follow-up surveys only for the control group. Controls for strata are included.

C Additional Results

Figure C1. Ethiopia: Quantile Treatment Effects on Quality of Mental Simulation



(a) Simulation Index



(b) Specificity and Emotionality Sub-Indices

Notes: The figure shows quantile treatment effects on the index and sub-indices of quality of mental simulation in the Ethiopia experiment. The indices are defined in the note to Table 2 of the paper. The outcomes are based on office-coded measures. 90 percent confidence intervals are shown.

Table C1. Ethiopia: PCA Index for Mental Simulation Quality

	(1)	(2)	(3) Sub-indices	(4)
	Simulation Index	Specificity	Emotionality	Frequency of Positives
<i>Panel A: Main Specification</i>				
GME Treatment	0.154** (0.062)	0.045 (0.061) [0.180]	0.133** (0.059) [0.032]	0.162** (0.063) [0.032]
<i>Panel B: PDS Lasso</i>				
GME Treatment	0.147** (0.061)	0.034 (0.060) [0.235]	0.127** (0.058) [0.046]	0.142** (0.062) [0.046]
Control Mean	-0.08	-0.02	-0.07	-0.08
Observations	1332	1332	1332	1329

Notes: This table shows the impact of the GME treatment on the quality of mental simulation, including the specificity, emotionality and frequency of positive images. Panel A shows ITT effects controlling for randomization strata (gender and prime treatment), while Panel B adds controls chosen using PDS Lasso. The table reproduces Table 2 of the main paper, but the outcome measures here are built from a Principal Component Analysis (PCA) that includes both respondents' answers and back-coded values. Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C2. GME Treatment Effect on Mental Simulation Quality with PDS Lasso Controls

Panel A: Ethiopia						
	(1)		(2)	(3)	(4)	
	Simulation Index		Sub-indices			
			Specificity	Emotionality	Frequency of Positives	
GME Treatment	0.092*** (0.030)		0.026 (0.044) [0.227]	0.132*** (0.048) [0.010]	0.112*** (0.038) [0.010]	
Mean DV in Control	0.00		0.00	-0.01	0.01	
N in Control	665		665	665	664	
N in Treatment	667		667	667	665	

Panel B: Colombia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Simulation Index, by Domain			Sub-indices (Business Scenarios)		
	All Scenarios	Business Scenarios	Non-business Scenarios	Specificity	Emotionality	Frequency of Use
<i>Panel B.1: Traditional vs Control</i>						
Traditional Training	-0.063 (0.072)	-0.157* (0.088)	-0.045 (0.070)	-0.130 (0.079) [0.238]	-0.122 (0.080) [0.238]	-0.071 (0.084) [0.238]
<i>Panel B.2: GME vs Traditional</i>						
GME Training	0.023 (0.060)	0.160** (0.072)	-0.009 (0.058)	0.114* (0.066) [0.093]	0.167** (0.065) [0.031]	0.066 (0.068) [0.145]
<i>Panel B.3: GME vs Control</i>						
GME Training	0.001 (0.061)	0.041 (0.073)	-0.007 (0.061)	0.027 (0.068) [1.000]	0.047 (0.067) [1.000]	0.021 (0.071) [1.000]
Mean DV in Control	0.00	0.00	0.00	0.00	0.00	0.00
Mean DV in Traditional	0.00	-0.10	0.00	-0.10	-0.10	0.00
N in Control	550	390	550	390	390	390
N in Traditional	656	456	656	454	454	456
N in GME	1140	839	1140	835	835	838

Notes. This table reproduces Table 2 from the main paper with the inclusion of controls chosen using PDS Lasso. Outcomes are defined in the note to Table 2. Further details on the outcomes are reported in the main text (Section 4), Appendix D.5 for Ethiopia and E.5 for Colombia. All the regressions control for randomization strata and, for Colombia, survey wave fixed effects. Parentheses report robust standard errors for Ethiopia, and clustered standard errors (at the household level) for Colombia. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C3. GME Treatment Effects on Economic Outcomes, Mental Health and Welfare with PDS Lasso Controls

Panel A: Ethiopia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Economic Outcomes			Mental Health & Welfare		
	Employment	Earnings	Consumption	Food Security	Life Satisfaction	WHODAS Score
GME Treatment	0.080 (0.049) [0.065]	0.121* (0.063) [0.042]	0.056 (0.043) [0.088]	0.217*** (0.048) [0.001]	0.134*** (0.050) [0.011]	0.165*** (0.047) [0.002]
Mean DV in Control	-0.00	0.00	0.00	0.00	-0.00	-0.00
N in Control	698	698	698	698	698	698
N in Treatment	697	697	697	697	697	697
Panel B: Colombia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Economic Outcomes			Mental Health & Welfare		
	Earnings (Pre-COVID19)	Earnings (COVID19)	Operating Business (Pre-COVID19)	Operating Business (COVID19)	Life Satisfaction	Kessler
<i>Panel B.1: Traditional vs Control</i>						
Traditional Training	-0.222** (0.096) [0.193]	-0.144* (0.071) [0.193]	0.001 (0.032) [1.000]	-0.022 (0.031) [1.000]	0.040 (0.083) [1.000]	-0.012 (0.070) [1.000]
<i>Panel B.2: GME vs Traditional</i>						
Traditional Training	0.191** (0.083) [0.100]	0.141** (0.060) [0.100]	0.020 (0.027) [0.987]	0.048 (0.025) [0.212]	-0.003 (0.065) [0.987]	0.017 (0.057) [0.987]
<i>Panel B.3: GME vs Control</i>						
Traditional Training	-0.007 (0.066) [1.000]	0.040 (0.057) [1.000]	0.031 (0.029) [1.000]	0.033 (0.028) [1.000]	0.040 (0.077) [1.000]	0.011 (0.064) [1.000]
Mean DV in Control	0.00	0.00	0.70	0.70	0.00	0.00
Mean DV in Traditional	-0.20	-0.10	0.70	0.60	0.10	0.00
N in Control	323	539	333	552	248	545
N in Traditional	380	642	392	659	304	647
N in GME	665	1115	679	1145	521	1133

Notes. This table reproduces Table 3 from the main paper with the inclusion of controls chosen using PDS Lasso. Outcomes are defined in the note to Table 3. Further details on the outcomes are reported in the main text (Section 4), Appendix D.5 for Ethiopia and E.5 for Colombia. All the regressions control for randomization strata and, for Colombia, survey wave fixed effects. Parentheses report robust standard errors for Ethiopia, and clustered standard errors (at the household level) for Colombia. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C4. Characterizing the Causal Chain: Beliefs and Actions with PDS Lasso Controls

Panel A: Ethiopia							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Beliefs				Actions		
	Intent to Stay	Probability of Work Permit	Probability of Finding Work	Time to Find Work	Reservation Wage	Job Search	Hours Worked
GME Treatment	0.171*** (0.053) [0.003]	0.189*** (0.054) [0.003]	0.186*** (0.054) [0.003]	-0.092* (0.049) [0.033]	0.009 (0.051) [0.140]	-0.099** (0.042) [0.019]	0.098** (0.050) [0.033]
Mean DV in Control	-0.00	-0.00	-0.00	-0.00	0.00	-0.00	-0.00
N in Control	698	698	698	698	698	698	698
N in Treatment	697	697	697	697	697	697	697
Panel B: Colombia							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Business Actions						
	New Financing: Any	New Financing: Amount	Introduced New Product(s)	Investment: Any Business Asset	Investment: Marketing	Worked in Business	Employees: Any
<i>Panel B.1: Traditional vs Control</i>							
Traditional Training	-0.065* (0.036) [0.347]	-0.149** (0.075) [0.347]	0.046 (0.030) [0.347]	0.017 (0.031) [0.638]	-0.003 (0.037) [1.000]	-0.017 (0.031) [0.638]	-0.024 (0.030) [0.638]
<i>Panel B.2: GME vs Traditional</i>							
GME Training	0.061** (0.030) [0.168]	0.131** (0.062) [0.168]	-0.001 (0.026) [0.982]	-0.002 (0.025) [0.982]	0.033 (0.031) [0.395]	0.041* (0.025) [0.194]	0.017 (0.026) [0.695]
<i>Panel B.3: GME vs Control</i>							
GME Training	0.015 (0.033) [0.678]	0.024 (0.070) [0.678]	0.055** (0.028) [0.476]	0.034 (0.028) [0.496]	0.058* (0.034) [0.476]	0.017 (0.029) [0.678]	0.033 (0.027) [0.496]
Mean DV in Control	0.30	0.00	0.20	0.50	0.30	0.50	0.60
Mean DV in Traditional	0.30	-0.10	0.30	0.50	0.20	0.50	0.50
N in Control	327	327	379	555	251	549	498
N in Traditional	382	382	456	658	306	651	561
N in GME	669	669	807	1143	526	1133	1018

Notes. This table reproduces Table 4 from the main paper with the inclusion of controls chosen using PDS Lasso. Outcomes are defined in the note to Table 4. Further details on the outcomes are reported in the main text (Section 4), Appendix D.5 for Ethiopia and E.5 for Colombia. All the regressions control for randomization strata and, for Colombia, survey wave fixed effects. Parentheses report robust standard errors for Ethiopia, and clustered standard errors (at the household level) for Colombia. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C5. Ethiopia: Main Outcomes by Baseline Trauma

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Overall Index	Mental Experiencing	Economic Outcomes			Mental Health & Welfare		
		Simulation Index	Employment	Earnings	Consumption	Food Security	Life Satisfaction	WHODAS Score
GME Treatment=1 × High Trauma=0	0.125*** (0.028)	0.080** (0.033) [0.063]	0.094 (0.059) [0.057]	0.148** (0.069) [0.217]	0.080 (0.055) [0.210]	0.202*** (0.053) [0.043]	0.072 (0.056) [0.001]	0.122** (0.052) [0.001]
GME Treatment=1 × High Trauma=1	0.271*** (0.066)	0.166* (0.085) [0.030]	0.300** (0.147) [0.091]	0.112 (0.151) [0.049]	-0.080 (0.100) [0.104]	0.348** (0.154) [0.001]	0.512*** (0.132) [0.125]	0.394*** (0.152) [0.001]
High Trauma	-0.237*** (0.053)	-0.116* (0.067) [0.082]	-0.032 (0.106) [0.296]	0.015 (0.119) [0.296]	-0.125 (0.091) [0.115]	-0.439*** (0.123) [0.001]	-0.443*** (0.109) [0.001]	-0.373*** (0.129) [.]
p-value: GME Treatment=1 × High Trauma=0 = GME Treatment=1 × High Trauma=1	0.04	0.34	0.19	0.83	0.16	0.37	0.00	0.09
Control Mean	-0	0	-0	0	0	0	-0	-0
N in Control	698	665	698	698	698	698	698	698
N in Treatment	697	667	697	697	697	697	697	697

Notes. This table shows heterogeneous treatment effects by baseline trauma. "High Trauma" is our pre-registered dummy which takes value one if the respondent scores above the threshold of 31 in the PCL-5 scale for trauma symptoms, which is suggestive of PTSD (asked at baseline). We run an interacted model: $y_{it} = \alpha + \beta_1 \cdot GME_t + \beta_2 \cdot HighTrauma_t + \beta_3 \cdot GME_t \cdot HighTrauma_t + \delta_i + \epsilon_{it}$. The coefficients reported in the Table are β_1 ("GME Treatment=1 × High Trauma=0"), $\beta_1 + \beta_3$ ("GME Treatment=1 × High Trauma=1") and β_2 ("High Trauma"). At the bottom of the table we include a test of whether β_1 equals $\beta_1 + \beta_3$. Outcomes are defined in the notes to Tables 2 and 3 of the paper. Further details on the outcomes are reported in Appendix D.5. All the regressions control for strata. Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C6. Ethiopia: Main Outcomes by Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Overall Index	Mental Experiencing	Economic Outcomes			Mental Health & Welfare		
		Simulation Index	Employment	Earnings	Consumption	Food Security	Life Satisfaction	WHODAS Score
GME Treatment=1 × Female=0	0.137*** (0.039)	0.078* (0.042) [0.034]	0.158* (0.086) [0.128]	0.215* (0.119) [0.074]	0.112* (0.064) [0.235]	0.171** (0.074) [0.001]	0.116 (0.072) [0.062]	0.047 (0.059) [0.002]
GME Treatment=1 × Female=1	0.151*** (0.034)	0.104** (0.044) [0.069]	0.088 (0.067) [0.069]	0.069* (0.040) [0.069]	-0.000 (0.077) [0.072]	0.268*** (0.069) [0.035]	0.147** (0.074) [0.086]	0.269*** (0.079) [0.153]
Female	-0.127*** (0.037)	-0.126*** (0.043) [0.008]	-0.364*** (0.074) [0.001]	-0.368*** (0.074) [0.001]	0.269*** (0.075) [0.002]	0.050 (0.076) [0.161]	0.119 (0.076) [0.091]	-0.430*** (0.074) [0.001]
p-value: GME Treatment=1 × Female=0 = GME Treatment=1 × Female=1	0.78	0.68	0.52	0.24	0.26	0.34	0.77	0.03
Control Mean	-0	0	-0	0	0	0	-0	-0
N in Control	698	665	698	698	698	698	698	698
N in Treatment	697	667	697	697	697	697	697	697

Notes. This table shows heterogeneous treatment effects by gender. We run an interacted model: $y_{i,t} = \alpha + \beta_1 \cdot GME_i + \beta_2 \cdot Female_i + \beta_3 \cdot GME_i \cdot Female_i + \delta_s + \epsilon_{i,t}$. The coefficients reported in the Table are β_1 ("GME Treatment=1 × Female=0"), $\beta_1 + \beta_3$ ("GME Treatment=1 × Female=1") and β_2 ("Female"). At the bottom of the table we include a test of whether β_1 equals $\beta_1 + \beta_3$. Outcomes are defined in the notes to Tables 2 and 3 of the paper. Further details on the outcomes are reported in Appendix D.5. All the regressions control for strata. Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C7. Colombia: Main Outcomes by Baseline Trauma

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(7)
	Overall Index	Mental Experiencing	Economic Outcomes				Mental Health & Welfare	
		Simulation Index (Business Domain)	Earnings (Pre-COVID19)	Earnings (COVID19)	Operating Business (Pre-COVID19)	Operating Business (COVID19)	Life Satisfaction	Kessler
<i>Panel A: Traditional vs Control</i>								
Traditional Training=1 × High Trauma=0	-0.091* (0.052)	-0.200* (0.116) [0.646]	-0.209* (0.121) [0.443]	-0.100 (0.089) [0.476]	-0.000 (0.040) [1.000]	-0.027 (0.039) [1.000]	-0.037 (0.096) [1.000]	-0.063 (0.081) [1.000]
Traditional Training=1 × High Trauma=1	-0.084 (0.082)	-0.249 (0.171) [0.476]	-0.437** (0.202) [0.476]	-0.253* (0.133) [1.000]	-0.029 (0.070) [1.000]	-0.045 (0.068) [1.000]	0.202 (0.192) [1.000]	0.103 (0.154) [1.000]
High Trauma	-0.094 (0.069)	0.137 (0.141) [1.000]	0.129 (0.114) [1.000]	0.023 (0.106) [1.000]	0.022 (0.060) [1.000]	0.003 (0.059) [1.000]	-0.148 (0.181) [1.000]	-0.647*** (0.132) [0.001]
p-value: Trad.=1 × High trauma=0 = Trad.=1 × High trauma=1	0.937	0.809	0.325	0.335	0.719	0.816	0.266	0.336
<i>Panel B: GME vs Traditional</i>								
GME Training=1 × High Trauma=0	0.103** (0.042)	0.154* (0.092) [0.364]	0.199* (0.105) [0.364]	0.109 (0.075) [0.368]	0.012 (0.032) [0.368]	0.064** (0.031) [0.364]	0.013 (0.081) [0.769]	0.008 (0.066) [0.592]
GME Training=1 × High Trauma=1	0.182** (0.071)	0.361** (0.161) [0.364]	0.359* (0.202) [0.364]	0.194 (0.126) [0.385]	0.088 (0.058) [0.719]	0.095* (0.056) [0.364]	-0.015 (0.133) [0.769]	0.101 (0.128) [0.769]
High Trauma	-0.106 (0.068)	0.066 (0.152) [0.719]	-0.154 (0.199) [0.592]	-0.150 (0.121) [0.395]	-0.012 (0.053) [0.769]	-0.021 (0.052) [0.719]	0.075 (0.119) [0.699]	-0.504*** (0.116) [0.001]
p-value: GME=1 × High trauma=0 = GME=1 × High trauma=1	0.342	0.264	0.485	0.564	0.249	0.621	0.859	0.521
<i>Panel C: GME vs Control</i>								
GME Training=1 × High Trauma=0	0.033 (0.043)	0.005 (0.095) [1.000]	0.013 (0.084) [1.000]	0.027 (0.075) [1.000]	0.027 (0.036) [1.000]	0.047 (0.035) [1.000]	-0.040 (0.087) [1.000]	-0.057 (0.076) [1.000]
GME Training=1 × High Trauma=1	0.119* (0.071)	0.125 (0.143) [1.000]	-0.063 (0.116) [1.000]	0.006 (0.112) [1.000]	0.069 (0.062) [1.000]	0.046 (0.061) [1.000]	0.198 (0.188) [1.000]	0.213 (0.139) [1.000]
High Trauma	-0.124* (0.068)	0.157 (0.142) [1.000]	0.098 (0.102) [1.000]	-0.057 (0.106) [1.000]	0.022 (0.058) [1.000]	0.005 (0.059) [1.000]	-0.207 (0.180) [1.000]	-0.698*** (0.132) [0.001]
p-value: GME=1 × High trauma=0 = GME=1 × High trauma=1	0.292	0.482	0.599	0.871	0.553	0.988	0.248	0.086
Mean DV in Control	0.00	0.10	0.00	0.00	0.70	0.60	0.00	0.00
Mean DV in Traditional	-0.10	-0.10	-0.20	-0.10	0.70	0.60	0.00	0.00
N in Control	467	326	270	457	279	467	214	463
N in Traditional	583	397	335	569	344	581	270	573
N in GME	989	733	572	961	585	988	450	976

Notes. This table shows heterogeneous treatment effects by baseline trauma. "High Trauma" is our pre-registered dummy which takes value one if the respondent scores above the threshold of 33 in the IES-R scale for trauma symptoms, which is suggestive of PTSD (asked at baseline). We run an interacted model: $y_{i,t} = \alpha + \beta_1 \cdot \text{Trad}_i + \beta_2 \cdot \text{HighTrauma}_i + \beta_3 \cdot \text{Trad}_i \cdot \text{HighTrauma}_i + \delta_i + \delta_{i,t} + \epsilon_{i,t}$, where Trad_i is Traditional vs Control in Panel A, GME vs Traditional in Panel B and GME vs Control in Panel C. The coefficients reported in the Table within each panel are β_1 ("Treatment=1 × High Trauma=0"), $\beta_1 + \beta_3$ ("Treatment=1 × High Trauma=1") and β_2 ("High Trauma"). At the bottom of each panel we include a test of whether β_1 equals $\beta_1 + \beta_3$. Outcomes and specifications are defined in the notes to Tables 2 and 3 of the paper. Further details on the outcomes are reported in Appendix E.5. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C8. Colombia: Main Outcomes by Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(7)
	Overall Index	Mental Experiencing	Economic Outcomes				Mental Health & Welfare	
		Simulation Index (Business Domain)	Earnings (Pre-COVID19)	Earnings (COVID19)	Operating Business (Pre-COVID19)	Operating Business (COVID19)	Life Satisfaction	Kessler
<i>Panel A: Traditional vs Control</i>								
Traditional Training=1 × Female=0	0.092 (0.060)	-0.003 (0.127) [0.117]	0.049 (0.131) [0.042]	-0.021 (0.096) [0.117]	0.114** (0.048) [0.137]	0.039 (0.047) [0.203]	0.179 (0.131) [0.651]	0.091 (0.098) [0.470]
Traditional Training=1 × Female=1	-0.182*** (0.053)	-0.266** (0.119) [0.874]	-0.414*** (0.133) [0.713]	-0.229** (0.099) [0.756]	-0.081* (0.042) [0.117]	-0.064 (0.040) [0.470]	-0.051 (0.106) [0.281]	-0.082 (0.095) [0.470]
Female	-0.009 (0.057)	0.215* (0.124) [0.187]	-0.022 (0.123) [0.756]	-0.196** (0.095) [0.123]	0.071 (0.048) [0.241]	0.036 (0.046) [0.470]	0.039 (0.136) [0.754]	-0.249** (0.104) [0.117]
p-value: Trad.=1 × Female=0 = Trad.=1 × Female=1	0.001	0.125	0.013	0.130	0.002	0.094	0.170	0.199
<i>Panel B: GME vs Traditional</i>								
GME Training=1 × Female=0	-0.019 (0.048)	0.003 (0.111) [0.005]	-0.076 (0.109) [0.005]	-0.058 (0.083) [0.005]	-0.023 (0.039) [0.229]	0.031 (0.038) [0.136]	-0.059 (0.092) [0.573]	-0.033 (0.084) [0.483]
GME Training=1 × Female=1	0.183*** (0.046)	0.310*** (0.098) [0.869]	0.379*** (0.118) [0.483]	0.279*** (0.084) [0.483]	0.050 (0.036) [0.485]	0.059* (0.034) [0.483]	0.036 (0.088) [0.483]	0.053 (0.078) [0.573]
Female	-0.267*** (0.055)	-0.031 (0.121) [0.666]	-0.481*** (0.139) [0.004]	-0.402*** (0.098) [0.001]	-0.113*** (0.041) [0.015]	-0.055 (0.040) [0.229]	-0.193* (0.099) [0.100]	-0.384*** (0.088) [0.001]
p-value: GME=1 × Female=0 = GME=1 × Female=1	0.003	0.037	0.005	0.005	0.167	0.567	0.452	0.451
<i>Panel C: GME vs Control</i>								
GME Training=1 × Female=0	0.088 (0.054)	0.033 (0.115) [1.000]	0.002 (0.114) [1.000]	-0.035 (0.086) [0.993]	0.094** (0.045) [1.000]	0.069 (0.043) [1.000]	0.106 (0.125) [1.000]	0.040 (0.094) [1.000]
GME Training=1 × Female=1	0.024 (0.043)	0.055 (0.098) [1.000]	-0.015 (0.081) [1.000]	0.094 (0.076) [1.000]	-0.017 (0.036) [0.437]	0.006 (0.036) [0.559]	-0.006 (0.095) [1.000]	-0.011 (0.086) [1.000]
Female	-0.008 (0.057)	0.227* (0.123) [0.437]	-0.030 (0.122) [1.000]	-0.193** (0.095) [0.437]	0.066 (0.047) [0.795]	0.033 (0.046) [1.000]	0.018 (0.134) [1.000]	-0.246** (0.104) [0.437]
p-value: GME=1 × Female=0 = GME=1 × Female=1	0.350	0.887	0.907	0.260	0.051	0.250	0.470	0.687
Mean DV in Control	0.00	0.00	0.00	0.00	0.70	0.70	0.00	0.00
Mean DV in Traditional	-0.10	-0.10	-0.20	-0.10	0.70	0.60	0.10	0.00
N in Control	552	390	323	539	333	552	248	545
N in Traditional	661	456	380	642	392	659	304	647
N in GME	1147	839	665	1115	679	1145	521	1133

Notes. This table shows heterogeneous treatment effects by gender. We run an interacted model: $y_{i,t} = \alpha + \beta_1 \cdot \text{Trad}_i + \beta_2 \cdot \text{Female}_i + \beta_3 \cdot \text{Trad}_i \cdot \text{Female}_i + \delta_i + \epsilon_{i,t}$, where Trad_i is Traditional vs Control in Panel A, GME vs Traditional in Panel B and GME vs Control in Panel C. The coefficients reported in the Table within each panel are β_1 ("Treatment=1 × Female=0"), $\beta_1 + \beta_3$ ("Treatment=1 × Female=1") and β_2 ("Female"). At the bottom of each panel we include a test of whether β_1 equals $\beta_1 + \beta_3$. Outcomes and specifications are defined in the notes to Tables 2 and 3 of the paper. Further details on the outcomes are reported in Appendix E.5. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C9. Ethiopia: Job Search by Baseline Work Status

Baseline Sample:	(1)	(2)	(3)	(4)	(5)	(6)
	Job Search		Reservation Wage		Hours Worked	
	No Work	Work	No Work	Work	No Work	Work
GME Treatment	-0.045 (0.046) [0.358]	-0.299*** (0.108) [0.019]	0.075 (0.057) [0.295]	-0.185 (0.153) [0.295]	0.136*** (0.047) [0.019]	0.045 (0.185) [0.519]
Mean DV in Control	-0.01	0.05	-0.05	0.23	-0.22	0.97
N in Control	571	127	571	127	571	127
N in Treatment	549	148	549	148	549	148

Notes. This table shows ITT effects of the GME treatment on work-related outcome variables. “No work” is the sample of respondents who were not working at baseline, and “Work” is the sample of respondents who were working at baseline. Job search is an index aggregating the number of hours spent, the number of calls made, the number of channels used, and the money spent in searching for any job. Reservation wage is respondents’ reservation wage conditional on not being given a formal work permit. Hours worked is the total number of hours spent working in the past week. Further details on the outcomes are reported in the main text (section 4) and Appendix D.5. All the regressions control for randomization strata. Parentheses report robust standard errors. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C10. Ethiopia: Non-Standardized Labor Market Impacts

	Employment	Earnings	Wage Earnings	Business Earnings
	(1)	(2)	(3)	(4)
<i>Panel A: Main Specification</i>				
GME Treatment	0.05** (0.022) [0.049]	113.30** (49.850) [0.049]	87.65* (48.464) [0.049]	25.65* (13.286) [0.049]
<i>Panel B: PDS Lasso</i>				
GME Treatment	0.03 (0.020) [0.114]	95.90* (49.702) [0.114]	80.79* (47.474) [0.114]	23.63* (12.765) [0.114]
Control Mean	0.21	214.82	217.48	-2.66
Observations	1395	1395	1395	1395

Notes. This table shows ITT estimates of the GME treatment on employment, total earnings, wage earnings, and self-employment earnings over the past seven days. Employment (column 1) is a dummy variable that takes the value of one when a respondent has worked over the past seven days, in any activity. Earnings (column 2) is the sum of earnings from wage employment (column 3) and earnings from self-employment (column 4). Earnings from wage employment and from self-employment are each winsorized at the 99th percentile. All earnings values are unconditional in Ethiopian birr per week. The variables in Columns 1 and 2 correspond to the non-standardized versions of main economic outcomes of Table 3, Panel A, Columns 1 and 2. Panel A shows the results controlling for randomization strata. Panel B shows the result obtained adding controls using PDS Lasso. Robust standard errors in parentheses, sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C11. Ethiopia: Detailed Labor Market Outcomes

	Work Status (Extensive Margin)					Hours Worked (Intensive Margin)				
	Any	Wage	Own business	Casual	Unpaid	Total	Wage	Own business	Casual	Unpaid
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>Panel A: Main Specification</i>										
GME Treatment	0.124**	0.095*	0.085	0.061	-0.066	0.145**	0.111**	0.116*	0.061	-0.035
	(0.054)	(0.055)	(0.061)	(0.056)	(0.047)	(0.056)	(0.056)	(0.070)	(0.057)	(0.052)
	[0.117]	[0.180]	[0.198]	[0.236]	[0.198]	[0.112]	[0.152]	[0.180]	[0.236]	[0.308]
<i>Panel B: PDS Lasso</i>										
GME Treatment	0.080	0.060	0.084	0.040	-0.054	0.098**	0.071	0.097	0.058	-0.030
	(0.049)	(0.050)	(0.061)	(0.055)	(0.044)	(0.050)	(0.051)	(0.067)	(0.058)	(0.051)
	[0.495]	[0.495]	[0.495]	[0.495]	[0.495]	[0.495]	[0.495]	[0.495]	[0.495]	[0.495]
Control Mean	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00
Observations	1395	1395	1395	1395	1395	1395	1395	1395	1395	1395

Notes. This table shows ITT estimates of the GME treatment on work status (extensive margin) and hours worked (intensive margin). Work status is a standardized dummy variable that takes the value of one when a respondent has worked over the past seven days, in any activity (column 1) or specific activities (columns 2 through 5). Hours worked refers to the standardized, unconditional number of hours worked, from any activity (column 6) or specific activities (columns 7 through 10). Outcomes in Columns 1 and 6 correspond to those presented in the main Tables 3 and 4. Panel A shows the results controlling for randomization strata, and Panel B adding controls chosen using PDS Lasso. Robust standard errors in parentheses, sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C12. Additional Outcomes Related to Beliefs and Actions

Panel A: Ethiopia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Beliefs			Self-reported Preferences		
	Self-efficacy	Optimism	Aspirations	Risk Aversion	Trust	Patience
GME Treatment	0.027 (0.054) [0.260]	0.165*** (0.053) [0.012]	0.110** (0.053) [0.064]	0.058 (0.054) [0.130]	0.106** (0.054) [0.064]	0.107** (0.051) [0.064]
Mean DV in Control	0.00	-0.00	0.00	-0.00	-0.00	-0.00
N in Control	698	698	698	698	698	698
N in Treatment	697	697	697	697	697	697

Panel B: Colombia						
	(1)	(2)	(3)	(4)	(5)	(6)
	Beliefs		Actions (Non-business Domain)			
	Self-efficacy	Safety Net	COVID19 Mitigation	Worked for Wage	Savings: Any	Savings: Amount
<i>Panel B.1: Traditional vs Control</i>						
Traditional Training	-0.052 (0.072) [1.000]	-0.009 (0.069) [1.000]	-0.074 (0.067) [1.000]	0.025 (0.033) [1.000]	0.013 (0.029) [1.000]	-0.081 (0.055) [1.000]
<i>Panel B.2: GME vs Traditional</i>						
GME Training	-0.035 (0.063) [1.000]	0.001 (0.059) [1.000]	0.084 (0.058) [1.000]	-0.015 (0.027) [1.000]	0.031 (0.023) [1.000]	0.023 (0.049) [1.000]
<i>Panel B.3: GME vs Control</i>						
GME Training	-0.049 (0.066) [1.000]	-0.002 (0.063) [1.000]	-0.003 (0.063) [1.000]	0.013 (0.029) [1.000]	0.048* (0.025) [0.518]	-0.050 (0.068) [1.000]
Mean DV in Control	0.00	0.00	0.00	0.50	0.80	0.00
Mean DV in Traditional	0.00	0.00	-0.10	0.50	0.80	-0.10
N in Control	545	545	551	547	318	318
N in Traditional	649	649	657	647	384	384
N in GME	1134	1133	1142	1131	648	648

Notes. This table shows ITT estimates of the treatments on additional intermediate outcomes. For Ethiopia (Panel A), Self-efficacy is the score on the Generalized Self-Efficacy scale, a 10-item psychometric scale to assess self-beliefs (Schwarzer and Jerusalem, 1995). Optimism is the standardized value of the Revised Life Orientation Test, a 10-item psychometric scale to assess levels of optimism versus pessimism in life (Scheier et al., 1994). Aspiration measures respondents’ aspirations based on three questions: “Do you expect to work for pay in a non-family enterprise (including your own business) in the future?”, “Do you expect to work for pay in a family enterprise in the future?” and “Do you see yourself continuing your education in the future?”. Self-reported preferences are based on unincentivized survey questions adapted from the Global Preference Survey (Falk et al., 2018). For Colombia (Panel B), Self-efficacy is an index aggregating answers to three scales: the Brief Resilient Coping Scale (Sinclair and Wallston, 2004), and adapted subset of statements from the Brief Resilience Scale (Smith et al., 2008) and a Self-efficacy scale (Chen et al., 2001). Safety nets is a standardized index aggregating answers to three questions about perceptions of whether people felt they had enough savings for the first two months of the lockdown, had enough cash to cover expenses for the coming week, or could easily find 200,000 pesos within a month for an unexpected need. COVID19 mitigation is a standardized index that indicates whether participants adapted their decision-making during the COVID-19 crisis. It aggregates yes/no answers to a series of questions about pandemic-related adjustments in the business, such as following government COVID19 guidelines, adopting hygiene protocols, or contacting suppliers. Worked for wage is an indicator of whether the respondent works in a wage work. Savings: any is an indicator for any savings. Savings: amount is the respondent’s amount of savings (unconditional). Questions about savings were only asked in one follow-up survey and refer to the pre-COVID19 period. Further details on the outcomes are reported in the main text (Section 4), Appendix D.5 for Ethiopia, and E.5 for Colombia. All the regressions control for randomization strata and, for Colombia, survey wave fixed effects. Parentheses report robust standard errors for Ethiopia, and clustered standard errors (at the household level) for Colombia. FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C13. Ethiopia: Outcomes by Social Desirability Scale

	Panel A: Simulation Outcomes				Panel B: Economic Outcomes					
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)	(6)
	Sub-indices				Economic Outcomes, Mental Health and Welfare				Mental Health & Welfare	
	Simulation Index	Specificity	Emotionality	Frequency of Positives	Employment	Earnings	Consumption	Food Security	Life Satisfaction	WHODAS Score
GME Treatment	0.091*** (0.030)	0.029 (0.044) [0.701]	0.130*** (0.048) [0.030]	0.115*** (0.039) [0.027]	0.126** (0.054) [0.059]	0.141** (0.062) [0.059]	0.055 (0.050) [0.196]	0.218*** (0.051) [0.001]	0.135*** (0.051) [0.032]	0.157*** (0.049) [0.008]
Social Desirability	0.008 (0.020)	-0.001 (0.031) [0.766]	-0.014 (0.031) [0.701]	0.040 (0.026) [0.444]	0.054 (0.037) [0.144]	0.069 (0.044) [0.140]	-0.035 (0.034) [0.216]	-0.077** (0.037) [0.077]	0.125*** (0.036) [0.006]	0.066* (0.038) [0.111]
GME Treatment x Social Desirability	-0.015 (0.029)	-0.030 (0.042) [0.701]	-0.060 (0.046) [0.472]	0.044 (0.037) [0.472]	0.039 (0.052) [0.265]	-0.130* (0.072) [0.108]	0.005 (0.044) [0.439]	0.020 (0.047) [0.318]	0.021 (0.052) [0.318]	-0.091* (0.047) [0.092]
p-value: GME Treatment + GME Treatment x Social Desirability = 0	0.07	0.99	0.29	0.00	0.04	0.90	0.35	0.00	0.03	0.33
Control Mean	0	0	-0	0	-0	0	0	0	-0	-0
N in Control	665	665	665	664	698	698	698	698	698	698
N in Treatment	667	667	667	665	697	697	697	697	697	697

	Panel C: Beliefs and Actions						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Beliefs					Actions	
	Intent to Stay	Probability of Work Permit	Probability of Finding Work	Time to Find Work	Reservation Wage	Job Search	Hours Worked
GME Treatment	0.179*** (0.053) [0.006]	0.191*** (0.054) [0.005]	0.202*** (0.054) [0.005]	-0.088* (0.050) [0.106]	0.030 (0.054) [0.399]	-0.095** (0.042) [0.054]	0.147*** (0.056) [0.043]
Social Desirability	-0.059 (0.038) [0.137]	-0.093** (0.039) [0.049]	-0.076** (0.038) [0.073]	0.080** (0.033) [0.046]	-0.018 (0.042) [0.425]	0.079** (0.032) [0.046]	0.038 (0.037) [0.249]
GME Treatment x Social Desirability	0.098* (0.052) [0.082]	0.071 (0.052) [0.176]	0.004 (0.052) [0.546]	0.052 (0.044) [0.203]	-0.003 (0.056) [0.546]	-0.002 (0.040) [0.546]	0.036 (0.054) [0.393]
p-value: GME Treatment + GME Treatment x Social Desirability = 0	0.00	0.00	0.01	0.62	0.74	0.12	0.03
Control Mean	-0	-0	-0	-0	0	-0	-0
N in Control	698	698	698	698	698	698	698
N in Treatment	697	697	697	697	697	697	697

Notes. This table shows the impact of the GME treatment on all the main outcomes shown in Tables 2, 3 and 4 of the paper, interacted with a score for social desirability. Social desirability is based on the Marlowe-Crowne Social Desirability Scale (Crowne Marlowe, 1960), which measures the individual propensity for social desirability bias. All the outcomes are defined in the notes to Tables 2, 3 and 4. All regressions control for randomization strata (gender and prime treatment). Robust standard errors in parentheses, sharpened q-values in square brackets (computed within each panel). *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C14. Ethiopia: Outcomes by Rule Orientation Scale

	Panel A: Simulation Outcomes				Panel B: Economic Outcomes					
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(5)	(6)
	Sub-indices				Economic Outcomes				Mental Health & Welfare	
Simulation Index	Specificity	Emotionality	Frequency of Positives	Employment	Earnings	Consumption	Food Security	Life Satisfaction	WHODAS Score	
GME Treatment	0.086*** (0.030)	0.025 (0.044) [0.279]	0.125*** (0.048) [0.021]	0.111*** (0.039) [0.018]	0.121** (0.054) [0.044]	0.147** (0.064) [0.042]	0.060 (0.050) [0.230]	0.227*** (0.050) [0.001]	0.137*** (0.051) [0.037]	0.161*** (0.049) [0.009]
Rule Orientation	0.060*** (0.021)	0.069** (0.032) [0.053]	0.152*** (0.032) [0.001]	-0.042 (0.028) [0.138]	0.091** (0.039) [0.042]	0.022 (0.032) [0.336]	0.040 (0.029) [0.187]	-0.102*** (0.039) [0.037]	-0.055 (0.037) [0.161]	-0.002 (0.037) [0.507]
GME Treatment x Rule Orientation	0.002 (0.030)	0.004 (0.043) [0.449]	-0.052 (0.046) [0.222]	0.056 (0.039) [0.138]	-0.056 (0.054) [0.243]	-0.092* (0.055) [0.120]	-0.107** (0.042) [0.037]	-0.012 (0.051) [0.502]	-0.036 (0.050) [0.336]	-0.056 (0.048) [0.230]
p-value: GME Treatment + GME Treatment x Rule Orientation = 0	0.04	0.65	0.28	0.00	0.40	0.46	0.48	0.00	0.18	0.12
Control Mean	0	0	-0	0	-0	0	0	0	-0	-0
N in Control	665	665	665	664	698	698	698	698	698	698
N in Treatment	667	667	667	665	697	697	697	697	697	697

	Panel C: Beliefs and Actions						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Beliefs					Choices	
Intent to Stay	Probability of Work Permit	Probability of Finding Work	Time to Find Work	Reservation Wage	Job Search	Hours Worked	
GME Treatment	0.178*** (0.053) [0.005]	0.186*** (0.054) [0.005]	0.207*** (0.054) [0.004]	-0.094* (0.050) [0.118]	0.029 (0.054) [0.493]	-0.101** (0.042) [0.045]	0.145** (0.056) [0.036]
Rule Orientation	-0.018 (0.035) [0.493]	0.056 (0.040) [0.190]	-0.067* (0.037) [0.124]	-0.078* (0.041) [0.118]	0.043 (0.042) [0.300]	0.123*** (0.035) [0.005]	0.050 (0.039) [0.223]
GME Treatment x Rule Orientation	0.021 (0.048) [0.493]	0.034 (0.054) [0.488]	0.025 (0.051) [0.493]	0.109* (0.056) [0.118]	-0.022 (0.063) [0.525]	-0.051 (0.044) [0.253]	-0.056 (0.056) [0.300]
p-value: GME Treatment + GME Treatment x Rule Orientation = 0	0.01	0.00	0.00	0.83	0.93	0.02	0.26
Control Mean	-0	-0	-0	-0	0	-0	-0
N in Control	698	698	698	698	698	698	698
N in Treatment	697	697	697	697	697	697	697

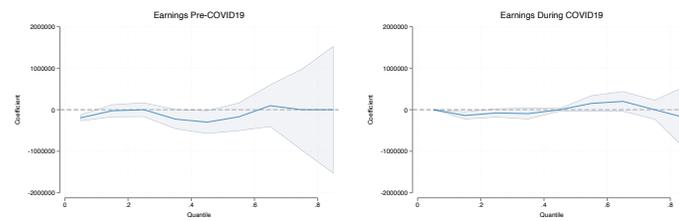
Notes. This table shows the impact of the GME treatment on all the main outcomes shown in Tables 2, 3 and 4 of the paper, interacted with the respondent's score on a "Rule Orientation" scale, which captures the extent to which they care about following rules and regulation. All the outcomes are defined in the notes to Tables 2, 3 and 4. All regressions control for randomization strata (gender and prime treatment). Robust standard errors in parentheses, sharpened q-values in square brackets (computed within each panel). *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C15. Colombia: Mental Simulation Outcomes in Sample with Business at Baseline

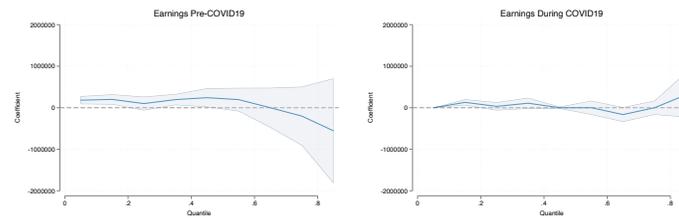
	(1)	(2)	(3)	(4)	(5)	(6)
	Simulation Index, by Domain			Sub-indices (Business Scenarios)		
	All Scenarios	Business Scenarios	Non-business Scenarios	Specificity	Emotionality	Frequency of Use
<i>Panel A: Traditional vs Control</i>						
Traditional Training	-0.068 (0.102)	-0.131 (0.111)	-0.045 (0.103)	-0.053 (0.111) [0.974]	-0.145 (0.104) [0.974]	-0.074 (0.102) [0.974]
<i>Panel B: GME vs Traditional</i>						
GME Training	0.066 (0.086)	0.181* (0.095)	0.005 (0.083)	0.086 (0.091) [0.164]	0.189** (0.088) [0.104]	0.142* (0.085) [0.104]
<i>Panel C: GME vs Control</i>						
GME Training	0.017 (0.095)	0.052 (0.098)	-0.013 (0.096)	0.057 (0.097) [1.000]	0.035 (0.090) [1.000]	0.047 (0.090) [1.000]
Mean DV in Control	0.00	0.00	0.00	0.00	0.00	0.00
Mean DV in Traditional	0.00	-0.10	0.00	0.00	-0.10	0.00
N in Control	278	241	278	241	241	241
N in Traditional	341	285	341	285	284	284
N in GME	595	513	595	512	509	509

Notes. This table reproduces Panel B of Table 2 from the main paper only in the sample of people with a business at baseline. Regression specifications and outcomes are defined in the note to Table 2. Panel A shows the effect of the traditional treatment versus control, Panel B compares the GME treatment to the traditional treatment, and panel C looks at the impact of the GME treatment with respect to the control. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Figure C2. Quantile Treatment Effects on Earnings Levels



(a) Traditional vs Control



(b) GME vs Traditional

Notes: The Figure shows quantile treatment effects on earnings (in levels, winsorised at the 99th percentile) in the Colombia experiment. Figure (a) shows the comparison between the traditional training and the control, while Figure (b) shows the comparison between the GME training and the traditional one. Figures on the left-hand side use as dependent variable non-standardized earnings levels pre-Covid, and figures on the right-hand side non-standardized earnings levels during Covid.

Table C16. Colombia: Business Practices

	Trad Training		GME Treat		t-test:
	mean	sd	mean	sd	difference p-value
<i>Marketing Practices:</i>					
Examined A Competitor'S Prices	65.33	47.67	70.69	45.46	0.12
Examined A Competitor'S Products	70.00	45.90	78.61	40.93	0.01***
Asked For Consumer Feedback On New Products	50.67	50.08	59.86	48.76	0.01**
Asked A Former Consumer Why They Stopped Buying	34.67	47.67	43.88	49.46	0.01**
Asked A Supplier About Product Sales	61.07	48.84	60.59	48.70	0.89
Introduced A Special Offer	69.33	46.19	72.22	44.73	0.39
Advertised	79.33	40.56	81.39	38.84	0.48
<i>Costing and Record-Keeping Practices:</i>					
Keep Written Business Records	57.33	49.54	48.26	49.82	0.01**
Record Every Purchase And Sales	66.67	47.22	70.28	45.42	0.29
Use Records To Examine Cash Balance At Any Time	64.43	47.95	61.79	48.33	0.46
<i>Financial Planning Practices:</i>					
Set A Sales Target For The Next Year	66.67	47.22	50.97	49.74	0.00***
Compares Sales To Target Monthly	70.47	45.69	65.21	47.57	0.13
Budgets Likely Costs	48.67	50.07	38.82	48.58	0.01***
<i>HR practices:</i>					
Training Provision Exists	35.57	47.95	33.89	47.28	0.63
Evaluated Employee Performance	39.60	48.99	36.81	48.18	0.44
Provides Employees Feedback	42.95	49.58	40.97	49.14	0.59
Observations	150		244		

Notes. The table shows business owners' engagement in business practices in the Colombia experiment. Data come from a survey conducted in 2021 (two years after the intervention), which reached around 50% of the original sample. The response rate was not differential by treatment (GME vs Traditional). The number of observations in the table is even lower as questions on business practices were only asked to people who had an operating business at the time of the survey. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Supplementary Materials – Experimental Context

Learning to See the World’s Opportunities: Memory, Mental Experiencing, and the Economic Lives of the Vulnerable

Nava Ashraf, Gharad Bryan, Alexia Delfino, Emily A. Holmes,
Leonardo Iacovone, Christian Johannes Meyer, Ashley Pople

Table of Contents

D	Details of Experiment 1 (Ethiopia)	2
D.1	Context	2
D.2	Intervention and Experimental Design	4
D.3	Timeline of Activities and Data Collection	4
D.4	Balance, Take up, and Attrition	5
D.5	Measurement and Construction of Outcome Variables	5
D.6	Cost Effectiveness and Scalability	8
D.7	Figures and Tables	9
E	Details of Experiment 2 (Colombia)	15
E.1	Context	15
E.2	Intervention and Experimental Design	16
E.3	Timeline of Activities and Data Collection	17
E.4	Balance, Take up, and Attrition	17
E.5	Measurement and Construction of Outcome Variables	18
E.6	Cost Effectiveness and Scalability	20
E.7	Figures and Tables	21

D Details of Experiment 1 (Ethiopia)

This appendix contains additional contextual and implementation details for Experiment 1 (Ethiopia).

D.1 Context

Ethiopia hosts a large number of refugees, reflecting enduring conflict in the broader Horn of Africa region. At the time of data collection in 2022, almost 880,000 registered refugees from 26 countries lived in Ethiopia. Most refugees come from South Sudan (46 percent), Somalia (29 percent), Eritrea (19 percent), and Sudan (5 percent). Ethiopia has, *de jure*, one of the most progressive refugee policy frameworks in Africa, which gives widespread rights to refugees including paid employment and self employment, access to education, and access to social and financial services. There remains, however, a significant gap between policy and implementation. At the time of data collection, formal work permits had not been granted to any refugees, and *de facto* access to economic opportunities was limited to a small number of urban refugees engaging in informal work (UNHCR, 2021, 2022).

This project focuses on one such urban refugee population: registered refugees from Eritrea located in Addis Ababa, Ethiopia’s capital city. In Ethiopia, a limited number of refugees are granted permission to live in urban areas outside of camps if they fulfill certain criteria, including specific protection needs or the ability to name a sponsor who can provide for them.³¹ Despite this policy not legally affording refugees the right to work or operate their own businesses, many refugees still take up informal economic opportunities. Thus, this population of refugees serves as an important test case for broader economic integration efforts by the government and illustrates the challenges of economic self-reliance in an urban setting, including the ability to find informal work or start a business without a permit. According to the UNHCR (2021), “[m]ost of the registered urban refugees are not able to meet their basic needs with the current income that they receive either from informal work or remittances” (p. 13). The Eritrean refugees we work with tend to have more access to economic opportunities than other refugee populations due their linguistic, ethnic, and cultural proximity with the host community. In Addis Ababa, networks among Eritrean refugees and with the host community are the main source of economic opportunities.³²

We partnered with the Ethiopian federal government’s refugee agency, the Ethiopian Refugee & Returnee Service (RRS), to implement the study. We worked closely with RRS to determine an appropriate study sample to study the economic lives of vulnerable urban refugees and agreed to focus on the Eritrean refugee population in Addis Ababa. Since economic behavior is our key focus, we restricted our sample to registered refugees aged 18 to 50 with at least junior high school education (Ethiopian grade 7 and above).³³ RRS provided us with a de-identified list of all registered refugees that satisfied these criteria ($N = 36,136$), from which we drew a random sample stratified by gender ($N=2,000$). Before survey enumerators approached refugees in our sample, we worked with RRS caseworkers and refugee community leaders to disseminate information about our study. After this initial approach, a professional survey firm administered informed consent and conducted a baseline survey on 1,652 refugees.³⁴ D1 summarizes the characteristics of our sample along the margins of economic vulnerability and trauma.

³¹See Woldetsadik et al. (2019) for a detailed discussion of the “out of camp” policy (OCP) from a legal perspective.

³²A 22-year old Eritrean refugee quoted by Betts et al. (2019) provides an illustrative account: “I am working at a metal workshop that does welding and repairing for doors, and houses and fences. The owner of the workshop is Ethiopian. I got connected to him through my Eritrean refugee friend who has contacts with Ethiopians... There are five employees at this workshop and all of them are Eritrean refugees” (p. 9).

³³The sampling frame is based on a centralized UNHCR-RRS database. Our sample excludes Eritrean asylum seekers not yet granted refugee status, refugees that fled from Tigray between 2020 and 2022, or Ethiopian internally displaced people. In Addis Ababa, 77 percent of refugees have completed primary education (based on the 2023 World Bank Socio-Economic Survey of Refugees in Ethiopia), so our sample is relevant to a large share of the population.

³⁴Potential respondents were approached in random order. If a respondent could not be found or did not consent to participate, they were replaced with someone from the same stratum drawn from a separate replacement frame.

The median refugee in our study left Eritrea five years prior to data collection and has lived in Addis Ababa for three years. A significant majority, 94 percent, expressed intentions to leave at some point in the future. Our survey elicits subjective expectations over the likelihood of staying in Ethiopia and expected income under various scenarios (no formal work permit, 1 percent chance of formal work permit, 50 percent chance, and 99 percent chance). At baseline, the median respondent indicates a 40 percent probability of staying in Ethiopia for the three years following the survey if not given a formal work-permit. Assuming a near-certain permit with 99 percent chance, the median respondent indicates a 60 percent probability of staying. This suggests that the uncertainty related to the legal right to work can partly explain the reluctance to stay in Ethiopia. The weak intent to stay underscores a tendency for envisioning a vague, distant positive future (e.g., migrating to a third country) over planning for an uncertain near-term future.

The avoidance of planning concretely for the near future is consistent with the low rate of economic activity observed in the sample. Only 19 percent of the respondents were economically active at the time of the baseline survey. This breaks down to 12 percent engaged in wage work and 7 percent in other activities including self-employment. 39 percent of respondents indicate that they are unemployed but looking for work. In comparison, 59 percent of people in the general population of Addis Ababa in the same age range and with the same education are economically active. Conditional on being employed, the average wage in our sample is slightly less than half of the general population (about USD 98 compared to about USD 189). There is a pronounced gender gap in the economic lives of our sample. Among men, 16 percent were involved in wage work compared to just 8 percent of women. This gap extends to income, where working men reported an average monthly income of 3,830 birr (about USD 68), compared to working women who reported an average monthly income of 2,555 birr (about USD 45).

Mental health in our sample is poor relative to the general population. We find that 13 percent of our sample is above the cut-off for probable PTSD, relative to 4 percent in the general population (Koenen et al., 2017). These findings are in line with Betts et al. (2019), who find that Eritrean refugees in Addis Ababa face significantly more mental health problems than the host community. In our study, over half of the participants (52 percent) reported direct exposure to at least one traumatic event (and significantly higher rates when looking at traumatic events directly experienced and those witnessed first-hand, as in Table 1 of the paper). This rate is notably higher compared to the general Ethiopian population, where only 35.8 percent report having directly experienced the same (Girma et al., 2022).³⁵ We find a wide spectrum of traumatic experiences within our sample, many of which are plausibly related to forced displacement: 20.7 percent of our sample report having directly experienced physical assault, 13.4 percent report having directly experienced severe human suffering, and 12.8 percent report having directly experienced physical combat or exposure to a war-zone in the military or as a civilian. Among women in our sample, 7.6 percent report having directly experienced severe sexual assault or threat of assault (e.g., rape, attempted rape). Most respondents reported that their most traumatic experience happened more than two years prior to data collection.

In summary, Eritrean refugees in Addis Ababa face significant challenges in economic integration. Even if informal livelihood opportunities are accessible, economic activity remains limited. Forcibly displaced populations do not have the time to pre-experience positive images and desirable outcomes about their life at destination, and many feel to be caught up in the past. We hypothesize that in our sample, the lack of these images may be one factor that explains the low willingness to invest into a life in Addis Ababa and the negative view of what their stay in Ethiopia may bring.

³⁵We use the Life Events Checklist (LEC-5) for DSM-5, which has demonstrated good construct validity among Ethiopian adults (Girma et al., 2022).

D.2 Intervention and Experimental Design

We randomly assigned 900 eligible individuals (50 percent) to the GME intervention and 900 to a control group without intervention. Treatment assignment was stratified by gender and a cross-randomized priming treatment not discussed in this paper. [Figure 1](#) in the main body of the paper illustrates our design using a Consolidated Standards of Reporting Trials (CONSORT) chart.

The GME curriculum consists of four one-hour sessions that demonstrate how mental experiencing can be used: to imagine and possibly reframe a potential life in Addis Ababa in the immediate term; to build a safe mental space as a foundation for making economic decisions; to reflect on specific actions that can be taken to access economic opportunities; and to think about specific plans and steps that can be taken towards improving economic outcomes. Each session features two to three GME exercises guided by the trainer and implemented using a pre-recorded audio by a clinical psychologist, lasting 5–7 minutes. These recordings ensure standardization and high quality. Trainers use a scripted guide (prepared by researchers) for the rest of the session, which includes moments for participants to reflect on their experiences and problem-solve challenges faced in practicing positive mental experiences.

All GME exercises follow the same structure: First, the trainer gives a brief introduction and prepares participants (e.g., posture, breathing). Then the trainer guides the exercise using a script prepared by the researchers. Finally, debriefing questions are asked to encourage reflection and motivate the subsequent activities. While we do not share information about new economic opportunities in the intervention, participants are encouraged to consider simple ways to leverage opportunities that already exist in Addis Ababa (e.g., asking acquaintances about potential jobs).

To build trust in the program, trainers were refugees from the same community without specialized training. They received weekly training and supervision from clinical psychologists and psychiatrists at Addis Ababa University, led by Dr. Benyam Worku Dubale. Using non-specialized trainers allows us to test an intervention with scale-up potential in similar contexts. Sessions took place in a community center rented from a local NGO. A comprehensive risk management protocol was in place to ensure participant, trainer, and enumerator safety during the project duration. This included a referral system for participants showing signs of severe mental illness.

D.3 Timeline of Activities and Data Collection

[Figure D1](#) shows the timeline of the intervention and data collection. The primary data sources for the Ethiopia experiment are a baseline survey and an endline survey conducted approximately 60 days after the end of the intervention. The baseline survey was collected on a rolling basis between April and July 2022. For treated participants, the first GME session was delivered two weeks after the baseline survey. The remaining three GME sessions were delivered on a weekly basis. The endline data collection was completed in October 2022. All baseline and endline surveys were conducted in person.

Given the sensitive context and vulnerability of the population, we collaborated with a survey firm to recruit and train Eritrean refugees as enumerators. This group of enumerators was different from the group of refugees implementing the GME sessions. We decided to work with refugee community members for three reasons: first, these refugees could communicate effectively with other refugees; second, our qualitative work and government partners suggested this approach might reduce social desirability concerns; and third, we anticipated their involvement would increase trust within the surveyed community. The project was presented to respondents as an independent academic research project. Protocols to minimize the risk of infection from COVID-19 were in place and were monitored by the survey firm.

D.4 Balance, Take up, and Attrition

Balance

Table D1 shows that treatment assignment is balanced at baseline. There is no significant difference between the treatment and control groups at baseline on relevant demographic, economic, and health characteristics. We also fail to reject the null in a joint orthogonality test of the significance of all the variables shown in Table D1 on treatment assignment ($F(18, 1631)=0.89, p=0.60$).

Program Take-up

In terms of program take-up, 75 percent of participants assigned to the GME treatment attended at least one session. On average, participants attended two sessions, i.e. half of the four-session program. Table D2 shows the main observable correlates of take-up. Older and more traumatized participants are more likely to attend at least one sessions, while women and respondents with higher expenditures are less likely to engage in the program.

Survey Attrition

We surveyed 84 percent of baseline respondents in our endline survey, so overall attrition is low. Table D3 shows that very few observables predict participation to the endline survey. Respondents who are younger at baseline attrited more than others (significant at 10 percent), and so did respondents with a better economic situation (total expenditures is significant at 1 percent and the frequency of savings is significant at 5 percent). No other demographic or employment characteristics are related with participation into the endline.

The survey attrition rate by treatment is summarized in the table below. A regression of attrition on the GME treatment indicator tells us that attrition among GME-trainees is 5.6 percentage points lower than in the control group ($p = 0.002$). However, notice that our endline contains exactly 50% of respondents from the treatment and the control groups, respectively, which corresponds to the treatment assignment shares in the overall sample (but only baseline participants could be invited to the intervention.)

Survey Attrition Rate, by Treatment

Treatment Status	(1) Overall	(2) Baseline	(3) Endline	(4) Attrition rate (1)-(3)	(5) Attrition rate (2)-(3)
GME Treatment	900	798	697	0.225	0.126
Control	900	854	698	0.224	0.183
Full sample	1,800	1,652	1,395	0.225	0.156

Table D4 mitigates concerns related to differential attrition by showing that the observable characteristics of participants in the endline survey are balanced between treatment groups. The only exception is total (unconditional) income, which is higher among the treated compared to the control group, but this is only significant at the 10 percent level. In our PAP report (a link can be found in Section 4), to further address differential attrition, we show that our main results are broadly robust using Lee Bounds to correct for differential attrition between the treatment and control groups.

D.5 Measurement and Construction of Outcome Variables

This section provides further details on the construction of the main outcomes as specified in our PAP. We comment below on specific cases in which an outcome was included in our pre-analysis, but it is shown

among the “secondary” outcomes in the paper.

We collected three families of primary outcomes: mental simulation, economic outcomes and mental and physical health. We construct our indices within each family following [Kling et al. \(2007\)](#): (i) all variables are first consistently signed (e.g. higher value associated with higher ability or welfare); and (ii) each component of the index is then standardized by subtracting the control group mean and dividing by the control group standard deviation. In the case where there are multiple subscales, we take the average of the standardized components. We provide details on the construction of the outcome variables for each family in the following paragraphs.

Mental Simulation

Our overall mental simulation index is constructed from an adapted version of the Prospective Imagery Task ([MacLeod et al., 1993](#); [Stöber, 2000](#); [Holmes et al., 2008](#)) and the Autobiographical memory test ([McNally et al., 1995](#)). In both the baseline and endline surveys, we present participants with two versions of this test: one oriented to the future and one oriented to the past. For both versions, we present participants with four different scenarios (specific to the version given) and ask them to imagine a specific memory or future event that is linked to the given scenario within 45 seconds. The scenarios include one negative, one positive and two neutral. For each memory or future event they mention, we then ask participants to reproduce it in their mind and report the vividness and emotional intensity of it, both on 5-points scales. We further ask the surveyor for the emotional affect reflected in the response.

Following our PAP, for both versions of the test, we coded the answers given by participants in the back office in terms of three measures: i) specificity, ii) emotionality and iii) positivity. Research assistants were blind to treatment and had to code the answers based on criteria given by the research team. Both specificity and emotionality were rated on a scale from 1 (“No memory at all” or “No emotion at all”) to 5 (“Very specific” or “Extremely strong emotions”). Positivity was rated on a scale from 1 (“Very negative”) to 4 (“Very positive”). Our detailed questions are listed in [Table D5](#) below.

We gave detailed guidelines to research assistants and examples on how to place recordings on different points of the scales. We followed the psychological literature ([Griffith et al., 2009](#)) to build the criteria for defining “specific” answers, which mean that they refer to events which happened on a particular day and lasted less than 24 hours. By contrast, non-specific memories do not refer to any specific event but rather to doing an action in general.

Overall we construct three different indices for each of the two versions of the scale (past and future), which we then combined into three aggregate indices of specificity, emotional intensity, and positivity.

Economic outcomes

We focus on two main sets: (1) Employment outcomes, (2) income, consumption, and welfare.

1. Employment outcomes: Engagement in employment and/or self-employment (extensive and intensive margin) and job search.
 - Engagement in employment and/or self-employment is measured as a dummy equal to 1 if a person is involved in any of the following activities i) has a non-agricultural or non-fishing business for yourself or for your household, or ii) is in casual, part-time, or temporary labor, or iii) is in wage or salary work, or iv) is in an unpaid apprenticeship. If respondents work in a specific activity, we ask for the number of hours worked. Our analysis focuses on a dummy for any economic activity and the total number of hours spent. We focus on the aggregate

across all activities, but we also report results for each activity separately for completeness in the appendix.

- Job search is measured with a battery of questions adapted from the Ethiopian Socioeconomic Survey (ESS) / Living Standard Measurement Study (LSMS) ([Central Statistics Agency of Ethiopia, 2020](#)) and previous literature in this context: first, we ask whether the person looked for ways to start one of the following activities: i) non-agricultural or non-fishing business for yourself or for your household, ii) casual, part-time, or temporary labor, iii) wage or salary work, iv) unpaid apprenticeship. If they did, we ask for the number of hours spent searching, the number of calls made, money spent, channels used, length of search and chances that they will start the activity within 3 months. As main outcome, we build an index of effort in search by aggregating in an index the hours, number of calls made, number of channels used and money spent.

Hours worked and job search are presented in the paper among “secondary” outcomes.

2. Income, consumption, and welfare: earnings, savings, consumption, food security

We adapt a number of survey module from the ESS/LSMS to measure baseline and endline living standards of study participants. Closely following the ESS/LSMS lets us compare the study sample to the broader population in the host community of Addis Ababa. For this outcome family, we collect data on six dimensions: (1) earnings from economic activities, (2) Food consumption (at home and away from home), dietary diversity, and non-food consumption, (3) Food security, (4) savings, (5) incoming and outgoing transfers, (6) support from various assistance programs.

- Our food security index is the total score (sum) of the four-item version of the Food Insecurity Experience Scale (FIEFS). It includes reverse-coded questions about i) not having enough food, ii) number of days with less preferred food, iii) number of days with limited food variety and iv) number of days no eating for full day in the past week.
- Our total earnings index includes earnings from wage work (in the last seven days) and profits from self-employment (in the last twelve months).

For conciseness, we do not report results on savings, incoming and outgoing transfers and assistance. These results can be found in our PAP report.

Mental and physical health

We use two broad measures of well-being and physical health: (1) the Cantril Self-Anchoring Striving Scale and the (2) World Health Organization Disability Assessment Schedule (WHODAS) Schedule 2.0.

We use the Cantril Self-Anchoring Striving Scale ([Cantril, 1965](#)) as a simple and widely-used self-reported measure of well-being. We follow Gallup’s implementation: we ask respondents to imagine, using a visual aid, a ladder with steps numbered from zero at the bottom to 10 at the top, where the top represents the best possible life for respondents and the bottom the worst possible life. We then ask respondents two questions. (a) “On which step of the ladder would you say you personally feel you stand at this time?” (b) “On which step do you think you will stand about five years from now?”. Our life satisfaction index is the standardized average of the answers to the two questions.

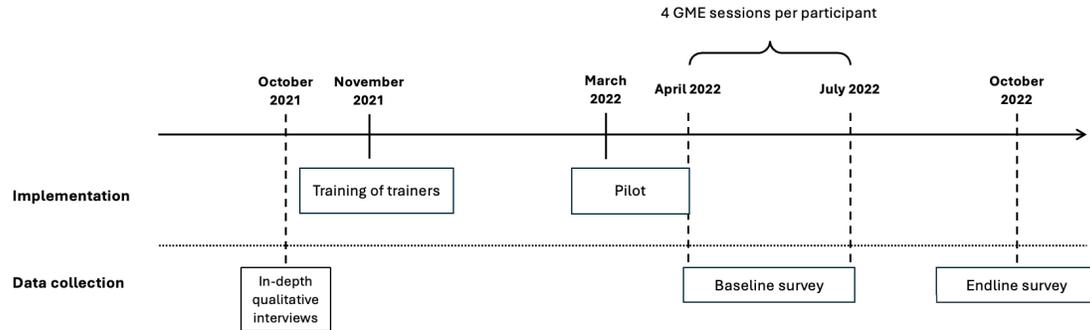
The World Health Organization Disability Assessment Schedule (WHODAS) is a widely used measure of disability and functional impairment in accordance with the International Classification of Functioning, Disability and Health. We use the 12-item self-reported version and score it using the WHO’s simple additive method (the score ranges from 12 to 60). Our WHODAS outcome variable is standardized and reverse-coded, so that higher values indicated lower disability.

D.6 Cost Effectiveness and Scalability

Table D6 provides more detailed data on program costs. Dividing the total cost of USD 152,180 by 664 actual beneficiaries (who took up at least one session of the GME intervention) results in average costs of USD 229 per beneficiary. We estimate the marginal costs to deliver the program to an additional beneficiary to be approximately USD 58, though we caution that specific features of the implementation infrastructure (such as our risk management framework) will not scale linearly. While the implementation of the intervention was researcher-led and supervised, we believe that the program is appropriate to be scaled up within the programming of governments and development partners. When doing so, it will be important to ensure a high quality of training and an effective risk management framework and referral system that ensures safety of participants and staff.

D.7 Figures and Tables

Figure D1. Timeline of Implementation and Data Collection



Notes: This figure illustrates the timeline of data collection and intervention delivery for the Ethiopia trial. The baseline survey was conducted on a rolling basis during April and July 2022. For treated participants, the first GME session was delivered two weeks after the baseline survey. The remaining three GME sessions were delivered on a weekly basis.

Table D1. Ethiopia: Baseline Balance

Variable	(1) Control	(2) Treatment	(3) Difference
<i>Demographics</i>			
Age	28.285 (6.685)	28.688 (7.182)	0.403 (0.341)
Female	0.499 (0.500)	0.500 (0.500)	0.001 (0.025)
Years of Schooling	10.561 (3.150)	10.528 (3.106)	-0.033 (0.154)
Household Size	1.794 (1.578)	1.878 (1.729)	0.085 (0.081)
Years in Addis	2.847 (1.954)	3.006 (2.333)	0.160 (0.106)
<i>Employment and economic outcomes</i>			
Engaged in Work	0.189 (0.391)	0.207 (0.405)	0.018 (0.020)
Hours Worked	7.948 (20.534)	8.330 (20.631)	0.381 (1.013)
Job Search Hours	4.102 (13.171)	4.986 (14.953)	0.884 (0.693)
Total Expenditure	4,391.381 (2,039.233)	4,360.888 (2,664.041)	-30.494 (116.282)
Food Security	16.960 (3.926)	17.218 (3.901)	0.258 (0.193)
Frequency of Savings	0.296 (0.457)	0.313 (0.464)	0.017 (0.023)
Total Income	282.336 (974.281)	326.947 (1,051.613)	44.612 (49.844)
Bank Account	0.952 (0.214)	0.949 (0.221)	-0.003 (0.011)
Weekly Wage	137.429 (488.720)	139.254 (493.654)	1.825 (24.180)
<i>Trauma and mental health</i>			
Number of Traumatic Events	4.770 (2.791)	4.860 (2.917)	0.089 (0.140)
PTSD Score	15.007 (13.933)	15.162 (13.970)	0.155 (0.687)
Life Satisfaction	6.729 (1.582)	6.660 (1.594)	-0.069 (0.078)
Overall Health	0.726 (0.111)	0.722 (0.108)	-0.004 (0.005)
Observations	854	798	1,652

Notes: The table presents a balance test at baseline for the Ethiopia experiment. Columns 1 and 2 show the mean and standard errors for the control and GME treatment. The last column show the difference between treatment arms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Engaged in work is a dummy that is one if respondents are engaged in any economic activity in the past week. Hours worked refers to the unconditional number of hours worked in the past week. Total expenditures refers to total individual consumption expenditures in the past week, in Ethiopian birr. Food security is the total score (sum) of the four-item version of the Food Insecurity Experience Scale (FIEFS). Frequency of savings indicates the frequency in which respondents say they could save 600 Ethiopian birr. Total income is the unconditional weekly income from wage employment and self-employment. Bank account is an indicator that is one if respondents have a registered account with a financial institution. Weekly wage is the unconditional weekly income from wage work only. Number of traumatic events refers to the number of events from the Life Events Checklist for DSM-5 (LEC-5) personally experienced or witnessed. PTSD Score refers to the score on the PTSD Checklist for DSM-5 (PCL-5). Life satisfaction refers to the Cantril Self-Anchoring Striving Scale. Overall health refers to the overall score on the World Health Organization Disability Assessment Schedule (WHODAS).

Table D2. Ethiopia: Intervention Take-up

Variable	(1) No Take-up	(2) Take-up	(3) Difference
<i>Demographics</i>			
Age	27.446 (6.352)	29.109 (7.401)	1.664*** (0.582)
Female	0.673 (0.470)	0.441 (0.497)	-0.232*** (0.040)
Years of Schooling	10.584 (3.280)	10.508 (3.047)	-0.076 (0.253)
Household Size	1.807 (1.715)	1.903 (1.734)	0.096 (0.141)
Years in Addis	2.931 (1.694)	3.032 (2.514)	0.101 (0.190)
<i>Employment and economic outcomes</i>			
Engaged in Work	0.203 (0.403)	0.208 (0.406)	0.005 (0.033)
Hours Worked	8.168 (20.272)	8.384 (20.768)	0.216 (1.681)
Job Search Hours	4.045 (14.341)	5.304 (15.129)	1.259 (1.216)
Total Expenditures	4,841.984 (3,929.763)	4,197.832 (2,045.594)	-644.153*** (215.824)
Food Security	18.025 (3.616)	16.945 (3.958)	-1.080*** (0.315)
Frequency of Savings	0.351 (0.479)	0.300 (0.459)	-0.051 (0.038)
Total Income	305.953 (1,140.837)	334.063 (1,020.522)	28.110 (85.665)
Bank Account	0.950 (0.217)	0.948 (0.222)	-0.003 (0.018)
Weekly Wage	112.450 (431.520)	148.339 (513.027)	35.888 (40.196)
<i>Trauma and mental health</i>			
Number of Traumatic Events	4.436 (2.849)	5.003 (2.928)	0.568** (0.237)
PTSD Score	13.460 (12.467)	15.738 (14.408)	2.278** (1.135)
Life Satisfaction	6.807 (1.451)	6.611 (1.637)	-0.196 (0.130)
Overall Health	0.732 (0.106)	0.719 (0.109)	-0.013 (0.009)
Observations	202	598	1,654

Notes: The table presents characteristics of participants that attended at least one session of the intervention (Column 2) and those that did not (Column 1). Columns 1 and 2 show the mean and standard errors for each group. Column 3 shows the difference between both groups, with significance indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variables are defined in the note to table D1.

Table D3. Ethiopia: Survey Attrition at Endline

Variable	(1) Did not drop	(2) Dropped	(3) Difference
	<i>Treatment Status</i>		
GME Treatment	0.500 (0.500)	0.393 (0.489)	-0.107*** (0.034)
	<i>Demographics</i>		
Age	28.619 (7.018)	27.724 (6.392)	-0.895* (0.470)
Female	0.497 (0.500)	0.514 (0.501)	0.017 (0.034)
Years of Schooling	10.578 (3.094)	10.366 (3.311)	-0.212 (0.212)
Household Size	1.853 (1.678)	1.735 (1.505)	-0.118 (0.112)
Years in Addis	2.905 (2.192)	3.027 (1.878)	0.123 (0.146)
	<i>Employment and economic outcomes</i>		
Engaged in Work	0.197 (0.398)	0.198 (0.400)	0.001 (0.027)
Hours Worked	8.025 (20.440)	8.716 (21.326)	0.691 (1.397)
Job Search Hours	4.747 (14.188)	3.354 (13.337)	-1.393 (0.955)
Total Expenditures	4,312.418 (2,050.704)	4,725.307 (3,592.882)	412.889*** (160.001)
Food Security	17.006 (3.887)	17.514 (4.042)	0.508* (0.266)
Frequency of Savings	0.292 (0.455)	0.370 (0.484)	0.077** (0.031)
Total Income	296.698 (994.441)	342.898 (1,105.554)	46.200 (68.729)
Bank Account	0.953 (0.211)	0.934 (0.249)	-0.020 (0.015)
Weekly Wage	130.952 (463.406)	178.255 (619.102)	47.303 (33.317)
	<i>Trauma and mental health</i>		
Number of Traumatic Events	4.812 (2.841)	4.821 (2.914)	0.009 (0.194)
PTSD Score	15.169 (14.006)	14.607 (13.636)	-0.562 (0.947)
Life Satisfaction	6.681 (1.586)	6.774 (1.599)	0.093 (0.108)
Overall Health	0.725 (0.108)	0.718 (0.119)	-0.008 (0.007)
Observations	1,395	257	1,652

Notes: The table presents baseline characteristics of respondents that were found at endline (Column 1) and those that were not found at endline (Column 2). Columns 1 and 2 show the mean and standard errors. The last column shows the difference between columns 1 and 2, with significance indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variables are defined in the note to table [D1](#).

Table D4. Ethiopia: Baseline Characteristics of Respondents at Endline, by Treatment

Variable	(1) Control	(2) Treatment	(3) Difference
<i>Demographics</i>			
Age	28.348 (6.718)	28.890 (7.301)	0.541 (0.376)
Female	0.497 (0.500)	0.496 (0.500)	-0.001 (0.027)
Years of Schooling	10.613 (3.094)	10.542 (3.095)	-0.071 (0.166)
Household Size	1.804 (1.589)	1.902 (1.762)	0.099 (0.090)
Years in Addis	2.785 (1.952)	3.024 (2.405)	0.239** (0.117)
<i>Employment and economic outcomes</i>			
Engaged in Work	0.182 (0.386)	0.212 (0.409)	0.030 (0.021)
Hours Worked	7.640 (20.336)	8.410 (20.552)	0.770 (1.095)
Job Search Hours	4.163 (12.812)	5.331 (15.429)	1.168 (0.761)
Total Expenditures	4,343.772 (2,052.544)	4,281.020 (2,049.853)	-62.753 (109.838)
Food Security	16.933 (3.885)	17.079 (3.891)	0.146 (0.208)
Bank Account	0.957 (0.203)	0.950 (0.219)	-0.007 (0.011)
Frequency of Savings	0.282 (0.450)	0.303 (0.460)	0.020 (0.024)
Weekly Wage	125.642 (442.982)	136.270 (483.255)	10.628 (24.822)
Total Income	251.339 (891.976)	342.122 (1,086.174)	90.783* (53.214)
<i>Trauma and mental health</i>			
Life Satisfaction	6.717 (1.586)	6.646 (1.585)	-0.071 (0.085)
Overall Health	0.727 (0.110)	0.723 (0.106)	-0.004 (0.006)
Number of Traumatic Events	4.716 (2.752)	4.908 (2.927)	0.192 (0.152)
PTSD Score	15.159 (14.141)	15.179 (13.880)	0.020 (0.750)
Observations	698	697	1,395

Notes: The table presents baseline characteristics of respondents at endline, by treatment status. Columns 1 and 2 show the mean and standard errors for the control and GME treatment. The last column show the difference between treatment arms, with significance indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variables are defined in the note to table D1.

Table D5. Measurement: Mental Simulation in Ethiopia Study

Reference Period	Scenarios	Rating Criteria for Each Scenario
	Respondents are presented with the following scenarios in randomized order. For each scenario, they are asked to think of a specific <i>event that has happened</i> .	
Past	1) A relative gets sick (negative) 2) One of your friends gets a job (positive) 3) You take a health test (neutral) 4) You plan your next day (neutral)	Specificity (1 “No memory at all” to 5 “Very specific”)
	Respondents are presented with the following scenarios in randomized order. For each scenario, they are asked to think of a specific <i>event that could happen in the future</i> .	
Future	1) Someone insults you on the street (negative) 2) Someone gives you a winning lottery ticket (positive) 3) You ask someone for a job (neutral) 4) You go out for a walk in your neighborhood (neutral)	Emotionality (1 “No emotion at all” to 5 “Extremely strong emotions”)
		Positivity (1 “Very negative” to 4 “Very positive”)

Notes: The questions for each scenario were self-assessed and rated by treatment-blind research assistants based on audio recordings. The table presents an English translation of the Tigrinya questionnaire. The original questionnaire is available from the authors upon request.

Table D6. Ethiopia: Detailed Cost Estimates

Category	Costs (USD)	Comments
Program administration and staff costs	58,157	Project steering and coordination. Staff time of refugee expert and psychiatrist.
Targeting costs	21,056	Community outreach campaign
Staff training	3,671	Government inception workshop plus training
Participant training	6,000	Training venue hire and related costs
Implementation and program material costs	53,057	Production of curriculum materials (esp. video) and implementation of training
User costs	none	
Averted costs	none	
Monitoring costs	10,240	Risk management framework and referral system
Total program costs	152,180	

Notes: The table presents a detailed cost estimates following the J-PAL costing template for cost effectiveness analysis. We assume an ETB/USD exchange rate of 51.76 for 2022 taken from the World Bank WDI database. See www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea for more details on the template.

E Details of Experiment 2 (Colombia)

This appendix contains additional details for Experiment 2 (Colombia).

E.1 Context

Decades of civil conflict in Colombia have internally displaced more than seven million people (IDMC, 2017). With over 8 million registered victims of conflict, the 2016 peace deal brought socio-economic inclusion to the forefront of policy priorities. Since 2015, Colombia has also faced a significant influx of Venezuelan refugees and migrants due to their country’s humanitarian crisis. By October 2022, the Colombian government reported 2.89 million Venezuelan refugees, migrants, and asylum seekers (R4V, 2023). This unprecedented inflow has added socio-economic challenges in many communities already struggling with post-conflict recovery.

As in other post-conflict contexts, self-employment is emphasized as an effective tool for promoting recovery, restoring individual economic autonomy and boosting community development. Micro and small-scale firms represented more than 87 percent of firms in Colombia in 2016 (OECD, 2018) and are increasingly supported by the government, for example through vocational and skills training programs provided by the National Training Service (SENA).

Governmental and non-governmental actors in our context increasingly recognized the challenges of supporting the economic livelihoods of traumatized populations. Our meetings with public and private organizations highlighted that survivors of conflict and displacement were either not taking up programs or performing worse than the general population. At our project’s inception, the government provided ad-hoc psychological support, but demand was low. Most high-trauma, low-income households were unwilling to prioritize mental health without first addressing their economic difficulties. The government sought to design and test a program that blended economic skills with psychological support, appealing to those motivated to improve their economic lives but requiring help with post-trauma symptoms. The key idea on the ground was that a combined program would trigger a virtuous cycle between psychological health and economic outcomes.³⁶

To deliver the intervention, we partnered with the local government of Bogotá, specifically the District Department of Social Integration (SDIS) and District Department of Economic Development. Our government partners recruited participants through a multi-channel media campaign by advertising an entrepreneurship program promoting soft skills, including through community centers and social media platforms.³⁷ As part of their mandate, the SDIS were required to train numerous groups that face economic and social challenges (e.g., victims of conflict and Venezuelan migrants, low-income youth, LGBTQ groups). The resulting sample thus features vulnerability along multiple dimensions and represents groups that – according to our field partners – struggle with training and taking up of economic opportunities.

Our final sample includes 1,967 participants. Both trauma exposure and symptoms are very high in the sample. At baseline, 83 percent report exposure to traumatic past experiences, which is higher than estimates of exposure in general populations, ranging between 61 percent and 70 percent. Moreover, the types of trauma reported by participants are strikingly different from other settings: 48 percent of

³⁶This concept has also been proposed by researchers as an effective way to boost the demand and supply of mental health services in low-resource contexts and increase the returns to economic programs (Ridley et al., 2020).

³⁷Interested applicants were required to complete a short application form, online or in person, which was used to screen for eligibility. Of 3,553 applicants, 2,337 met the predefined criteria. Eligibility required demonstrating entrepreneurship potential by having a business or plans to start one within three months and being able to describe the business or idea and classify it by sector. To account for self-funded transportation costs, applicants needed to report non-zero income or business sales in the past six months. Additional criteria included being literate, being over 18 years old, and providing three points of contact.

participants report having lived or witnessed war-related trauma, including death, displacement and torture, and 61 percent report surviving or witnessing assault. As a benchmark, the WHO World Mental Health survey finds that exposure to these same types of trauma is 13.1 percent and 22.9 percent in general populations across more than 20 countries (Koenen et al., 2017). The share of participants positively screening for PTSD is also high in our Colombia sample, with 22 percent being above the cut-off for probable PTSD.

Vulnerability also extends to the economic domain. Participants are low-income, with 67 percent earning less than the minimum wage (compared to 51 percent overall in Colombia). The average monthly income at baseline is COP \$693,077 (approximately USD \$200), which is around 80 percent of the Colombian monthly minimum wage. The average household has 3.5 people (compared to 3.1 in the country). While 55 percent of the sample has an existing business, the remaining share has only a business idea. In terms of demographics, 58 percent of participants are women, the average participant is 32 years old and has completed secondary schooling.

In summary, the aspiring micro-entrepreneurs in our Bogotá sample face significant challenges in improving their psychological wellbeing and economic livelihoods. As in our Ethiopia setting, many individuals tend to envision a vague, distant, and positive future, but avoid planning for it and struggle to identify concrete steps to move forward.

E.2 Intervention and Experimental Design

The experiment in Colombia evaluates whether integrating GME into standard economic support programs can enhance their effectiveness for vulnerable populations. Accordingly, our design randomly assigns eligible applicants into two main treatments: a traditional business training (“traditional training”) and a business training plus GME (“GME treatment”). To explore the net effect of receiving a program, we also randomly assign part of the sample to a no intervention group, which does not receive any training (“control”). Given restrictions in government funds, we assigned 956 participants to the GME treatment (48 percent), 558 to the traditional treatment (28 percent) and 453 to the control (23 percent). Figure 2 in the main body of the paper illustrates our design using a Consolidated Standards of Reporting Trials (CONSORT) chart.

Treatment assignment was stratified by subdivision, sex, age, entrepreneurship status (having an existing business, an idea or both), sales for business owners or income for people without a business.³⁸

A key challenge was designing both treatment arms to be nearly identical, differing only in the inclusion of GME exercises. Both training programs consisted of ten three-hour sessions covering the same entrepreneurship themes in the same order. The sessions followed the journey of an aspiring entrepreneur, with themes including product development, customer experience, marketing, competition, savings, finance, productivity, and employee management.³⁹ In both curricula, we motivated the relevance of the content of each session through relatable real-life stories.

Both treatment arms also had an identical delivery format. Both curricula were implemented in weekly group sessions, with class sizes of 15 to 25 participants and venues available in the same neighborhoods across Bogotá. Two trainers led each session: one with expertise in psychology or social work, and the other in entrepreneurship. To minimize the impact of trainer heterogeneity, trainers were required to follow a standardized scripted manual and presentation deck prepared by the research team. All trainers were recruited through the same call using a well-defined selection criteria. Trainers received over three weeks of training. They also rehearsed before each week’s session in group workshops led by four supervisors who

³⁸Participants living in the same address and formerly homeless affiliated with the same shelter were assigned to the same treatment status to avoid spillovers. For our analysis, we aggregate the data to household/center level.

³⁹The curricula drew on materials from existing entrepreneurship training programs, including the International Labour Organization’s Start and Improve Your Business program.

had co-designed and piloted the curricula together with the research team. For all activities, trainers were kept separate by treatment arm. Participants who attended at least seven sessions received a certificate. Both programs were implemented by the local government of Bogotá, specifically the District Department of Social Integration (SDIS) and District Department of Economic Development.

The key feature that distinguishes the GME training from the traditional one is the inclusion of three to four mental experiencing exercises in each session. In the traditional curriculum, we replace the GME exercises with group discussions, role plays and written work of the same content and time length. For example, in the GME curriculum, participants were encouraged to imagine their product or service in detail and how they felt towards the product, before imagining the day in the life of their business and noting the emotions that arise. In contrast, participants in the traditional curriculum were asked to think about their product and write down the ways in which their business idea matches their passions and skills.⁴⁰

The design of the entrepreneurship training was informed by extensive qualitative data collection in our sample. We also incorporated ongoing feedback from various government ministries throughout the curriculum development process. Key collaborators included the lead psychologist team at the National Victims Unit (*Unidad para las Víctimas*), the High Council for the Rights of Victims, Peace, and Reconciliation (*Alta Consejería para los Derechos de las Víctimas, la Paz y la Reconciliación*), and the National Training Service (*Servicio Nacional de Aprendizaje*).

E.3 Timeline of Activities and Data Collection

The intervention was delivered in two waves, the first taking place from July to September 2019 and the second from September to December 2019. Figure E1 shows the timeline of the intervention and data collection. We conducted an extensive qualitative study and two pilots to refine the curriculum and RCT design in the year prior to implementation. Before starting the training within both waves, we first screened interested applicants to determine eligibility, randomized eligible participants into the three treatment arms and conducted a baseline survey. For participants in both waves, we conducted the first follow-up phone survey 6 to 8 months and the second follow-up phone survey 12 to 14 months after the end of the training program. We had originally scheduled an in-person follow-up survey for early March 2020, but it was cancelled in response to the Covid-19 pandemic and the stringent nationwide lockdown in Colombia. Thus we had to change plans and implemented two phone surveys commencing in May and November 2020.

Given the sensitive context and vulnerability of the population, we used surveyors from Innovations for Poverty Action (IPA) Colombia, an independent nonprofit research organization. IPA Colombia communicated to respondents that they were conducting a survey on entrepreneurs in Bogotá in collaboration with the local government partner. We needed to cite our government partner to encourage participation in a context with some of the lowest trust rates globally, but there was no mention of the program itself. All respondents were compensated COP \$10,000 (\$2.8) for their time, irrespective of whether they completed the survey.

E.4 Balance, Take up, and Attrition

Balance

Table E1 shows that treatment assignment is balanced on observables in the sample who accepted to participate in the training, which is the one used for the analysis, but balance also holds for the full sample of randomized participants. We fail to reject the null of a joint orthogonality test of the significance of all the variables shown in Table E1 on treatment assignment ($F(15,1818)=0.65$, $p=0.83$).

⁴⁰As a second example, in the GME group participants imagined their product or service from the perspective of their target customer, using empathy to understand customer needs. In contrast, the traditional training asked participants to complete a table summarizing customer needs and listing how their product addressed them.

Take-Up

The take up of each treatment group is substantial, with about 67% of individuals attending at least one session in the GME training across both waves, conditional on confirming attendance. By way of comparison, attendance rates in the [McKenzie and Woodruff \(2014\)](#) review of business training programs in developing countries range from 39 to 92%, with a mean of 64%. Of those attending at least one session, 48% attended seven or more of the ten sessions in the GME treatment and qualified for a certificate. Similarly, 63% attended at least one session in the traditional training across both waves. Of those attending at least one session, 53% attended seven or more sessions in the traditional training. Take up rates are 58% and 55% in the GME and traditional group, respectively, if we do not condition on confirmed attendance. Our sample is highly resource-constrained and mobile. In collaboration with our government partners, we developed several strategies to maximize compliance among those who were randomized to both training treatments. For instance, participants were assigned to venues that matched their location preferences to the best extent possible. Our supervisors ensured that participants were reminded on a weekly basis over WhatsApp groups, text messages and phone calls. Participants who attended seven or more sessions received a certificate signed off by the government partner (SDIS) and the World Bank, and invited to attend a graduation ceremony in the case of the second wave.

Additionally, [Table E2](#) shows which baseline characteristics are correlated with take up. We find that older people, those who have a business and with had a larger number of traumatic experiences are more likely to participate in at least one session, while women and people with stronger trauma symptoms and/or distress are less likely to take-up the training.

Attrition

We surveyed 74% of all participants in our first follow-up survey and 63% of participants in our second follow-up survey. [Table E3](#) shows that selection into the surveys is related with a few observable characteristics. For instance, participants with a slightly higher income before the training are less likely to participate in the follow-up surveys than their peers with lower pre-intervention income. Individuals who took part to the baseline survey or attended more sessions of the training are more likely to also participate in the follow-up.

When testing whether attrition is differential across treatment groups, we find that participation in both the midline and endline surveys is the same between participants from our GME and traditional training groups. 76% and 75% of participants from our GME and traditional training groups responded to our midline survey, and 64% and 65% participated in the endline survey, respectively. However, 70% and 57% of the no-intervention group participants were interviewed in the first and second follow-up surveys, respectively. Overall, the likelihood of replying to any of the two follow-up surveys is 4 percentage points lower in the control than any of the other two treatment arms (statistically significant at the 10%).

Despite this small difference in response rate between the training arms and the control, [Table E4](#) shows that the observable characteristics of participants in the midline or endline are balanced between treatment groups. This is reassuring evidence that, on observables, the samples of respondents look similar across treatment arms. In our PAP reports ([linked in Section 4](#)), to further deal with attrition, we use Lee Bounds to correct for differential attrition between the treatments and the control.

E.5 Measurement and Construction of Outcome Variables

This section provides further details on the construction of the main outcomes, following the structure and definitions specified in our PAP. We comment below on specific cases in which an outcome was included in our pre-analysis, but it is shown among the “secondary” outcomes in the paper.

We collected three families of primary outcomes: mental simulation, economic outcomes and psychological

resilience. We construct our indices within each family following [Kling et al. \(2007\)](#): (i) all variables are first consistently signed (e.g. higher value associated with higher ability or welfare); and (ii) each component of the index is then standardized by subtracting the control group mean and dividing by the control group standard deviation. In the case where there are multiple subscales, we take two additional steps: (iii) the sum of the standardized components is taken and (iv) the sum is standardized again using the control group mean and standard deviation. For the psychological scales, we first sum the individual response items within a scale prior to standardizing the indices. We provide details on the construction of the outcome variables for each family in the following paragraphs.

Mental Simulation

Our overall mental simulation index is constructed from an adapted version of the Spontaneous Use of Imagery Scale ([Reisberg et al., 2003](#)) and the Prospective Imagery Task ([MacLeod et al., 1993](#); [Stöber, 2000](#); [Holmes et al., 2008](#)). Our overall index of mental experiencing quality combines three indices:

- 1 Frequency of use. The scale consists of eight statements designed to assess the propensity of an individual to use mental simulation in either business-related or non-business scenarios. For example, we ask respondents to consider “When I need to go to a meeting, I picture the route in my mind before going”, and state the extent to which they agree on a Likert scale from 1 to 5. We build one index on the frequency of use of mental simulation from this scale.
- 2-3 Specificity and emotionality. The scale asks respondents to imagine three positive and three negative scenarios, either related to their business or not. For example, respondents are asked to imagine a scenario in which “the COVID-19 pandemic is over, and you are struggling to make ends meet” or “the COVID-19 pandemic is over, and your business is doing well”. Respondents are then asked to assess the vividness and emotional intensity associated with each image. We build two indices from this scale: one for specificity, and one for emotionality (both on a scale from 1 to 5).

Details about the questions used in our scales are reported in [Table E5](#).

Economic Outcomes

We distinguish between two time periods: (1) economic activity prior to the Covid-19 induced lockdown (i.e. before 24 March 2020) and (2) economic activity during the lockdown. All the main continuous variables in this family (i.e., income, revenues, savings, hours and funding) are winsorized at the 99th percentile and transformed using an inverse hyperbolic sine transformation.

For the pre-COVID period, we have two main measures of economic outcomes, both based on participants’ recall of the months prior to the national lockdown:

- Business status: this is a dummy equal to one if the person had a business pre-COVID, and 0 otherwise.
- Earnings: this is an index which includes sales and take-home income, conditional on owning a business, or simply take-home income if no business exists at the time of the survey. We ask about income in a typical week in the month prior to the start of the lockdown, and about the best month of sales in the six months prior to the start of the lockdown.

For the period during COVID-19, we construct our outcomes as follows:

- Business status: this is a dummy equal to 1 if the person has a business which is NOT permanently closed, 0 if the person has no business or the business has permanently closed (since March 2020);

- Earnings: this index is constructed as the one for the pre-COVID period, but refers to the sales in the month prior to the survey and the income in the week before;
- Investment: this index measures whether a person acquired a new asset or significantly improve an existing asset for their business, out of a given list of possible investment categories. This index does not clearly distinguish between the pre-COVID and COVID period;
- Savings and Perceived Safety nets: we measure savings in the pre-Covid period and an index of perceptions of informal support networks during the pandemic (perceived safety nets). These perceptions include whether the respondent thinks to have enough savings for the first two months of the lockdown, whether they would be able to easily obtain 200,000 pesos within the next month and whether they have enough cash to cover expenses for the following week.
- COVID19 Mitigation: this index tries to capture how entrepreneurs adapt their business to the pandemic and the associated government restrictions. This index includes two sub-indices. One sub-index is for the behavioral response that people had to COVID-19 pandemic, for instance we ask participants whether they set up a safe work environments, identified alternative supply chains or diversified their products. The other sub-index is for “safe” working hours, defined as the proportion of total hours worked during which social distancing, frequent hand washing, the use of face masks or home working were adhered to.

To reflect our theoretical framework, we include investment, savings, perceived safety nets and COVID19 mitigation among “secondary outcomes” in the paper.

Psychological resilience

We explore whether the treatment has effects on mental well-being during the pandemic by building two indices.

The first index reflects psychological distress, as measured by the Kessler K6 non-specific distress scale (Kessler et al., 2002).

The second index captures psychological resilience, defined as the ability to respond well in the face of adversity. We use three scales for this index: the Brief Resilient Coping Scale (Sinclair and Wallston, 2004), and adapted subset of statements from the Brief Resilience Scale (Smith et al., 2008) and a Self-efficacy scale (Chen et al., 2001).

To reflect our theoretical framework, we include psychological resilience (called “self-efficacy”) among “secondary outcomes” in the paper.

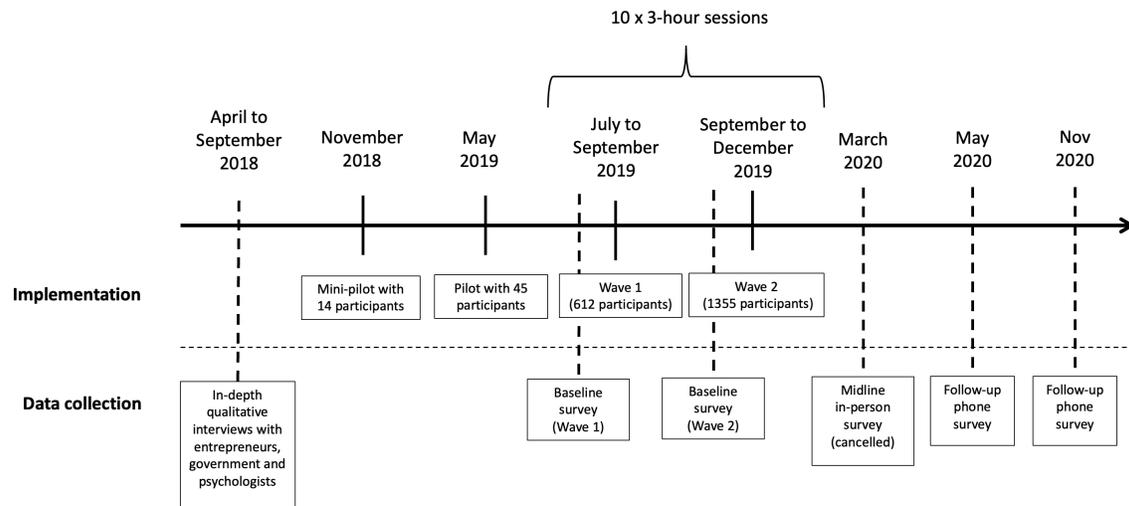
E.6 Cost Effectiveness and Scalability

Table E6 provides more detailed data on program costs, following the J-PAL costing template for cost-effectiveness analysis. Dividing the total cost by actual beneficiaries (who took up at least one session of either the GME or traditional training interventions) results in average costs of USD 532 per beneficiary. We report the costs to deliver the program to the beneficiaries that took up the program. We estimate the marginal costs to deliver the program to an additional beneficiary to be approximately USD 146, though we caution that specific features of the implementation infrastructure may not scale up linearly.

A key aspect of scalability in Colombia is the ability to deliver the program through non-specialized trainers and in a group-based settings. We believe that the program can be scaled up within the programming of governments and development partners. In novel contexts, the effectiveness of GME will likely hinge on carefully adapting the GME exercises to the specific situation faced by the population of interest. This can be done through qualitative work in collaboration with mental health professionals and other relevant experts.

E.7 Figures and Tables

Figure E1. Timeline of Implementation and Data Collection



Notes: This figure illustrates the timeline of data collection and intervention delivery for the Colombia trial. The baseline survey and intervention were conducted in two separate waves in July-September 2019 and September-December 2019. Participants in both waves were invited to the follow-up surveys in 2020.

Table E1. Colombia: Baseline Balance in Analysis Sample

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Means			Pairwise differences		
	GME	Traditional	Control	GME-Trad	GME-C	Trad-C
<i>Demographics</i>						
Female	0.574 (0.495)	0.579 (0.494)	0.571 (0.496)	-0.006 (0.847)	0.003 (0.934)	0.008 (0.809)
Age (18-28)	0.591 (0.492)	0.573 (0.495)	0.629 (0.484)	0.019 (0.522)	-0.038 (0.201)	-0.056* (0.090)
Age (29-45)	0.245 (0.431)	0.249 (0.433)	0.272 (0.446)	-0.004 (0.869)	-0.027 (0.312)	-0.023 (0.443)
Age (46-59)	0.096 (0.294)	0.119 (0.325)	0.094 (0.292)	-0.024 (0.197)	0.002 (0.927)	0.025 (0.224)
Years of Education	13.157 (3.315)	13.035 (3.590)	13.366 (3.135)	0.122 (0.587)	-0.209 (0.333)	-0.331 (0.189)
<i>Business and assets</i>						
Only Business Owner	0.245 (0.431)	0.245 (0.431)	0.236 (0.425)	0.000 (0.995)	0.009 (0.724)	0.009 (0.757)
Only Have a Business Idea	0.477 (0.500)	0.477 (0.500)	0.494 (0.501)	-0.000 (0.987)	-0.017 (0.569)	-0.017 (0.621)
Have Business and Business Idea	0.278 (0.448)	0.275 (0.447)	0.270 (0.444)	0.002 (0.924)	0.008 (0.764)	0.006 (0.853)
Income \leq min wage	0.665 (0.472)	0.677 (0.468)	0.641 (0.480)	-0.011 (0.679)	0.024 (0.398)	0.036 (0.265)
Income $>$ min wage	0.238 (0.426)	0.221 (0.416)	0.248 (0.432)	0.016 (0.502)	-0.010 (0.688)	-0.027 (0.349)
Assets Owned	11.661 (3.839)	11.894 (3.962)	11.835 (3.782)	-0.233 (0.354)	-0.175 (0.498)	0.058 (0.841)
Household Size	3.452 (1.506)	3.609 (1.607)	3.650 (1.713)	-0.157 (0.120)	-0.197* (0.077)	-0.041 (0.745)
Save Monthly (Y/N)	0.549 (0.498)	0.535 (0.499)	0.580 (0.494)	0.014 (0.661)	-0.031 (0.359)	-0.045 (0.230)
Had Access to Credit (Y/N)	0.153 (0.360)	0.150 (0.358)	0.138 (0.345)	0.003 (0.909)	0.015 (0.531)	0.012 (0.644)
<i>Trauma and mental health</i>						
Number of Traumatic Events	2.247 (2.330)	2.342 (2.367)	2.308 (2.539)	-0.090 (0.513)	-0.056 (0.710)	0.040 (0.812)
High Trauma Symptoms (IES $>$ 33)	0.242 (0.429)	0.228 (0.420)	0.231 (0.422)	0.014 (0.609)	0.010 (0.718)	-0.004 (0.912)
Trauma Symptoms (IES)	17.645 (22.273)	17.142 (22.005)	17.785 (20.958)	0.503 (0.724)	-0.140 (0.923)	-0.643 (0.691)
Distress Score (Kessler)	13.098 (4.102)	13.187 (3.980)	13.138 (3.970)	-0.090 (0.728)	-0.040 (0.882)	0.050 (0.868)
Observations	795	461	415	1,257	1,211	877

Notes: The table presents a balance test at baseline for all three treatment arms, considering only the sample of people who accepted to participate in the training (i.e. our analysis sample as specified in our PAP). Columns 1, 2, and 3 show the mean and standard errors for the GME treatment, traditional treatment, and control groups, respectively. The following 3 columns show the differences between treatment arms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. “Income \leq min wage” indicates that the respondent was earning less on a monthly basis than the Colombian minimum wage at the time of eligibility screening for the project. “Assets Owned” is a variable which counts the number of assets that the person owns (out of a given list). The number of traumatic events is calculated using a contextually-relevant trauma history checklist. Trauma symptoms are measured using the Impact of Event Score Revised scale. The “Distress score” is a measure of anxiety, depression and general distress computed using the Kessler K-6 scale.

Table E2. Colombia: Intervention Take-up

Variable	(1) No Take-up	(2) Take-up	(3) Difference
<i>Demographics</i>			
Female	0.611 (0.488)	0.553 (0.497)	-0.058** (0.028)
Age (18-28)	0.623 (0.485)	0.565 (0.496)	-0.058** (0.027)
Age (29-45)	0.255 (0.436)	0.236 (0.425)	-0.020 (0.397)
Age (46-59)	0.070 (0.255)	0.124 (0.330)	0.054*** (0.001)
Years of Education	13.019 (3.656)	13.111 (3.362)	0.091 (0.675)
<i>Business and assets</i>			
Only Business Owner	0.221 (0.415)	0.261 (0.439)	0.039* (0.084)
Only Have a Business Idea	0.522 (0.500)	0.464 (0.499)	-0.058** (0.030)
Have Business and Business Idea	0.257 (0.437)	0.274 (0.446)	0.017 (0.465)
Income \leq min wage	0.650 (0.477)	0.677 (0.468)	0.027 (0.286)
Income $>$ min wage	0.244 (0.430)	0.233 (0.423)	-0.011 (0.641)
Assets Owned	11.654 (3.918)	11.759 (3.863)	0.105 (0.661)
Household Size	3.600 (1.659)	3.486 (1.518)	-0.114 (0.250)
Save Monthly (Y/N)	0.523 (0.500)	0.551 (0.498)	0.028 (0.363)
Had Access to Credit (Y/N)	0.146 (0.354)	0.154 (0.361)	0.008 (0.723)
<i>Trauma and Mental Health</i>			
Number of Traumatic Events	1.875 (2.390)	2.421 (2.274)	0.546*** (0.000)
High Trauma Symptoms (IES $>$ 33)	0.275 (0.447)	0.217 (0.412)	-0.059** (0.029)
Trauma Symptoms (IES)	20.269 (24.340)	16.296 (21.157)	-3.974*** (0.006)
Distress Score (Kessler)	13.776 (4.607)	12.834 (3.797)	-0.941*** (0.000)
Observations	615	806	1,854

Notes: The table presents characteristics of participants that attended at least one session of the intervention (Column 2) and those that did not (Column 1). Columns 1 and 2 show the mean and standard errors for each group. Column 3 shows the difference between both groups, with significance indicated as follows: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variables are defined in the note to table E1.

Table E3. Colombia: Survey Attrition at Midline and Endline

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Participated in Midline			Participated in Endline		
	No	Yes	Diff	No	Yes	Diff
<i>Treatments</i>						
GME Treatment	0.446 (0.498)	0.485 (0.500)	0.039 (0.177)	0.457 (0.499)	0.486 (0.500)	0.029 (0.257)
Traditional Treatment	0.273 (0.446)	0.277 (0.447)	0.003 (0.897)	0.263 (0.441)	0.283 (0.450)	0.019 (0.394)
Control	0.278 (0.449)	0.239 (0.427)	-0.039 (0.122)	0.278 (0.449)	0.232 (0.422)	-0.047** (0.038)
Baseline Participation	0.672 (0.470)	0.866 (0.341)	0.194*** (0.000)	0.716 (0.451)	0.875 (0.330)	0.159*** (0.000)
Total Attendance	2.407 (3.241)	3.756 (3.699)	1.350*** (0.000)	2.577 (3.250)	3.892 (3.752)	1.315*** (0.000)
Wave 1 Participant	0.359 (0.480)	0.317 (0.466)	-0.042 (0.126)	0.318 (0.466)	0.332 (0.471)	0.014 (0.549)
<i>Demographics</i>						
Female	0.585 (0.493)	0.571 (0.495)	-0.014 (0.613)	0.558 (0.497)	0.584 (0.493)	0.026 (0.311)
Age (18-28)	0.593 (0.492)	0.596 (0.491)	0.003 (0.908)	0.619 (0.486)	0.583 (0.493)	-0.036 (0.146)
Age (29-45)	0.276 (0.448)	0.246 (0.431)	-0.031 (0.231)	0.241 (0.428)	0.259 (0.439)	0.018 (0.417)
Age (46-59)	0.058 (0.234)	0.115 (0.320)	0.058*** (0.000)	0.082 (0.274)	0.113 (0.316)	0.031** (0.037)
Years of Education	13.794 (3.390)	13.021 (3.329)	-0.772*** (0.001)	13.483 (3.338)	13.034 (3.353)	-0.449** (0.022)
<i>Business and assets</i>						
Only Business Owner	0.198 (0.399)	0.257 (0.437)	0.058** (0.013)	0.221 (0.415)	0.255 (0.436)	0.034 (0.119)
Only Have a Business Idea	0.485 (0.500)	0.480 (0.500)	-0.005 (0.863)	0.474 (0.500)	0.485 (0.500)	0.010 (0.688)
Have Business and Business Idea	0.317 (0.466)	0.262 (0.440)	-0.054** (0.040)	0.304 (0.461)	0.259 (0.439)	-0.045* (0.053)
Income \leq min wage	0.580 (0.494)	0.688 (0.463)	0.108*** (0.000)	0.616 (0.487)	0.688 (0.464)	0.072*** (0.003)
Income $>$ min wage	0.289 (0.454)	0.219 (0.414)	-0.070*** (0.006)	0.270 (0.445)	0.217 (0.412)	-0.053** (0.016)
Assets Owned	12.328 (4.005)	11.632 (3.813)	-0.697*** (0.010)	12.209 (4.021)	11.572 (3.771)	-0.637*** (0.006)
Household Size	3.278 (1.544)	3.608 (1.592)	0.330*** (0.002)	3.431 (1.606)	3.593 (1.577)	0.162* (0.083)
Save Monthly (Y/N)	0.539 (0.499)	0.555 (0.497)	0.016 (0.635)	0.563 (0.497)	0.548 (0.498)	-0.015 (0.597)
Had Access to Credit (Y/N)	0.161 (0.368)	0.146 (0.353)	-0.015 (0.535)	0.138 (0.345)	0.153 (0.360)	0.015 (0.460)
<i>Trauma and mental health</i>						
Number of Traumatic Events	1.559 (2.016)	2.515 (2.455)	0.956*** (0.000)	1.826 (2.184)	2.538 (2.463)	0.712*** (0.000)
High Trauma Symptoms (IES $>$ 33)	0.207 (0.406)	0.242 (0.429)	0.036 (0.203)	0.225 (0.418)	0.240 (0.427)	0.014 (0.558)
Trauma Symptoms (IES)	15.369 (21.283)	18.063 (21.994)	2.694* (0.066)	16.911 (21.663)	17.812 (21.975)	0.901 (0.481)
Distress Score (Kessler)	13.246 (4.130)	13.105 (4.012)	-0.141 (0.615)	13.249 (4.056)	13.081 (4.025)	-0.169 (0.476)
Observations	399	1,273	1,672	589	1,083	1,672

Notes: The table shows observable differences between people who participated or not in the midline survey (Columns 1, 2 and 3) and endline survey (Columns 4, 5 and 6). Columns 1, 2 and 4, 5 show means and standard errors, while Columns 3 and 6 show differences in means and their standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Variables are defined in the note to table E1.

Table E4. Colombia: Differences across Treatments in the Baseline Characteristics of Respondents at Midline and Endline

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Midline differences			Endline differences		
	GME-Trad	GME-PC	Trad-PC	GME-Trad	GME-PC	Trad-PC
<i>Demographics</i>						
Female	-0.044 (0.187)	-0.013 (0.705)	0.030 (0.432)	-0.038 (0.283)	-0.027 (0.474)	0.011 (0.794)
Age (18-28)	0.031 (0.352)	-0.001 (0.983)	-0.031 (0.416)	0.061* (0.088)	-0.013 (0.737)	-0.074* (0.080)
Age (29-45)	-0.026 (0.366)	-0.049 (0.108)	-0.023 (0.498)	-0.018 (0.564)	-0.047 (0.166)	-0.029 (0.442)
Age (46-59)	-0.033 (0.139)	-0.011 (0.605)	0.021 (0.413)	-0.050** (0.042)	0.009 (0.685)	0.059** (0.033)
Years of Education	-0.129 (0.600)	-0.184 (0.449)	-0.055 (0.845)	0.003 (0.992)	-0.235 (0.384)	-0.237 (0.444)
<i>Business and assets</i>						
Only Business Owner	-0.003 (0.912)	0.021 (0.492)	0.024 (0.479)	-0.003 (0.927)	0.020 (0.555)	0.022 (0.544)
Only Have a Business Idea	0.012 (0.715)	-0.022 (0.533)	-0.034 (0.384)	-0.006 (0.869)	-0.009 (0.814)	-0.003 (0.942)
Have Business and Business Idea	-0.006 (0.836)	0.001 (0.972)	0.007 (0.835)	0.012 (0.699)	-0.010 (0.758)	-0.023 (0.547)
Income ≤ min wage	-0.010 (0.744)	0.014 (0.659)	0.025 (0.501)	0.038 (0.257)	0.012 (0.729)	-0.026 (0.517)
Income > min wage	0.018 (0.514)	0.005 (0.867)	-0.013 (0.686)	-0.019 (0.525)	0.006 (0.854)	0.025 (0.481)
Assets Owned	-0.333 (0.227)	-0.217 (0.448)	0.117 (0.717)	-0.204 (0.488)	-0.156 (0.609)	0.048 (0.888)
Household size	-0.132 (0.233)	-0.232* (0.070)	-0.100 (0.483)	-0.111 (0.345)	-0.199 (0.153)	-0.088 (0.564)
Save Monthly (Y/N)	0.017 (0.632)	0.002 (0.955)	-0.015 (0.723)	0.022 (0.569)	-0.012 (0.778)	-0.033 (0.463)
Had Access to Credit (Y/N)	0.003 (0.893)	0.013 (0.628)	0.009 (0.750)	-0.011 (0.682)	0.001 (0.960)	0.013 (0.696)
<i>Trauma and mental health</i>						
Number of Traumatic Events	-0.104 (0.522)	-0.084 (0.637)	0.020 (0.919)	-0.095 (0.584)	-0.112 (0.567)	-0.017 (0.938)
High Trauma Symptoms (IES>33)	0.012 (0.706)	0.012 (0.705)	0.001 (0.982)	0.012 (0.705)	0.002 (0.953)	-0.010 (0.792)
Trauma Symptoms (IES)	0.046 (0.977)	-0.403 (0.807)	-0.448 (0.809)	0.109 (0.949)	-0.561 (0.753)	-0.670 (0.734)
Distress Score (Kessler)	-0.271 (0.348)	-0.120 (0.692)	0.150 (0.659)	-0.184 (0.544)	-0.085 (0.801)	0.099 (0.788)
Observations	969	921	656	832	777	557

Notes: The table shows pairwise differences in means for baseline variables between each treatment group at midline (Columns 1, 2 and 3) and endline (Columns 4, 5 and 6). * p<0.10, ** p<0.05, *** p<0.01. Variables are defined in the note to table E1.

Table E5. Measurement: Mental Simulation

Mental Experiencing Questions	
Indexes	Question(s)
Frequency of use	<p>“On a scale of 1 to 5, how much do you agree with this statement?”, where numbers 1 to 5 meant respectively: Strongly disagree; Disagree; Neither agree nor disagree; Agree; Strongly agree. The statements used were as follows:</p> <ul style="list-style-type: none"> • When I need to go to a meeting, I picture the route in my mind before going. • When I think about a customer using my product or service, I imagine the customer’s experience through pictures and sensations in my mind. • When I think about the day ahead, I create mental pictures of all the tasks I must do. • When I am faced with difficult situations, I mentally experience the actions I could take and the consequences of those actions before reacting. • When I think about the type of business I want to have, I live the experience of running that business in my mind. • When I feel overwhelmed, I find a mental place or time where I feel safe and calm. • When someone is upset with me, I live that person’s experience in my mind to understand what might have caused the situation. • When I buy an asset for my business, an image of owning the asset pops up in my mind before buying it.
Emotionality and Specificity	<p>Now I am going to ask you to imagine a couple of scenarios. When you imagine each one of them, please close your eyes, and let me know when you are done imagining it. Afterwards, I will ask you questions about the image.</p> <p>Question on emotionality. For each of the following scenarios, “What is the intensity of the emotion produced in you by this image?. Use a scale from 1 to 5 where 1 means “no emotion at all”, 2 “little, but weak emotions”, 3 “moderate emotions”, 4 “strong emotions” and 5 “extremely strong emotions”.”</p> <p>Question on specificity. For each of the following scenarios, “Using a scale for the mental image where 1 means “no image at all”, 2 means “vague and dim”, 3 means “moderately clear and vivid”, 4 means “reasonably clear and vivid” and 5 means “perfectly clear and vivid”, how detailed is this image from 1 to 5?”</p> <ul style="list-style-type: none"> • I want you to imagine that the COVID-19 pandemic is over, and you save enough money to buy an asset you really want. • I want you to imagine the COVID-19 pandemic is over and you spend quality time with your family and friends. • I want you to imagine that the COVID-19 pandemic is over, and your business is doing well. • I want you to imagine that the COVID-19 pandemic is over, and you are struggling to make ends meet. • I want you to imagine that the COVID-19 pandemic is over, and you have had a serious disagreement with someone close to you. • I want you to imagine that the COVID-19 pandemic is over and your business closes.

Table E6. Colombia: Detailed Cost Estimates

Category	Costs (USD)	Comments
Program administration and staff costs	295,615	Project steering and coordination. Staff time of project manager, supervisors and trainers.
Targeting costs	15,633	Community outreach campaign and application screening
Staff training	7,036	Training venues and materials; back-up trainers
Participant training	46,200	Training venue hire and materials
Implementation and program material costs	50,653	Production of curriculum materials and implementation of training, incl. graduation ceremony and fair
User costs	13,291	Transport subsidy
Averted costs	none	
Monitoring costs	none	
Total program costs	428,429	

Notes: The table presents a detailed cost estimates following the J-PAL costing template for cost effectiveness analysis. We assume a COP/USD exchange rate of 3403 and 2019 costs are adjusted to 2022 USD using a CPI-based inflation adjustment factor of 1.169, derived from the World Bank WDI database. See www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea for more details on the template.

Supplementary References

- Betts, A., L. Fryszler, O. Sterck, and N. Omata (2019). Refugee Economies in Addis Ababa – Towards Sustainable Opportunities for Urban Communities. Refugee Studies Centre Report, Oxford Department of International Development, Oxford, UK.
- Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.
- Central Statistics Agency of Ethiopia (2020). Ethiopia Socioeconomic Survey (ESS) / Living Standards Measurement Study (LSMS) 2018-2019 Survey Report. Technical report, Central Statistics Agency of Ethiopia and World Bank, Addis Ababa.
- Chen, G., S. M. Gully, and D. Eden (2001). Validation of a new general self-efficacy scale. *Organizational research methods* 4(1), 62–83.
- Girma, E., A. Ametaj, M. Alemayehu, B. Milkias, M. Yared, S. Misra, A. Stevenson, K. C. Koenen, B. Gelaye, and S. Teferra (2022). Measuring traumatic experiences in a sample of Ethiopian adults: Psychometric properties of the life events checklist-5. *European Journal of Trauma & Dissociation* 6(4), 100298.
- Griffith, J. W., J. A. Sumner, E. Debeer, F. Raes, D. Hermans, S. Mineka, R. E. Zinbarg, and M. G. Craske (2009). An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. *Memory* 17(6), 609–623.
- Holmes, E. A., T. J. Lang, M. L. Moulds, and A. M. Steele (2008). Prospective and positive mental imagery deficits in dysphoria. *Behaviour Research and Therapy* 46(8), 976–981.
- IDMC (2017). GRID 2017: Global Report on Internal Displacement. Technical report, Internal Displacement Monitoring Centre (IDMC), Geneva, Switzerland.
- Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. Normand, E. E. Walters, and A. M. Zaslavsky (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological medicine* 32(6), 959–976.
- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica* 75(1), 83–119.
- Koenen, K. C., A. Ratanatharathorn, L. Ng, K. A. McLaughlin, E. J. Bromet, D. J. Stein, E. G. Karam, A. Meron Ruscio, C. Benjet, K. Scott, and et al. (2017). Posttraumatic stress disorder in the World Mental Health Surveys. *Psychological Medicine* 47(13), 2260–2274.
- MacLeod, A. K., G. S. Rose, and J. M. G. Williams (1993). Components of hopelessness about the future in parasuicide. *Cognitive Therapy and Research* 17(5), 441–455.
- McKenzie, D. and C. Woodruff (2014). What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer* 29(1), 48–82.
- McNally, R. J., N. B. Lasko, M. L. Macklin, and R. K. Pitman (1995). Autobiographical memory disturbance in combat-related posttraumatic stress disorder. *Behaviour research and therapy* 33(6), 619–630.
- OECD (2018). *SMEs in Public Procurement : Practices and Strategies for Shared Benefits*. Éditions OCDE, Paris.
- R4V (2023). R4V Movements Report: Fourth Quarter 2022. Technical report, Interagency Coordination Platform for Refugees and Migrants (R4V).

- Reisberg, D., D. Pearson, and S. Kosslyn (2003). Intuitions and introspections about imagery: The role of imagery experience in shaping an investigator's theoretical views. *Applied Cognitive Psychology* 17, 147–160.
- Ridley, M., G. Rao, F. Schilbach, and V. Patel (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science* 370(6522).
- Sinclair, V. G. and K. A. Wallston (2004). The development and psychometric evaluation of the Brief Resilient Coping Scale. *Assessment* 11(1), 94–101.
- Smith, B. W., J. Dalen, K. Wiggins, E. Tooley, P. Christopher, and J. Bernard (2008). The brief resilience scale: assessing the ability to bounce back. *International journal of behavioral medicine* 15, 194–200.
- Stöber, J. (2000). Prospective cognitions in anxiety and depression: Replication and methodological extension. *Cognition & Emotion* 14(5), 725–729.
- UNHCR (2021). Ethiopia Country Refugee Response Plan. Country Refugee Response Plan 2020 to 2021, United Nations High Commissioner for Refugees (UNHCR), Geneva.
- UNHCR (2022). Ethiopia Country Refugee Response Plan. Country Refugee Response Plan January 2022 to December 2022, United Nations High Commissioner for Refugees (UNHCR), Geneva.
- Woldetsadik, T. K., F. Mulatu, and J. Edosa (2019). Ethiopia's Refugee Policy Overhaul: Implications on the Out of Camp Regime and Rights to Residence, Movement and Engagement in Gainful Employment. *Journal of Ethiopian Human Rights Law* 4(1), 124–186.