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††

December 19, 2024

Abstract

Many of the world's poor have experienced trauma. We argue that memories of this trauma interfere with the process of future simulation, diminishing the ability to see how actions today can improve outcomes tomorrow. We introduce guided mental experiencing (GME) – an intervention in which participants mentally simulate pathways between their actions and desired economic outcomes – as a response, and study GME's impact in two RCTs. In a population of Eritrean refugees in Ethiopia, GME increases the ability of refugees to see a positive future, increases their intent to stay in Ethiopia, increases labor force participation and improves self-reported welfare. In a population that has experienced violence and poverty in Colombia, a traditional entrepreneurial training program reduces the ability to imagine a future in business and worsens economic outcomes. Integrating GME into entrepreneurial training restores future thinking and removes these negative economic effects. The largest gains accrue to the most traumatized participants in our samples.

*We thank Stefan Dercon, Xavier Gabaix, Karla Hoff, Supreet Kaur, Asim Khwaja, Eliana La Ferrara, David Laibson, 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, 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
**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

Traumatic experiences are common, and they disproportionately affect the poor. More than seventy percent of the world's population has experienced a trauma, and by 2030 more than sixty percent of the world's poor are expected to live in fragile states, characterized by conflict, violence, and the trauma that inevitably follows (Kessler et al., 2017). Displacement is particularly traumatic, and there are currently 110 million forcibly displaced people worldwide, 75 percent of whom live in low- and middle-income countries. Working to alleviate poverty increasingly means working with these trauma-affected populations, in which anxiety, depression, and post-traumatic stress disorder (PTSD) are, unsurprisingly, more prevalent (Pozuelo et al., 2023). Evidence from psychology suggests that trauma affects one's ability to imagine the future, plan, and learn new skills (Holmes et al., 2017). Yet prominent anti-poverty programs, such as business training and cash transfers, rely heavily on the very cognitive capacities which are impaired by traumatic experiences.

That investing in the future imposes a significant cognitive cost has long been recognized by economists. As Böhm-Bawerk wrote in 1889, "Provision for the future makes no inconsiderable demands on our intellectual strength... the future we must anticipate and picture. 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".[1] A long tradition in neuroscience, psychiatry and psychology has studied these "double anticipations" and their relationship with motivation. This literature shows that imagining the future draws on past experience, uncovering a strong link between memory and future simulation (Schacter et al., 2007). Recent economics literature has modeled the impact of memory on simulation (Bordalo et al., 2020; Malmendier and Wachter, 2021; Bordalo et al., 2024), but this literature has not studied the specific role of trauma, which psychologists believe significantly distorts the simulation process, and has not studied whether the negative impacts of past experience on mental experiencing and choice can be ameliorated.[2]

We study the interaction between trauma and memory, drawing out the implications for anti-poverty programs. We argue that trauma interferes with mental simulation, reducing a survivor's ability to see the world's opportunities. Failure to account for this distortion reduces the effectiveness of traditional programs, and may even lead them to cause harm. We design and test a new intervention that aims to address the memory distortions created by trauma and restore the ability to see a (positive) future. Guided Mental Experiencing (GME) trains participants to create projections into the future with specificity and emotions, and to identify pathways for achieving

[1] The Positive Theory of Capital (1889), pp. 244 (Böhm-Bawerk, 1889)

[2] By 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., 2015; Holmes et al., 2016). This process is also known in the psychology literature as mental imagery, visualization, mental simulation, mentalization or imagination.

outcomes in a safe space.[3] We study the impact of GME in two RCTs conducted with different vulnerable populations: Eritrean refugees in Addis Ababa, Ethiopia; and post-conflict and displaced persons in Bogotá, Colombia. We tailor a GME program for each of our study populations and show that, as hypothesized, GME improves the quality of memory recall and future simulation – quantified using novel measurement tools that we develop. GME also leads to more forward-looking behavior and improved economic outcomes relative to traditional programs, which are either underutilized or have negative impacts in our populations.

To explain how trauma interferes with economic decision making and to motivate our GME intervention, we provide a conceptual framework that draws on the neuroscience and psychology literature on memory, trauma and mental experiencing (Kahana 2012; Cohen and Kahana 2022; Holmes et al. 2020; Bordalo et al. 2024).[4] The literature describes how memory recall depends on context, which includes both internal (e.g., emotions) and external states (e.g., time and place). Experiences are encoded in memory along with their context (forming an "experience bank"), and are more likely to be recalled when the retrieval context is similar to the encoding context. Emotion at the time of experiencing is a key part of context.

Trauma impacts the experience bank in two ways. First, trauma memories carry very negative emotions. Because memory recall is stochastic, these negative memories may be spontaneously recalled, even when they are not relevant to choice, triggering negative flashbacks. Second, the negative emotions from trauma strongly affect context, and hence any memories created while mood remains low will encode part of the negative experience (Talmi et al. 2019). Mood congruent recall means this process continues for some time: negative mood leads to recall of negative memories, which maintains negative affect, which passes to more memories in the memory bank (Cohen and Kahana, 2022). Hence trauma creates two types of new memories: direct trauma memories and memories with negatively distorted emotional valence. If enough of these memories are created, then trauma has two additional impacts: avoidance, where trauma survivors avoid recalling memories and simulating the future so as not to encounter the pain of negative recall; and distortion, where negative memories interfere with any simulations the survivor makes, crowding out more relevant domains specific memories and negatively affecting anticipations of the future (Bordalo et al., 2024). These predictions account for three key diagnostic criteria for PTSD: intrusive memories, avoidance and distorted cognition about the self and future (American Psychiatric Association et al., 2013).

This account of the impact of trauma has important implications for the design of poverty alleviation programs. Work permits are often advocated as removing a constraint that causes poverty among refugees. However, if the targeted workers cannot imagine a positive outcome from

[3]Our intervention is distinct from either goal setting or planning, in that it specifically attempts to create multisensorial experiences, as well as from therapies that directly address mental health challenges and symptoms.

[4]We concentrate on the role of memory in mediating the impact of trauma because it is thought to be a primary pathway, but it is not the only theorized mechanism through which trauma can affect economic decision-making (Brewin and Holmes, 2003).

job search, they will not take advantage of the opportunity. A similar argument applies to any program that provides access to an opportunity. Business training programs, which help self-employed participants run small businesses, also require significant forward thinking; they may not work if they do not directly address trauma's constraints. Indeed, our model of trauma suggests that a training program may make matters worse. Training usually involves taking an entrepreneur through the steps of running a business, and this runs the risk that these steps all become distorted by the negative emotions from participants' trauma, building even greater barriers to their success. This account suggests that trauma's negative economic impacts come from negative emotional coding of the tasks required to get ahead.[5] Our intervention - GME - is a natural response to this problem.

GME encourages individuals to imagine a set of vivid and realistic future experiences, while learning how to tolerate the emergence of negative emotions raised by distorted memories. As more realistic experiences are simulated and encoded in memory, negatively distorted memories are less likely to be recalled, and motivation is restored. GME draws on the literature on mental experiencing (or visualization/imagery) which shows that, relative to other approaches like Cognitive Behavioral Therapy (CBT), mental experiencing is able to create more vivid and emotional simulation (Holmes and Mathews, 2005). Vivid and emotional mental experiences of the tasks required for an action facilitate their memory consolidation, encourage learning, and have also been shown to be more motivational when used as the basis for simulating the future (Rösch et al., 2022).

We test the efficacy of GME in two randomized controlled trials, and argue the results are consistent with the theory discussed. In our first trial, we worked with a random sample of 1,652 Eritrean refugees in Addis Ababa, Ethiopia. These refugees were nominally entitled to receive a work permit (a grant of access) but had not received it at the time of our program. Government officials raised the possibility that a permit may not be helpful because, despite the fact that they could currently work informally, few choose to do so, suggesting the lack of permit was not a key constraint. We collaborated with the government and a local psychologist to design a GME program that teaches this population to build positive mental experiences about their current home (Addis Ababa), to envision a safe mental space to manage negative mental experiences that crop up, and to mentally experience concrete steps towards finding economic opportunities in Addis. We designed the program to be scalable, and delivered it with help of trainers from the same refugee community.

Two months after the intervention, we see a significant increase in the quality of mental experiencing, as measured by the vividness and emotional intensity of memories and imagined future scenarios. This is consistent with GME induced memories crowing out recall of negatively distorted memories, reducing avoidance. This change in simulation translates into impacts on targeted behaviors. Relative to a control group, recipients of GME are more likely to report that

[5] A large empirical literature documents negative impacts of trauma on motivation (Holmes and Mathews, 2005, 2010; Holmes et al., 2006, 2008; Mathews et al., 2013; Renner et al., 2019).

they intend to stay in Addis Ababa, to be employed and to work more hours. Their cognitions are less distorted. The effect on employment is large: the likelihood of being employed among refugees assigned to the GME program increases by 5 percentage points over a control group mean of 21 percent. Treated participants also significantly increase their working hours by about three hours per week, relative to a control group mean of about 8 hours per week. Importantly, this increase in hours worked is driven by hours from wage work as opposed to casual work or unpaid apprenticeships. We also see an improvement in self-reported food security and wellbeing.[6]

In our second trial, we work with a sample of 1,967 aspiring entrepreneurs in Bogotá, Colombia in collaboration with the Bogotá City Council. The sample was identified by two local departments responsible for re-skilling groups that often struggle to engage with and benefit from traditional training programs. A large share of our sample consists of individuals who were victims of violence during the Colombian internal armed conflict and migrants who recently fled the Venezuelan crisis. They represent a second important population with respect to our hypotheses, as the government reported that attempts to improve economic outcomes for these groups in Colombia had shown limited success. At the same time, entrepreneurial livelihoods are the most likely source of income for this population. For this trial, we designed and delivered two versions of an entrepreneurial training program. The first takes a more traditional form, by delivering skills thought to be important for business success.[7] The second combined these basic skills with GME exercises designed to create a set of new realistic imaginings across a range of business-focused actions. A key innovation of this trial is an attempt to keep the two programs comparable: the traditional program taught a lesson and then completed it with a pedagogical device to increase retention and understanding, for example think-pair-share. The GME-enhanced program substituted out the pedagogical device with a GME exercise on the topic of the lesson. The trial also includes a control group that received nothing.

We show several results consistent with our theory. First, we find that the traditional training program reduced earnings among our sample compared to the control, consistent with our concerns about harmful effects. Second, we document a key mechanism for the loss in earnings: those in the traditional training program see a reduction in the quality of their mental experiencing related to business activities, consistent with our hypothesis that traditional training may spread negative emotional context to more business activities, leading to avoidance of specific thinking about business actions. Third, we show that the GME training program removes these effects: we see no reduction in the quality of mental experiencing and no reduction in earnings compared to those who received no intervention. Notably, these impacts are stronger for participants who score

[6]We argue that these results are unlikely to be driven by social desirability bias as we see no interaction of the GME treatment with an endline measure of the Marlowe-Crowne scale (Crowne and Marlowe, 1960; Dhar et al., 2022). We also see a decrease in hours spent searching for work, which this is consistent with a reduction due to becoming more effective in finding employment.

[7]The curriculum draws on widely used business training programs, such as the International Labour Organization's Start and Improve Your Business or the International Finance Corporation's Business Edge programs.

higher on a baseline measure of trauma experience. Consistent with the empirical evidence showing that women are more vulnerable to emotional tagging, results are also more pronounced for women.

While our results from Colombia align with our theory, an important caveat remains. The GME treatment overcame the problems with the traditional treatment, but it did not lead to positive outcomes on average. This is, perhaps, not surprising, given the modest evidence on business training programs and the relevance of selection effects in assessing their impacts (McKenzie, 2021; McKenzie et al., 2023; McKenzie and Woodruff, 2014). Since training programs often target mixed populations with varying trauma histories, negative effects on trauma-affected individuals may be masked by positive outcomes for others. Our GME intervention has the potential to eliminate these negative effects while preserving benefits for less-affected individuals

Our paper contributes to three key areas of literature. First, it highlights the role of vividly imagining future scenarios, linking this ability to inter-temporal decision-making and belief formation (Becker and Mulligan, 1997; Gabaix and Laibson, 2017; Bordalo et al., 2024). Empirically, Alan and Ertac (2018) demonstrate that teaching students to envision their future selves improves patience and performance, while John and Orkin (2021) find that visualizing water chlorination yields significant behavioral changes. We also relate to work on imperfect recall and its impact on economic forecasting and beliefs (Taubinsky et al., 2024; Cenzon, 2023). Second, our findings contribute to research on the economic consequences of conflict (Voors et al., 2012; Callen et al., 2014; Moya, 2018), offering perspective inspired by theories of memory and recall (Mullainathan, 2002; Gennaioli and Shleifer, 2010; Bordalo et al., 2016, 2024; Malmendier and Wachter, 2021; Kahana, 2020). Lastly, we design and test a trauma-informed training tool—GME—targeting job search and entrepreneurship. Recent evidence highlights the value of tailoring interventions to individual needs (Caria et al., 2024; Carranza and McKenzie, 2024), and our approach emphasizes the importance of incorporating trauma-awareness to enhance impact in fragile settings.

The paper proceeds as follows: Section 2 provides the conceptual framework of economic decision-making in the face of trauma and clarifies how guided mental experiencing can address the impacts of trauma on decision-making. Section 3 provides a mathematical version of the same argument building on Bordalo et al. (2024). Section 4 describes our approach in designing trauma-informed programs in the two settings of our randomized controlled trials. Section 5 describes the intervention and evaluation design of the two experiments, as well as the measurement of mental experiencing quality and downstream economic outcomes. Section 6 describes the results. Section 7 provides a discussion of these results, and Section 8 concludes.

2 Memory, trauma and economic decisions

We provide a memory-based explanation for the impact of trauma and describe why GME is predicted to help. Our basic argument is simple. Trauma causes a memory distortion whereby a large set of memories from a broad range of settings and contexts become associated with strong negative emotions. These memories interfere with the recall of more relevant non-distorted memories. Because we simulate the future by remembering the past, a trauma victim will need to draw on this set of negatively tagged memories to think about the future, leading to two primary effects. First, trauma survivors try to avoid thinking about the future, and second when they do think about the future they do so using negative memories thus making the future seem negative. We summarize this by saying that trauma survivors cannot see the world's opportunities. We also explore the implications of this model of trauma for development programming. We argue that many programs will be less effective for those who have experienced trauma, and may even backfire, but that GME can be combined with other programs to improve outcomes.

We make our argument in two ways. The remainder of this section provides a verbal discussion, and relates the argument to the existing literature in psychology and neuroscience. Section 3 provides a more formal mathematical model that builds on Bordalo et al. (2024). The sections are to some extent exchangeable and could be read in any order.

2.1 The Psychological Impacts of Trauma

Following the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association et al., 2013), we define a traumatic event as direct or witnessed exposure to actual or threatened death, serious injury, or sexual violence. Trauma can also occur indirectly, such as learning that a violent or accidental event happened to a close friend or family member, or through extreme exposure to the disturbing details of such events.

Traumatic events and their symptoms should be conceptualized as being on a continuum, with Post-Traumatic Stress Disorder (PTSD) diagnosed on the severe end of the spectrum. However, many individuals may experience trauma-related symptoms that affect their daily lives, even if they do not meet the full diagnostic criteria for PTSD (Insel et al., 2010). Exposure to trauma typically leads to four primary psychological impacts:

(a) Re-experiencing and interference: such as via intrusive sensory-based memories of the traumatic event(s) and flashbacks

(b) Avoidance: such as attempts to push away memories of the trauma and avoid their recall though avoidance of internal (e.g. emotional states) or external reminders of the events (e.g. places or people)

(c) Distortion: such as negative beliefs about the self, the world and the future

(d) Arousal: such being on edge, hyper-vigilance, and poor sleep

Our framework concentrates on the first three of these primary impacts, as we believe they are the most important for understanding the decision making impacts of trauma. It is also important to note that experiencing trauma has wider impacts than PTSD, such as depressed mood. People with depression can also have intrusive memories of negative past experiences and avoidance, for which the mechanisms of memory outlined below can share relevance. However, unlike depression treatments (such as CBT) which focus on verbal cognition and negative thought, our GME approach is trauma-driven and focused on imagery-based cognition and prospection.

2.2 Memory as a Pathway for Trauma's Impact

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). Here we argue that a simple model of memory encoding and recall that builds on the literature on retrieved context (Howard and Kahana, 2002; Kahana, 2012, 2020; Wachter and Kahana, 2023; Bordalo et al., 2024, 2023) can explain both this re-experiencing, as well as the avoidance and distortion effects noted above.

We think of episodic memories (that is, memories of events in our life, see Tulving 2002) as being stored in a bank of past experiences. The bank contains mental representations that encode both details of the experience (content) and the *context* at the time of encoding. Context is a key construct and it determines the probability of recall, as well as the experience of recall. Context describes an individual's mental state, and includes both (mental representations of) external factors such as place and time, and internal factors such as emotional state and current pre-occupations. Context both affects, and is affected by, memory recall through three interacting mechanisms:[8]

1. As already noted, context at the time of encoding forms part of the memory that is encoded. So, for example, if you were obsessed with One Direction when you bought your first car, then thinking about your first car may lead you to recall One Direction;

2. Context similarity determines the likelihood of recall, with the probability of recall increasing when the recall context is more similar to the encoding context. For example, when people are in a positive mood they are more likely to recall positive memories (mood congruent recall,

[8] Our emphasis on context draws on the large literature on retrieved context theory. See Cohen and Kahana (2022) for an application to emotional disorder that we have been inspired by; see Bordalo et al. (2020), Bordalo et al. (2023) and Bordalo et al. (2024) for evidence showing that the memories that we draw on are those we see as more similar to the context in which we are making a decision.

Cohen and Kahana, 2022), and when visiting a childhood home, people are more likely to recall memories from their childhood (Wachter and Kahana, 2023);[9]

3. Retrieval (or imagination) is an internal experience in itself, such that recalling a negative memory (and its potential embellishment) creates a new negative memory and also leads to a negative emotional state.

Emotions are a particularly important part of context and affect both encoding and recall. Emotions are encoded with a memory, and when a person recalls a memory they re-experience some of the emotions stored with it. For example, subjects shown a neutral object become emotionally aroused if that object was previously encountered in an emotional context (Ventura-Bort et al., 2016) (mechanisms 1 and 3). Emotion also changes recall. Emotional memories are more likely to be retrieved, and recalling one emotional item increases the likelihood of recalling others with similar emotional valence (mechanism 2, see Talmi et al., 2019 for a review). Both these effects are strengthened by our tendency to pay more attention to emotional content (e.g., Chun et al., 2011).

These three mechanisms, combined with the role of emotion, help explain the memory impacts of trauma, which we conceptualize as the spread of negative emotions across the experience bank:

1. Trauma is, by definition, a strongly negatively emotional event, and it will create an initial set of negative memories, specific to the trauma context.

2. Trauma will also create a negative emotional state that spreads to both new and old memories:

    - Negativity will be encoded as part of the context of other temporally proximate experiences, even if they are otherwise neutral or positive events (mechanism 1); and

    - The recall of a previously positive memory after the trauma will lead to the encoding of a new memory with a negative emotional tag (mechanisms 1 and 3).

3. Negative emotion will lead to the recall of other negative memories, including re-experiencing the trauma (mood congruent recall, mechanism 2), which will maintain negative mood.

    - This negative mood will then spread further to new and old memories, as in step 2.

This spread of negative emotions throughout the memory system can explain the primary psychological impacts of trauma. Re-experiencing and interference of trauma memories can be accounted for by the extra attention we pay during emotional events, which means that trauma memories are stored with great detail, and the spread of negative emotion from this trauma to other nearby contexts means they are likely to be triggered in a growing range of settings. This negative

[9] Retrieved context theory started as an explanation for the empirical regularity that recalling one element from a list increases the likelihood of recall of other nearby elements. This can be captured by mechanisms 1 and 2 together, so long as time is part of context.

emotional spread impacts choice and cognition through the mechanism described in Bordalo et al. (2024), who show that different experiences - not necessarily belonging to the same domain - interfere with each other in recall. Through the lens of this model, trauma can bias the recall process in a wide range of domains by taking the priority over the recall of domain-specific experiences that would be (more) relevant for a given decision.[10]

Avoidance is accounted for by the aversive nature of recalling negative memories, and the understandable wish to avoid recalling negative experiences and unwanted emotions. Studies suggest that this avoidance can become very broad, evidencing the spread of negative emotions through the memory bank. After trauma, individuals often avoid both *external* context reminders, such as specific places or people that are linked by sight or sound to sensory memories of the trauma, and *internal* context reminders, reporting feeling numb rather than sad. A particularly well evidenced aspect of this numbing effect is the development of an "overgeneral memory," whereby people focus only on the general situation (e.g. "during the war") rather than specific event (e.g. "the morning my neighbor was shot"). For example, McNally et al. (1995) found in a sample of 32 combat veterans, that those with PTSD showed a marked reduction in the ability to retrieve specific autobiographical memories. Similar results are reported by Kleim and Ehlers (2008), see also Moore and Zoellner (2007). If this were just an attempt to avoid specific trauma memories then we might expect that these impacts affect only negative events, but evidence shows reduced specificity in the recall of *positive* events, as if all specific recall is being shut down and avoided (Kleim et al., 2014). Indeed, those who have experienced past trauma may struggle to generate mental images at all (even future images), with studies showing deficits in various memory processes (Douglas Bremner et al., 1995; Jenkins et al., 1998; Buckley et al., 2000).[11]

Negative emotional spread also creates distorted thinking. There is rich evidence to suggest that when people think about the future (prospective thinking) they do so by recalling and recombining memories from the experience bank (Schacter et al., 2007; Addis et al., 2007; Addis, 2020; Schacter et al., 2017; Bulley and Schacter, 2023). If these memories are negative so will be the resulting thoughts and projections about the future. That prospective thinking relies on episodic memory is suggested by a series of facts: the ability to remember the past and to think about the future both develop around the same time in children (Suddendorf and Redshaw, 2013); they are both disrupted by mental ill health and physical damage to the brain (Klein et al., 2002; Williams et al., 1996; Hassabis et al., 2007); aging reduces the specificity of reports on both memories of the past and imaginings of the future (Addis et al., 2008); and imagining the future and reflecting on the past use the same neural systems, with fMRI studies showing a strong correspondence between brain areas that show activity when participants are asked to remember the past and imagine the future (Addis et al., 2007; Thakral et al., 2020; Hassabis et al., 2007). Studies of the impact of

[10] A recent treatment for post-traumatic intrusive memories is one that uses another memory task to interfere with these intrusions, known as an imagery-competing task intervention (Iyadurai et al., 2023; Ramineni et al., 2023).

[11] Bremner (2006) shows changes to key brain areas like the hippocampus, amygdala, and medial prefrontal cortex.

trauma also document how memory affects thinking about the future. For example, Kleim et al. (2014) found PTSD patients were less able than controls to generate specific details of cued future events. Critically, this effect was particularly strong when they were asked to imagine positive future events, see also Rahman and Brown (2021) for a review.[12]

2.3 Trauma affects Decision-Making through its Link with Memory

Trauma's memory distortions directly affect economic decision-making because, as we noted above, the way we simulate the future is by remembering the past, and what we simulate changes choices (Bordalo et al., 2024). Simulation is indeed crucial to think about the future, and both the likelihood attached to future events and the strength with which we can imagine them depend on the relationship between a recalled memory and the target event. Because of the strong link between past memory and episodic future thinking, the ability to mentally project into the future is deeply affected by the contents of experience bank, memory recall processes and, thus, trauma. An "overgeneral" recall style (a form of avoidance intended to prevent the re-experiencing of trauma memories) means that it is hard for people to pre-experience specific events in their future. For someone after trauma, the future can thus seem unclear, confusing and simply hard to imagine. This deficit in future thinking applies especially to positive events. If the easiest thing to recall is negative events, and one uses the past to predict the future, then a trauma history increases the sense of a negative future. After trauma – particularly traumatic events involving threat to life – the future can feel foreshortened – for example that one has less years left to live and making a plan is futile. In summary, trauma is predicted to have three key impacts on future thinking:

- Re-experiencing and interference: When negative memories intrude into a trauma victim's thinking they hijack attention, suck the individual into the past and dramatically interfere with the ability to think clearly. The interference created by trauma memories also runs the risk of further spread of negative emotion to new contexts in the memory system. This happens because trauma creates a stock of salient non-domain specific experiences that can impede the recall of other memories, even those within the domain of choice and that would be more useful for thinking about the future (Bordalo et al., 2024).
- Avoidance: Inability to imagine specific events in the future (overgeneral memory) means trauma survivors find it hard to simulate the future, have limited information with which to make decisions and perceive a foreshortened future that is not worth planning for. They find it hard to make concrete plans and decisions, and to pre-experience possible future events or simulate alternative outcomes.

[12] Trauma also impairs working memory, worsening cognitive bandwidth (Scott et al., 2015). The psychological impacts of trauma summarized in this section can be thought of as a micro-foundation of such constraints: attempts to push away traumatic memories (through avoidance and overgeneral memory) absorb cognitive resources. Yet, as discussed, there are further direct impacts of trauma on cognition which do not go through bandwidth constraints.

- Distortion: Predominantly imagining negative future events makes people expect more negative outcomes. Mental projections into the future may lack details or specific plans, and even if positive, their lack of clarity can fail to evoke strong positive emotions. What seems to be particularly impaired after trauma is the ability to imagine a positive future, even when deliberately trying to imagine positive future events. Again, this effect can be explained by trauma's negative emotions spreading across domains and interfering during recall.

We summarize the impact of these effects, combined with the fact that we imagine the future by remembering the past, as: trauma survivors struggle to see the world's opportunities. They will struggle to predict the precise implications of their actions (avoidance), when they do try to make precise predictions they run the risk of re-experiencing and interference, with a further spread of negative emotions, and if they are successful in making precise predictions, they will be overly pessimistic (distortion).

2.4 Implications for Economic Development Programs

In the context of economic development programs, the impact of trauma on memory reduces the ability for forward thinking and decision-making that many development training programs require, opens the person to the risk of unwanted negative emotions, and contributes to lower motivation in general. Indeed, a large literature documents negative impacts of trauma on motivation (e.g., Goenjian et al., 2001). As a result, survivors of trauma will not be able to take best advantage of the opportunities they may have. Thus, programs that increase availability of opportunities, for example loans, cash handouts or work permits, and business training are predicted to be less effective in trauma affected populations.

In addition, to the extent that training programs require attendees to think about the specific context of their future actions, due to retrieval-context theory (Cohen and Kahana, 2022), left unguided participants risk binding negative emotions to those new contexts. This will further (inadvertently) induce problematic distortions and avoidance in these domains. For instance, the possibility of potential losses or potentially negative experiences (like being rejected from a loan, or not having any customers) can cue negative memories. This mechanism raises the possibility of backfiring. This echoes key issues raised by the literature on trauma informed programming, that is the need to avoid reigniting past trauma when providing services to vulnerable individuals (e.g., Center for Substance Abuse Treatment, 2014), and the possibility that trauma survivors may respond differently from others.

Finally, since trauma can impair a person's ability to imagine future events, especially positive ones that foster motivation, trauma survivors often struggle to plan for desirable outcomes or compare alternatives. Economic programs that encourage adaptive future thinking and help people mentally pre-experience positive events could provide substantial support.

2.5 The Role of Guided Mental Experiencing (GME)

The role of memory in imagining the future implies that trauma in the past can have ongoing impacts on people's choices about their future. However, it also opens the possibility that the negative impacts of trauma could be mitigated by changing the set of memories that are used in simulating the future and reshaping the mental pathways that are used in the retrieval process.

Broadly, there are two main classes of approaches that could be taken to overcome or mitigate the effects of trauma on the brain:

i) The first is a *treatment-based approach* to reduce symptoms, by focusing on the past trauma. The leading evidence-based therapies for this are trauma-focused CBT techniques and Eye movement desensitization and reprocessing (EMDR) therapy (Bisson and Olff, 2021). These include the recall and/or re-scripting of negative intrusive memories in a safe space, in order to lessen their emotional content, reduce re-experiencing of trauma, reduce avoidance and mitigate interference of these symptoms in daily life.

ii) The second approach can be described as a *skills building approach* and focuses instead on the future. Such an approach actively promotes an alternative way of imagining the future, by building a set of alternative (including positive) mental experiences. This creates more direct pathways to envisage alternative futures (Pearson et al., 2015; Holmes et al., 2007, 2019). This adaptive skills building can be done without discussing traumatic past memories, which is an advantage since many trauma survivors are reluctant to do so and furthermore would need to be done by trained mental health professionals (for which there is a resource bottleneck). Nevertheless, as we will describe, the adaptive skills building approach we use has the benefit that it strengthens people against the effects of being disrupted by trauma memories by building up a rich more adaptive prospective memory bank. This latter skills building approach we have termed Guided Mental Experiencing (GME). The term mental experiencing (ME) encompasses the processes described in the psychology literature variously as mental imagery, visualization, mental simulation, imagination, or mentalization. It is guided in that GME is instructed and taught in the first instance, with the aim that the individual learns how to use ME flexibly and thus transfers skills into to their daily life.[13]

Three features distinguish GME as a tool that can be easily woven into economic interventions

[13]We have three main reasons to prefer GME for our objectives. First, a treatment approach such as re-scripting or exposure to trauma must be done in a clinical setting, for which there are severe resource bottlenecks. Second, even if therapeutic approaches are available in low-income settings, low-demand for mental health services and social norms may represent important barriers to potential beneficiaries' engagement into these programs due to, for example, stigma. Third, we have described how trauma affects both episodic and prospective memories. The main focus of therapy is removing very salient negative experiences from memory, which should in turn mitigate distortions, re-experiencing and – perhaps in the longer term – enable trauma survivors to start imagining the future again. GME instead directly targets future thinking aiming to more rapidly boost this skill. Overall, creating a set of positive alternative prospective memories strikes us as the best option to build a scalable and acceptable intervention.

to help mitigate trauma-related episodic memory problems and improve prospective memory:[14]

- GME focuses on promoting mental experiencing about future economic outcomes, including positive ones, and the pathways of actions that could materialize these intended outcomes. This directly redress the inability to imagine the future, step by step, which is a first-order skill needed for economic decision making and for taking advantage of economic opportunities.

- GME promotes more and rich positive images in the memory bank that can compete with the retrieval of negative memories induced by trauma, i.e. which are sensorially rich, detailed and emotional. This helps reduce likelihood of the individual having their attention hijacked by intrusive traumatic memories by actively creating an alternative set of mental simulations (Pearson et al., 2015) which can capture attention. As multi-sensorial mental experiences (whether real or imagined) are stored in the brain's "memory bank", we aim to create more direct pathways to envisage alternative futures without touching traumatic past memories.

- GME enhances the quality of mental experiences by increasing the specificity and emotionality of mental scenarios, which in turn supports memory encoding and retention. This helps redress the "overgeneral" memory effect of trauma, and hence murky or unspecific mental experiencing about the future. Studies on recall discussed above show that emotional memories have a retrieval advantage by being strongly tied to the recall context. Research on episodic future thinking implies that more vivid and positive emotional recalls are more motivational.

In principle, it would be possible to try to create a set of positive experiences in multiple ways. For example, interventions that show participants the positive experiences of people like themselves would fit in our framework. However, GME is predicted to perform better than alternative more traditional and verbal ways to create such a pathway because: i) mental experiences (sensory representations) of positive experiences carry more emotion than does verbal thinking (Holmes et al., 2008) and ii) adding more positive emotional and detailed memories to the prospective memory bank makes them more likely to be recalled and more likely to change behavior (Renner et al., 2017). These properties of mental experiencing follow because mental experiencing is known to make use of the same brain processes as actually experiencing in-vivo (Pearson et al., 2015). In summary, GME helps create a set of positive, strongly emotional, and vivid (prospective) memories that are tied to important domains of economic action.

3 A model of memory 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 $u(a)$,

[14] A guided approach helps create a safe place for the aforementioned tasks and can be done in a group setting.

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 randomnly 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.[15]

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

$$r(e, a; E) = \frac{s(e, a)}{\sum_{e'\in E} s(e', a)}, \tag{1}$$

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 (1) 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

[15] We think of this as concentrating on the set of risky but potentially worthwhile actions.

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 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,

$$S(A, E) = \{a \in A : V(A) + U^p(a, E) > 0\} \tag{2}$$

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.[16] The simulation process then defines a set of actions that the DM will take, which depends on E

$$\mathcal{A}(A, E) = \{a \in S(A, E) : U^s(a, E) > c\}. \tag{3}$$

3.1 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 $\tilde{E}^R(A)$, 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

[16]It would not change anything if we allowed the DM to draw multiple memories.

interfere with the recall of relevant memories. Denote e^R to be any relevant memory, then from (1)

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

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

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

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 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 $\tilde{u} << 0$. For any memory e recalled while $\tilde{u}$ is the emotional state, a new negatively distorted memory $\tilde{e}$ will be created with

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

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 $u(\tilde{e}) << u(e)$.

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 (2) 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

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

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 $\tilde{A}(A, \cdot)$ is defined in (3). 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).

3.2 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.

3.3 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.

4 Designing Trauma-Informed Programs: Two Examples

This section describes our two populations, and the specifics of the GME programs delivered to each. The first subsection characterizes past trauma exposure and symptoms in the target populations, and describes how this trauma affects economic activity. The second subsection explains how we designed our interventions to take into account people's traumatic experiences.

4.1 Trauma and Economic Activity among the Vulnerable

Setting 1: Eritrean Refugees in Addis Ababa, Ethiopia

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.[17] 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.[18]

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).[19] 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.[20] Table 1 summarizes the characteristics of our sample along the margins of economic vulnerability and trauma.

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

[17] See Woldetsadik et al. (2019) for a detailed discussion of the "out of camp" policy (OCP) from a legal perspective.

[18] A 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).

[19] The 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.

[20] Potential 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.

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).[21] 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. This rate is notably higher compared to the general Ethiopian population, where only 35.8 percent report similar experiences (Girma et al., 2022).[22] 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.

[21] We use the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) with a cut-off of 31.

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

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.

Setting 2: Vulnerable Groups in Bogotà, Colombia

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 Servicio Nacional de Aprendizaje (SENA).

However, governmental and non-governmental actors increasingly recognize 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.[23]

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.[24] As part of their mandate, the SDIS were required to train

[23]This 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).

[24]Interested applicants were required to complete a short application form, online or in person, which was used

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 group is labeled "subdivision" in the experiment). 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 participants report having lived or witnessed war-related trauma, including death, displacement and torture, and 61 percent report suriving 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.[25]

Vulnerability also extends to the economic domain. As shown in Table 1, 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 Bogotà 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.

4.2 Guided Mental Experiencing in Program Design

Building on the insight that effective economic programs for traumatized populations should address the cognitive barriers caused by trauma, we designed two new training curricula that teach the use of GME in economic decision-making. The GME exercises in the curricula drew on existing

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.

[25] We use the Impact of Event Scale-Revised with a cut-off of 33 for symptoms of PTSD (Creamer et al., 2003).

imagery-based cognitive therapies (for instance, Holmes et al., 2019; Hackmann et al., 2011) and the practical experience working with people post-trauma of Professor Emily Holmes – one of the leading clinical psychologists on mental imagery globally and co-author of this study. To tailor the exercises to the economic domain, the research team conducted extensive qualitative research with target populations and clinical psychologists in Colombia and Ethiopia. These interviews revealed the memories, beliefs, and behaviors that aid or hinder economic progress, informing case studies inspired by real individuals identified during data collection. Before implementing the study, we piloted the curricula with our target populations in both settings.[26] Throughout the curriculum development process, we consulted government partners – mandated to develop and deliver programs that benefit vulnerable populations – as well as other stakeholders.[27]

The curricula implemented in Colombia and Ethiopia share similar features in the way in which GME is taught, and differ only in their content and delivery/duration, as outlined in the next section. Given the limited intent to stay and economic activity among our refugees, the training in Ethiopia aims to showcase how mental experiencing can be used to support all livelihood forms. In particular, we focus on imagining a potential life in Addis Ababa in the short term and in practicing the steps that can be taken to access economic opportunities in this context. The curriculum in Colombia focused on leveraging mental experiencing in the context of entrepreneurship.

Three Key Features of GME Training

The GME training curricula in both Colombia and Ethiopia share three key features. First, the GME training programs promote mental experiencing about future economic outcomes and the pathways of actions that could materialize these intended outcomes. Our curricula focus specifically on teaching participants to project the future and pre-experience the utility associated with alternate future outcomes. Once participants define a future scenario, they are taught to mentally practice the steps needed to achieve it. Our qualitative research emphasized the importance of promoting concrete action pathways for our target populations. Many conflict survivors aspiring to start a business struggled to articulate how they would make their first sale—a form of avoidance. For instance, some could not imagine selling to customers, while others envisioned exporting products without a clear plan for their first sale. Actively practicing future scenarios is expected to enhance the concreteness and realism of outcomes, increasing participants' likelihood of taking action. Additionally, while mentally rehearsing their plans in a safe space, participants identify potential obstacles and strategies to overcome them, rather than merely imagining positive solutions.

Participants apply GME to economic decisions that are widely applicable to the economic lives

[26] In Colombia, we piloted the curriculum in November 2018 and May 2019 to refine the sessions to the needs of the target population. We piloted all four sessions of the Ethiopia curriculum in January and February 2022.

[27] For instance, in Colombia, we worked closely with the lead psychologist team at the National Victims Unit (Unidad para las Victimas), the National Training Service (Servicio Nacional de Aprendizaje or SENA) and Alta Consejería for Victims' Rights within the local government, in addition to our implementing partners.

of the poor. For example, participants mentally practiced conducting a customer survey and a plan to become more productive and less wasteful in Colombia. A session on savings taught participants to first calculate their weekly savings contribution and then imagine their savings contribution accumulating over weeks, months and years to appreciate the long-term motivational power of small habits. To motivate the importance of simulating alternative pathways into the future, participants subsequently visualized the consequences of experiencing a personal negative shock in the instance where they had saved versus not. In Ethiopia, participants use GME to identify a step or action that they can take towards generating income in the near term. Participants were encouraged to implement these plans as homework and report back on their experiences, with the hope that these reflections should serve to improve the quality of mental experiences over time.

The GME exercises are also structured to address self-limiting beliefs associated with traumatic experiences and economic choices. For instance, our qualitative work in Colombia showcased how a "victim mentality" was often associated with low self-confidence and limited personal initiative in seeking economic opportunities. In Addis Ababa, nearly all refugees aspire to leave Ethiopia, despite the slim chances of achieving this through formal resettlement.[28] Such expectations limit the willingness to invest into a life in Addis Ababa. Given the high likelihood that most will remain there in the short term, our curriculum encourages participants to envision their lives and livelihoods in Addis Ababa over the next six months.

Second, our GME training programs are designed to improve the quality of mental experiencing through increased specificity and emotionality of mental scenarios. As described in Section 2, traumatic experiences can produce an "overgeneral" memory and hence murky or unspecific mental experiencing about the future. To counteract these negative effects of trauma, the GME exercises in our curricula encourage participants to produce highly detailed images about future outcomes and the pathways of action to attain them. Like experiences in real life, mental experiences produce an emotional response within participants, which carries information about the utility associated with the content of their simulated scenarios. The GME exercises in our curricula actively sought to amplify positive emotions generated by the mental experiencing, to boost motivation. Moreover, emotions favor the encoding of the newly created mental experiences into memory.

The following excerpts from both manuals show how participants are encouraged to create specific and emotional mental images, and to encode them in memory for future retrieval:

- Using all your senses, what does the most **vivid image** of a good life in Addis Ababa look like? I want you to create a picture in your head. (...) I would like you to capture this image, as if you were taking a photograph, so that you can mentally store it and use whenever you need to remember it.

[28] See Subsection 3.1 and Betts et al. (2023). In 2021, 158,294 refugees from Eritrea were registered in Ethiopia, yet UNHCR submitted only 803 cases for formal resettlement.

- Think about this scene for a moment. I would like for you to paint an image or a video in your head, as if you were an artist or a film director trying to capture this moment. What do you see? What do you smell? How do you feel right now?
- What are you feeling? It is a good feeling, or a bad feeling? (...) Do you notice any sensation in your body? What would you need to see in the image to make it more powerful? On a scale from 0 to 100, grade how persuasive is your image and your emotion in this moment. If you have less than 80 percent, what else do you have to do to feel motivated to save? It should be as vivid as possible.

Third, our GME training programs focus on promoting more positive images that can compete with the retrieval of negative memories induced by trauma. Our GME programs intend to weaken the interference of traumatic memories by actively creating an alternative set of mental simulations stored in the experience bank. These alternative mental images need to be powerful enough to counteract trauma by being sensorially rich and emotionally charged. By making positive and adaptive images richer and more detailed, they are more likely to be held in attention. The programs also provide techniques for tolerating negative images, albeit in slightly different ways across the two settings. For instance, our training in Ethiopia was largely focused on creating positive mental images to challenge negative thinking. Nevertheless, participants were encouraged to accept negative thoughts that cropped up and leverage a mental safe space when these thoughts became overwhelming. The safe space was a dedicated place in one's mind where one could feel safe and calm, and where one could return to whenever a current or future situation felt threatening or concerning. In Colombia, the training went a step further and asked participants to simulate both negative and positive future scenarios. The goal is to reduce avoidance and teach participants how they can effectively face negative emotions in a controlled way without being overwhelmed by them. In different scenario simulations, participants learnt how to turn negative emotions into feelings that motivate them to be more productive and do better for their lives and business.

5 Evaluating the Programs

5.1 Trial Designs

Experiment I: Ethiopia

We randomly assign 50 percent of the sample to the GME intervention and the other 50 percent to a control group without intervention. Treatment assignment was stratified by gender, and a cross-randomized priming treatment not discussed in this paper.

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.[29]

Experiment II: Colombia

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 treatment" from hereon) and a business training plus GME ("GME treatment" from hereon). 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" from hereon). 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). 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.[30]

A key challenge was designing both treatment arms to be nearly identical, differing only in the

[29] This included a referral system for participants showing signs of severe mental illness. Of the 24 identified as needing additional mental health evaluation, 17 agreed to seek care and were referred to a local service provider, which assessed their needs in collaboration with clinical psychologists from Addis Ababa University.

[30] Participants 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.

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.[31] 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. For all activities, trainers were kept separate by treatment arm. Participants who attended at least seven sessions received a certificate.

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.[32]

5.2 Measurement

We conduct in-person baseline surveys in both studies, followed by phone or in-person follow-up surveys. In Ethiopia, we conducted baseline surveys between May and July 2022. In-person follow-up surveys were conducted approximately three months after the baseline and approximately six weeks after the intervention. In Colombia, the baseline survey was conducted in two waves between July to September 2019 and September to December 2019. Two rounds of phone follow-up surveys were conducted in May and November 2020 (approximately seven and thirteen months after the intervention).[33] In our analysis, we pool the latter two follow-up surveys. Our outcome measures follow our pre-analysis plans ("PAP" from hereon) registered in the AEA RCT Registry.[34]

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

[32] As 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.

[33] In Ethiopia, the average time between baseline and endline was 92 days. In Colombia, an in-person survey was initially scheduled to commence in early March 2020, but then cancelled in response to the COVID-19 pandemic.

[34] Ethiopia RCT ID: AEARCTR-0008934 (doi:10.1257/rct.8934-1.1); Colombia RCT ID: AEARCTR-0004695 (doi:10.1257/rct.4695-1.7). Reports that detail deviations from each PAP are available upon request.

We hypothesize that our intervention will first impact the quality of mental experiencing, which then in turn improves downstream economic outcomes, especially among people with trauma. The remainder of this section summarizes how we measure each outcome group, with a particular focus on the novel instruments we used for mental experiencing quality. Appendix subsection A.3 and subsection B.3 contain more details on measurement in Ethiopia and Colombia respectively.

Trauma

We measure both trauma exposure and symptoms in our samples. For the former, we conduct at baseline a contextually-relevant trauma history checklist that captures whether respondents experienced or witnessed a traumatic event, and its type (e.g., assault, natural disaster, torture).

To proxy for the level of trauma, we use scales that assess subjective distress caused by the most traumatic event ever lived in the past month (at baseline only). In Ethiopia, we use the PCL-5 (Weathers et al., 2013), which is a 20-item self-report measure that assesses the 20 DSM-5 symptoms of PTSD. 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 describe participants as "high trauma" when their scores are above practitioners' thresholds for a provisional PTSD diagnosis. A score of 31 for the PCL-5 and a score of 33 for the IES-R are the thresholds above which post-traumatic stress symptoms are considered of probable clinical concern.

Mental Experiencing

Our GME programs aimed at improving the quality of mental experiencing through increased specificity and emotionality of mental scenarios. To this goal, we adapt two main scales widely used in the psychology literature to economic domains and our field settings: the Prospective Imagery Task (PIT) (Stöber, 2000) and the Autobiographical Memory Test (Griffith et al., 2009).

Our main instrument presents participants with different scenarios – which can have positive, negative, or neutral sentiment – and asks them to imagine each one in their minds using all their senses. For each scenario, participants are asked to imagine either a related past or future event. For example, in Ethiopia, we ask participants to imagine a future scenario where "you ask a friend to recommend you for a job" and a past memory where "a friend of yours gets fired from a job". In Colombia, we similarly ask future scenarios with either positive sentiment (e.g., "the COVID-19 pandemic is over, and you save enough money to buy an asset you really want") or negative sentiment (e.g., "the COVID-19 pandemic is over, and your business closes").[35]

[35] In Ethiopia, we ask participants to imagine one positive, one negative and two neutral scenarios for past memories, and the same mix for future scenarios. In Colombia, we ask participants to imagine three positive and three negative scenarios, divided into business-specific and non-business scenarios, only about the future.

For each scenario, the quality of respondents' mental experiences is scored according to its level of specificity and emotionality. In Ethiopia, this is done by treatment-blind research assistants, who coded participants' descriptions of their mental experience (i.e. audio recordings) following a pre-defined coding scheme. Coders are also asked to assess the scenarios' emotional valence (from very negative to very positive). In both Colombia and Ethiopia, respondents also describe their mental experiencing according to its level of specificity and emotionality (on a Likert scale from 1 to 5).[36] In Colombia, we additionally elicit the extent to which participants report spontaneously using mental experiencing during daily life using the Spontaneous Use of Imagery Scale (SUIS) (Nelis et al., 2019) adapted to the entrepreneurship context.

We construct an overall index for the quality of mental experiencing by aggregating these sub-components, as per our pre-analysis plans, across both past and future scenarios. In Ethiopia, we average the office-coded scores on specificity, emotionality, and frequency of positive scenarios (among neutral ones). In the absence of audio recordings in Colombia, we aggregate measures of self-reported specificity and emotionality from the scenarios and frequency of use from the SUIS.

Appendix C.1 provides support for the validity of our measures, including correlations of our mental experiencing indices with baseline economic outcomes and trauma. This evidence is aligned with the key patterns highlighted in Section 2.

Economic Outcomes

We then measure the impact of the GME interventions on economic decision making. Whereas the GME intervention in Colombia sought to target entrepreneurship behaviors, the GME intervention in Ethiopia supported more general income-generating activities. Hence, in Ethiopia, we measure i) the respondent's current economic activity; ii) the number of hours worked in any job in the past week; and iii) a job search index. Given the widespread intentions to leave Addis Ababa, we also capture the intentions to stay in Ethiopia using a battery of structured questions.

In Colombia, we measure effects on business-relevant economic outcomes. We collect data at the height of the COVID-19 pandemic in 2020, after a national lockdown was initiated and severely restricted economic activity for our sample. Hence, we distinguish between economic outcomes in the months prior to and after the lockdown imposed on the 24^{th} of March 2020. For both periods, we construct two primary economic measures: i) an earnings index combining take-home income and business sales, with the latter conditional on a business operating; and ii) business status indicating whether a business is active, defined as open or only temporarily closed during the lockdown. We construct three additional economic outcomes measured at the time of the surveys during the pandemic: i) an investment index capturing newly acquired or significantly improved business assets; ii) a safety nets index combining savings accrued before the lockdown and perceptions of

[36] Self-reports and back-office coded measures are positively correlated in the Ethiopia sample (see Figure C2).

informal support networks during the lockdown; iii) an index of COVID-specific actions, ranging from diversifying products to adhering to government COVID-19 guidelines on creating a safe work environment. We further consider downstream impacts on more general welfare outcomes, including mental wellbeing in both study contexts. In Ethiopia, we also measure food security, consumption expenditures, and physical wellbeing.

6 Results

6.1 Ethiopia: Results for a Direct GME Program

This section presents the main results of our experiment with Eritrean refugees in Addis Ababa, Ethiopia. We first show that the intervention is effective in increasing the quality of mental experiencing and improves economic outcomes among participants, and then consider heterogeneous treatment effects as specified in our pre-analysis plan.

Empirical Strategy

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

$$y_{is} = \alpha + \beta \cdot GME_i + \delta_s + \epsilon_{is} \tag{4}$$

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. We also augment this specification with a set of individual-level controls X_i chosen using Post-Double Selection Lasso (Belloni et al., 2014). 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) The Online Appendix reproduces our main analysis with enumerator fixed effects (in Tables C9 and C10).

Balance and Attrition. Treatment assignment was balanced on baseline observables (Table A2). Overall attrition was low in the follow-up survey (at 15 percent). The treatment group was 6 percentage points more likely to reply to the survey than the control. Nevertheless, we find no differences in observables among endline respondents between treatment and control (Table A5). Online Appendix A.2 contains further details about attrition and related robustness exercises.

Results on Mental Experiencing

Our first set of results shows that the GME intervention significantly increased the quality of mental experiencing among participants (Table 2). Participants assigned to the GME treatment increased their overall quality of mental experiencing quality by 0.09 standard deviations, using office-coded audio recordings of GME scenarios. This overall impact is driven by a significant increase in the emotionality of reported scenarios (a 0.13 standard deviation increase) and in the frequency of positive scenarios (a 0.11 standard deviation increase).[37] This is consistent with the focus of the program on generating positive and emotional images.

Our mental experiencing quality measures are based on encoded audio recordings to mitigate concerns about participants' reporting or social desirability bias. Nevertheless, one may still worry that the types words that the participants use – on which treatment-blinded research assistants base their coding – can still be influenced by this bias. To address this, Table C2 presents estimates on mental experiencing 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 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 experiencing questions. Overall, our results indicate that our intervention was efficacious in increasing the quality of mental experiencing among a highly vulnerable and traumatized population.

Results on Economic Outcomes

In our theoretical framework, GME can create new experiences that compete with trauma memories for recall, reducing distortion and avoidance, ultimately improving decision-making. In line with the latter prediction, Table 3 provides evidence for the positive impacts of GME on downstream economic outcomes. Results in Columns (1) to (4) show impact of the GME treatment on participants' willingness to invest in their stay in Addis Ababa, an important determinant in the decision making of these refugees and a key feature in our intervention. The GME treatment increases participants' intent to stay in Addis Ababa in the following three years by 0.17 standard deviations. This corresponds to an increase of 3.9 percentage points from a control group mean of about 42.9 percent of respondents who indicate that they intent to remain in Addis Ababa.

Consistent with increased willingness to stay in Addis Ababa, we find that treated participants take concrete actions to improve their current economic livelihoods. Participants are 0.12 standard deviations more likely to be working, a substantial increase of 5 percentage points over a control group mean of 21 percent. Treated participants also significantly increase their working hours by

[37] These results are based on quality assessments by treatment-blind research assistants, as outlined in subsection 5.2. Results are robust, albeit noisier, when using principal component analysis (PCA) indices that combine externally-coded measures and self-reported measures of mental experiencing quality (Table C1).

0.14 standard deviations or about three hours per week, relative to a control group mean of about 8 hours per week. This increase in hours worked is driven by a significant increase in hours from wage work as opposed to casual work or unpaid apprenticeships (Table C7).[38]

Both the intensive and extensive effects mainly come from participants who were actively looking for a job at baseline (40 percent of the sample). In line with that, we find that the GME treatment significantly decreases job search activity in our sample (Table 3, Column 2). This suggests that the treatment improves participants' effectiveness at looking for economic opportunities. As a result of more hours spent working, income from labor (both wage and self employment) also increases. While the effect on income is large – 0.43 standard deviations, or a 52.7 percent increase in weekly income – 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 C8 for non-standardized labor market outcomes).

The GME treatment has additional positive impacts on downstream economic outcomes in our sample (Table 3, Columns 5 to 9). Reported food security, based on a four-item version of the commonly used Food Insecurity Experience Scale (FIEFS), increases substantially by 0.22 standard deviations. The coefficient on the treatment dummy for total consumption expenditure is positive, but not statistically significant. Life satisfaction measured by the Cantril ladder increases by 0.13 standard deviations. Finally, the GME treatment significantly increases a combined measure of physical and mental health based on the WHO Disability Assessment Schedule (WHODAS) by 0.16 standard deviations. The results on targeted behaviors and downstream outcomes are robust when we interact the treatment indicator with the participants' social desirability score (Table C3).[39]

Heterogeneous Treatment Effects

We examine heterogeneous treatment effects by baseline trauma exposure and gender. Our theoretical framework suggests that GME can mitigate the impact of trauma on economic outcomes, so we anticipate larger treatment effects for individuals with higher baseline trauma symptoms.

We define respondents that are above the PCL-5 cut-off for probable PTSD at baseline as "high trauma." We find that the GME treatment has relatively larger positive treatment effects for those participants on some dimensions (Table C4 and Figure 3a, left panel). For hours worked per week, control group participants in the "high trauma" group worked 0.19 standard deviations or 3.8 hours less than their peers. The GME treatment reverses this gap by significantly increasing hours worked among participants in the "high trauma" group by 0.51 standard deviations or 10.3 hours.

[38] In Table C7, 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.

[39] We also estimate treatment effects on other pre-specified behavioral outcomes. We do not find any effects on risk taking, self-efficacy and reservation wages. We find strong treatment effects on optimism (0.17 standard deviations), which aligns with our finding that the GME treatment boosts positive mental experiencing.

For reported life satisfaction, we similarly find that participants in the "high trauma" group are 0.44 standard deviations less satisfied, but the GME treatment reverses this gap by significantly increasing life satisfaction in the "high trauma" group by 0.51 standard deviations.

For female participants, results are mixed (Table C5 and Figure 3a, right panel). Female participants in the control group fare worse than men on most dimensions of economic activity, but treatment effects on targeted behaviors are also generally weaker for women. The only outcome on which we find significantly stronger effects for women relative to men is the WHODAS score, where the overall treatment effects are driven by female participants.

A confounding channel through which the intervention may affect labor market outcomes is by changing norms related to the acceptability of working. As discussed in subsection 4.1, many Eritrean refugees in Addis Ababa are not formally allowed to work, although informal employment is widespread. In our sample, we find no effect of the GME treatment on the share of participants who report being inactive because "they are not allowed to work" (also not differentially by gender or baseline trauma). To further provide evidence on whether the treatment changed perceptions on work permissions, we collected participants' scores in a "rule orientation" scale (Fine et al., 2016), which captures one's attitudes towards rules and regulation. We do not find heterogeneous treatment effects by the participants' scores in this scale, suggesting that the treatment does not only induce participants who tend to be more tolerant of violating rules to work more (Table C6).

In sum, we find evidence that the GME treatment improves both mental experiencing and economic outcomes among refugees, with some positive effects being concentrated in participants with higher baseline trauma burden. We now turn to our second experiment, in which we test whether GME can increase the effectiveness of standard economic programs for vulnerable populations.

6.2 Colombia: Results for a Program Adapted Using GME

This section presents the main results of the experiment in Colombia. We first show that the traditional training decreases both the quality of mental experiencing and earnings, but the GME-adapted program is able to overcome these negative effects. We then discuss heterogeneous treatment effects, which will shed light on possible mechanisms.

Empirical Strategy

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

$$y_{isw} = \alpha + \beta \cdot T_i + \delta_s + \delta_w + \epsilon_{isw} \tag{5}$$

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.[40] We report in the Appendix specifications which also include a set of household-level controls X_i chosen using Post-Double Selection Lasso (Belloni et al., 2014). 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.[41] The comparison with the no-intervention group provides 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).

Balance and Attrition. Treatment assignment was balanced on baseline observables (Table B1). Overall, 78 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. While the control group was around 5 percentage points less likely to reply to the surveys than any of the treatment arms, we find little observable differences among follow-up respondents between the control and the treatment groups (Table B5). The Online Appendix A.2 contains further details about attrition and shows robustness exercises to deal with it.

Results on Mental Experiencing

Table 4 shows the main results of the Colombia experiment on mental experiencing quality. Column (1) presents treatment effects on an overall index which combines how frequently participants use mental experiencing in their daily lives, as well as measures of specificity and emotionality. Columns (2) and (3) of Table 4 split the index in a business-related component (including only answers to business-related scenarios) and a non-business one. While we expect the treatments to create a wedge in the quality of mental experiencing between experimental groups in the business domain, whether these improvements spill over into other domains remains an empirical question.

Panel A of Table 4 compares the mental experiencing quality of participants in the traditional training with those in the no intervention group. Column (2) shows that assignment to the

[40] We 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.

[41] As specified in our PAP, due to insufficient sample size and our implementation partner's preferences, we do not have two subdivisions in the pure control group for one of the experimental waves. In order to use the full available sample, we prefer running separate regressions comparing the GME treatment to each of the other two arms.

traditional training worsens mental experiencing quality in the business domain (by 0.16 SD). Columns (4) to (6) unpack this effect by showing that participants trained in a conventional way are both less likely to use mental experiencing (Column 4) and have lower specificity and emotionality associated with business scenarios they are asked to imagine (last two columns). Coefficients on the traditional treatment dummy are indeed negative in all the dimensions of mental experiencing quality we asked about (but not statistically significant).

While the traditional training crowds-out participants' mental experiencing in the business domain, this negative effect does not extend to other domains. Table 4 (Column 3) shows that there are no differences in our mental experiencing index between the treatment and control groups in non-business scenarios (not statistically significant coefficient of -0.05 SD). As a result, the treatment effect on the overall mental experiencing index which combines both business and non-business scenarios is also negative, but not statistically significant (Column 1). Overall, these results suggest that teaching business concepts in a traditional way – mostly verbal – crowds out the quality and use of mental experiencing in the targeted domain.

Table 4 (Panel B) compares the mental experiencing quality of participants in the GME training with those in the traditional training. Column (2) shows that GME-trainees score 0.18 SD higher in mental experiencing quality in the business domain than their peers in the traditional program. This positive wedge is concentrated on measures of mental experiencing quality rather than use, in particular specificity (0.13 SD, p=0.05) and emotionality (0.17 SD, p=0.03). However, there is no positive gap between the two groups in non-business scenarios.

When looking at the net impact of the GME training compared to control (Table 4, Panel C), we find that there is a positive, but small and not statistically significant result on both business and non-business mental experiencing quality. Overall, the evidence in Table 4 indicates that the traditional training reduces the quality of mental experiencing among participants. The GME training is able to restore mental experiencing quality, but does not lead to a net improvement compared to the control.

These result can be understood through the lens of our conceptual framework. Participation into training entails two different forces. On the one hand, participants are asked to think about the future, even when they usually tend to avoid it. This can lead to negative emotions and, through context retrieval, a risk that traumatic memories get drawn from the experience bank. As a reaction, participants post-training may become even more likely to avoid planning for the future and thinking concretely about it. On the other hand, a training may enable participants to simulate new (more positive) scenarios in their minds and counteract trauma interference, restoring motivation and concrete steps to invest in the future. The results of Table 4 suggest that the former force prevail in the traditional program. After the training, the decline in mental experiencing quality is a symptom of participants' refraining from concretely and vividly imagining future business scenarios. The GME training counteracts this decline by helping participants reshape the types of scenarios recalled during the program. Following the logic of our theory, we then expect motivation and forward-looking actions in the traditional training to be worse on average than in the GME training. The comparison between the GME training and the control is instead ambiguous.

There are two points worth discussing before looking at downstream economic outcomes. First, one may wonder why the results on mental experiencing differ in our second experiment from those in Ethiopia. We acknowledge that differences in the sample or setting may contribute to this. However, we believe that there is also a key difference in the design of the two experiments which may explain why the GME training does not have a net positive impact in Colombia. The treatment in the refugees sample only consists of GME exercises, which are meant to limit the insurgence of negative intrusive recall. In the Colombia training, GME exercises only represent a third of the contents to which participants are exposed. This leaves room for thinking about the future outside of the guided and safe environment offered by our mental experiencing intervention.

A second point is about the domain-specific effects that we find in Colombia. A caveat of this result is that business scenarios were only asked to individuals with a business at the time of the surveys, and thus we are considering only approximately half our sample for this result.[42] This feature of our measures means that there are two possible reasons underlying the treatment effects in business-specific scenarios relative to non-business scenarios. First, GME may be domain-specific and thus can only be trained for a specific type of decision-making. According to this explanation, the sub-sample of participants with a business may have an advantage in learning because their experience bank already contains images within the same domain, which are faster to be retrieved and used during GME exercises. Second, it may be the case that entrepreneurs have different baseline characteristics which makes them more likely to learn mental experiencing (e.g, lower costs of mental simulation because of higher income). To distinguish between these alternative explanations, we check whether the business sub-sample improves its mental experiencing outside the business domain, which would indicate that this group has lower costs of learning mental experiencing in general. We do not find support for this effect, given that also the business sub-sample does not improve non-business mental experiencing in the GME training (see Table C14). We now turn to the discussion of treatment effects on downstream economic outcomes.

Results on Economic Outcomes

According to our conceptual framework, a decline in mental experiencing quality should be associated with weaker ability to imagine and plan for the future, with consequent lower motivation and

[42]Having a business at the time of the surveys is endogenous to treatment status. However, results on the comparison between GME and the traditional training are unchanged when restricting our sample to households that had a business at baseline (51 percent: coeff=0.181, p=0.058). We find no differences in baseline business status across treatment arms in both follow-up surveys, lending support to the assumption of monotonicity: 84.5 percent of households in the GME treatment that had a business at baseline also had a business in the first follow-up survey, compared to 81 percent in the traditional training group ($p = 0.29$). These proportions shift to 75.5 percent in the GME treatment and 73.7 percent in the traditional training in the second follow-up survey ($p = 0.66$).

worse economic choices. We predict economic outcomes to be better in the GME training compared to the traditional, while the overall net impact of the GME training is ambiguous. Our results are in line with these predictions.

Table 5 presents the main results on economic outcomes in the Colombia experiment. First, we find that the traditional training backfires. Compared to the no intervention group, participants assigned to the traditional program score 0.16 SD lower ($p < 0.05$) on an overall index of economic outcomes in the months after training and before the COVID-19 pandemic (Table 5, Panel A, Column 1). A similar negative effect persists in the months during the pandemic, but the effect becomes smaller and not significant (Column 2).

In contrast, Panel B shows that embedding the GME treatment within the business curriculum is efficacious in avoiding this negative earning effect. For both the pre-COVID and COVID periods, participants in the GME treatment have significantly higher aggregate economic outcomes than their peers in the traditional treatment. This evidence suggests that the GME component of our curricula brings a large positive value, which is able to compensate the negative effects brought by the traditional training contents.

The wedge in outcomes between the GME and the traditional training is driven by earnings, which include both income from any source and sales in the business. Given the very skewed distribution of earnings in the sample, Columns (3) and (5) of Table 5 use an inverse-hyperbolic sine (IHS) transformation standardized with respect to the no intervention group. According to this measure, the traditional program decreases earnings by 0.14 SD to 0.22 SD in the two periods considered with respect to not receiving any course. While the GME training fully restores the equality of earnings with the control, Panel C of Table 5 shows that the GME does not lead to a net improvement.

As the IHS transformation averages effects at the intensive and extensive margin, we look at quantile treatment effects on the level of earnings to assess the distributional impact of the training. Figure C4 shows that the decline in earnings in the traditional training mainly happens below the median. With respect to the control, trainees in the traditional program experience a significant decrease in earnings in percentiles 5 and 10, pointing to a larger share of participants with no earnings at all or with very low financial inflows. In both pre-COVID and COVID periods, there is also a significant negative effect on earnings at intermediate percentiles (between 20 and 50).[43] The gap in earnings between the two trainings is not driven by a differential likelihood of opening a business in the few months right after the intervention (Column (4) of Table 5). 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 (Column (6)), perhaps pointing to higher resilience or motivation to take concrete actions for their business.

[43] Table C17 shows quantile treatment effects. The earnings gap between the traditional and GME training before COVID is concentrated on the left tail. In the COVID period, distributional results are instead more mixed.

Column (7) of Table 5 shows that there is no impact of any of the trainings on an index of safety nets, which includes savings, perceived ease of finding money for current expenses and working hours in a safe environment. Similarly, neither the traditional nor the GME treatment affect the types of actions that respondents take in response to the pandemic shock (e.g., contacting suppliers, seeking government support) or their willingness to invest in items for their business or work.

Panel C of Table 5 shows the comparison in economic outcomes between the GME training and the control group. In line with the effects discussed for mental experiencing quality, we find that there are no significant effects of the GME training compared to the no-intervention group. Even if coefficients tend to be positive in sign for most of the indexes, none of them is above the threshold for statistical significance. We provide some explanation for this result in the discussion.

Heterogeneous Treatment Effects

Figure 3b shows heterogeneity in treatment effects by baseline trauma. Our measure of trauma is an indicator that takes value of one if a person reports to be a victim of the conflict or a recent Venezuelan migrant, or if she scores above a threshold of 33 in the Impact of Event Score scale. We label as "high trauma" participants with value one on this indicator, and "low trauma" those who neither report to be post-conflict victims/migrants nor have high trauma symptoms.[44] The figures shows two sets of results: the comparison between the traditional and control (with empty squares), and the comparison between GME and traditional (with full diamonds). Light blue coefficients report treatment effects for people with low trauma, whereas orange coefficients report effects for people with high trauma (scoring 1 in the aforementioned indicator). Table C18 reports the regression output corresponding to the coefficients shown in Figure 3b.

Figure 3b shows that our treatment effects are mitigated in the low-trauma group, for both comparisons. For this group, we find no significant effects on either mental experiencing quality or economic outcomes for either the GME or traditional training. In contrast, our results are driven by the high-trauma group. Both the negative impact of the traditional training and the compensating effect of the GME training are larger among high-trauma participants. For instance, mental experiencing quality in business scenarios decreases by 0.37 SD among traumatized participants assigned to the traditional training (compared to an average effect of -0.16 SD), but it is also completed restored in the GME training. Coefficients are similarly large for the effects on earnings.

Figure 4 further shows that our results on both mental experiencing and economic outcomes are driven by female participants. Treatment effects on female participants are stronger across all key variables of interest, except for starting a 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,

[44] We obtain similar results, albeit noisier, when defining "high trauma" only using the IES threshold.

suggesting that gender-specific factors beyond trauma burden—such as differences in the emotional tagging of memories —are important in shaping training outcomes.

These heterogeneous treatment effects confirm the key role that traumatic experiences and gender play in shaping people's returns from participation into economic programs. They also provide support to the idea that the traditional training exacerbates the economic costs of trauma in our setting. As participants are pushed to think about future choices, the risk of recalling past negative experiences seems to bring them back to a learnt behavior: shutting down their mental images and emotions. As the future becomes even less concrete and vivid in people's minds, their motivation to act is lowered, and their economic outcomes are worse.

We provide suggestive evidence for this motivational channel in the next paragraphs.

Suggestive Evidence on Mechanisms

Why does the traditional training lead to a decline in economic outcomes? We explore the aforementioned motivational channel: participants in the traditional training lose motivation to put effort in their business. We zoom in on different types of actions that trainees can take to start or grow their business: targeted investments, introducing new products, getting funds, expanding inputs (i.e. working hours, employees, raw materials). We believe these outcomes are good indicators for motivation because they only depend on people's willingness to take initiative for their business (e.g., they do not rely on demand by clients) and, thus, differences may emerge right after the intervention.

Figure 5 shows treatment effects on this set of business-oriented actions. Light blue coefficients show the comparison between the traditional training and the control and GME trainings pooled together, while dark blue coefficients show the comparison between the GME and control group. The figure shows that participants assigned to the traditional training – compared to the other two experimental groups – are less likely to invest time, money or effort in their business. For instance, compared to the GME training, the traditional trainees score 0.12 SD lower on a measure of funding obtained for the business and are 5 percentage points less likely to invest an hour of work in the business. This evidence is in line with the idea that the traditional training discourages people from even trying to invest in their future business. In contrast, the GME training does not only restore motivation, but also improves it with respect to the control along some dimensions. For instance, GME participants are more likely to invest in business items, to introduce new products or to get new inputs for their business compared to their peers who did not receive any intervention.

Taken together, these results point at changes in motivation – driven by a differential use and effectiveness in mental simulation – as one of the main channels through which the two treatment groups had an impact on participants' behavior and choices.[45]

[45] Data collected from approximately 50 percent of the original sample in 2021 (approximately two years after the intervention) suggestively show that GME-trained business owners are more likely to engage in marketing activities

7 Discussion

The Net Impact of the GME Program in Colombia

In the Colombia RCT, the GME training overcomes the drawbacks of the traditional program, but fails to achieve a net positive impact relative to the control group (Table 5). This outcome likely stems from two competing effects: i) the emergence of negative emotions potentially recalling bad experiences and negatively tagging business-related activities (through mood congruent retrieval) and ii) the counteracting effect of learning how to generate positive mental experiences.

This theoretical account has implications for interpreting the effects of training programs across different settings. As most training programs target diverse participants, including those with and without trauma, negative effects may be obscured by averages. Our GME intervention has the potential to mitigate these negative effects while maintaining benefits for non-traumatized populations. This possibility depends on how generalizable the risk of negative recall is across settings and populations. Figure C5 provides suggestive evidence on the generalizability of potentially negative impact of traditional training on traumatized populations. Considering all entrepreneurship training interventions from the meta-analysis in McKenzie et al. (2023) and without access to microdata, we approximate trauma levels in study samples by categorizing countries based on the 2019 OECD index of "Violence against Women" (Organisation for Economic Cooperation and Development, 2020) and their share of female participants. The intuition is that female participants in training programs conducted in contexts with high gender-based violence are more likely to have experienced trauma and violence in the past. In line with our own results, we find that training programs delivered to women in high gender-based violence contexts have lower average treatment effects and larger confidence intervals (even if sample sizes are similar across categories). Notably, out of all the studies with a negative point estimate (9 out of 21), 7 studies have a female share above 50 percent and are conducted in countries with high violence against women. This finding suggests that traditional training programs, designed without considering participants' past experiences, may indeed be less effective or even harmful to some beneficiaries.

We also acknowledge that the COVID-19 pandemic may have affected our results, as there were only a few months between the end of the program and the onset of the pandemic. This limited time window may have been insufficient for entrepreneurs to apply their new knowledge and generate positive effects. None of our courses was designed to guide entrepreneurs through periods of wider economic turmoil. 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. We believe that integrating GME within a successful program tailored to current economic circumstances may lead to positive net effects, presenting an important area for future research.

which may expose them to negative feedback and emotions (e.g., seek feedback from customers and competitors). In contrast, traditional trainees engage more in budgeting and record keeping.

8 Conclusion

As of today, more than 100 million people worldwide have been forcibly displaced by war, conflict and human rights violations (UNHCR, 2024). Yet, we still have a limited understanding about how these traumatic experiences shape the choices and investment these populations living in fragile settings make for their future. The potential for differential impacts of development programs based on individual's past experiences and traumatic histories highlights the importance of designing trauma-informed interventions.

In this paper, we designed and evaluated two trauma-informed training programs aimed at restoring the ability to imagine the future among vulnerable populations. Our approach is grounded in neuroscientific and psychology research demonstrating that traumatic experiences can significantly impede the ability to imagine the future, either through avoidance or excessive negative future thinking. Consistent with this hypothesis, we find that our training disproportionately benefits those who have experienced high baseline trauma.

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.

Figures

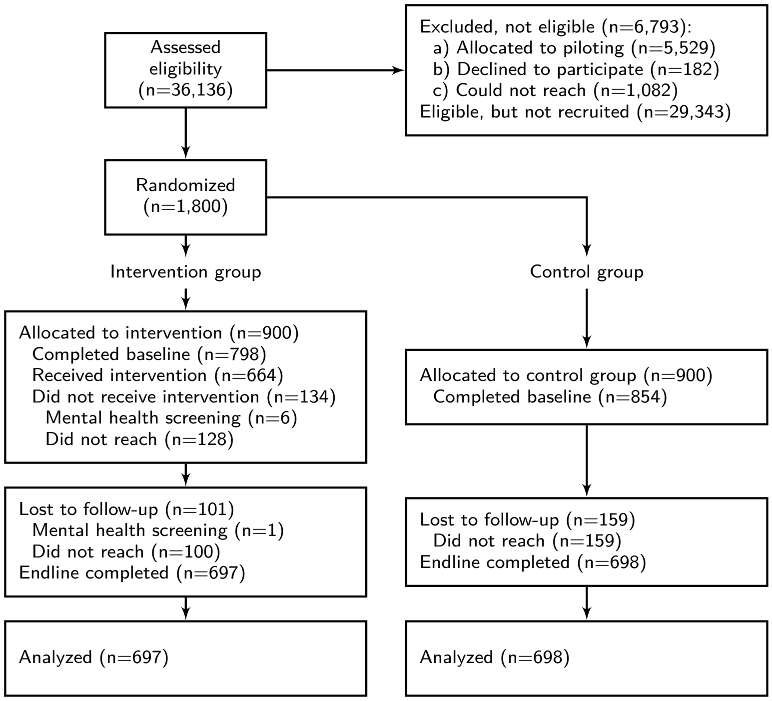


Figure 1. Ethiopia Trial Flow Diagram

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.

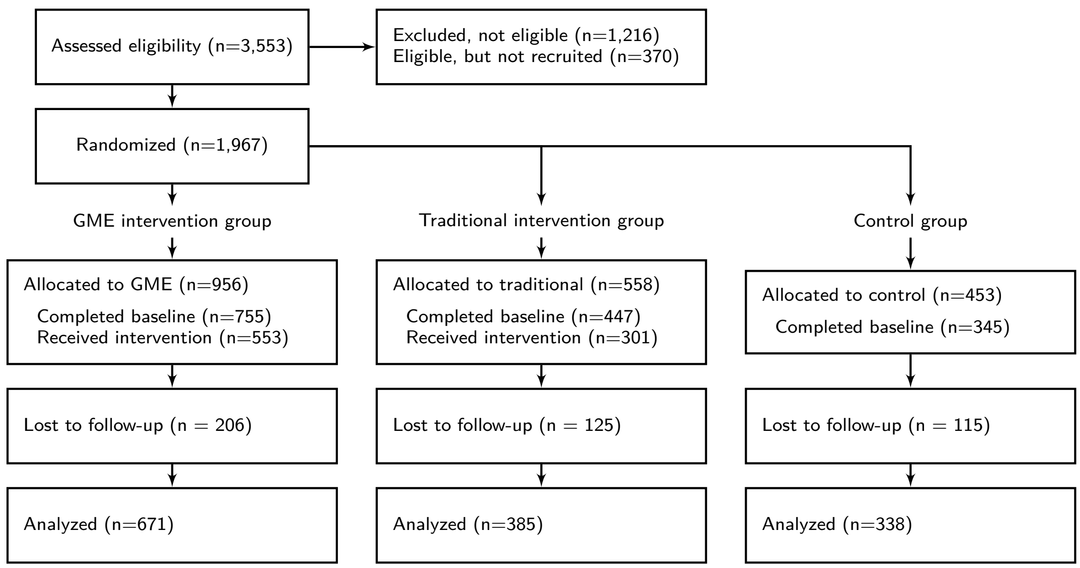
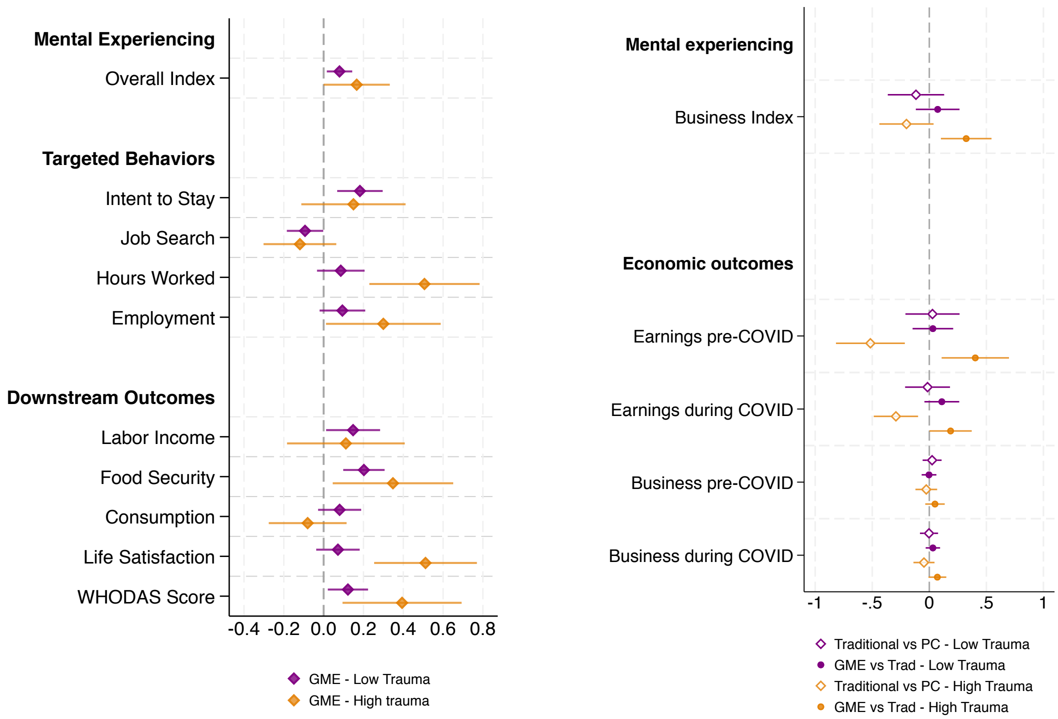


Figure 2. Colombia Trial Flow Diagram

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.

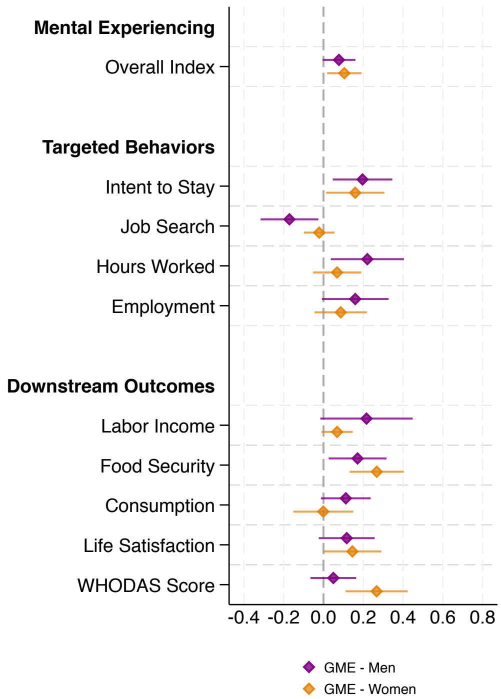


(a) Ethiopia

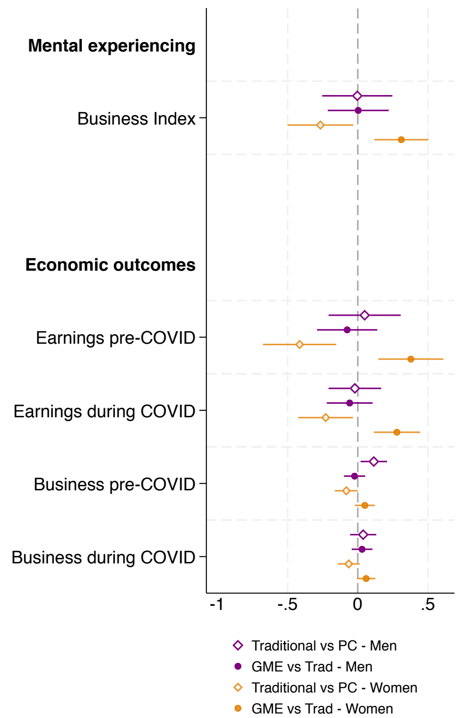
(b) Colombia

Figure 3. 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 on the Impact of Event Score (IES-R) at baseline, or if they report to have been a victim of the civil conflict, or a recent Venezuelan migrant. Outcomes and regression specifications are defined as in Tables 4 and 5. The figure in panel (b) reports 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 people with high (in orange) or low (in purple) trauma. 90 percent confidence intervals are shown for all coefficient estimates.



(a) Ethiopia



(b) Colombia

Figure 4. 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 Tables 4 and 5. 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.

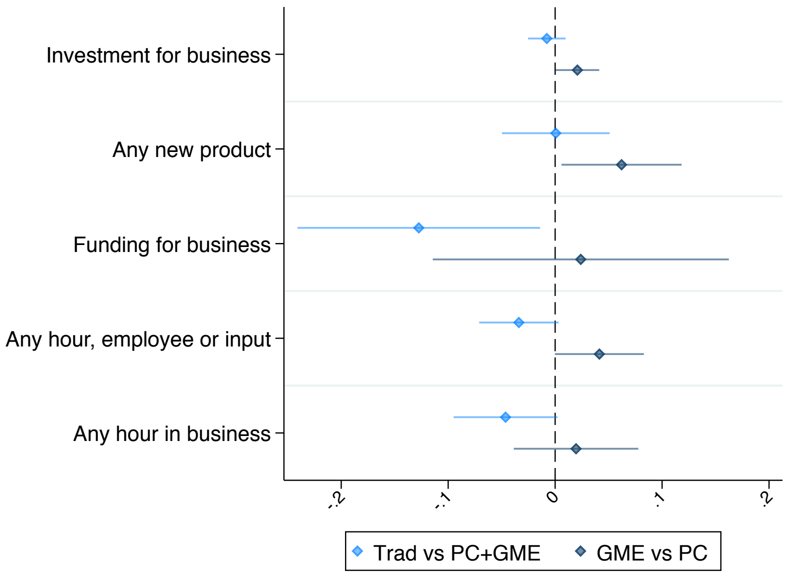


Figure 5. Colombia: a Decline in Motivation in the Traditional Training

Notes: The figure shows results on measures of entrepreneurial action. "Investment for business" is an indicator for having invested in business assets or tools in the last six months, "Any new product" is an indicator for having introduced a new product in the business, "Funding for business" is the standardized unconditional amount of funding obtained for the business, "Any hour, employees or input" is an indicator for having invested any amount of hours, having any employees or having bought new inputs for the business. The light blue coefficients show the effect of the traditional training compared to both the GME and control (pooled). The dark blue coefficients show the comparison between the GME and the control groups. The regression specification is defined as in Tables 4 and 5.

Tables

Table 1. Economic Vulnerability and Trauma in Our Samples

	Experimental Samples		Comparison Samples	
	(1)	(2)	(3)	(4)
	Ethiopia	Colombia	Ethiopia†	Colombia‡
Panel A: Socioeconomic characteristics and economic activity				
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	
Panel B: Post traumatic stress disorder$^{(c)}$				
At risk of PTSD (fraction of sample above threshold)	0.13	0.22	0.04¶	
Panel C: Exposure to trauma (lived or witnessed)				
Any trauma (fraction of sample)	0.98	0.83	0.61** to 0.70*	
Number of traumas				
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**	
Panel D: Exposure to specific types of trauma				
Assault, incl. sexual, weapon (fraction of sample)	0.70	0.61	0.23*	
War, torture, death, displacement (fraction of sample)	0.76	0.48	0.13*	
Life-threatening illness and accidents (fraction of sample)	0.92	0.51	0.34*	
Panel E: Worst trauma lived or witnessed				
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. Ethiopia: GME Treatment Increases The Quality of Mental Experiencing

	Overall Index	Sub-indices		
		Specificity	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Panel A: Main Specification				
GME Treatment	0.091***	0.030	0.132***	0.113***
	(0.030)	(0.044)	(0.048)	(0.039)
		[0.197]	[0.010]	[0.010]
Panel B: PDS Lasso				
GME Treatment	0.092***	0.026	0.132***	0.119***
	(0.030)	(0.044)	(0.048)	(0.038)
		[0.227]	[0.007]	[0.006]
Control Mean	0.00	0.00	-0.01	0.01
Observations	1332	1332	1332	1329

Notes. This table shows the impact of the GME treatment on the quality of mental experiencing, 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 with PDS Lasso. 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. Positivity refers to the frequency of positive images appearing when respondents are presented with neutral scenarios. Further details on the outcomes are reported in Appendix A.3. All the indexes are based on treatment-blind coding done in the back-office, based on participants' recordings.
Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 3. Ethiopia: GME Treatment Improves Targeted Behaviors and Downstream Outcomes

	Targeted Behaviors				Downstream Outcomes			
	Intent to Stay	Job Search	Hours Worked	Employment	Food Security	Consumption	Life Satisfaction	WHODAS Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Main Specification								
GME Treatment	0.178***	-0.097**	0.145**	0.124**	0.219***	0.056	0.131**	0.158***
	(0.053)	(0.042)	(0.056)	(0.054)	(0.051)	(0.050)	(0.052)	(0.049)
	[0.004]	[0.017]	[0.012]	[0.017]	[0.001]	[0.034]	[0.012]	[0.004]
Panel B: PDS Lasso								
GME Treatment	0.171***	-0.099**	0.098**	0.080	0.217***	0.056	0.134***	0.165***
	(0.053)	(0.042)	(0.050)	(0.049)	(0.048)	(0.043)	(0.050)	(0.047)
	[0.003]	[0.016]	[0.030]	[0.046]	[0.001]	[0.070]	[0.010]	[0.002]
Control Mean	-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

Notes. This table shows the effect of the GME treatment on behaviors targeted by the intervention (Columns 1 to 4) and further downstream outcomes (Columns 5 to 9). Panel A shows ITT effects controlling for randomization strata (gender and prime treatment), while Panel B adds controls chosen with PDS Lasso. 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. 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 refers to the standardized number of hours worked in any kind of job. Employment is a standardized dummy equal to 1 if respondents had been engaged in any type of work in the past week. The Food security index is 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. Consumption is an index of total expenditures, excluding those linked with outgoing and ongoing transfers such as remittances and other transfers. 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. Further details on the outcomes are reported in Appendix A.3. Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 4. Colombia: Traditional Training Reduces the Quality of Mental Experiencing, GME Training Restores it

	Overall index	Business index	Non business index	Sub-indices in business domain: Frequency of use	Specificity	Emotionality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Traditional vs Control						
Traditional training	-0.069	-0.157*	-0.049	-0.073	-0.129	-0.122
	(0.075)	(0.089)	(0.074)	(0.086)	(0.080)	(0.081)
				[0.249]	[0.249]	[0.249]
Panel B: GME vs Traditional						
GME treatment	0.026	0.178**	-0.006	0.090	0.133*	0.167**
	(0.063)	(0.074)	(0.061)	(0.072)	(0.068)	(0.065)
				[0.083]	[0.054]	[0.034]
Panel C: GME vs Control						
GME treatment	0.014	0.046	0.004	0.023	0.037	0.044
	(0.066)	(0.076)	(0.064)	(0.073)	(0.070)	(0.069)
				[1.000]	[1.000]	[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	-0.10
N in Control	550	390	550	390	390	390
N in Traditional	656	456	656	456	454	454
N in GME	1140	839	1140	838	835	835

Notes. This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the quality of mental experiencing. Panel A shows the effect of the traditional training compared to the control. Panel B shows the effect of the GME training compared to the traditional. Panel C shows the effect of the GME training compared to the control. The overall index in Column (1) is aggregates the measures of frequency of use, specificity and emotionality. The business (non-business) index is constructed considering only scenarios in (outside) the business domain. Specificity refers to the clarity of the images respondents are able to create. Emotionality refers to the strength of the emotions these images entail. Frequency of use refers to the frequency with which respondents use mental experiencing. All indices are based on self-reports. Further details on the outcomes are reported in Appendix B.3. All regressions control for randomization strata and survey wave fixed effects. Clustered standard errors are in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 5. Colombia Downstream Outcomes: Traditional Training Backfires, GME Training is Trauma Robust

	Overall indices		Pre-covid sub-indices		Covid sub-indices				
	Pre-covid	Covid	Earnings	Business	Earnings	Business	Safety nets	Covid actions	Investment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Traditional vs Control									
Traditional training	-0.157**	-0.090	-0.222**	0.001	-0.144**	-0.022	0.014	-0.074	-0.048
	(0.078)	(0.070)	(0.096)	(0.032)	(0.071)	(0.031)	(0.068)	(0.067)	(0.066)
			[0.167]	[1.000]	[0.167]	[0.909]	[1.000]	[0.832]	[0.909]
Panel B: GME vs Traditional									
GME treatment	0.134**	0.140**	0.191**	0.020	0.141**	0.048*	0.024	0.084	0.066
	(0.066)	(0.059)	(0.083)	(0.027)	(0.060)	(0.025)	(0.057)	(0.058)	(0.053)
			[0.082]	[0.349]	[0.082]	[0.107]	[0.402]	[0.174]	[0.207]
Panel C: GME vs Control									
GME treatment	0.011	0.075	-0.007	0.031	0.040	0.033	0.047	-0.003	0.062
	(0.063)	(0.060)	(0.066)	(0.029)	(0.057)	(0.028)	(0.061)	(0.063)	(0.060)
			[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Mean DV in Control	0.00	0.00	0.00	0.70	0.00	0.70	0.00	0.00	0.00
Mean DV in Traditional	-0.10	-0.10	-0.20	0.70	-0.10	0.60	0.00	-0.10	-0.10
N in Control	346	552	323	333	539	552	546	551	564
N in Traditional	407	660	380	392	642	659	650	657	667
N in GME	704	1147	665	679	1115	1145	1134	1142	1159

Notes. This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on downstream economic outcomes. Panel A shows the effect of the traditional training compared to the control. Panel B shows the effect of the GME training compared to the traditional. Panel C shows the effect of the GME training compared to the control. The overall indices are constructed from the aggregation of the sub-indices listed in Columns (3) to (9). The outcome variable in Columns (3) and (5) is the inverse-hyperbolic sine of earnings (income and sales). In Columns (4) and (6), the outcome is an indicator for having an operating business. The index of safety nets includes both actual (e.g., savings) and perceived safety nets during the Covid pandemic. The index of Covid actions averages different measures that respondents may have taken in response to the pandemic. The investment index is the standardized share of different category investments that a person made. Further details on the outcomes are reported in Appendix B.3. All regressions control for randomization strata and survey wave fixed effects. Clustered standard errors are in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

References

Addis, D. R. (2020). Mental time travel? A neurocognitive model of event simulation. *Review of Philosophy and Psychology 11*(2), 233–259.

Addis, D. R., A. T. Wong, and D. L. Schacter (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia 45*(7), 1363–1377.

Addis, D. R., A. T. Wong, and D. L. Schacter (2008). Age-related changes in the episodic simulation of future events. *Psychological science 19*(1), 33–41.

Alan, S. and S. Ertac (2018). Fostering Patience in the Classroom: Results from Randomized Educational Intervention. *Journal of Political Economy 126*(5), 1865 – 1911.

American Psychiatric Association, D., A. P. Association, et al. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*, Volume 5. American psychiatric association Washington, DC.

Becker, G. S. and C. B. Mulligan (1997). The Endogenous Determination of Time Preference*. *The Quarterly Journal of Economics 112*(3), 729–758.

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.

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.

Betts, A., L. Fryszer, 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.

Betts, A., N. Omata, J. Siu, and O. Sterck (2023). Refugee mobilities in East Africa: understanding secondary movements. *Journal of Ethnic and Migration Studies 49*(11), 2648–2675.

Bisson, J. I. and M. Olff (2021). Prevention and treatment of PTSD: the current evidence base.

Bordalo, P., G. Burro, K. Coffman, N. Gennaioli, and A. Shleifer (2024). Imagining the future: memory, simulation, and beliefs. *Review of Economic Studies*, rdae070.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2016). Stereotypes*. *The Quarterly Journal of Economics 131*(4), 1753–1794.

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

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

Böhm-Bawerk, E. v. (1889). *The positive theory of capital.* London and New York, Macmillan and co.

Bremner, J. D. (2006). Traumatic stress: effects on the brain. *Dialogues in clinical neuroscience 8*(4), 445–461.

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. and E. A. Holmes (2003). Psychological theories of posttraumatic stress disorder. *Clinical psychology review 23*(3), 339–376.

Buckley, T. C., E. B. Blanchard, and W. Neill (2000). Information processing and ptsd: A review of the empirical literature. *Clinical Psychology Review 20*(8), 1041–1065.

Bulley, A. and D. L. Schacter (2023). Episodic future thinking, memory, and decision-making. *Memory in Science for Society: There is Nothing as Practical as a Good Theory*, 123.

Callen, M., M. Isaqzadeh, J. D. Long, and C. Sprenger (2014). Violence and Risk Preference: Experimental Evidence from Afghanistan. *American Economic Review 104*(1), 123–48.

Cantril, H. (1965). *The pattern of human concerns.* New Brunswick, NJ: Rutgers University Press.

Caria, A. S., G. Gordon, M. Kasy, S. Quinn, S. O. Shami, and A. Teytelboym (2024). An adaptive targeted field experiment: Job search assistance for refugees in Jordan. *Journal of the European Economic Association 22*(2), 781–836.

Carranza, E. and D. McKenzie (2024). Job training and job search assistance policies in developing countries. *Journal of Economic Perspectives 38*(1), 221–244.

Center for Substance Abuse Treatment, T. (2014). Trauma-informed care in behavioral health services.

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.

Cenzon, J. (2023). Credit Market Experiences and Macroeconomic Expectations: Evidence and Theory.

Chen, G., S. M. Gully, and D. Eden (2001). Validation of a new general self-efficacy scale. *Organizational research methods 4*(1), 62–83.

Chun, M. M., J. D. Golomb, and N. B. Turk-Browne (2011). A taxonomy of external and internal attention. *Annual review of psychology 62*, 73–101.

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.

Douglas Bremner, J., P. Randall, T. M. Scott, S. Capelli, R. Delaney, G. McCarthy, and D. S. Charney (1995). Deficits in short-term memory in adult survivors of childhood abuse. *Psychiatry Research 59*(1), 97–107.

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.

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.

Gabaix, X. and D. Laibson (2017). Myopia and Discounting. Technical report, National Bureau of Economic Research, Cambridge, MA.

Gennaioli, N. and A. Shleifer (2010). What Comes to Mind. *Quarterly Journal of Economics 125*(4), 1399–1433.

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.

Goenjian, A. K., L. Molina, A. M. Steinberg, L. A. Fairbanks, M. L. Alvarez, H. A. Goenjian, and R. S. Pynoos (2001). Posttraumatic stress and depressive reactions among Nicaraguan adolescents after Hurricane Mitch. *American journal of psychiatry 158*(5), 788–794.

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.

Hackmann, A., J. Bennett-Levy, and E. A. Holmes (2011). *Oxford guide to imagery in cognitive therapy*. OUP Oxford.

Hassabis, D., D. Kumaran, and E. A. Maguire (2007). Using imagination to understand the neural basis of episodic memory. *Journal of neuroscience 27*(52), 14365–14374.

Hassabis, D., D. Kumaran, S. D. Vann, and E. A. Maguire (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences 104*(5), 1726–1731.

Holmes, E. and A. Mathews (2010). Mental imagery in emotion and emotional disorders. *Clinical psychology review 30*, 349–62.

Holmes, E. A., A. Arntz, and M. R. Smucker (2007). Imagery rescripting in cognitive behaviour therapy: Images, treatment techniques and outcomes. *Journal of Behavior Therapy and Experimental Psychiatry 38*(4), 297–305.

Holmes, E. A., S. E. Blackwell, S. Burnett Heyes, F. Renner, and F. Raes (2016). Mental Imagery in Depression: Phenomenology, Potential Mechanisms, and Treatment Implications. *Annual review of clinical psychology 12*(1), 249–280.

Holmes, E. A., A. Ghaderi, E. Eriksson, K. O. Lauri, O. M. Kukacka, M. Mamish, E. L. James, and R. M. Visser (2017). 'I can't concentrate': a feasibility study with young refugees in Sweden on developing science-driven interventions for intrusive memories related to trauma. *Behavioural and Cognitive Psychotherapy 45*(2), 97–109.

Holmes, E. A., S. A. Hales, K. Young, and M. Di Simplicio (2019). *Imagery-Based Cognitive Therapy for Bipolar Disorder and Mood Instability.* New York: The Guilford Press.

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.

Holmes, E. A. and A. Mathews (2005). Mental imagery and emotion: A special relationship? *Emotion 5*(4), 489.

Holmes, E. A., A. Mathews, T. Dalgleish, and B. Mackintosh (2006). Positive interpretation training: Effects of mental imagery versus verbal training on positive mood. *Behavior therapy 37*(3), 237–247.

Holmes, E. A., A. Mathews, B. Mackintosh, and T. Dalgleish (2008). The causal effect of mental imagery on emotion assessed using picture-word cues. *Emotion 8*(3), 395.

Holmes, E. A., R. C. O'Connor, V. H. Perry, I. Tracey, S. Wessely, L. Arseneault, C. Ballard, H. Christensen, R. C. Silver, I. Everall, T. Ford, A. John, T. Kabir, K. King, I. Madan, S. Michie, A. K. Przybylski, R. Shafran, A. Sweeney, C. M. Worthman, L. Yardley, K. Cowan, C. Cope, M. Hotopf, and E. Bullmore (2020). Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry 7*, 547–560.

Howard, M. W. and M. J. Kahana (2002). A distributed representation of temporal context. *Journal of mathematical psychology 46*(3), 269–299.

IDMC (2017). GRID 2017: Global Report on Internal Displacement. Technical report, Internal Displacement Monitoring Centre (IDMC), Geneva, Switzerland.

Insel, T., B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow, and P. Wang (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders.

Iyadurai, L., J. Highfield, M. Kanstrup, A. Markham, V. Ramineni, B. Guo, T. Jaki, J. Kingslake, G. M. Goodwin, C. Summers, et al. (2023). Reducing intrusive memories after trauma via an imagery-competing task intervention in COVID-19 intensive care staff: a randomised controlled trial. *Translational Psychiatry 13*(1), 290.

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.

Jenkins, M., P. Langlais, D. Delis, and R. Cohen (1998). Learning and memory in rape victims with posttraumatic stress disorder. *The American journal of psychiatry 155*, 278–9.

John, A. and K. Orkin (2021). Can Simple Psychological Interventions Increase Preventive Health Investment? *Journal of the European Economic Association, Conditionally accepted.*

Kahana, M. J. (2012). *Foundations of human memory.* OUP USA.

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology 71*, 107–138.

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. and A. Ehlers (2008). Reduced autobiographical memory specificity predicts depression and posttraumatic stress disorder after recent trauma. *Journal of consulting and clinical psychology 76*(2), 231.

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.

Klein, S. B., J. Loftus, and J. F. Kihlstrom (2002). Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition 20*(5), 353–379.

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.

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.

Mathews, A., V. Ridgeway, and E. A. Holmes (2013). Feels like the real thing: Imagery is both more realistic and emotional than verbal thought. *Cognition and Emotion 27*(2), 217–229. PMID: 22780220.

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. (2021). Small business training to improve management practices in developing countries: re-assessing the evidence for 'training doesn't work'. *Oxford Review of Economic Policy 37*, 276–301.

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.

McKenzie, D., C. Woodruff, K. Bjorvatn, M. Bruhn, J. Cai, J. Gonzalez-Uribe, S. Quinn, T. Sonobe, and M. Valdivia (2023). Training Entrepreneurs. *VoxDevLit 1*.

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*.

Moore, S. A. and L. A. Zoellner (2007). Overgeneral autobiographical memory and traumatic events: an evaluative review. *Psychological bulletin 133*(3), 419.

Moya, A. (2018). Violence, psychological trauma, and risk attitudes: Evidence from victims of violence in Colombia. *Journal of Development Economics 131*, 15–27.

Mullainathan, S. (2002). A Memory-Based Model of Bounded Rationality*. *The Quarterly Journal of Economics 117*(3), 735–774.

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.

OECD (2018). *SMEs in Public Procurement : Practices and Strategies for Shared Benefits*. Éditions OCDE, Paris.

Organisation for Economic Cooperation and Development (2020). *Gender, Institutions and Development (Edition 2019) Database*.

Pearson, J., T. Naselaris, E. A. Holmes, and S. M. Kosslyn (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in cognitive sciences 19*(10), 590–602.

Pozuelo, J. R., R. Bradenbrink, M. F. Stierna, and O. Sterck (2023). Depression, violence and socioeconomic outcomes among refugees in East Africa: evidence from a multicountry representative survey. *BMJ Ment Health 26*(1).

R4V (2023). R4V Movements Report: Fourth Quarter 2022. Technical report, Interagency Coordination Platform for Refugees and Migrants (R4V).

Rahman, N. and A. D. Brown (2021). Mental time travel in post-traumatic stress disorder: current gaps and future directions. *Frontiers in Psychology 12*, 624707.

Ramineni, V., P. Millroth, L. Iyadurai, T. Jaki, J. Kingslake, J. Highfield, C. Summers, M. B. Bonsall, and E. A. Holmes (2023). Treating intrusive memories after trauma in healthcare workers: a Bayesian adaptive randomised trial developing an imagery-competing task intervention. *Molecular Psychiatry 28*(7), 2985–2994.

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.

Renner, F., J. L. Ji, A. Pictet, E. A. Holmes, and S. E. Blackwell (2017). Effects of engaging in repeated mental imagery of future positive events on behavioural activation in individuals with major depressive disorder. *Cognitive Therapy and Research 41*, 369–380.

Renner, F., F. Murphy, J. Ji, T. Manly, and E. Holmes (2019). Mental imagery as a "motivational amplifier" to promote activities. *Behaviour Research and Therapy 114*.

Ridley, M., G. Rao, F. Schilbach, and V. Patel (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science 370*(6522), eaay0214.

Rösch, S. A., D. F. Stramaccia, and R. G. Benoit (2022). Promoting farsighted decisions via episodic future thinking: A meta-analysis. *Journal of Experimental Psychology: General 151*(7), 1606.

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 K. K. Szpunar (2017). Escaping the past: Contributions of the hippocampus to future thinking and imagination. *The hippocampus from cells to systems: Structure, connectivity, and functional contributions to memory and flexible cognition*, 439–465.

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.

Scott, J. C., G. E. Matt, K. M. Wrocklage, C. Crnich, J. Jordan, S. M. Southwick, J. H. Krystal, and B. C. Schweinsburg (2015). A quantitative meta-analysis of neurocognitive functioning in posttraumatic stress disorder. *Psychological bulletin 141*(1), 105.

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.

Suddendorf, T. and J. Redshaw (2013). The development of mental scenario building and episodic foresight. *Annals of the New York Academy of Sciences 1296*(1), 135–153.

Talmi, D., L. J. Lohnas, and N. D. Daw (2019). A retrieved context model of the emotional modulation of memory. *Psychological review 126*(4), 455.

Taubinsky, D., L. Butera, M. Saccarola, and C. Lian (2024). Beliefs About the Economy are Excessively Sensitive to Household-Level Shocks: Evidence from Linked Survey and Administrative Data. Technical report, National Bureau of Economic Research.

Thakral, P. P., K. P. Madore, D. R. Addis, and D. L. Schacter (2020). Reinstatement of event details during episodic simulation in the hippocampus. *Cerebral Cortex 30*(4), 2321–2337.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual review of psychology 53*(1), 1–25.

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.

UNHCR (2024). Global Trends: Forced displacement in 2023. Technical report, United Nations High Commissioner for Refugees (UNHCR), Copenhagen, Denmark.

Ventura-Bort, C., A. Löw, J. Wendt, F. Dolcos, A. O. Hamm, and M. Weymar (2016). When neutral turns significant: brain dynamics of rapidly formed associations between neutral stimuli and emotional contexts. *European Journal of Neuroscience 44*(5), 2176–2183.

Voors, M. J., E. E. M. Nillesen, P. Verwimp, E. H. Bulte, R. Lensink, and D. P. V. Soest (2012). Violent Conflict and Behavior: A Field Experiment in Burundi. *American Economic Review 102*(2), 941–964.

Wachter, J. A. and M. J. Kahana (2023). A Retrieved-Context Theory of Financial Decisions*. *The Quarterly Journal of Economics 139*(2), 1095–1147.

Weathers, F. W., B. T. Litz, T. M. Keane, P. A. Palmieri, B. P. Marx, P. P. Schnurr, et al. (2013). The ptsd checklist for dsm-5 (pcl-5).

Weiss, D. and C. Marmar (1997). The Impact of Event Scale—Revised. W: Wilson J., Keane T.(red.). Assessing psychological trauma and PTSD: A handbook for practitioners.

Williams, J. M. G., N. C. Ellis, C. Tyers, H. Healy, G. Rose, and A. K. Macleod (1996). The specificity of autobiographical memory and imageability of the future. *Memory & cognition 24*, 116–125.

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.

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 Details of Experiment 1 (Ethiopia) **2**

A.1 Timeline of Activities and Data Collection 2
A.2 Balance, Take up, and Attrition 2
A.3 Measurement and Construction of Outcome Variables 3
A.4 Cost Effectiveness and Scalability 5
A.5 Appendix A Figures and Tables 6

B Details of Experiment 2 (Colombia) **13**

B.1 Timeline of Activities and Data Collection 13
B.2 Balance, Take up, and Attrition 13
B.3 Measurement and Construction of Outcome Variables 14
B.4 Cost Effectiveness and Scalability 16
B.5 Appendix B Figures and Tables 17

C Additional Results **24**

C.1 Mental Experiencing Measures: Validation 24
C.2 Experiment 1 (Ethiopia) 27
C.3 Experiment 2 (Colombia) 37

D Pre-Analysis Plan (PAP) Reports **48**

A Details of Experiment 1 (Ethiopia)

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

A.1 Timeline of Activities and Data Collection

Figure A1 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.

A.2 Balance, Take up, and Attrition

Balance

Table A2 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 A2 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 A3 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 A4 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 Table A1 below. The attrition rate between baseline and endline in the group assigned to the GME treatment is 5.6 percentage points lower than in the control group ($p = 0.002$).[46] 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.)

Table A1. Survey Attrition Rate, by Treatment

	(1)	(2)	(3)	(4)	(5)
Treatment Status	Overall	Baseline	Endline	Attrition rate	
				(1)-(3)	(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 A5 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 36 percent higher among the treated compared to the control group, but this is only significant at the 10 percent level. In Supplementary materials, 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.

A.3 Measurement and Construction of Outcome Variables

We collected three families of primary outcomes: mental experiencing, economic outcomes and psychological outcomes. We construct our indices within each family following Kling, Liebman, and Katz (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.

Family 1: Mental Experiencing

Our overall mental experiencing 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.

[46] We obtain a p-value from an OLS regression of attrition on the GME treatment indicator.

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 A6 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.

Family 2: Economic outcomes

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

1. Employment outcomes: Job search, engagement in employment and/or self-employment (extensive and intensive margin).

    - 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 outcomes, we use a dummy variable for any search (equal to 1 if a person searched for any of the activities) and 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.
    - 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.

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 seven dimensions: (1) income from economic activities, (2) Food consumption (at home and away from home), dietary diversity, and non-food consumption, (3) Food security, (4) savings, (5) housing, (6) incoming and outgoing transfers, (7) 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 income index includes earnings from wage work (in the last seven days) and profits from self-employment (in the last twelve months).

Family 3: 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 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.

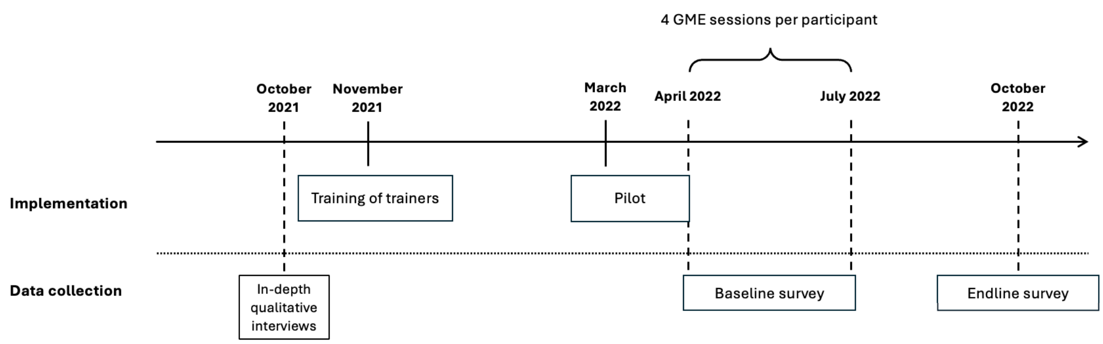
A.4 Cost Effectiveness and Scalability

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 C8). 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.

Table A7 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.

A.5 Appendix A Figures and Tables

Figure A1. 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 A2. Ethiopia: Baseline Balance

Variable	(1) Control	(2) Treatment	(3) Difference
	Demographics		
Age	28.285	28.688	0.403
	(6.685)	(7.182)	(0.341)
Female	0.499	0.500	0.001
	(0.500)	(0.500)	(0.025)
Years of Schooling	10.561	10.528	-0.033
	(3.150)	(3.106)	(0.154)
Household Size	1.794	1.878	0.085
	(1.578)	(1.729)	(0.081)
Years in Addis	2.847	3.006	0.160
	(1.954)	(2.333)	(0.106)
	Employment and economic outcomes		
Engaged in Work	0.189	0.207	0.018
	(0.391)	(0.405)	(0.020)
Hours Worked	7.948	8.330	0.381
	(20.534)	(20.631)	(1.013)
Job Search Hours	4.102	4.986	0.884
	(13.171)	(14.953)	(0.693)
Total Expenditure	4,391.381	4,360.888	-30.494
	(2,039.233)	(2,664.041)	(116.282)
Food Security	16.960	17.218	0.258
	(3.926)	(3.901)	(0.193)
Frequency of Savings	0.296	0.313	0.017
	(0.457)	(0.464)	(0.023)
Total Income	282.336	326.947	44.612
	(974.281)	(1,051.613)	(49.844)
Bank Account	0.952	0.949	-0.003
	(0.214)	(0.221)	(0.011)
Weekly Wage	137.429	139.254	1.825
	(488.720)	(493.654)	(24.180)
	Trauma and mental health		
Number of Traumatic Events	4.770	4.860	0.089
	(2.791)	(2.917)	(0.140)
PTSD Score	15.007	15.162	0.155
	(13.933)	(13.970)	(0.687)
Life Satisfaction	6.729	6.660	-0.069
	(1.582)	(1.594)	(0.078)
Overall Health	0.726	0.722	-0.004
	(0.111)	(0.108)	(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 A3. Ethiopia: Intervention Take-up

Variable	(1) No Take-up	(2) Take-up	(3) Difference
	Demographics		
Age	27.446	29.109	1.664***
	(6.352)	(7.401)	(0.582)
Female	0.673	0.441	-0.232***
	(0.470)	(0.497)	(0.040)
Years of Schooling	10.584	10.508	-0.076
	(3.280)	(3.047)	(0.253)
Household Size	1.807	1.903	0.096
	(1.715)	(1.734)	(0.141)
Years in Addis	2.931	3.032	0.101
	(1.694)	(2.514)	(0.190)
	Employment and economic outcomes		
Engaged in Work	0.203	0.208	0.005
	(0.403)	(0.406)	(0.033)
Hours Worked	8.168	8.384	0.216
	(20.272)	(20.768)	(1.681)
Job Search Hours	4.045	5.304	1.259
	(14.341)	(15.129)	(1.216)
Total Expenditures	4,841.984	4,197.832	-644.153***
	(3,929.763)	(2,045.594)	(215.824)
Food Security	18.025	16.945	-1.080***
	(3.616)	(3.958)	(0.315)
Frequency of Savings	0.351	0.300	-0.051
	(0.479)	(0.459)	(0.038)
Total Income	305.953	334.063	28.110
	(1,140.837)	(1,020.522)	(85.665)
Bank Account	0.950	0.948	-0.003
	(0.217)	(0.222)	(0.018)
Weekly Wage	112.450	148.339	35.888
	(431.520)	(513.027)	(40.196)
	Trauma and mental health		
Number of Traumatic Events	4.436	5.003	0.568**
	(2.849)	(2.928)	(0.237)
PTSD Score	13.460	15.738	2.278**
	(12.467)	(14.408)	(1.135)
Life Satisfaction	6.807	6.611	-0.196
	(1.451)	(1.637)	(0.130)
Overall Health	0.732	0.719	-0.013
	(0.106)	(0.109)	(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 A2.

Table A4. Ethiopia: Survey Attrition at Endline

Variable	(1) Did not drop	(2) Dropped	(3) Difference
	Treatment Status		
GME Treatment	0.500	0.393	-0.107***
	(0.500)	(0.489)	(0.034)
	Demographics		
Age	28.619	27.724	-0.895*
	(7.018)	(6.392)	(0.470)
Female	0.497	0.514	0.017
	(0.500)	(0.501)	(0.034)
Years of Schooling	10.578	10.366	-0.212
	(3.094)	(3.311)	(0.212)
Household Size	1.853	1.735	-0.118
	(1.678)	(1.505)	(0.112)
Years in Addis	2.905	3.027	0.123
	(2.192)	(1.878)	(0.146)
	Employment and economic outcomes		
Engaged in Work	0.197	0.198	0.001
	(0.398)	(0.400)	(0.027)
Hours Worked	8.025	8.716	0.691
	(20.440)	(21.326)	(1.397)
Job Search Hours	4.747	3.354	-1.393
	(14.188)	(13.337)	(0.955)
Total Expenditures	4,312.418	4,725.307	412.889***
	(2,050.704)	(3,592.882)	(160.001)
Food Security	17.006	17.514	0.508*
	(3.887)	(4.042)	(0.266)
Frequency of Savings	0.292	0.370	0.077**
	(0.455)	(0.484)	(0.031)
Total Income	296.698	342.898	46.200
	(994.441)	(1,105.554)	(68.729)
Bank Account	0.953	0.934	-0.020
	(0.211)	(0.249)	(0.015)
Weekly Wage	130.952	178.255	47.303
	(463.406)	(619.102)	(33.317)
	Trauma and mental health		
Number of Traumatic Events	4.812	4.821	0.009
	(2.841)	(2.914)	(0.194)
PTSD Score	15.169	14.607	-0.562
	(14.006)	(13.636)	(0.947)
Life Satisfaction	6.681	6.774	0.093
	(1.586)	(1.599)	(0.108)
Overall Health	0.725	0.718	-0.008
	(0.108)	(0.119)	(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 A2.

Table A5. Ethiopia: Baseline Characteristics of Respondents at Endline, by Treatment

Variable	(1) Control	(2) Treatment	(3) Difference
Demographics			
Age	28.348	28.890	0.541
	(6.718)	(7.301)	(0.376)
Female	0.497	0.496	-0.001
	(0.500)	(0.500)	(0.027)
Years of Schooling	10.613	10.542	-0.071
	(3.094)	(3.095)	(0.166)
Household Size	1.804	1.902	0.099
	(1.589)	(1.762)	(0.090)
Years in Addis	2.785	3.024	0.239**
	(1.952)	(2.405)	(0.117)
Employment and economic outcomes			
Engaged in Work	0.182	0.212	0.030
	(0.386)	(0.409)	(0.021)
Hours Worked	7.640	8.410	0.770
	(20.336)	(20.552)	(1.095)
Job Search Hours	4.163	5.331	1.168
	(12.812)	(15.429)	(0.761)
Total Expenditures	4,343.772	4,281.020	-62.753
	(2,052.544)	(2,049.853)	(109.838)
Food Security	16.933	17.079	0.146
	(3.885)	(3.891)	(0.208)
Bank Account	0.957	0.950	-0.007
	(0.203)	(0.219)	(0.011)
Frequency of Savings	0.282	0.303	0.020
	(0.450)	(0.460)	(0.024)
Weekly Wage	125.642	136.270	10.628
	(442.982)	(483.255)	(24.822)
Total Income	251.339	342.122	90.783*
	(891.976)	(1,086.174)	(53.214)
Trauma and mental health			
Life Satisfaction	6.717	6.646	-0.071
	(1.586)	(1.585)	(0.085)
Overall Health	0.727	0.723	-0.004
	(0.110)	(0.106)	(0.006)
Number of Traumatic Events	4.716	4.908	0.192
	(2.752)	(2.927)	(0.152)
PTSD Score	15.159	15.179	0.020
	(14.141)	(13.880)	(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 A2.

Table A6. Measurement: Mental Experiencing in Ethiopia Study

<table>
<tr><th>Reference
Period</th><th>Scenarios</th><th>Questions for
Each Scenario</th></tr>
<tr><td>Past</td><td>Respondents are presented with the following scenarios in randomized order. For each scenario, they are asked to think of a specific event that has happened.

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)</td><td rowspan="2">Specificity
(1 “No memory at all’
to 5 ”Very specific”)

Emotionality
(1 “No emotion at all’
to 5 ”Extremely strong emotions”)

Positivity
(1 “Very negative’
to 4 ”Very positive”)</td></tr>
<tr><td>Future</td><td>Respondents are presented with the following scenarios in randomized order. For each scenario, they are asked to think of a specific event that could happen in the 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)</td></tr>
</table>

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 A7. 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.

B Details of Experiment 2 (Colombia)

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

B.1 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 B1 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.

B.2 Balance, Take up, and Attrition

Balance

Table B1 shows that treatment assignment is balanced on observables for the full sample of randomized participants. Results are the same when limiting to only the sample who accepted to participate in the training, which is the one used for the analysis (Table B2). We also fail to reject the null of a joint orthogonality test of the significance of all the variables shown in Table B1 on treatment assignment (F(15,1818)=0.65, p=0.83).

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 B3 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 B4 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 pure 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 B5 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. Tables C15 and C16 show that the results are robust when adding inverse-probability weighting to our regressions. In Supplementary materials, to further deal with attrition, we use Lee Bounds to correct for differential attrition between the treatments and the control.

B.3 Measurement and Construction of Outcome Variables

We collected three families of primary outcomes: mental experiencing, 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.

Family 1: Mental Experiencing

Our overall mental experiencing 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 experiencing 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 experiencing 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 B6.

Family 2: Economic Outcomes

We construct economic indices to capture downstream treatment effects and 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 99^{th} 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. As this index does not clearly distinguish between the pre-COVID and COVID period, we include it within the COVID period outcomes;
- Safety nets: this index includes two sub-indexes, one for savings in the pre-Covid period (actual safety nets) and one for their 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.
- Covid actions: 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.

Family 3: Psychological resilience

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

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

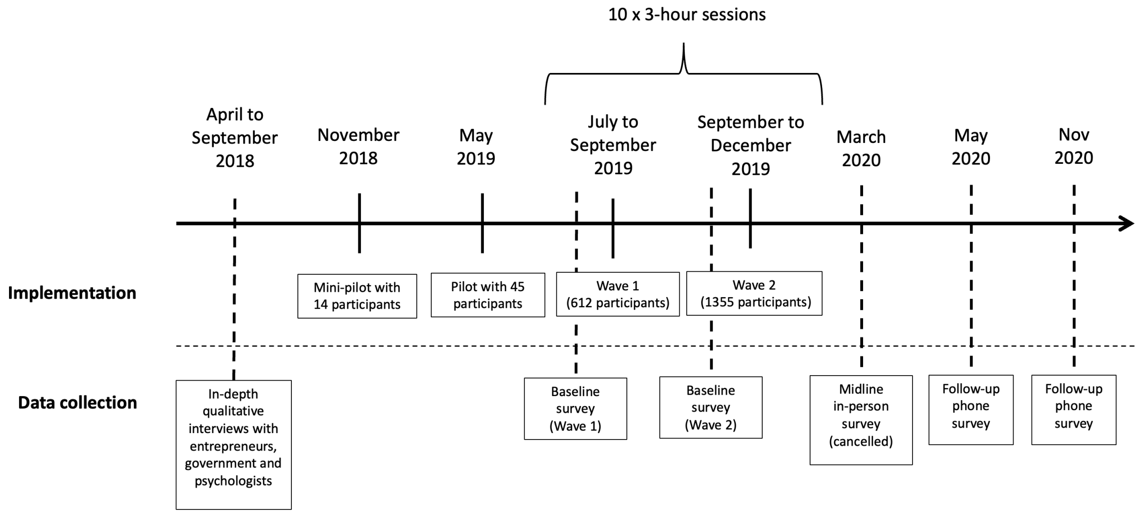
B.4 Cost Effectiveness and Scalability

Table B7 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.

B.5 Appendix B Figures and Tables

Figure B1. 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 B1. Colombia: Baseline Balance

	(1)	(2)	(3)	(4)	(5)	(6)
		Means			Pairwise differences	
Variable	GME	Traditional	Control	GME-Trad	GME-C	Trad-C
			Demographics			
Female	0.578	0.580	0.571	-0.002	0.007	0.009
	(0.494)	(0.494)	(0.496)	(0.951)	(0.804)	(0.782)
Age (18-28)	0.604	0.571	0.629	0.033	-0.025	-0.058*
	(0.489)	(0.495)	(0.484)	(0.219)	(0.382)	(0.069)
Age (29-45)	0.240	0.250	0.272	-0.011	-0.033	-0.022
	(0.427)	(0.434)	(0.446)	(0.642)	(0.209)	(0.450)
Age (46-59)	0.092	0.113	0.094	-0.021	-0.002	0.019
	(0.289)	(0.317)	(0.292)	(0.202)	(0.891)	(0.339)
Years of Education	13.132	13.046	13.366	0.086	-0.234	-0.320
	(3.344)	(3.666)	(3.135)	(0.690)	(0.273)	(0.196)
			Business and assets			
Only Business Owner	0.242	0.245	0.236	-0.003	0.006	0.009
	(0.428)	(0.430)	(0.425)	(0.895)	(0.825)	(0.757)
Only Have a Business Idea	0.491	0.488	0.494	0.003	-0.003	-0.006
	(0.500)	(0.500)	(0.501)	(0.901)	(0.925)	(0.850)
Have Business and Business Idea	0.267	0.266	0.270	0.002	-0.003	-0.004
	(0.443)	(0.442)	(0.444)	(0.948)	(0.916)	(0.881)
Income $\leq$ min wage	0.667	0.663	0.641	0.004	0.026	0.022
	(0.472)	(0.473)	(0.480)	(0.884)	(0.364)	(0.483)
Income > min wage	0.240	0.230	0.248	0.010	-0.009	-0.018
	(0.427)	(0.421)	(0.432)	(0.673)	(0.734)	(0.511)
Assets Owned	11.681	11.794	11.835	-0.114	-0.155	-0.041
	(3.853)	(3.915)	(3.782)	(0.632)	(0.542)	(0.885)
Household Size	3.484	3.579	3.650	-0.095	-0.165	-0.070
	(1.556)	(1.599)	(1.713)	(0.325)	(0.137)	(0.567)
Save Monthly (Y/N)	0.552	0.531	0.580	0.020	-0.028	-0.048
	(0.498)	(0.500)	(0.494)	(0.507)	(0.396)	(0.186)
Had Access to Credit (Y/N)	0.156	0.143	0.138	0.014	0.018	0.004
	(0.364)	(0.350)	(0.345)	(0.522)	(0.433)	(0.861)
			Trauma and mental health			
Number of Traumatic Events	2.123	2.295	2.308	-0.163	-0.174	-0.002
	(2.296)	(2.422)	(2.539)	(0.209)	(0.234)	(0.991)
High Trauma Symptoms (IES > 33)	0.244	0.231	0.231	0.012	0.012	-0.000
	(0.430)	(0.422)	(0.422)	(0.635)	(0.669)	(0.996)
Trauma Symptoms (IES)	17.945	17.650	17.785	0.295	0.159	-0.136
	(22.442)	(22.592)	(20.958)	(0.831)	(0.912)	(0.932)
Distress Score (Kessler)	13.170	13.232	13.138	-0.062	0.032	0.094
	(4.213)	(4.021)	(3.970)	(0.803)	(0.906)	(0.748)
Observations	906	531	415	1,439	1,323	948

Notes: The table presents a balance test at baseline for all three treatment arms. 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 with significance indicated as follows: * $p<0.10$, ** $p<0.05$, *** $p<0.01$. "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 B2. Colombia: Baseline Balance among Those who Accepted Training

	(1)	(2)	(3)	(4)	(5)	(6)
		Means			Pairwise differences	
Variable	GME	Traditional	Control	GME-Trad	GME-C	Trad-C
			Demographics			
Female	0.574	0.579	0.571	-0.006	0.003	0.008
	(0.495)	(0.494)	(0.496)	(0.847)	(0.934)	(0.809)
Age (18-28)	0.591	0.573	0.629	0.019	-0.038	-0.056*
	(0.492)	(0.495)	(0.484)	(0.522)	(0.201)	(0.090)
Age (29-45)	0.245	0.249	0.272	-0.004	-0.027	-0.023
	(0.431)	(0.433)	(0.446)	(0.869)	(0.312)	(0.443)
Age (46-59)	0.096	0.119	0.094	-0.024	0.002	0.025
	(0.294)	(0.325)	(0.292)	(0.197)	(0.927)	(0.224)
Years of Education	13.157	13.035	13.366	0.122	-0.209	-0.331
	(3.315)	(3.590)	(3.135)	(0.587)	(0.333)	(0.189)
			Business and assets			
Only Business Owner	0.245	0.245	0.236	0.000	0.009	0.009
	(0.431)	(0.431)	(0.425)	(0.995)	(0.724)	(0.757)
Only Have a Business Idea	0.477	0.477	0.494	-0.000	-0.017	-0.017
	(0.500)	(0.500)	(0.501)	(0.987)	(0.569)	(0.621)
Have Business and Business Idea	0.278	0.275	0.270	0.002	0.008	0.006
	(0.448)	(0.447)	(0.444)	(0.924)	(0.764)	(0.853)
Income $\leq$ min wage	0.665	0.677	0.641	-0.011	0.024	0.036
	(0.472)	(0.468)	(0.480)	(0.679)	(0.398)	(0.265)
Income $>$ min wage	0.238	0.221	0.248	0.016	-0.010	-0.027
	(0.426)	(0.416)	(0.432)	(0.502)	(0.688)	(0.349)
Assets Owned	11.661	11.894	11.835	-0.233	-0.175	0.058
	(3.839)	(3.962)	(3.782)	(0.354)	(0.498)	(0.841)
Household Size	3.452	3.609	3.650	-0.157	-0.197*	-0.041
	(1.506)	(1.607)	(1.713)	(0.120)	(0.077)	(0.745)
Save Monthly (Y/N)	0.549	0.535	0.580	0.014	-0.031	-0.045
	(0.498)	(0.499)	(0.494)	(0.661)	(0.359)	(0.230)
Had Access to Credit (Y/N)	0.153	0.150	0.138	0.003	0.015	0.012
	(0.360)	(0.358)	(0.345)	(0.909)	(0.531)	(0.644)
			Trauma and mental health			
Number of Traumatic Events	2.247	2.342	2.308	-0.090	-0.056	0.040
	(2.330)	(2.367)	(2.539)	(0.513)	(0.710)	(0.812)
High Trauma Symptoms (IES > 33)	0.242	0.228	0.231	0.014	0.010	-0.004
	(0.429)	(0.420)	(0.422)	(0.609)	(0.718)	(0.912)
Trauma Symptoms (IES)	17.645	17.142	17.785	0.503	-0.140	-0.643
	(22.273)	(22.005)	(20.958)	(0.724)	(0.923)	(0.691)
Distress Score (Kessler)	13.098	13.187	13.138	-0.090	-0.040	0.050
	(4.102)	(3.980)	(3.970)	(0.728)	(0.882)	(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). 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. Variables are defined in the note to table B1.

Table B3. Colombia: Intervention Take-up

Variable	(1) No Take-up	(2) Take-up	(3) Difference
		Demographics	
Female	0.611	0.553	-0.058**
	(0.488)	(0.497)	(0.028)
Age (18-28)	0.623	0.565	-0.058**
	(0.485)	(0.496)	(0.027)
Age (29-45)	0.255	0.236	-0.020
	(0.436)	(0.425)	(0.397)
Age (46-59)	0.070	0.124	0.054***
	(0.255)	(0.330)	(0.001)
Years of Education	13.019	13.111	0.091
	(3.656)	(3.362)	(0.675)
		Business and assets	
Only Business Owner	0.221	0.261	0.039*
	(0.415)	(0.439)	(0.084)
Only Have a Business Idea	0.522	0.464	-0.058**
	(0.500)	(0.499)	(0.030)
Have Business and Business Idea	0.257	0.274	0.017
	(0.437)	(0.446)	(0.465)
Income $\leq$ min wage	0.650	0.677	0.027
	(0.477)	(0.468)	(0.286)
Income $>$ min wage	0.244	0.233	-0.011
	(0.430)	(0.423)	(0.641)
Assets Owned	11.654	11.759	0.105
	(3.918)	(3.863)	(0.661)
Household Size	3.600	3.486	-0.114
	(1.659)	(1.518)	(0.250)
Save Monthly (Y/N)	0.523	0.551	0.028
	(0.500)	(0.498)	(0.363)
Had Access to Credit (Y/N)	0.146	0.154	0.008
	(0.354)	(0.361)	(0.723)
		Trauma and Mental Health	
Number of Traumatic Events	1.875	2.421	0.546***
	(2.390)	(2.274)	(0.000)
High Trauma Symptoms (IES>33)	0.275	0.217	-0.059**
	(0.447)	(0.412)	(0.029)
Trauma Symptoms (IES)	20.269	16.296	-3.974***
	(24.340)	(21.157)	(0.006)
Distress Score (Kessler)	13.776	12.834	-0.941***
	(4.607)	(3.797)	(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 B1.

Table B4. Colombia: Survey Attrition at Midline and endline

	(1)	(2)	(3)	(4)	(5)	(6)
	Participated in Midline			Participated in Endline		
Variable	No	Yes	Diff	No	Yes	Diff
			Treatments			
GME Treatment	0.446	0.485	0.039	0.457	0.486	0.029
	(0.498)	(0.500)	(0.177)	(0.499)	(0.500)	(0.257)
Traditional Treatment	0.273	0.277	0.003	0.263	0.283	0.019
	(0.446)	(0.447)	(0.897)	(0.441)	(0.450)	(0.394)
Control	0.278	0.239	-0.039	0.278	0.232	-0.047**
	(0.449)	(0.427)	(0.122)	(0.449)	(0.422)	(0.038)
Baseline Participation	0.672	0.866	0.194***	0.716	0.875	0.159***
	(0.470)	(0.341)	(0.000)	(0.451)	(0.330)	(0.000)
Total Attendance	2.407	3.756	1.350***	2.577	3.892	1.315***
	(3.241)	(3.699)	(0.000)	(3.250)	(3.752)	(0.000)
Wave 1 Participant	0.359	0.317	-0.042	0.318	0.332	0.014
	(0.480)	(0.466)	(0.126)	(0.466)	(0.471)	(0.549)
			Demographics			
Female	0.585	0.571	-0.014	0.558	0.584	0.026
	(0.493)	(0.495)	(0.613)	(0.497)	(0.493)	(0.311)
Age (18-28)	0.593	0.596	0.003	0.619	0.583	-0.036
	(0.492)	(0.491)	(0.908)	(0.486)	(0.493)	(0.146)
Age (29-45)	0.276	0.246	-0.031	0.241	0.259	0.018
	(0.448)	(0.431)	(0.231)	(0.428)	(0.439)	(0.417)
Age (46-59)	0.058	0.115	0.058***	0.082	0.113	0.031**
	(0.234)	(0.320)	(0.000)	(0.274)	(0.316)	(0.037)
Years of Education	13.794	13.021	-0.772***	13.483	13.034	-0.449**
	(3.390)	(3.329)	(0.001)	(3.338)	(3.353)	(0.022)
			Business and assets			
Only Business Owner	0.198	0.257	0.058**	0.221	0.255	0.034
	(0.399)	(0.437)	(0.013)	(0.415)	(0.436)	(0.119)
Only Have a Business Idea	0.485	0.480	-0.005	0.474	0.485	0.010
	(0.500)	(0.500)	(0.863)	(0.500)	(0.500)	(0.688)
Have Business and Business Idea	0.317	0.262	-0.054**	0.304	0.259	-0.045*
	(0.466)	(0.440)	(0.040)	(0.461)	(0.439)	(0.053)
Income $\leq$ min wage	0.580	0.688	0.108***	0.616	0.688	0.072***
	(0.494)	(0.463)	(0.000)	(0.487)	(0.464)	(0.003)
Income $>$ min wage	0.289	0.219	-0.070***	0.270	0.217	-0.053**
	(0.454)	(0.414)	(0.006)	(0.445)	(0.412)	(0.016)
Assets Owned	12.328	11.632	-0.697***	12.209	11.572	-0.637***
	(4.005)	(3.813)	(0.010)	(4.021)	(3.771)	(0.006)
Household Size	3.278	3.608	0.330***	3.431	3.593	0.162*
	(1.544)	(1.592)	(0.002)	(1.606)	(1.577)	(0.083)
Save Monthly (Y/N)	0.539	0.555	0.016	0.563	0.548	-0.015
	(0.499)	(0.497)	(0.635)	(0.497)	(0.498)	(0.597)
Had Access to Credit (Y/N)	0.161	0.146	-0.015	0.138	0.153	0.015
	(0.368)	(0.353)	(0.535)	(0.345)	(0.360)	(0.460)
			Trauma and mental health			
Number of Traumatic Events	1.559	2.515	0.956***	1.826	2.538	0.712***
	(2.016)	(2.455)	(0.000)	(2.184)	(2.463)	(0.000)
High Trauma Symptoms (IES>33)	0.207	0.242	0.036	0.225	0.240	0.014
	(0.406)	(0.429)	(0.203)	(0.418)	(0.427)	(0.558)
Trauma Symptoms (IES)	15.369	18.063	2.694*	16.911	17.812	0.901
	(21.283)	(21.994)	(0.066)	(21.663)	(21.975)	(0.481)
Distress Score (Kessler)	13.246	13.105	-0.141	13.249	13.081	-0.169
	(4.130)	(4.012)	(0.615)	(4.056)	(4.025)	(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 B1.

Table B5. Colombia: Differences across Treatments in the Baseline Characteristics of Respondents at Midline and Endline

	(1)	(2)	(3)	(4)	(5)	(6)
	Midline differences			Endline differences		
Variable	GME-Trad	GME-PC	Trad-PC	GME-Trad	GME-PC	Trad-PC
		Demographics				
Female	-0.044	-0.013	0.030	-0.038	-0.027	0.011
	(0.187)	(0.705)	(0.432)	(0.283)	(0.474)	(0.794)
Age (18-28)	0.031	-0.001	-0.031	0.061*	-0.013	-0.074*
	(0.352)	(0.983)	(0.416)	(0.088)	(0.737)	(0.080)
Age (29-45)	-0.026	-0.049	-0.023	-0.018	-0.047	-0.029
	(0.366)	(0.108)	(0.498)	(0.564)	(0.166)	(0.442)
Age (46-59)	-0.033	-0.011	0.021	-0.050**	0.009	0.059**
	(0.139)	(0.605)	(0.413)	(0.042)	(0.685)	(0.033)
Years of Education	-0.129	-0.184	-0.055	0.003	-0.235	-0.237
	(0.600)	(0.449)	(0.845)	(0.992)	(0.384)	(0.444)
		Business and assets				
Only Business Owner	-0.003	0.021	0.024	-0.003	0.020	0.022
	(0.912)	(0.492)	(0.479)	(0.927)	(0.555)	(0.544)
Only Have a Business Idea	0.012	-0.022	-0.034	-0.006	-0.009	-0.003
	(0.715)	(0.533)	(0.384)	(0.869)	(0.814)	(0.942)
Have Business and Business Idea	-0.006	0.001	0.007	0.012	-0.010	-0.023
	(0.836)	(0.972)	(0.835)	(0.699)	(0.758)	(0.547)
Income $\leq$ min wage	-0.010	0.014	0.025	0.038	0.012	-0.026
	(0.744)	(0.659)	(0.501)	(0.257)	(0.729)	(0.517)
Income > min wage	0.018	0.005	-0.013	-0.019	0.006	0.025
	(0.514)	(0.867)	(0.686)	(0.525)	(0.854)	(0.481)
Assets Owned	-0.333	-0.217	0.117	-0.204	-0.156	0.048
	(0.227)	(0.448)	(0.717)	(0.488)	(0.609)	(0.888)
Household size	-0.132	-0.232*	-0.100	-0.111	-0.199	-0.088
	(0.233)	(0.070)	(0.483)	(0.345)	(0.153)	(0.564)
Save Monthly (Y/N)	0.017	0.002	-0.015	0.022	-0.012	-0.033
	(0.632)	(0.955)	(0.723)	(0.569)	(0.778)	(0.463)
Had Access to Credit (Y/N)	0.003	0.013	0.009	-0.011	0.001	0.013
	(0.893)	(0.628)	(0.750)	(0.682)	(0.960)	(0.696)
		Trauma and mental health				
Number of Traumatic Events	-0.104	-0.084	0.020	-0.095	-0.112	-0.017
	(0.522)	(0.637)	(0.919)	(0.584)	(0.567)	(0.938)
High Trauma Symptoms (IES>33)	0.012	0.012	0.001	0.012	0.002	-0.010
	(0.706)	(0.705)	(0.982)	(0.705)	(0.953)	(0.792)
Trauma Symptoms (IES)	0.046	-0.403	-0.448	0.109	-0.561	-0.670
	(0.977)	(0.807)	(0.809)	(0.949)	(0.753)	(0.734)
Distress Score (Kessler)	-0.271	-0.120	0.150	-0.184	-0.085	0.099
	(0.348)	(0.692)	(0.659)	(0.544)	(0.801)	(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 B1.

Table B6. Measurement: Mental Experiencing

Mental experiencing Questions	
Indexes	**Question(s)**
Frequency of use	"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: • 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	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. **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"." **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?" • 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.

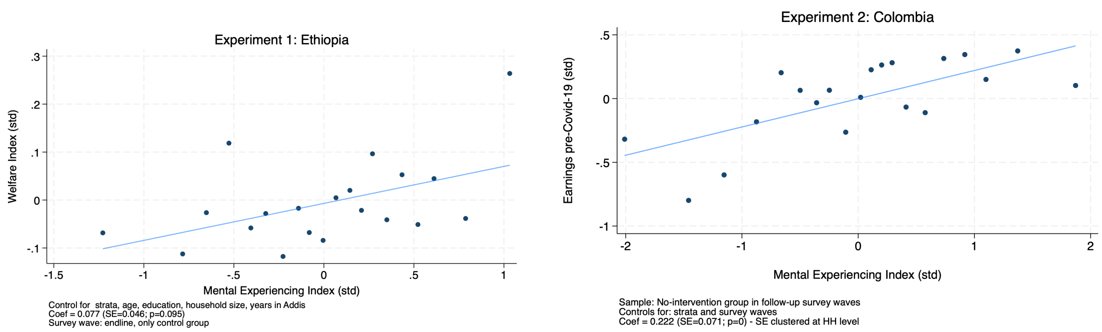
Table B7. Colombia: Detailed cost estimates

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	

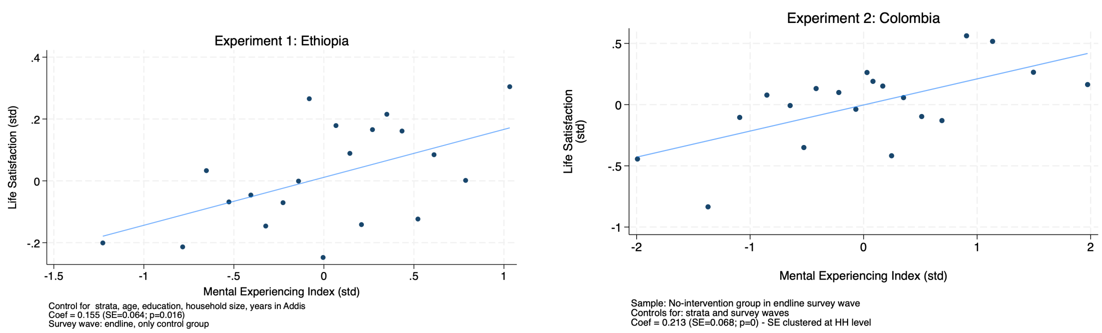
Notes: The table presents a detailed cost estimates following the J-PAL costing template for cost effectiveness analysis. We assume a COP/USD exchange rate of 3403 and 2019 costs are adjusted to 2022 USD using a CPI-based inflation adjustment factor of 1.169, derived from the World Bank WDI database.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

C Additional Results

C.1 Mental Experiencing Measures: Validation



(a) Mental Experiencing and Economic Welfare



(b) Mental Experiencing and Life Satisfaction

Figure C1. Mental Experiencing and Economic Outcomes

Notes: The figure shows binned scatter plots for the correlation between our mental experiencing index and different measures of economic outcomes. 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 for Colombia. For Ethiopia, mental experiencing measures are back-coded, and include specificity, emotionality and positivity. For Colombia, mental experiencing 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 is for the pre-Covid 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.

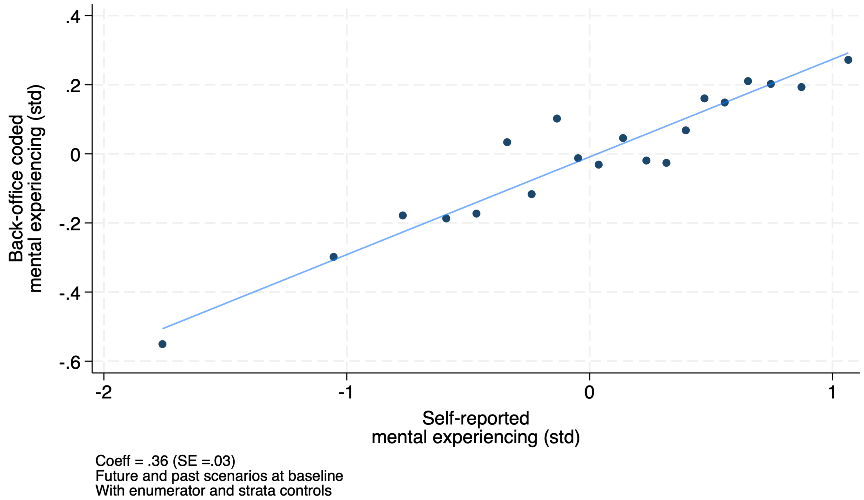
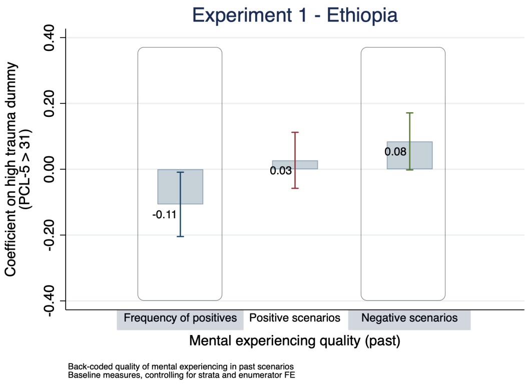
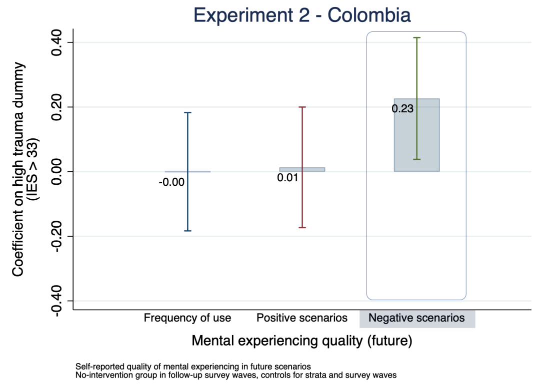


Figure C2. Mental Experiencing: Correlation between Back-Coded and Self-Reported Measures

Notes: The figure shows a binned scatter plot for the correlation between back-coded and self-reported measures of mental experiencing from the Ethiopia experiment. 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. Results are unchanged when removing the enumerator's fixed effects.



(a) Ethiopia



(b) Colombia

Figure C3. Mental Experiencing Quality and Trauma

Notes: The figure reports coefficients on an indicator variable for "High trauma" symptoms for different measures of mental experiencing: 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 experiencing 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.2 Experiment 1 (Ethiopia)

Table C1. Ethiopia: PCA Index for Mental Experiencing Quality

	Overall Index	Sub-indices		
		Specificity	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Panel A: Main Specification				
GME Treatment	0.154**	0.045	0.133**	0.162**
	(0.062)	(0.061)	(0.059)	(0.063)
Panel B: PDS Lasso				
GME Treatment	0.147**	0.034	0.127**	0.142**
	(0.061)	(0.060)	(0.058)	(0.062)
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 experiencing, 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) using mental experiencing variables that include 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. Ethiopia: Mental Experiencing Quality Interacted with Social Desirability Score

	Overall Index	Sub-indices		
		Specification	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Main Specification				
GME Treatment	0.091**	0.029	0.130**	0.115**
	(0.030)	(0.044)	(0.048)	(0.039)
Social Desirability	0.008	-0.001	-0.014	0.040
	(0.020)	(0.031)	(0.031)	(0.026)
GME Treatment × Social desirability	-0.015	-0.030	-0.060	0.044
	(0.029)	(0.042)	(0.046)	(0.037)
p-value: GME Treat. + GME Treat. x Social Desirability = 0	0.069	0.992	0.290	0.002
Control Mean	0.00	0.00	-0.01	0.01
Observations	1332	1332	1332	1329

Notes. This table shows the impact of the GME treatment on the quality of mental experiencing, 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 outcome measures for mental experiencing quality in this table are defined as in Table 2. Social desirability is based on the Marlowe–Crowne Social Desirability Scale, which measures the individual propensity for social desirability bias. The score was measured at endline. Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C3. Ethiopia: Economic Outcomes Interacted with Social Desirability Score

	Targeted Behaviors				Downstream Outcomes			
	Intent to Stay	Job Search	Hours Worked	Employment	Food Security	Consumption	Life Satisfaction	WHODAS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Specification								
GME Treatment	0.179***	-0.095**	0.147***	0.082	0.218***	0.046	0.135***	0.157***
	(0.053)	(0.042)	(0.056)	(0.052)	(0.051)	(0.031)	(0.051)	(0.049)
	[0.005]	[0.046]	[0.026]	[0.126]	[0.001]	[0.131]	[0.026]	[0.006]
Social Desirability	-0.059	0.079**	0.038	0.047	-0.077**	-0.088***	0.125***	0.066*
	(0.038)	(0.032)	(0.037)	(0.036)	(0.037)	(0.024)	(0.036)	(0.038)
	[0.126]	[0.033]	[0.224]	[0.162]	[0.069]	[0.003]	[0.005]	[0.106]
GME Treatment x Social Desirability	0.098*	-0.002	0.036	0.025	0.020	0.004	0.021	-0.091*
	(0.052)	(0.040)	(0.054)	(0.051)	(0.047)	(0.031)	(0.052)	(0.047)
	[0.078]	[0.390]	[0.271]	[0.331]	[0.331]	[0.387]	[0.331]	[0.078]
GME Treat.+ GME Treat. x Social Desirability = 0	0.00	0.12	0.03	0.16	0.00	0.24	0.03	0.33
Control Mean	-0	-0	-0	-0	0	-0	-0	-0
Observations	1395	1395	1395	1395	1395	1395	1395	1395

Notes. This table reproduces Table 3 of the main paper, which shows the effect of the GME treatment on employment and wellbeing outcomes. In this table, the treatment indicator GME is interacted with the score on the Marlowe–Crowne Social Desirability Scale, which measures the individual propensity for social desirability bias. The score was measured at endline. Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C4. Ethiopia: Economic Outcomes by Baseline Trauma

	Targeted Behaviors				Downstream Outcomes			
	Intent to Stay	Job Search	Hours Worked	Employment	Food Security	Consumption	Life Satisfaction	WHODAS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Specification								
GME Treatment	0.182***	-0.093**	0.086	0.062	0.202***	0.062*	0.072	0.122**
	(0.058)	(0.047)	(0.061)	(0.056)	(0.053)	(0.035)	(0.056)	(0.052)
	[0.010]	[0.074]	[0.201]	[0.260]	[0.002]	[0.117]	[0.224]	[0.036]
High Trauma	0.044	0.028	-0.186**	0.027	-0.439***	-0.016	-0.443***	-0.373***
	(0.107)	(0.076)	(0.079)	(0.107)	(0.123)	(0.063)	(0.109)	(0.129)
	[0.523]	[0.523]	[0.036]	[0.523]	[0.003]	[0.523]	[0.002]	[0.013]
GME Treatment x High Trauma	-0.033	-0.026	0.420***	0.125	0.146	-0.093	0.440***	0.272*
	(0.146)	(0.104)	(0.154)	(0.153)	(0.163)	(0.080)	(0.143)	(0.161)
	[0.523]	[0.523]	[0.017]	[0.382]	[0.356]	[0.248]	[0.010]	[0.123]
GME Treat.+ GME Treat. High Trauma = 0	0.26	0.20	0.00	0.19	0.02	0.67	0.00	0.01
Control Mean	-0	-0	-0	-0	0	-0	-0	-0
Observations	1395	1395	1395	1395	1395	1395	1395	1395

Notes. This table reproduces Table 3 of the main paper, which shows the effect of the GME treatment on employment and wellbeing outcomes. In this table, the treatment indicator GME is interacted with an indicator for "High Trauma", 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). Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C5. Ethiopia: Economic Outcomes by Gender

	Targeted Behaviors				Downstream Outcomes			
	Intent to Stay	Job Search	Hours Worked	Employment	Food Security	Consumption	Life Satisfaction	WHODAS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Specification								
GME Treatment	0.199***	-0.173**	0.219**	0.047	0.171**	0.064	0.116	0.047
	(0.076)	(0.074)	(0.094)	(0.037)	(0.074)	(0.045)	(0.072)	(0.059)
	[0.021]	[0.033]	[0.033]	[0.152]	[0.033]	[0.144]	[0.100]	[0.267]
Female	1.065***	-0.610***	-0.512***	-0.367***	-0.167***	0.048	-0.268***	-0.899***
	(0.061)	(0.060)	(0.067)	(0.029)	(0.060)	(0.035)	(0.059)	(0.050)
	[0.001]	[0.001]	[0.001]	[0.001]	[0.016]	[0.145]	[0.001]	[0.001]
GME Treatment x Female	-0.040	0.151*	-0.151	-0.020	0.097	-0.031	0.029	0.220**
	(0.107)	(0.084)	(0.112)	(0.049)	(0.102)	(0.063)	(0.103)	(0.099)
	[0.339]	[0.071]	[0.145]	[0.339]	[0.221]	[0.339]	[0.339]	[0.037]
GME Treat.+GME Treat.xFemale = 0	0.03	0.59	0.27	0.40	0.00	0.46	0.05	0.00
Control Mean	-0	-0	-0	0	0	-0	-0	-0
Observations	1395	1395	1395	1395	1395	1395	1395	1395

Notes. This table reproduces Table 3 of the main paper, which shows the effect of the GME treatment on employment and wellbeing outcomes. In this table, the treatment indicator GME is interacted with an indicator for being a female respondent. 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: Economic Outcomes by Rule Orientation Scale

	Targeted Behaviors				Downstream Outcomes			
	Intent to Stay	Job Search	Hours Worked	Employment	Food Security	Consumption	Life Satisfaction	WHODAS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Specification								
GME Treatment	0.178***	-0.101**	0.145**	0.082	0.227***	0.050	0.137***	0.161***
	(0.053)	(0.042)	(0.056)	(0.053)	(0.050)	(0.032)	(0.051)	(0.049)
	[0.006]	[0.036]	[0.030]	[0.235]	[0.001]	[0.235]	[0.030]	[0.006]
Rule Orientation	-0.018	0.123***	0.050	-0.004	-0.102***	0.001	-0.055	-0.002
	(0.035)	(0.035)	(0.039)	(0.038)	(0.039)	(0.019)	(0.037)	(0.037)
	[0.636]	[0.005]	[0.350]	[0.636]	[0.030]	[0.636]	[0.253]	[0.636]
GME Treatment x Rule Orientation	0.021	-0.051	-0.056	-0.032	-0.012	-0.032	-0.036	-0.056
	(0.048)	(0.044)	(0.056)	(0.052)	(0.051)	(0.026)	(0.050)	(0.048)
	[0.636]	[0.350]	[0.398]	[0.636]	[0.636]	[0.350]	[0.568]	[0.350]
GME Treat.+ GME Treat. x Rule Orientation = 0	0.01	0.02	0.26	0.49	0.00	0.66	0.18	0.12
Control Mean	-0	-0	-0	-0	0	-0	-0	-0
Observations	1395	1395	1395	1395	1395	1395	1395	1395

Notes. This table reproduces Table 3 of the main paper, which shows the effect of the GME treatment on employment and wellbeing outcomes. In this table, the treatment indicator GME is interacted with the respondent's score on a "Rule Orientation" scale, which captures the extent to which they care about following rules and regulation. 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. Ethiopia: Detailed Labor Market Outcomes

	Work Status (Extensive Margin)					Hours Worked (Intensive Margin)					Labor Income
	Any	Wage	Own business	Casual	Unpaid	Total	Wage	Own business	Casual	Unpaid	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Main Specification											
GME Treatment	0.124**	0.095*	0.085	0.061	-0.066	0.145**	0.111**	0.116*	0.061	-0.035	0.143**
	(0.054)	(0.055)	(0.061)	(0.056)	(0.047)	(0.056)	(0.056)	(0.070)	(0.057)	(0.052)	(0.063)
Panel B: PDS Lasso											
GME Treatment	0.080	0.060	0.075	0.040	-0.054	0.098**	0.071	0.097	0.058	-0.030	0.121*
	(0.049)	(0.050)	(0.060)	(0.055)	(0.044)	(0.050)	(0.051)	(0.067)	(0.058)	(0.051)	(0.063)
Control Mean	-0.00	-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	1395

Notes. This table shows intention-to-treat estimates of the GME treatment on work status (extensive margin), hours worked (intensive margin), and labor income over the past seven days. 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). Labor income refers to the standardized, unconditional sum of income over the past 7 days from self-employment or wage employment. Panel A shows ITT effects controlling for randomization strata (gender and prime treatment), while Panel B adds controls chosen with PDS Lasso. Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C8. Ethiopia: Non-Standardized Labor Market Impacts

	Any work	Total Income	Wage Income	Business Income
	(1)	(2)	(3)	(4)
Panel A: Main Specification				
GME Treatment	0.05**	113.30**	87.65*	25.65*
	(0.022)	(49.850)	(48.464)	(13.286)
Panel B: PDS Lasso				
GME Treatment	0.03	95.90*	80.79*	23.63*
	(0.020)	(50.123)	(47.912)	(12.779)
Control Mean	0.21	214.82	217.48	-2.66
Observations	1395	1395	1395	1395

Notes. This table shows intention-to-treat estimates of the GME treatment on employment, total labor income, wage employment income, and self-employment income 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. Total labor income (column 2) is the sum of income from wage employment (column 3) and income from self-employment (column 4). Income from wage employment and from self-employment are each winsorized at the 99th percentile. All income values are unconditional in Ethiopian birr per week. Total income per month in the control group is approximately USD 16.60 (4 $\times$ 214.82 ETB / 51.76 ETB:USD. Average exchange rate for 2022 taken from the World Bank World Development Indicators database). Panel A shows ITT effects controlling for randomization strata (gender and prime treatment), while Panel B adds controls chosen with PDS Lasso. Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C9. Ethiopia: Treatment Increases The Quality of Mental Experiencing (With Enumerator Fixed Effects)

	Overall Index	Sub-indices		
		Specificity	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Panel A: Main Specification				
GME Treatment	0.082***	0.031	0.108**	0.109***
	(0.029)	(0.042)	(0.047)	(0.038)
		[0.181]	[0.022]	[0.014]
Panel B: PDS Lasso				
GME Treatment	0.076***	0.008	0.114**	0.110***
	(0.028)	(0.042)	(0.047)	(0.038)
		[0.394]	[0.016]	[0.011]
Control Mean	0.00	0.00	-0.01	0.01
Observations	1332	1332	1332	1329

Notes. This table reproduces Table 2 of the main paper, adding enumerators' fixed effects. Regression specifications and outcomes are defined in the note to Table 2. Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C10. Ethiopia: Treatment Improves Targeted Behaviors and Downstream Outcomes (With Enumerator Fixed Effects)

	Targeted Behaviors				Downstream Outcomes			
	Intent to Stay	Job Search	Hours Worked	Employment	Food Security	Consumption	Life Satisfaction	WHODAS Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Main Specification								
GME Treatment	0.150***	-0.086**	0.146***	0.122**	0.200***	0.037	0.098**	0.156***
	(0.044)	(0.039)	(0.056)	(0.054)	(0.048)	(0.051)	(0.048)	(0.047)
	[0.003]	[0.019]	[0.012]	[0.019]	[0.001]	[0.064]	[0.024]	[0.003]
Panel B: PDS Lasso								
GME Treatment	0.155***	-0.075*	0.097**	0.076	0.203***	0.054	0.108**	0.169***
	(0.045)	(0.041)	(0.049)	(0.048)	(0.046)	(0.043)	(0.047)	(0.045)
	[0.002]	[0.046]	[0.044]	[0.070]	[0.001]	[0.085]	[0.027]	[0.001]
Control Mean	-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

Notes. This table reproduces Table 3 of the main paper, adding enumerators' fixed effects. Regression specifications and outcomes are defined in the note to Table 3. Robust standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

C.3 Experiment 2 (Colombia)

Table C11. Colombia: Mental Experiencing Results with PDS Lasso

				Sub-indices in business domain		
	Overall index	Business index	Non business index	Frequency of use	Specificity	Emotionality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Traditional vs Control						
Traditional Training	-0.075	-0.157*	-0.049	-0.074	-0.130	-0.122
	(0.073)	(0.088)	(0.073)	(0.085)	(0.079)	(0.080)
				[0.238]	[0.238]	[0.238]
Panel B: GME vs Traditional						
GME Treatment	0.032	0.175**	-0.001	0.090	0.133**	0.167**
	(0.060)	(0.073)	(0.059)	(0.071)	(0.067)	(0.065)
				[0.080]	[0.052]	[0.031]
Panel C: GME vs Control						
GME Treatment	0.001	0.035	-0.007	0.020	0.027	0.050
	(0.061)	(0.074)	(0.061)	(0.073)	(0.068)	(0.067)
				[1.000]	[1.000]	[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	-0.10
N in Control	550	390	550	390	390	390
N in Traditional	656	456	656	456	454	454
N in GME	1140	839	1140	838	835	835

Notes. This table reproduces Table 4 from the main paper adding controls chosen with PDS Lasso. Regression specifications and outcomes are defined in the note to Table 4. 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.

Table C12. Colombia: Downstream Outcomes with PDS Lasso

	Overall Indices		Pre-covid sub-indices		Covid sub-indices				
	Pre-covid	Covid	Earnings	Business	Earnings	Business	Safety nets	Covid actions	Investment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Traditional vs Control									
Traditional training	-0.117	-0.074	-0.199**	0.019	-0.142**	-0.009	0.010	-0.074	-0.016
	(0.075)	(0.065)	(0.092)	(0.030)	(0.070)	(0.030)	(0.064)	(0.067)	(0.065)
			[0.176]	[1.000]	[0.176]	[1.000]	[1.000]	[0.789]	[1.000]
Panel B: GME vs Traditional									
GME training	0.098	0.109*	0.186**	0.010	0.141**	0.036	0.040	0.084	0.041
	(0.062)	(0.056)	(0.082)	(0.026)	(0.060)	(0.024)	(0.054)	(0.057)	(0.051)
			[0.090]	[0.430]	[0.090]	[0.221]	[0.340]	[0.221]	[0.340]
Panel C: GME vs Control									
GME training	0.014	0.081	0.005	0.028	0.029	0.030	0.060	-0.005	0.056
	(0.061)	(0.058)	(0.065)	(0.027)	(0.057)	(0.027)	(0.059)	(0.062)	(0.057)
			[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Mean DV in Control	0.00	0.00	0.00	0.70	0.00	0.70	0.00	0.00	0.00
Mean DV in Traditional	-0.10	-0.10	-0.20	0.70	-0.10	0.60	0.00	-0.10	-0.10
N in Control	346	552	323	333	539	552	546	551	564
N in Traditional	407	660	380	392	642	659	650	657	667
N in GME	704	1147	665	679	1115	1145	1134	1142	1159

Notes. This table reproduces Table 5 from the main paper adding controls chosen with PDS Lasso. Regression specifications and outcomes are defined in the note to Table 5. 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.

Table C13. Colombia: Wellbeing Outcomes

	Overall Index	Sub-indices	
	Wellbeing	Kessler	Resilience and self-efficacy
	(1)	(2)	(3)
Panel A: Traditional vs Control			
Traditional Training	-0.016	-0.012	-0.052
	(0.070)	(0.070)	(0.072)
Panel B: GME vs Traditional			
GME Treatment	0.004	0.017	-0.035
	(0.058)	(0.057)	(0.063)
Panel C: GME vs Control			
GME Treatment	0.003	0.011	-0.049
	(0.064)	(0.064)	(0.066)
Mean DV in Control	0.00	0.00	0.00
Mean DV in Traditional	0.00	0.00	0.00
N in Control	545	545	545
N in Traditional	649	647	649
N in GME	1134	1133	1134

Notes. This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on wellbeing. The outcome in Column (1) is an overall index, which combines the sub-indices from Columns (2) and (3). The index in Column (2) is the standardized score in the Kessler K6 Distress scale (reverse coded to that higher values indicate higher wellbeing). The index in Column (3) measures the respondent's resilience and self-efficacy. Panel A shows the effect of the traditional training compared to the control, Panel B compares the GME treatment to the traditional training and Panel C looks at the impact of GME with respect to the control. All regressions control for randomization strata and survey wave fixed effects. Clustered standard errors in parentheses, sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table C14. Colombia: Mental Experiencing Outcomes in Sample with Business at Baseline

	Overall index	Business index	Non business index	Sub-indices in Business Domain: Frequency of use	Specificity	Emotionality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Traditional vs Control						
Traditional training	-0.068	-0.131	-0.045	-0.053	-0.145	-0.074
	(0.102)	(0.111)	(0.103)	(0.111)	(0.104)	(0.102)
				[0.974]	[0.974]	[0.974]
Panel B: GME vs Traditional						
GME training	0.066	0.181*	0.005	0.086	0.189**	0.142*
	(0.086)	(0.095)	(0.083)	(0.091)	(0.088)	(0.085)
				[0.164]	[0.104]	[0.104]
Panel C: GME vs Control						
GME training	0.017	0.052	-0.013	0.057	0.035	0.047
	(0.095)	(0.098)	(0.096)	(0.097)	(0.090)	(0.090)
				[1.000]	[1.000]	[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 Table 4 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 4. 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.

Table C15. Colombia: Mental Experiencing Outcomes with Inverse Probability Weighting

				Subindices in the Business Domain		
	Overall index	Business index	Non business index	Frequency of use	Specificity	Emotionality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Traditional vs Control						
Traditional training	-0.052	-0.163*	-0.034	-0.046	-0.126	-0.153*
	(0.076)	(0.092)	(0.074)	(0.091)	(0.081)	(0.084)
				[0.258]	[0.223]	[0.223]
Panel B: GME vs Traditional						
GME training	0.005	0.174**	-0.028	0.094	0.136*	0.161**
	(0.065)	(0.077)	(0.063)	(0.073)	(0.070)	(0.068)
				[0.088]	[0.057]	[0.055]
Panel C: GME vs Control						
GME training	-0.005	0.025	-0.018	0.034	0.024	0.018
	(0.067)	(0.077)	(0.066)	(0.076)	(0.070)	(0.070)
				[1.000]	[1.000]	[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	-0.10
N in Control	550	390	550	390	390	390
N in Traditional	656	456	656	456	454	454
N in GME	1140	839	1140	838	835	835

Notes. This table reproduces Table 4 from the main paper adding weights for the probability of taking part to the midline and endline surveys. Regression specifications and outcomes are defined in the note to Table 4. 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.

Table C16. Colombia: Downstream Outcomes with Inverse Probability Weighting

	Overall Indices			Pre-covid sub-indices		Covid sub-indices					Wellbeing sub-indices	
	Pre-covid econ	Covid econ	Wellbeing	Earnings	Business	Earnings	Business	Safety nets	Covid actions	Investment	Kessler	Resilience and self-efficacy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Traditional vs Control												
Traditional training	-0.078	-0.077	-0.010	-0.139	0.033	-0.154**	-0.009	0.031	-0.078	-0.033	-0.006	-0.043
	(0.096)	(0.070)	(0.070)	(0.108)	(0.039)	(0.071)	(0.034)	(0.068)	(0.067)	(0.067)	(0.070)	(0.072)
Panel B: GME vs Traditional												
GME treatment	0.137**	0.141**	0.004	0.184**	0.021	0.152**	0.050*	0.019	0.081	0.072	0.017	-0.041
	(0.065)	(0.059)	(0.058)	(0.080)	(0.027)	(0.061)	(0.025)	(0.057)	(0.057)	(0.054)	(0.058)	(0.064)
Panel C: GME vs Control												
GME treatment	0.041	0.076	0.005	0.012	0.044	0.040	0.043	0.049	-0.019	0.075	0.015	-0.059
	(0.067)	(0.059)	(0.064)	(0.068)	(0.030)	(0.057)	(0.030)	(0.061)	(0.062)	(0.061)	(0.064)	(0.067)
Mean DV in Control	0.00	0.00	0.00	0.00	0.70	0.00	0.70	0.00	0.00	0.00	0.00	0.00
Mean DV in Traditional	-0.10	-0.10	0.00	-0.20	0.70	-0.10	0.60	0.00	-0.10	-0.10	0.00	0.00
N in Control	346	552	545	323	333	539	552	546	551	564	545	545
N in Traditional	407	660	649	380	392	642	659	650	657	667	647	649
N in GME	704	1147	1134	665	679	1115	1145	1134	1142	1159	1133	1134

Notes. This table reproduces Table 5 from the main paper adding weights for the probability of taking part to the midline and endline surveys. Regression specifications and outcomes are defined in the note to Table 5 and, for wellbeing outcomes, to Table C13. 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.

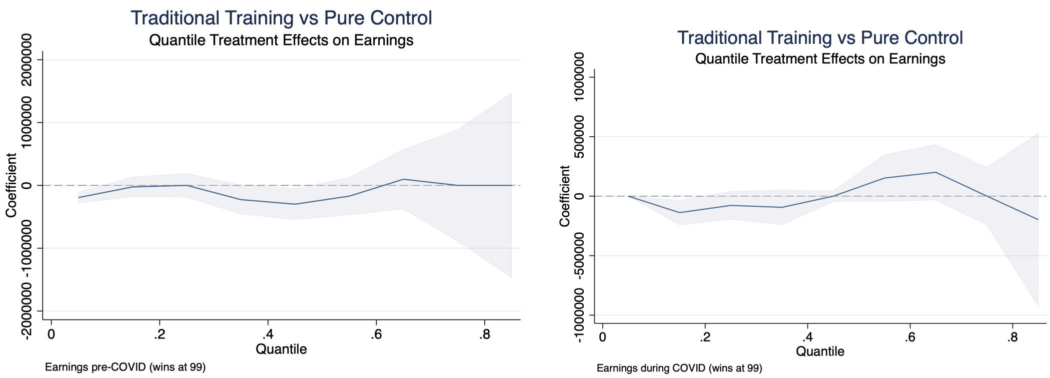
Table C17. Colombia: Quantile Treatment Effects on Earnings

	Earnings Index	
	Pre-covid	Covid
	(1)	(2)
Panel A: Traditional vs Control		
Q05: Traditional training	-0.072***	0.000
	(0.017)	(0.015)
Q10: Traditional training	-0.021	0.006
	(0.017)	(0.014)
Q25: Traditional training	-0.022	-0.012
	(0.021)	(0.014)
Q50: Traditional training	-0.033	0.000
	(0.025)	(0.017)
Q75: Traditional training	-0.022	0.093***
	(0.049)	(0.032)
Q90: Traditional training	0.029	0.093
	(0.208)	(0.100)
Q95: Traditional training	-0.160	-0.457**
	(0.499)	(0.232)
Panel B: GME vs Traditional		
Q05: GME treatment	0.045***	0.000
	(0.013)	(0.013)
Q10: GME treatment	0.024*	0.006
	(0.013)	(0.012)
Q25: GME treatment	0.022	0.020*
	(0.015)	(0.011)
Q50: GME treatment	0.015	0.000
	(0.019)	(0.013)
Q75: GME treatment	-0.016	-0.043
	(0.038)	(0.030)
Q90: GME treatment	-0.023	-0.008
	(0.188)	(0.085)
Q95: GME treatment	-0.160	0.027
	(0.224)	(0.130)
Panel C: GME vs Control		
Q05: GME treatment	0.000	0.000
	(0.012)	(0.015)
Q10: GME treatment	0.004	0.012
	(0.014)	(0.014)
Q25: GME treatment	0.000	0.008
	(0.016)	(0.012)
Q50: GME treatment	-0.018	0.000
	(0.020)	(0.013)
Q75: GME treatment	-0.034	0.032
	(0.045)	(0.026)
Q90: GME treatment	-0.083	0.116
	(0.224)	(0.094)
Q95: GME treatment	-0.298	-0.383
	(0.483)	(0.253)
Mean DV in Control	0.00	0.00
Mean DV in Traditional	0.00	-0.10
N in Control	323	539
N in Traditional	380	642
N in GME	665	1115

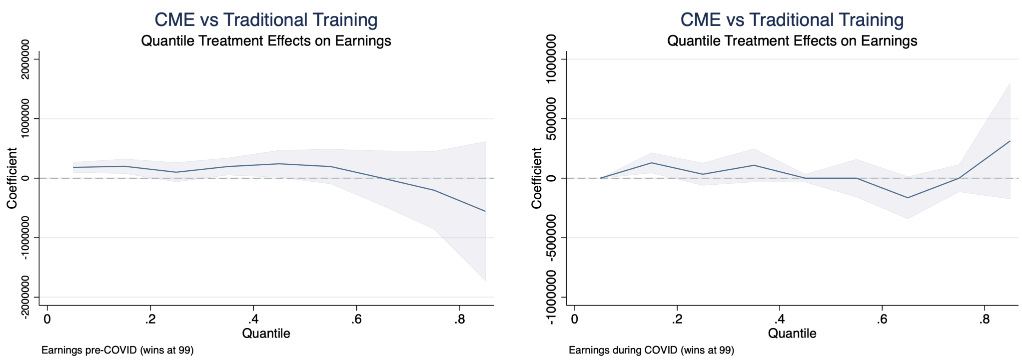
Notes. This table shows quantile treatment effects on the standardized earning indices in the Colombia experiment. This table shows the impact of the two treatment arms on pre and post-Covid earnings separately for different income percentiles in the sample. Panel A shows the effect of the traditional training 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. All regressions control for randomization strata and survey wave fixed effects. Clustered standard errors are in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Figure C4. 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 C18. Colombia: Heterogeneity by Baseline Trauma

	Mental Experiencing (business)	Earnings pre-COVID	Earnings during COVID	Business pre-COVID	Business during COVID
	(1)	(2)	(3)	(4)	(5)
Panel A: Traditional vs Control					
Traditional Training	-0.117	0.028	-0.015	0.024	-0.003
	(0.126)	(0.121)	(0.100)	(0.042)	(0.040)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
High trauma	0.079	0.200*	0.157	0.036	0.004
	(0.124)	(0.116)	(0.096)	(0.048)	(0.047)
	[1.000]	[0.558]	[0.558]	[1.000]	[1.000]
Traditional Training × High Trauma	-0.083	-0.544***	-0.277**	-0.051	-0.044
	(0.170)	(0.194)	(0.140)	(0.063)	(0.061)
	[1.000]	[0.087]	[0.502]	[1.000]	[1.000]
Trad + Trad × High Trauma	-0.200	-0.516	-0.293	-0.026	-0.047
p-value: Trad + Trad × High Trauma = 0	0.099	0.001	0.003	0.586	0.321
Panel B: GME vs Traditional					
GME Treatment	0.073	0.031	0.110	-0.003	0.031
	(0.098)	(0.091)	(0.078)	(0.033)	(0.032)
	[1.000]	[1.000]	[0.881]	[1.000]	[1.000]
High Trauma	-0.022	-0.399**	-0.127	-0.033	-0.054
	(0.125)	(0.156)	(0.105)	(0.045)	(0.042)
	[1.000]	[0.196]	[0.881]	[1.000]	[0.881]
GME Treatment × High Trauma	0.250*	0.372**	0.077	0.052	0.039
	(0.149)	(0.176)	(0.122)	(0.054)	(0.051)
	[0.683]	[0.329]	[1.000]	[1.000]	[1.000]
GME + GME × High Trauma	0.323	0.403	0.187	0.050	0.071
p-value: GME + GME × High Trauma = 0	0.004	0.008	0.048	0.253	0.076
Panel C: GME vs Control					
GME Treatment	-0.026	0.068	0.126	0.033	0.038
	(0.110)	(0.103)	(0.083)	(0.038)	(0.037)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
High Trauma	0.061	0.152	0.120	0.022	-0.007
	(0.123)	(0.106)	(0.094)	(0.048)	(0.047)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
GME Treatment × High Trauma	0.169	-0.162	-0.186	-0.004	-0.012
	(0.147)	(0.134)	(0.114)	(0.057)	(0.055)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
GME + GME × High Trauma	0.143	-0.094	-0.060	0.029	0.026
p-value: GME + GME × High Trauma	0.154	0.254	0.441	0.499	0.526
Mean DV in Control	0.00	0.00	0.00	0.70	0.70
Mean DV in Traditional	-0.10	-0.20	-0.10	0.70	0.60
N in Control	390	323	539	333	552
N in Traditional	456	380	642	392	659
N in GME	839	665	1115	679	1145

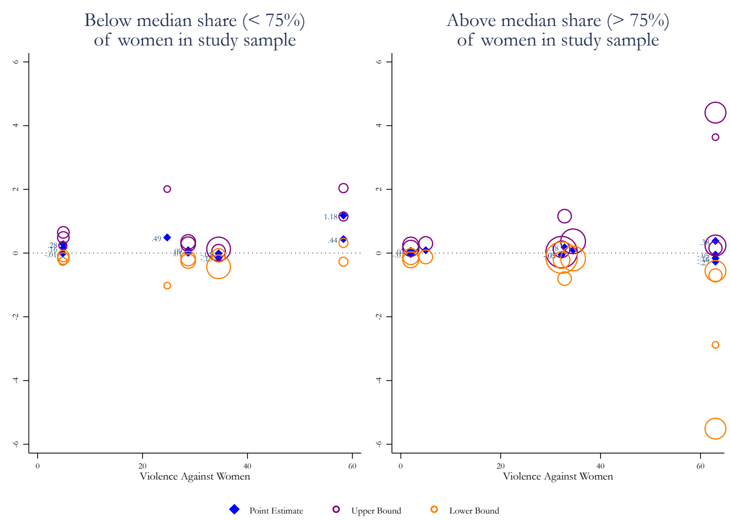
This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the main outcomes, interacted with an index for "High Trauma". Outcome variables and regression controls are defined as in the main tables of the paper (Tables 4 and 5). We categorize respondents as "High Trauma" if they score 33 or above on the Impact of Event Score (IES-R) at baseline, or if they report to have been a victim of the civil conflict, or a recent Venezuelan migrant. Panel A shows the effect of the traditional training 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.

Table C19. Colombia: Heterogeneity by Gender

	Mental Experiencing (business)	Earnings pre-COVID	Earnings during COVID	Business pre-COVID	Business during COVID
	(1)	(2)	(3)	(4)	(5)
Panel A: Traditional vs Control					
Traditional Training	-0.003	0.049	-0.021	0.114**	0.039
	(0.127)	(0.131)	(0.096)	(0.048)	(0.047)
	[0.644]	[0.547]	[0.585]	[0.090]	[0.312]
Female	0.215*	-0.022	-0.196**	0.071	0.036
	(0.124)	(0.123)	(0.095)	(0.048)	(0.046)
	[0.204]	[0.585]	[0.139]	[0.204]	[0.312]
Traditional Training × Female	-0.263	-0.464**	-0.208	-0.195***	-0.103*
	(0.171)	(0.186)	(0.138)	(0.063)	(0.061)
	[0.204]	[0.090]	[0.204]	[0.030]	[0.204]
Trad + Trad × Female	-0.266	-0.414	-0.229	-0.081	-0.064
p-value: Trad + Trad × Female = 0	0.026	0.002	0.022	0.052	0.108
Panel B: GME vs Traditional					
GME Treatment	0.003	-0.076	-0.058	-0.023	0.031
	(0.111)	(0.109)	(0.083)	(0.039)	(0.038)
	[0.837]	[0.470]	[0.470]	[0.470]	[0.470]
Female	-0.031	-0.481***	-0.402***	-0.113***	-0.055
	(0.121)	(0.139)	(0.098)	(0.041)	(0.040)
	[0.666]	[0.004]	[0.001]	[0.018]	[0.238]
GME Treatment × Female	0.307**	0.454***	0.337***	0.073	0.029
	(0.147)	(0.162)	(0.118)	(0.053)	(0.050)
	[0.067]	[0.017]	[0.017]	[0.238]	[0.470]
GME + GME × Female	0.310	0.379	0.279	0.050	0.059
p-value: GME + GME × Female = 0	0.002	0.001	0.001	0.164	0.076
Panel C: GME vs Control					
GME Treatment	0.033	0.002	-0.035	0.094**	0.069
	(0.115)	(0.114)	(0.086)	(0.045)	(0.043)
	[0.953]	[0.953]	[0.953]	[0.328]	[0.328]
Female	0.227*	-0.030	-0.193**	0.066	0.033
	(0.123)	(0.122)	(0.095)	(0.047)	(0.046)
	[0.328]	[0.953]	[0.328]	[0.425]	[0.911]
GME Treatment × Female	0.021	-0.017	0.129	-0.111*	-0.063
	(0.149)	(0.143)	(0.114)	(0.057)	(0.055)
	[0.953]	[0.953]	[0.483]	[0.328]	[0.483]
GME + GME × Female	0.055	-0.015	0.094	-0.017	0.006
p-value: GME + GME × Female = 0	0.579	0.857	0.218	0.631	0.869
Mean DV in Control	0.00	0.00	0.00	0.70	0.70
Mean DV in Traditional	-0.10	-0.20	-0.10	0.70	0.60
N in Control	390	323	539	333	552
N in Traditional	456	380	642	392	659
N in GME	839	665	1115	679	1145

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the main outcomes, interacted with an indicator for being a woman. Outcome variables and regression controls are defined as in the main tables of the paper (Tables 4 and 5). Panel A shows the effect of the traditional training 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 C5. Traditional Training Programs and Trauma: Generalizing their Potentially Weaker Impact in Vulnerable Samples



Notes: The Figure shows point estimates and the upper and lower bounds of the confidence intervals of the effect of 21 different entrepreneurship trainings on sales. Data come from McKenzie et al. (2023) and the coefficients shown come from the random effects meta-analysis performed by the authors. We categorized countries where the studies were conducted based on the 2019 OECD index of "Violence against Women" (OECD, 2019). This is the variable on the x-axis. We then distinguish between studies with a high or low share of female participants, splitting at the median (71%). The graph on the left-hand side shows studies with a below-median share of female participants (10 in total), and the graph on the right-hand side shows studies with an above-median share of female participants (11 in total, of which 8 with 100% women in the training). The intuition for this exercise is that female participants in training programs conducted in contexts with high violence against women are more likely to have experienced trauma and violence in the past. According to our evidence, we expect the training programs to have more mixed outcomes in these contexts, as they may average a positive impact on non-traumatized participants and a muted or even negative impact on traumatized participants.

D Pre-Analysis Plan (PAP) Reports

These reports describe how the analysis of the Ethiopia and Colombia experiments presented in the main body of the paper and online appendix align with the pre-analysis plans (PAP) registered with the American Economic Association RCT Registry.

If not found below, these reports are available upon request.

Learning to See the World's Opportunities:

Memory, Mental Experiencing, and the Economic Lives of the Vulnerable

Pre-Analysis Plan Report (Ethiopia)

Nava Ashraf* Gharad Bryan† Alexia Delfino‡ Emily A. Holmes§

Leonardo Iacovone¶ Christian Johannes Meyer‖ Ashley Pople**

This document describes how the analysis of the Ethiopia experiment presented in the main body of the paper and online appendix aligns with the pre-analysis plan registered with the American Economic Association RCT Registry as AEARCTR-0008934 (https://www.socialscienceregistry.org/trials/8934).

*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

‖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

Contents

1 Comparison Between Pre-Analysis Plan and Paper **3**

2 Priming Intervention **16**

2.1 Description of the Intervention . . . 16

2.2 Priming Treatment Effect . . . 16

2.2.1 Effect of Priming on Mental Experiencing at Baseline . . . 16

2.2.2 Effect of Priming on Mental Experiencing, Interaction with Gender . . . 17

2.2.3 Effect of Priming on Mental Experiencing, Interaction with PTSD . . . 18

2.3 Interaction of Priming and GME Treatment . . . 19

3 Baseline Patterns **20**

3.1 Relationship between Indices of Past/Future and Positive/Negative Mental Experiencing . . . 20

3.2 Trauma and Future Mental Experiencing . . . 21

3.3 Trauma, Risk Preferences and Priming . . . 22

4 Treatment Effects on Risk Taking as per PAP **23**

5 Additional Analysis on Mechanisms as per PAP **24**

6 Additional Treatment Effect Estimates **26**

7 Lee Bounds **28**

1 Comparison Between Pre-Analysis Plan and Paper

Table 1. Differences Between Registered Pre-Analysis Plan and Final Manuscript.

Pre-Analysis Plan	Final manuscript, adherence to PAP, and reasons	Additional Tables and Figures
1. Introduction We use a representative survey among refugees in Addis Ababa and a randomized trial of a psychological program on "positive imagery" to study the role psychological constraints such as trauma on the ability to imagine the present or future, and secondly on economic outcomes.	Implemented as specified.	
2. Description of the interventions		
2.1 Imagery-based intervention Positive imagery training, closely following Holmes et al. (2019)	Implemented as specified. The final manuscript refers to the imagery-based intervention as guided mental experiencing (GME), a more descriptive and intuitive label for our intervention.	
2.2 Priming intervention We cross-randomize a priming intervention that has been used in the literatures of psychology and economics. We adapt the protocol by Callen et al. (2014) to administer two primes just before asking subjects to make a risky investment choice.	The cross-randomized priming intervention is not discussed in the paper. We prefer to focus on the main treatment for expositional clarity, not least because the cross-randomized priming intervention was not administered in Colombia. Detailed results on the priming treatment are provided below.	Section 2 below provides a detailed description of the priming intervention.

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
3. Timeline of Implementation and Data Collection Baseline survey in May 2022. Four midline surveys 28, 35, 42, and 49 days after baseline. Endline approximatively 60 days after the baseline survey.	The timing of surveys is slightly different from the timing specified in the PAP due to logistical challenges during implementation. The baseline survey stretched from May through July 2022. In-person follow-up surveys were conducted approximately three months after the baseline and approximately six weeks after the intervention. The average time between time between baseline and endline was 92 days (86 days at the 25th percentile, 107 at the 75th percentile). The four midline surveys are not reported in the paper for expositional clarity.	Figure A5 in the Appendix of the paper provides an illustration of the experimental timeline.
4. Sample Our population of interest is recognized refugees from Eritrea, currently living in Addis Ababa, Ethiopia. As economic behavior is a key focus of our study, we restrict our sample to refugees aged 18 to 50 with at least junior high school education (Ethiopian grade 7 and above). This definition excludes Eritrean asylum seekers who have not been granted refugee status, refugees that fled from Tigray as a result of the ongoing conflict, or internally displaced people. We obtained a list covering the universe of all formally registered refugees that satisfied these criteria as of November 10, 2021 (N = 36,136) from the Government of Ethiopia's Refugees and Returnees Services (RRS). We then take a random sample stratified by gender and an additional one as replacement frame.	As specified.	

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
5. Measurement of primary outcomes		
5.1 Family 1: Risk taking Two unincentivized survey items adapted from the Global Preference Survey (Falk et al., 2018). One replicates exactly the GPS question on risk preferences and one contextualizes it within the past week. An incentivized measure of risk aversion using the Gneezy-Potters-Charness investment task (Charness and Gneezy, 2010; Gneezy and Potters, 1997). The three measures are combined into an index of risk aversion. Outcomes for each subpart of the index are reported for completeness.	As specified, but not reported in the paper for expositional clarity.	Results reported in section 4 below.
5.2 Family 2: Memory Two scales, both adapted to the Ethiopian context: 1) Word based autobiographical memory test (McNally et al., 1995), oriented towards the past and 2) Scenario based adapted version of the autobiographical memory test (McNally et al., 1995), oriented towards the past. For both scales, we will code the answers given by participants in the back-office in terms of three measures: i) their specificity, ii) emotional intensity and iii) sentiment (positive or negative). We will follow the psychological literature (Griffith et al., 2009) to build the criteria for defining "specific" answers, and have research assistants (blind to treatment) code the answers based on the criteria given. We will also ask the research assistants to code the answers in terms of the level of emotional intensity and sentiment (positive or negative). This step will lead to three different indexes for each of the scales above, which we will combine into three aggregate indexes of specificity, emotional intensity and sentiment of memories.	Done as specified. In the paper, we use indices defined from the scenario-based scale for comparability with our future-oriented measures of mental experiencing.	

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
We will check the correlation between the back-office measures with measures of: – Participants' self-reported vividness and emotional intensity – Surveyors' recorded specificity and sentiment – Specificity, emotional intensity and sentiment produced using text analysis	Text analysis of answers was not done given more detailed data available from back-office coding. Correlations between self-reported measures and back-coded indexes are reported in Appendix C.1 of the paper. The correlations between other different measures were not included in the paper for expositional clarity.	See subsection 3.1 below for correlations of different mental experiencing measures.
We will create the following indexes: • an index of the proportion of neutral scenarios/words that are coded as negative and an index of the speed with which people retrieve negative and an index of the speed of retrieval of positive scenarios/words • a vividness of negative images index and a vividness of positive images index using self reported vividness • a measure of specificity for positive and negative images using a combined index of surveyor and text analysis • we will create a separate measures of emotional intensity for positive and negative images using self reported emotional intensity Overall, this will give us 9 measures: positive speed, positive vividness, positive specificity, positive emotional intensity, negative frequency (positive frequency is the complementary share), negative speed, negative vividness, negative specificity and negative intensity.	Speed was not computed. For conciseness, not all these disaggregated indices are reported in the paper. Appendix Table C1 in the main manuscript shows results when combining self-reported and back-coded measures in an index built from principal component analysis.	See subsection 3.1 below for correlations of these different mental experiencing measures.

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
5.3 Family 3: Future Imagery One scale: Adapted version of the autobiographical memory test (Griffith et al., 2009), oriented to the future, adapted to the Ethiopian context. We will code the answers given by participants in the back-office in terms of their specificity, emotional intensity and sentiment (positive or negative). We will follow the psychological literature (Griffith et al., 2009) to build the criteria for defining ”specific” answers, and have research assistants code the answers based on the criteria given (blind to treatment). We will ask the research assistants to also code the answers in terms of the level of emotional intensity and sentiment (positive or negative). This will lead to three different indexes for the scale.	Done as specified.	
Together with the back-office measures, we will also correlate and construct measures based on: – Participants' self-reported vividness and emotional intensity – Surveyors' recorded valence – Text analysis, which will give use measures of specificity, emotional intensity and sentiment.	Text analysis of answers was not done given more detailed data available from back-office coding. Correlations between self-reported measures and back-coded indexes are reported in the Appendix of the paper. The correlations between other different measures were not included in the paper for expositional clarity.	See subsection 3.1 below for correlations of different mental experiencing measures.
We will also construct three overall indexes which combine specificity, emotional intensity and sentiment for both past and future events to characterize people's imagery.	As specified. This is our main mental experiencing index in the paper.	

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
We will create the following indices: • an index of the proportion of neutral scenarios that are coded as negative; • a vividness of negative images index and a vividness of positive images index using self reported vividness; • a measure of specificity for positive and negative images using text analysis; • a separate measures of emotional intensity for positive and negative images using self reported emotional intensity. Overall, this will give us 7 measures: positive vividness, positive specificity, positive emotional intensity, negative frequency (positive frequency is the complementary share), negative vividness, negative specificity and negative intensity.	Text analysis of answers was not done given more detailed data available from back-office coding. For conciseness, not all these disaggregated indices are reported in the paper. Appendix Table C1 in the main manuscript shows results when combining self-reported and back-coded measures in an index built from principal component analysis.	See subsection 3.1 below for correlations of these different mental experiencing measures.

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
6. Measurement of secondary outcomes		
6.1 Employment outcomes Job search: measures with a battery of questions adapted from the Ethiopian Socioeconomic Survey / Living Standard Measurement Study (LSMS) and previous literature on the subject. Dummy for any search, index of effort in search by aggregating hours, number of calls made, number of channels and money spent. 4 types of activities taken into account: 1 – non-agricultural or non-fishing business for yourself or household, 2 – casual part-time or temporary labour, 3 – wage work 4 – unpaid apprenticeship. Look at the length of search and beliefs on starting an activity to identify mechanisms. Engagement in employment and or self-employment: dummy equal to 1 if a person is involved in any of the four activities. Number of hours worked in specific activities. Aggregate across all activities and use for FDR corrections + report results for each separate activity.	As specified.	
6.2. Income, Consumption, and Welfare We adapt survey modules from the ESS/LSMS. 1 – Income from economic activities 2 – Food consumption, dietary diversity, non-food consumption. 3 – Food security 4 – Savings 5 – Housing 6 – Incoming and outgoing transfers 7 – Support from various assistance programs. We construct an overall, standardized index following Kling et al. (2007). The standardized index will be our main outcome and will be used for FDR corrections, the remaining outcomes will be reported for completeness and to aid interpretation.	The paper reports results on selected components of the index for clarity (income, consumption, food security). Data on housing quality was never collected due to modifications to the survey instrument during the final stages of instrument development.	Results on the pre-specified overall index and all subindices (except for housing) are reported in Table 12 below.
6.3 Well-being and Physical health 1 – Cantril Self-Anchoring Striving Scale 2 – WHO Disability Assessment Schedule (WHODAS) 2.0. We report the two dimensions separately through sub-indexes and construct an overall standardized index following Kling et al. (2007), used in FDR corrections.	As specified.	

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
6.4 Other measures To study the mechanisms of the intervention, we also measure and analyze: 1. Generalized self-efficacy (Schwarzer and Jerusalem, 1995) measured using a shortened 6-items scale (Romppel et al., 2013) 2. Revised Life Orientation Test (measured only at endline) 3. Expectations of employment, earnings and movement conditional on receiving with a certain probability a permit to work in Ethiopia 4. Reservation wage 5. Reaction to negative and positive priming on risk-taking	We test this mechanisms as specified, but results are not included in the paper for expositional clarity.	Results reported in Table 10 and section 5 below.
7. Robustness Checks • Social desirability bias: following Dhar et al. (2022), we use the Marlowe-Crowne Social Desirability Scale. We will present the interaction of the Marlowe–Crowne scale with the imagery intervention treatment status as a robustness check. • Respondents that are more concerned about rule violations (i.e. working without a permit) may under-report their engagement in economic activities. We measure the individual propensity to think about rules in a rule-oriented manner vs in a manner that recognizes exceptions using the Rule orientation scale (Fine et al., 2016). We will report and discuss potential differences in employment outcomes by the individual score on this scale as a robustness check.	Done as specified. Tables with heterogeneity along both dimensions are reported in the Appendix of the paper.	

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
8. Empirical Strategy We estimate the intent-to-treat effect of offering imagery training on our primary and secondary outcomes, compared to a control group who received no intervention. All specifications take the following form: $Y_i = \beta_0 + \beta_1 T_i + \sigma X_i + \epsilon_i$ where Y_i is the outcome of interest for individual i. T_i is a dummy variable indicating whether an individual was offered the imagery training. X_i is a vector of baseline controls and strata variables used in sampling and ϵ_i is the error term. β_1 measures the estimated treatment effect of the imagery training. For every primary outcome, we test the null hypothesis that the imagery training has no impact. We use robust standard errors because treatment assignment is at the individual level. Our main specification only controls for stratification variables in X_i (gender and priming).	Done as specified.	
9. Heterogeneity Analysis We expect our imagery program to have differential treatment effects for those individuals who show higher symptoms of post-traumatic stress at baseline. We focus on trauma and gender. We aim to discipline this exploratory analysis by using the generalized random forest algorithm by Chernozhukov et al. (2018). We measure trauma symptoms by using the PCL-5 scale: for PTSD, we construct a trauma dummy variable to indicate people with a PCL-5 score above 31. Given that we are not solely interested in populations with PTSD but also generally high symptoms of trauma, we construct a second dummy variable where we split the sample at the median PCL-5 score. Heterogeneous treatment effects will be estimated by interacting treatment status with the first trauma dummy, and in a separate regression, with a female dummy. We will further explore our results by conducting heterogeneity analysis with the second dummy for trauma described above.	The paper shows heterogeneity on trauma and gender as pre-specified. Heterogeneity by PTSD and gender is included in the paper's Appendix. Heterogeneity analysis based on generalized random forest is not included in the main paper for expositional clarity. Exploratory analysis using PCL-5 above median as alternative variable for trauma is also not reported in the paper for conciseness.	Heterogeneity analysis based on generalized random forests for main downstream outcomes and alternative specification of trauma is available upon request.

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
10. Multiple Hypotheses testing correction Following Benjamini et al. (2006), we will use false discovery rate corrections to account for multiple hypothesis testing across our primary and secondary outcome variables. For each hypothesis test, we will report two values: p-value from the Wald test and the false discovery rate q-value, taken across primary outcomes. We will do FDR corrections separately for primary and secondary outcomes, reflecting out belief that results in each domain are of separate interest.	We report q-values and p-values as specified.	
11. Covariate selection We first present all specifications controlling for our stratification variables, gender and prime treatment. To increase the precision of our estimates, we also present a second set of results including baseline covariates. Following Belloni et al. (2014), we adopt the "post-double-selection" method for selecting regressors, including first-order interaction and quadratic terms. We will correct for FDR within each method, but not across.	Done as specified.	
12. Balance We will test for balance on baseline variables across treatment and control groups. We will estimate a regression where y_i denotes the baseline variables for individual i, and control for gender. For each baseline variable, we report the p-value from the Wald test on the treatment dummy.	Table A2 in the appendix of the main manuscript presents a balance test. We report the means of selected baseline variables in control and treatment, as well as a test of the difference. We prefer this balance table for clarity and for comparability with the Colombia experiment.	
13. Testing for differential attrition To test for differential attrition, we will create a dummy variable for whether the individual's interview is missing in follow-up surveys and regress this dummy on the treatment dummies. If and only if we find significant differential attrition by treatment status, we will report Lee (2009) bounds. To adjust for attrition, we will also report the main tables with inverse probability weighting (IPW).	Attrition table done as specified. In the Paper, we show balance on observables for the respondents to the endline survey.	Table 13 and Table 14 below show Lee Bounds for our main results tables. Tables with IPW are available upon request.

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
14. Hypotheses formulated in the PAP (summary)		
14.1. Baseline Patterns		
14.1.1 Memory and Future Imagery Our first hypothesis is that memory is a source of thinking about the future (Schacter et al., 2007), hence we expect future imagery to be related to memory: • More (frequent, vivid/specific and emotionally intense) positive memories are positively correlated with more (frequent, vivid/specific and emotionally intense) positive imagined future scenarios. • More (frequent, vivid/specific and emotionally intense) negative memories are positively correlated with more (frequent, vivid/specific and emotionally intense negative) imagined future scenarios.	While the variables mentioned are discussed in the paper, the results of the correlations are not shown in the paper for conciseness.	We present these correlations below in subsection 3.1.
14.1.2 Trauma and Memory We expect people with stronger trauma symptoms to have a different memory database with respect to people with weaker trauma symptoms. We expect: • Memories among traumatized people to be more negative compared to people with weaker trauma symptoms. • Quality predictions are ambiguous. Positive and negative memories among traumatized people may also differ in vividness, specificity and emotional intensity compared to people with weaker trauma symptoms.	Tested as specified. Results are included in the Appendix of the paper.	

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
14.1.3 Trauma and Future Imagery We expect: • Future imagined scenarios among traumatized people will be more negative compared to people with weaker trauma symptoms. This implies that people with trauma may come up more frequently with negative future scenarios. • Quality predictions are ambiguous.	Tested as specified, but not included in the paper for expositional clarity.	Results included in Figure 1 below.
14.1.4 Trauma and Risk Preferences We expect people with stronger trauma symptoms to be more risk averse with respect to people with weaker trauma symptoms. In particular we expect more traumatized people to show more risk aversion compared to less traumatized people.	Tested as specified, but not included in the paper for expositional clarity.	Results included in Figure 2 below.
14.1.5 Trauma and Priming We expect the negative priming intervention to have the following effects: • People who are exposed to the negative prime invest less in the risky investment task compared to people exposed to the positive prime. • It's ambiguous whether more traumatized people respond more or less strongly to the negative prime, compared to less traumatized people exposed to the negative prime.	We do not include analysis of the priming intervention in the paper for expositional clarity, but	Analysis included in section 2 below. Table 2 shows the effect of priming on mental experiencing measures, and Tables 3 and 4 show heterogeneity by gender and trauma. Table 5 shows the interaction between the mental experiencing intervention and the priming manipulation on mental experiencing outcomes. In subsection 3.3, Figure 3 shows that the priming effect on risk is concentrated on people with high trauma.

Continuation of Table 1		
Pre-Analysis Plan	**Final manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
14.2. The impact of imagery training on outcomes		
14.2.1 Memory We expect the imagery training to have the following effects on memory: 1) Positive memories produced by people trained in imagery tend to be more vivid/specific and emotional, compared to the control group and 2) This treatment of imagery effect will be larger for traumatized people.	Tested as specified, hypothesis confirmed in the data. In the paper, we report indices that combine past and future scenarios.	
14.2.1 Future Imagery We expect the imagery training to have the following effects on future imagery: 1) Positive future imagined scenarios should be more positive, either in frequency or sentiment, 2) Imagined future scenarios produced by people trained in imagery tend to be more vivid/specific and emotional, compared to the control group, 3) It's ambiguous whether the training will increase the vividness/specificity of only positive scenarios, or both positive and negative ones, 4) These treatment effects will be stronger for traumatized people, 5) This treatment effect is larger for women.	Tested as specified, hypothesis confirmed in the data. In the paper, we report indices that combine past and future scenarios.	
14.2.2 Risk Taking We expect the imagery training to have the following effects on risk taking: 1) People trained in imagery invest more in the risky investment game, compared to the control group, 2) This treatment effect is larger for traumatized people, 3) This treatment effect is larger for women.	Tested as specified, but not included in the paper for expositional clarity.	Included in section 4 below.
14.3 Secondary outcomes We expect the imagery training to improve secondary outcomes: 1) People trained in imagery search more for an economic activity, have better employment outcomes, income, consumption, welfare and mental and physical health compared to the control group, 2) These treatment effects are larger for those who are traumatized at baseline, 3) This treatment effects are larger for women.	Tested as specified, and reported in the paper.	

2 Priming Intervention

2.1 Description of the Intervention

To shed light on the mechanism through which the treatment affects outcomes for participants, we cross-randomize a priming intervention that has been used in the literatures of psychology and economics. We adapt the protocol by Callen et al. (2014) to administer two primes just before asking subjects to make a risky investment choice. Our protocol is based on extensive piloting and qualitative interviews conducted with the population of interest.

We ask one half of respondents: "I would now like to speak with you about an experience in the past that made you happy or joyous. This could be anything, for example birth of child, marriage of a relative, or success in your job. Please try to think of this event and remember how you felt happy or joyous. If you feel comfortable, could you share with me what this event was?" (POSITIVE) and the other half of respondents a negative analogue (NEGATIVE): "I would now like to speak with you about an experience in the past that made you fearful or anxious. This could be anything, for example getting sick, experiencing violence, losing a loved person, losing a job, etc. Please try to think of this event and remember how you felt fearful or anxious. If you feel comfortable, could you share with me what this event was?". The priming was conducted at baseline and endline.

2.2 Priming Treatment Effect

2.2.1 Effect of Priming on Mental Experiencing at Baseline

Table 2. Priming and Mental Experiencing at Baseline

	Overall Index	Sub-indices		
		Specificity	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Trauma Priming	-0.016	-0.029	0.032	-0.054
	(0.027)	(0.040)	(0.042)	(0.037)
Control Mean	0.05	0.01	0.07	0.08
Observations	1631	1652	1652	1652

This table shows the impact of the priming treatment on the quality of mental experiencing, including the specificity, emotionality, and frequency of positive images. We show the ITT effect for an OLS regression. Specificity refers to the clarity of the images respondents remember. Emotionality refers to the strength of the emotions these images entail. Positivity refers to the frequency of positive images appearing when respondents are presented with scenarios. All regressions control for randomization strata (gender and prime treatment). Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

We find no significant effects of the priming intervention on our measures of mental experiencing at baseline (back-coded measures).

2.2.2 Effect of Priming on Mental Experiencing, Interaction with Gender

Table 3. Priming and Mental Experiencing at Baseline, by Gender

	Overall Index	Sub-indices		
		Specificity	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Trauma Priming	-0.017	0.004	0.019	-0.067
	(0.038)	(0.056)	(0.062)	(0.052)
Female	-0.051	-0.134**	0.078	-0.060
	(0.039)	(0.058)	(0.060)	(0.052)
Trauma Priming × Female	0.002	-0.065	0.026	0.026
	(0.055)	(0.081)	(0.085)	(0.074)
Control Mean	0.05	0.01	0.07	0.08
Observations	1631	1652	1652	1652

This table shows the impact of priming and gender on the quality of mental experiencing, including the specificity, emotionality, and frequency of positive images. We show the ITT effect for an OLS regression. Indices are described in the note to Table 2. All regressions control for randomization strata (gender and prime treatment). Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

The table indicates that female respondents are able to use significantly less specific past images than their male counterparts ($p < 0.05$). It also appears that there is no effect of the trauma priming by gender.

2.2.3 Effect of Priming on Mental Experiencing, Interaction with PTSD

Table 4. Priming and Mental Experiencing at Baseline, by PTSD

	Overall Index	Sub-indices		
		Specificity	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Trauma Priming	-0.033	-0.033	0.015	-0.081**
	(0.029)	(0.043)	(0.046)	(0.040)
PTSD	0.091	0.225**	0.095	-0.049
	(0.063)	(0.092)	(0.094)	(0.086)
Trauma Priming × PTSD	0.110	-0.005	0.102	0.198*
	(0.081)	(0.120)	(0.124)	(0.111)
Control Mean	0.05	0.01	0.07	0.08
Observations	1631	1652	1652	1652

This table shows the impact of priming and trauma on the quality of mental experiencing, including the specificity, emotionality, and frequency of positive images. We show the ITT effect for an OLS regression. Indices are described in the note to Table 2. All regressions control for randomization strata (gender and prime treatment). Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

This table indicates that priming has a significant negative effect on the positivity of memories ($p < 0.05$) for people with no PTSD risk. However, this effect is reversed for respondents with PTSD ($p < 0.1$). The negative priming thus has an ambiguous effect on the quality of memories.

2.3 Interaction of Priming and GME Treatment

Table 5. Priming and GME treatment

	Overall Index	Sub-indices		
		Specificity	Emotionality	Positivity
	(1)	(2)	(3)	(4)
Trauma Priming	-0.030	-0.013	-0.107	0.029
	(0.043)	(0.065)	(0.067)	(0.057)
GME Treatment	0.073*	0.007	0.057	0.159***
	(0.043)	(0.062)	(0.069)	(0.055)
Trauma Priming x GME	0.036	0.047	0.153	-0.092
	(0.061)	(0.088)	(0.097)	(0.078)
Control Mean	0.05	0.01	0.07	0.08
Observations	1332	1332	1332	1329

This table shows the impact of priming on the quality of mental experiencing, including the specificity, emotionality, and frequency of positive images. We show the ITT effect for an OLS regression. Indices are described in the note to Table 2. Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

As the table above illustrates, we do not find a significant interaction between the GME treatment and the priming treatment.

3 Baseline Patterns

3.1 Relationship between Indices of Past/Future and Positive/Negative Mental Experiencing

Table 6. Correlation between indices related to positive scenarios

Variables	P: Emotions	P: Vividness	P: Frequency	F: Emotions	F: Vividness	F: Frequency
P: Emotions	1.000					
P: Vividness	0.850***	1.000				
P: Frequency	0.157***	0.134***	1.000			
F: Emotions	0.251***	0.243***	0.008	1.000		
F: Vividness	0.221***	0.253***	0.006	0.869***	1.000	
F: Frequency	0.072***	0.075***	0.082***	0.116***	0.125***	1.000

Table 7. Correlation between indices related to negative scenarios

Variables	P: Emotions	P: Vividness	F: Emotions	F: Vividness
P: Emotions	1.000			
P: Vividness	0.629***	1.000		
F: Emotions	0.110***	0.103***	1.000	
F: Vividness	0.185***	0.257***	0.627***	1.000

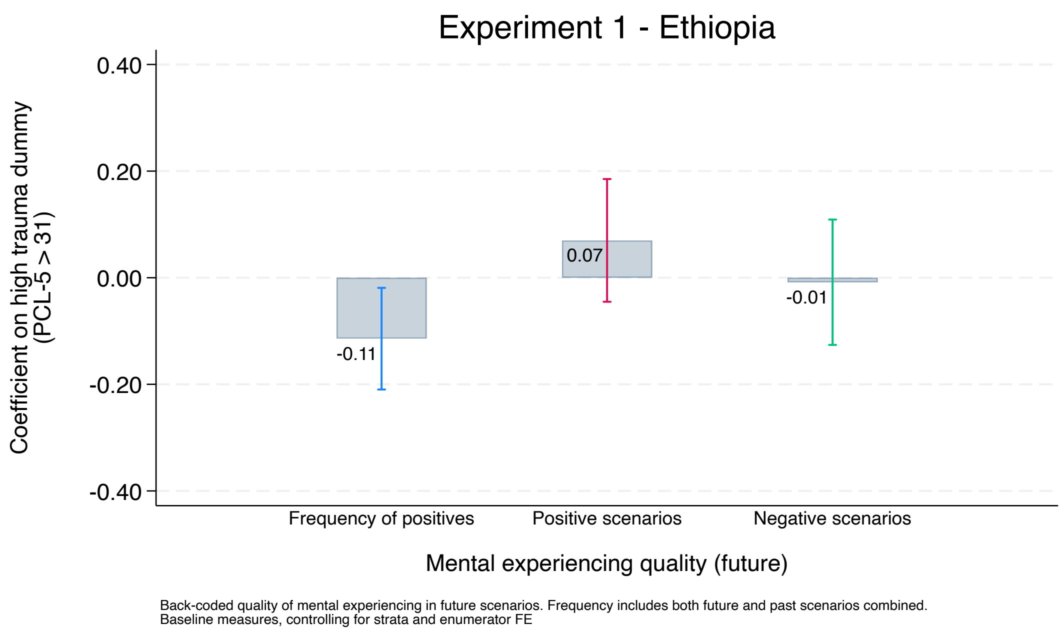
Table 8. Correlation between indices related to both positive and negative scenarios

Variables	P: Neg Emotions	P: Neg Vividness	P: Pos Emotions	P: Pos Vividness	F: Neg Emotions	F: Neg Vividness	F: Pos Emotions	F: Pos Vividness
P: Neg Emotions	1.000							
P: Neg Vividness	0.629***	1.000						
P: Pos Emotions	0.232***	0.154***	1.000					
P: Pos Vividness	0.187***	0.260***	0.850***	1.000				
F: Neg Emotions	0.110***	0.103***	0.078***	0.069***	1.000			
F: Neg Vividness	0.185***	0.257***	0.075***	0.130***	0.627***	1.000		
F: Pos Emotions	0.239***	0.269***	0.251***	0.243***	0.221***	0.272***	1.000	
F: Pos Vividness	0.228***	0.294***	0.221***	0.253***	0.221***	0.305***	0.869***	1.000

In Tables 6, 7 and 8, “P” denotes past memories, while “F” denotes future scenarios. As expected, the emotionality, vividness, and positivity of memories and future images within positive scenarios are positively correlated. Moreover, not only these correlations hold within the past or the future, but also between different time frames. This indicates that the quality components of memory and future mental experiencing are positively correlated, as hypothesized.

3.2 Trauma and Future Mental Experiencing

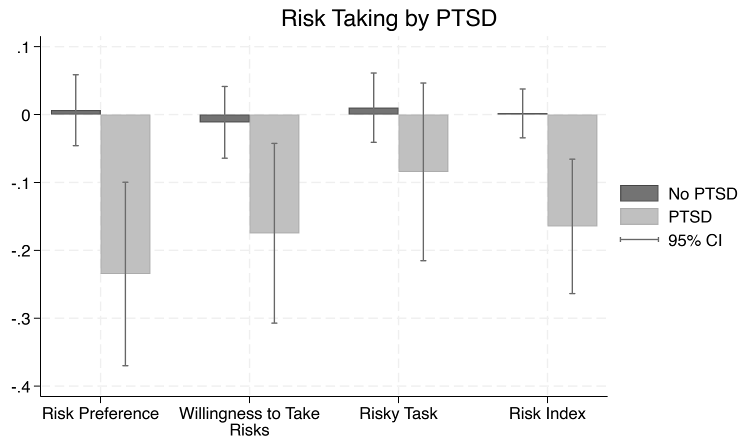
Figure 1. Trauma and the quality of future mental experiencing



In line with our hypotheses, respondents with PTSD symptoms appear to imagine positive future scenarios less frequently than respondents without PTSD symptoms. The quality of positive or negative future scenarios is similar to their non-traumatized counterparts.

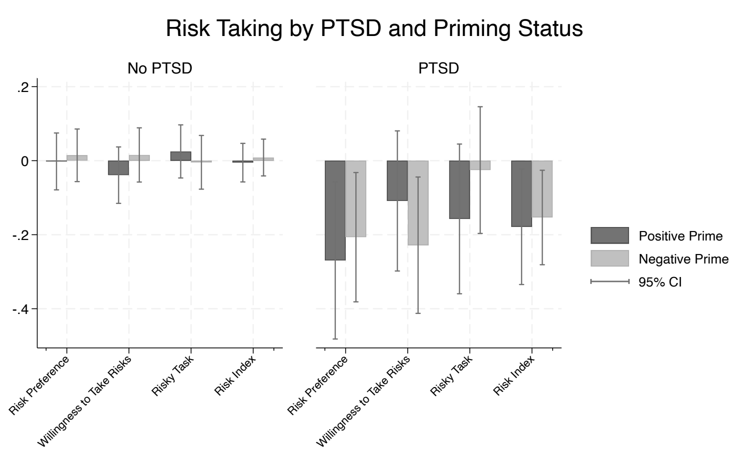
3.3 Trauma, Risk Preferences and Priming

Figure 2. Trauma and risk taking



In the graph above, risk preference is the self-declared (standardized) willingness to take risks ranked from 0 to 10. The willingness to take risks is the self-declared (standardized) willingness to take risks in the past 7 days, ranked from 0 to 10. Risky task is the standardized amount of birr the respondent declared they would be willing to invest in a risky investment opportunity. Finally, the risk index is a aggregated measure of the three other variables. In line with what we assumed in our PAP, respondents with PTSD are significantly more risk averse than those without PTSD.

Figure 3. Trauma, priming and risk taking



The figure above suggests that the effect of priming is significantly more pronounced among individuals with probable PTSD. Looking at the risk index as our best combined measure of risk preferences, we find that both the negative prime makes respondents slightly less - but not significantly - risk averse if they have PTSD. There is no such effect for people with no PTSD risk.

4 Treatment Effects on Risk Taking as per PAP

Table 9. GME Treatment Effect on Risk Taking

	(1) Risk Index	(2) Self Reported	(3) Self Reported (7d)	(4) Risk Task
GME Treatment	0.0127	0.0577	-0.0277	0.00823
	(0.0385)	(0.0541)	(0.0540)	(0.0520)
Observations	1395	1395	1395	1395
Mean DV	0.00637	0.0287	-0.0138	0.00422

This table shows the impact of the GME treatment on the three different risk preference measures (columns 2, 3, and 4) and an index that combines all three measures (column 1). Self-reported risk preference is the (standardized) willingness to take risks, self-reported on a scale from 0 to 10 using the survey question from the Global Preferences Survey Falk et al. (2018). Self-reported risk preference (7 day) is the same measure, but for a 7-day reference period. Risk task refers to the (standardized) amount of Ethiopian birr the respondent declared they would be willing to invest in a risky investment opportunity based on the Gneezy-Potters-Charness investment task (Charness and Gneezy, 2010; Gneezy and Potters, 1997). All regressions control for randomization strata (gender and prime treatment) and enumerator fixed effects. Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

This table shows that the GME treatment does not have a significant effect on the risk aversion of respondents. Therefore, risk aversion is likely not a mechanism explaining the improved downstream outcomes after the GME treatment.

5 Additional Analysis on Mechanisms as per PAP

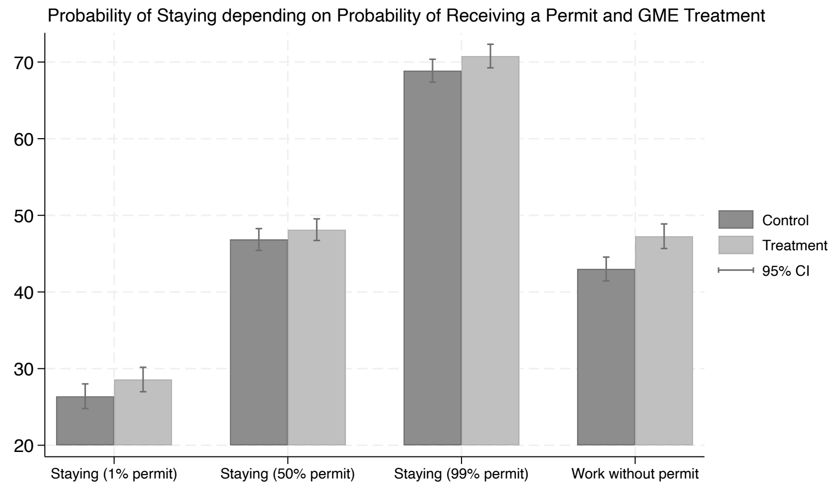
Table 10. Effect of GME Treatment on Risk Taking, Optimism, Aspirations, Self-Efficacy, and Reservation Wages

	Risk Index	Optimism	Aspirations	Self Efficacy	Reservation Wage
	(1)	(2)	(3)	(4)	(5)
GME Treatment	0.013	0.165***	0.099*	0.027	0.048
	(0.039)	(0.053)	(0.054)	(0.054)	(0.047)
Control Mean	-0.00	-0.00	-0.05	0.00	0.00
Observations	1395	1395	1395	1395	1395

This table shows the impact of the GME treatment on the intermediate outcomes likely to highlight the mechanisms at play. Risk Taking is the standardized index of risk preference measures described in Table 9. Optimism is the standardized value of the Revised Life Orientation Test (Scheier et al., 1994). Aspiration measures respondents' aspirations based on four 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?", "Do you see yourself continuing your education in the future?" and "What is the level of schooling you would like your eldest child to complete?". Self Efficacy is the standardized score of the generalized self-efficacy scale (Schwarzer and Jerusalem, 1995). The reservation wage index is an index based on declared reservation wages under various scenarios of receiving or not receiving a work permit in Ethiopia. All regressions control for randomization strata (gender and prime treatment) and enumerator fixed effects. Robust standard errors in parentheses, *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

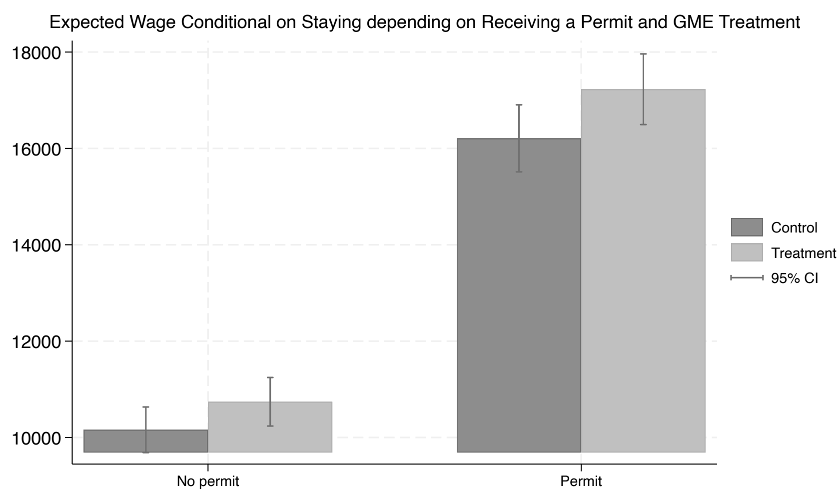
This table shows that the GME treatment has a strong, positive, and statistically significant ($p < 0.05$) effect on treated respondents' optimism. It also has a statistically significant ($p < 0.1$) positive effect on respondent's aspirations.

Figure 4. Treatment Effects on Probability of Staying



This graph suggests that the declared probability of staying is consistently higher for treated respondents compared to the control group, irrespective on the probability of obtaining a work permit. Moreover, the declared probability of working without a work permit is higher for the treated than for the control group. This evidence reinforces that one of main results of the intervention is an increase in intent to stay.

Figure 5. Treatment Effects on Expected Wage



This graph shows that the treatment seems to increase the expected wage for treated group compared to the control group, irrespective on receiving a permit or not. This is consistent with the other results we find on the actual increase in income from employment, as well as the treatment group's increased optimism.

6 Additional Treatment Effect Estimates

Table 11. Effect of GME Treatment on Labor Market Outcomes

	Any work	Own business	Casual work	Wage work	Unpaid work
	(1)	(2)	(3)	(4)	(5)
GME Treatment	0.124**	0.085	0.061	0.095*	-0.066
	(0.054)	(0.061)	(0.056)	(0.055)	(0.047)
Control Mean	-0.00	-0.00	0.00	-0.00	0.00
Observations	1395	1395	1395	1395	1395

This table shows the effect of the GME treatment on disaggregated employment outcomes. This is also shown in the paper in Appendix Table C7. Any work is a standardized dummy equal to 1 if the respondent worked in the past 7 days. Own business is a standardized dummy equal to 1 if the respondent had their own business in the past 7 days. Casual work is a standardized dummmy equal to 1 if the respondent worked under a casual contract in the past 7 day. Wage work is a standardized dummy equal to 1 if the respondent worked for wage income in the past 7 days. Finally, unpaid work is a standardized dummy equal to 1 if the respondent undertook any unpaid work in the past 7 days. All regressions control for randomization strata (gender and prime treatment). Robust standard errors in parentheses. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 11 indicates that the GME treatment has a significant ($p < 0.05$) positive effect on respondents getting any job. This effect appears to be mostly driven by respondents finding wage work, while the effect on owning a business and undertaking casual or unpaid work is insignificant.

Table 12. Effect of GME Treatment on Living Standards Index and Subindices

	Overall Index	Income	Consumption	Food Security	Savings	Transfers	Assistance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GME Treatment	0.102***	0.143**	0.056	0.219***	0.192***	0.042	-0.040
	(0.025)	(0.063)	(0.050)	(0.051)	(0.055)	(0.040)	(0.050)
Control Mean	0.00	0.00	0.00	0.00	-0.00	-0.00	0.00
Observations	1395	1395	1395	1395	1395	1395	1395

This table shows the effect of the GME treatment on the pre-specified wellbeing index and its subcomponents. The pre-analysis plan included a sub-index on housing characteristics; however, this measure was omitted from data collection due to modifications to the survey instrument during the final stage of instrument development. Column (1) presents the overall standardized index, columns (2) through (7) present the (standardized) sub-indices. Income refers to the standardized, unconditional sum of income over the past 7 days from self-employment or wage employment (same as reported in Table C7 of the appendix of the main manuscript). Consumption is an index of total expenditures, excluding those linked with outgoing and ongoing transfers such as remittances and other transfers (same as reported in Table 3 of the main manuscript). Food security is an index taken based on the Ethiopia Socioeconomic Survey (ESS) and includes reverse-coded questions about, e.g., not having enough food or the number of days with less preferred food (same as reported in Table 3 of the main manuscript). Savings is an index of total savings over the past 7 days. Transfers refers to an index of transfers to others households sent and received. Assistance refers to an index of any social assistance programs benefited from, e.g. Ethiopia's Urban Productive Safety Net. All regressions control for randomization strata (gender and prime treatment) and enumerator fixed effects. Robust standard errors in parentheses, sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 12 indicates that the positive welfare effects documented in the main manuscript are also reflected in the pre-registered, overall index of welfare (column 1).

7 Lee Bounds

Table 13. Lee Bounds for Table 2 of the Paper (Mental Experiencing Results)

	Mean Treatment Effect [N]	Lower/Upper Bounds [N]
Overall Index	0.091***	0.026/0.162***
	[1332]	[1287]
Specificity	0.030	-0.056/0.137***
	[1332]	[1287]
Emotionality	0.132***	0.001/0.241***
	[1332]	[1287]
Frequency of Positives	0.113***	0.059/0.221***
	[1329]	[1285]

Lee (2009) bounds for estimated treatment effects are shown in Column (2). The number of observations included when calculating the upper and lower bounds are shown in square brackets. Differential attrition is at 5.6 percentage points between the treatment and the control groups. *p<0.1, ** p<0.05, *** p<0.01.

Table 14. Lee Bounds for Table 3 of the Paper (Downstream Outcomes Results)

	Standardised Mean Treatment Eff. [N]	Diff. Attrition 3.2% [N]
Intent to Stay	0.178***	0.083/0.282
	[1395]	[1350]
Job Search	-0.097**	-0.254/-0.096
	[1395]	[1350]
Hours Worked	0.145**	-0.051/0.145**
	[1395]	[1350]
Employment	0.124**	-0.516***/0.168***
	[1395]	[1350]
Food Security	0.219***	0.172***/0.344***
	[1395]	[1350]
Consumption	0.056	-0.091/0.126**
	[1395]	[1350]
Life Satisfaction	0.131**	-0.002/0.268***
	[1395]	[1350]
WHODAS Score	0.158***	0.130**/0.313***
	[1395]	[1350]

Lee (2009) bounds for estimated treatment effects are shown in Column (2). The number of observations included when calculating the upper and lower bounds are shown in square brackets. Differential attrition is at 5.6 percentage points between the treatment and the control groups. *p<0.1, ** p<0.05, *** p<0.01.

Tables 13 and 14 show that our main results are broadly robust using Lee Bounds to correct for differential attrition between the treatment and control groups.

References

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies 81*(2), 608–650.

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.

Callen, M., M. Isaqzadeh, J. D. Long, and C. Sprenger (2014). Violence and Risk Preference: Experimental Evidence from Afghanistan. *American Economic Review 104*(1), 123–148.

Charness, G. and U. Gneezy (2010). Portfolio Choice and Risk Attitudes: An Experiment. *Economic Inquiry 48*(1), 133–146.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India. Technical report, National Bureau of Economic Research.

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.

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.

Gneezy, U. and J. Potters (1997). An Experiment on Risk Taking and Evaluation Periods. *The Quarterly Journal of Economics 112*(2), 631–645.

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., S. A. Hales, K. Young, and M. Di Simplicio (2019). *Imagery-based cognitive therapy for bipolar disorder and mood instability.* Imagery-based cognitive therapy for bipolar disorder and mood instability. New York, NY, US: The Guilford Press. Pages: viii, 184.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies 76*(3), 1071–1102.

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.

Romppel, M., C. Herrmann-Lingen, R. Wachter, F. Edelmann, H.-D. Düngen, B. Pieske, and G. Grande (2013). A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an intercultural non-clinical sample and a sample of patients at risk for heart failure. *GMS Psycho-Social-Medicine 10*(1). Medium: text/html Publisher: German Medical Science GMS Publishing House.

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.

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.

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.

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

Pre-Analysis Plan Report (Colombia)

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

This document describes how the analysis of the Colombia experiment presented in the main body of the paper and online appendix aligns with the pre-analysis plan registered with the American Economic Association RCT Registry as AEARCTR-0004695 (https://www.socialscienceregistry.org/trials/4695)

[*]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
[‖]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

Contents

1. Comparison Between Pre-Analysis Plan and Paper **3**

2. Mental Experiencing: Table with Indices Disaggregated by Valence **9**

3. Exploratory Analysis: Treatment Assignment as Instrument for Attendance **11**

4. Heterogeneity Analysis **13**

4.1 Pre-Registered Threshold for Baseline Trauma (IES > 33) . . . 13

4.2 Baseline Trauma and Well-Being Outcomes . . . 14

4.3 Alternative Threshold for Baseline Trauma (IES > 24) . . . 16

5. Robustness of Main Results Removing Winsorisation **18**

6. Dealing with Attrition: Lee Bounds (extra, not included in PAP) **19**

1. Comparison Between Pre-Analysis Plan and Paper

Table 1: Differences between final manuscript and pre-registered analysis plan

Pre-registered analysis plan	**Final Manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
1. **Introduction** The research goal of the study is to test whether visualisation can be taught and whether it improves psychological and economic resilience among vulnerable recipients.	Done as specified.	
2. **Description of the Interventions** We designed an entrepreneurship training overlaid with visualization techniques and a traditional training structured so as to reduce confounding differences - 2.a Teaching Visualisations We drew on mental imagery techniques developed by PI Emily Holmes and employed in clinical psychology. All visualization exercises followed a standardized structure: 1-Envisioning the future, 2-Mental practice of goals and sub steps, 3-Adopting the perspective of others	Implemented as specified. The final manuscript refers to the imagery intervention as guided mental experiencing (GME), as a more descriptive and intuitive label for our intervention.	
3. **Timeline of implementation and data collection** The intervention was implemented in two blocks: Wave 1 took place between July and September 2019 and Wave 2 took place between September and December 2019.	Done as specified	
4. **Sample** Our government partner, SDIS, recruited participants through a multi-channel media campaign, which were then screened according to predefined criteria. Randomisation into the three treatment arms was stratified based on some demographic variables. We doubled the sample size for the visualization treatment in order to meet the budget reporting requirements of our government partner.	Done as specified.	

Table 1: Differences between final manuscript and pre-registered analysis plan

Pre-registered analysis plan	Final Manuscript, adherence to PAP, and reasons	Additional Tables and Figures
5. **Analysis while blind to treatment assignment** Innovations for Poverty Action (IPA) Colombia provided the research team with all data omitting treatment status. We assessed whether the variables were poorly measured, contained very little variation or many large outliers. We then revised the PAP with an addendum submitted to the AEA RCT registry.	Done as specified.	We report the alignment with the addendum below.
6. **Primary outcomes**		
• **6.a Outcome Variables construction** Unless otherwise specified, we construct our indices within each family following Kling et al. (2007).	Done as specified	
• **6.b Family 1: Mental Imagery** *Adapted version of the Spontaneous Use of Imagery Scale* (Nelis et al., 2019): To measure the frequency of visualization usage, we adapted the Spontaneous Use of Imagery Scale *Adapted Prospective Imagery* Task (Stöber, 2000): To assess the quality of visualization in an entrepreneurship domain, we adapted the Prospective Imagery Task. We build four indices from this scale: (1) Two indices for positive scenarios: one for emotional valence and one for vividness; (2) Two indices for negative scenarios: one for emotional valence and one for vividness.	In the main paper, we do not distinguish between positive and negative valence, but we report one index for emotional valence (called "emotionality") and one index for vividness (called "specificity"). Moreover, in the main paper, we distinguish between an index in the business domain and outside the business domain.	Tables 3 and 4 reported below distinguish between positive and negative valence.
• **6.c Family 2: Psychological Resilience** We build two indices: (1) psychological resilience and (2) anxiety and depression.	Done as specified.	

Table 1: Differences between final manuscript and pre-registered analysis plan

Pre-registered analysis plan	Final Manuscript, adherence to PAP, and reasons	Additional Tables and Figures
• **6.d Family 3: Economic resilience** We distinguish between two time periods: (1) economic activity before the COVID-19-induced lockdown and (2) economic activity during the COVID-19 lockdown period.	Done as specified, except that the investment questions referred to investments potentially done during both periods. In the endline survey in November 2020, we asked about investment between June 2020 and the interview. The final index thus includes investment across both the pre-COVID and COVID periods. This is why the paper includes it within the COVID period.	
7.a **Empirical Strategy** For every primary outcome, we estimate the following specification at the individual level: $y_i = \beta_0 + \beta_1 \cdot \text{imagery}y_i + \theta \cdot m_i + \delta \cdot x_i + \eta + \mu_i$ We include wave and wave-sub direction fixed effects η. μ_i is a mean zero error term. For the imagery and "business-as-usual" treatment arms, we restrict the sample for analysis to people who confirmed their participation to the programs (before treatment assignment).We nonetheless collected outcomes on the full sample. Randomization occurred at the individual level for most of the sample. For all specifications, we use robust standard errors to correct for heteroskedasticity. We also estimate the intent-to-treat impact of the "business-as-usual" training relative to the pure control. We will also explore instrumenting using random assignment as an instrument for attendance.	Done as specified, except: Given that we have two different survey waves, we add a survey wave dummy and cluster the standard error at the household level. We decided to collapse the dataset at the household level and cluster the standard errors at the household level to account for those participants who live together and/or belong to the same formerly-homeless cluster, which were assigned to the same treatment status. The intent-to-treat impact of the "business-as-usual" training relative to the pure control is reported in Panel C in the main tables in the paper.	As exploratory analysis, the PAP also proposed to use an instrumental variable strategy to instrument attendance with treatment assignment. We do not report this analysis in the paper, but show it below in Tables 5 and 6.
7.b **Correction for multiple hypothesis testing** We use false discovery rate (FDR) corrections (Benjamini et al., 2006) to account for multiple comparisons across our indices of primary outcome variables within Families 1, 2 and 3. We will not correct for multiple hypothesis testing in our exploratory analysis.	Done as specified, and shown in the paper.	

Table 1: Differences between final manuscript and pre-registered analysis plan

<table>
<tr><th>Pre-registered analysis plan</th><th>Final Manuscript, adherence to PAP, and reasons</th><th>Additional Tables and Figures</th></tr>
<tr><td>7.c Covariates

In our analysis, we first present results with only our randomization strata dummy variables and fixed effects. To increase the precision of our estimates, following Belloni et al. (2014), we adopt the “post-double-selection” method for selecting regressors, including first-order interaction and quadratic terms.</td><td>Done as specified, and shown in the paper.</td><td></td></tr>
<tr><td>8. Heterogeneity Analysis

We expect our imagery program to have differential treatment effects for those individuals who show higher symptoms of post-traumatic stress at baseline. To conduct heterogeneity analysis, we run a fully interacted model, whereby the key coefficients are on the treatment, a dummy for high reported trauma symptoms at baseline and their interaction.

1. We give a score of zero in the Impact of Event Scale-Revised to all respondents who reported that they did not experience a traumatic event in the past.
2. We create a dummy variable for those participants with an Impact of Event Scale-Revised score of above 33 at baseline.
3. We include a dummy for high reported post-trauma symptoms and an interaction.
4. We correct for multiple hypothesis testing across indices within families.
5. We conduct a two-sided test to test the null hypothesis that there are no heterogenous treatment effects.</td><td>Construction of variable as specified. However, because we only have approximately a third of the sample above a score of 33 in the IES-R, the paper uses an alternative index for “high trauma” that provides a more balanced split of the sample. This index combines the pre-registered IES threshold (IES $>$ 33) with an indicator for whether a participant reports to be a recent Venezuelan migrant or a victim of the Colombian civil conflict (i.e. it is a dummy equal to 1 if a person has as IES score above 33 or is registered as victim/migrant). These were two key populations that were targeted for the program. This is the main trauma index used in the paper for heterogeneity.
In this report, we show robustness using i) the pre-registered threshold and ii) using an alternative threshold of 24 in the Impact of Event Scale-Revised (this was specified in our PAP addendum, see below).
For brevity, we do not show in the paper heterogeneity by trauma for the well-being outcomes, but it is shown in this report.</td><td>Table 7 shows a fully interacted model for the main outcomes of the paper using the pre-registered IES$>$ 33 threshold for high trauma.
Tables 8 and 9 show a fully interacted model for well-being outcomes using the different definitions for high trauma.
Tables 10 and 11 show a fully interacted model for the main and well-being outcomes of the paper using the $IES > 24$ threshold for high trauma.</td></tr>
</table>

Table 2: Differences between final manuscript and addendum to pre-registered analysis plan

PAP Addendum	Final Manuscript, adherence to PAP, and reasons	Additional Tables and Figures
A1. Transforming variables We will take a logarithmic transformation of all the monetary variables prior to standardising. We will impute the mid-point for responses that were provided in ranges. Ranges were an option for respondents who felt uncomfortable or unable to provide exact estimates of their income, revenues and savings. We will also multiply the income measures by four to convert them from weekly to comparable monthly measures.	Done as specified, using an Inverse Hyperbolic Syne transformation for all the mentioned variables.	
A2. Treating outliers We found that many distributions were highly skewed to the right. We will also winsorize at the 99th percentile and present results with and without winsorization.	As specified, we winsorized at the 99^{th} percentile and presented all results on all variables with these extreme values with winsorization.	In the main paper, we didn't include results without winsorization. This analysis can be found below in Table 12.
A3. Imagery and psychological resilience indexes: minimum detectable effect We found that imagery and psychological scales for resilience and self-efficacy tended to be skewed towards high values for most of the scale items. We assigned fake treatment statuses to respondents, with shares belonging to the three treatment arms which correspond to the actual shares of the field experiment. We then estimated treatment effects on the imagery index of the visualization training against one of the other treatment arms, controlling for strata to find the MDE.	Done as specified.	
A4. Treating standard errors For our primary specifications, we present both robust standard errors and randomization inference standard errors, following Young (2019).	Done as specified.	

Table 2: Differences between final manuscript and addendum to pre-registered analysis plan

PAP Addendum	**Final Manuscript, adherence to PAP, and reasons**	**Additional Tables and Figures**
A5. Heterogeneity analysis using baseline trauma When conducting our heterogeneity analysis, we use an impact of event score of 33, as pre-specified, threshold above which post-traumatic stress symptoms may be considered to be a *probable* clinical concern. To address potential power concerns, we will also use a threshold of 24 in our analysis, above which post-traumatic stress symptoms are suggestive of a clinical concern. We will treat this as a separate family of outcomes and correct for multiple hypotheses testing within the family.	This analysis is not shown in the paper, but can be found below.	Tables 10 and 11 show trauma heterogeneity on main outcomes and wellbeing outcomes with the IES $>$ 24 threshold.
A6. EXTRA NOT IN PAP - Dealing with attrition In the Colombia PAP, we did not specify a strategy to deal with attrition. Yet, in the Ethiopia PAP, we specified that we would use Lee Bounds (Lee, 2009) and Inverse Probability Weighting (IPW). For completeness, we report on these strategies here.	Tables with IPW are shown in the main appendix of the paper.	Table 13 reported below shows Lee Bounds.

2. Mental Experiencing: Table with Indices Disaggregated by Valence

Table 3: Results on Overall Mental Experiencing: Disaggregated Indices with Valence

	Overall	Subcomponents of Overall Index				
	Overall index	Frequency of use	Positive specificity	Positive emotionality	Negative specificity	Negative emotionality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Traditional vs Control						
Traditional training	-0.069	-0.020	-0.034	-0.051	-0.024	-0.106
	(0.075)	(0.073)	(0.075)	(0.071)	(0.068)	(0.071)
				[0.889]	[0.939]	[0.695]
Panel B: GME vs Traditional						
GME training	0.026	-0.072	0.021	0.024	0.056	0.098*
	(0.063)	(0.061)	(0.063)	(0.060)	(0.059)	(0.059)
				[0.858]	[0.516]	[0.405]
Panel C: GME vs Control						
GME training	0.014	-0.068	0.038	0.023	0.064	0.038
	(0.066)	(0.065)	(0.064)	(0.061)	(0.061)	(0.060)
				[1.000]	[1.000]	[1.000]
Mean DV in Control	0.00	0.00	0.00	0.00	0.00	0.00
Mean DV in Traditional	0.00	0.00	0.00	0.00	0.00	-0.10
N in Control	550	550	547	547	547	547
N in Traditional	656	656	653	653	654	654
N in GME	1140	1139	1135	1135	1136	1136

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the quality of mental experiencing. The overall index in Column (1) is aggregates the measures of frequency of use, specificity and emotionality. Specificity refers to the clarity of the images respondents are able to create, for positive (in Column (3)) or negative (in Column (5)) scenarios. Emotionality refers to the strength of the emotions these images entail, , for positive (in Column (4)) or negative (in Column (6)) scenarios. Frequency of use refers to the frequency with which respondents use mental experiencing. All indexes are based on self-reports, and computed on all domains across scenarios (related or not to business). All regressions control for randomization strata and survey wave fixed effects. Clustered standard errors are in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Column (1) of Table 3 replicates the main results shown in the paper (Column 1 of Table 4). Columns (3) to (6) split the overall mental experiencing index by positive or negative valence. Results are very similar across indices, with just slightly bigger coefficients on "negative emotionality", i.e. the intensity of emotions for negative scenarios. This effect is in line with one of the main goals of the program in Colombia, which was to reduce avoidance and train participants to tolerate negative emotions and leverage them for productivity.

Table 4: Results on Business-Related Mental Experiencing: Disaggregated Indices with Valence

	Overall			Subcomponents of Business Index				
	Overall index	Business index	Non-business index	Frequency of use	Positive specificity	Positive emotionality	Negative specificity	Negative emotionality
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Traditional vs Control								
Traditional training	-0.069	-0.157*	-0.049	-0.073	-0.086	-0.126	-0.096	-0.085
	(0.075)	(0.089)	(0.074)	(0.086)	(0.087)	(0.083)	(0.078)	(0.079)
				[0.653]	[0.653]	[0.653]	[0.653]	[0.653]
Panel B: GME vs Traditional								
GME training	0.026	0.178**	-0.006	0.090	0.073	0.091	0.101	0.162**
	(0.063)	(0.074)	(0.061)	(0.072)	(0.072)	(0.069)	(0.066)	(0.063)
				[0.264]	[0.327]	[0.264]	[0.264]	[0.058]
Panel C: GME vs Control								
GME training	0.014	0.046	0.004	0.023	0.039	-0.031	0.020	0.073
	(0.066)	(0.076)	(0.064)	(0.073)	(0.074)	(0.069)	(0.069)	(0.066)
				[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Mean DV in Control	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean DV in Traditional	0.00	-0.10	0.00	0.00	0.00	-0.10	0.00	0.00
N in Control	550	390	550	390	390	390	389	390
N in Traditional	656	456	656	456	452	453	452	450
N in GME	1140	839	1140	838	835	835	832	833

This table shows the same results as Table 3 above, but distinguishes between business and non-business related scenarios. Columns (4) to (8) only focus on business-related scenarios.

Columns (1) to (4) of Table 4 replicate the main results shown in the paper (Columns (1) to (4) of Table 4). Columns (5) to (8) split the business-related mental experiencing index by positive or negative valence. As discussed for the table above, results are very similar across indices. This confirms a general improvement in mental experiencing in both the positive and negative domains, in line with the GME exercises focusing on building realistic future scenarios.

3. Exploratory Analysis: Treatment Assignment as Instrument for Attendance

Table 5: Results on Mental Experiencing Instrumenting Attendance with Treatment Assignment

				Sub-indices in Business Domain		
	Overall index	Business index	Non business index	Frequency of use	Specificity	Emotionality
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Traditional vs Control						
Traditional Training	-0.011	-0.035	-0.006	-0.017	-0.028	-0.030
	(0.020)	(0.022)	(0.020)	(0.021)	(0.020)	(0.020)
GME Treatment	-0.004	0.007	-0.007	0.004	0.005	0.009
	(0.019)	(0.020)	(0.018)	(0.019)	(0.018)	(0.018)
p-value: Traditional vs GME	0.670	0.021	0.951	0.233	0.055	0.015
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	-0.10
N in Control	550	390	550	390	390	390
N in Traditional	656	456	656	456	454	454
N in GME	1140	839	1140	838	835	835

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the quality of mental experiencing. Attendance in each program is instrumented with the original treatment assignment. The overall index in Column (1) is aggregates the measures of frequency of use, specificity and emotionality. The business (non-business) index is constructed considering only scenarios in (outside) the business domain. Specificity refers to the clarity of the images respondents are able to create. Emotionality refers to the strength of the emotions these images entail. Frequency of use refers to the frequency with which respondents use mental experiencing. All indexes are based on self-reports. All regressions control for randomization strata and survey wave fixed effects. Clustered standard errors are in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 5 shows results from six different regressions of mental experiencing indices on two main independent variables for attendance to each program (Traditional and GME training). Attendance is instrumented with the original treatment assignment. The omitted group is the control group. The last row of the table ("p-value: Traditional vs GME") shows the p-value of a test of equality of the two coefficients shown in the table. The Table confirms the main results of the paper: the traditional training decreases mental experiencing quality, and the two programs create a wedge in mental experiencing quality in the business domain (which is statistically different to zero, $p = 0.02$).

Table 6: Results on Economic Outcomes Instrumenting Attendance with Treatment Assignment

	Overall Indices		Pre-covid sub-indices		Covid sub-indices				
	Pre-covid	Covid	Earnings	Business	Earnings	Business	Safety nets	Covid actions	Investment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Traditional vs Control									
Traditional Training	-0.036*	-0.021	-0.055**	0.003	-0.033*	-0.005	0.005	-0.022	-0.007
	(0.021)	(0.019)	(0.026)	(0.009)	(0.019)	(0.008)	(0.018)	(0.018)	(0.018)
GME Treatment	-0.002	0.016	-0.007	0.008	0.004	0.008	0.011	-0.000	0.011
	(0.018)	(0.017)	(0.019)	(0.008)	(0.016)	(0.008)	(0.017)	(0.017)	(0.017)
p-value: Traditional vs GME	0.052	0.018	0.030	0.468	0.021	0.056	0.671	0.145	0.184
Mean DV in Control	0.00	0.00	0.00	0.70	0.00	0.70	0.00	0.00	0.00
Mean DV in Traditional	-0.10	-0.10	-0.20	0.70	-0.10	0.60	0.00	-0.10	-0.10
N in Control	346	552	323	333	539	552	546	551	564
N in Traditional	407	660	380	392	642	659	650	657	667
N in GME	704	1147	665	679	1115	1145	1134	1142	1159

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on downstream economic outcomes. Attendance in each program is instrumented with the original treatment assignment. The overall indices are constructed from the aggregation of the sub-indices listed in Columns (3) to (9). The outcome variable in Columns (3) and (5) is the inverse-hyperbolic sine of earnings (income and sales). In Columns (4) and (6), the outcome is an indicator for having an operating business. The index of safety nets includes both actual (e.g., savings) and perceived safety nets during the Covid pandemic. The index of Covid actions averages different measures that respondents may have taken in response to the pandemic. The investment index is the standardized share of different category investments that a person made. All regressions control for randomization strata and survey wave fixed effects. Clustered standard errors are in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 6 shows results from nine different regressions of economic outcomes on two main independent variables for attendance to each program (Traditional and GME training). Attendance is instrumented with the original treatment assignment. The omitted group is the control group. The last row of the table ("p-value: Traditional vs GME") shows the p-value of a test of equality of the two coefficients shown in the table. The Table confirms the main results of the paper: the traditional training decreases earnings, and the two programs create a wedge in economic outcomes both in the pre-Covid and Covid period (which are statistically different to zero, $p = 0.05$ and $p = 0.018$ respectively).

4. Heterogeneity Analysis

4.1 Pre-Registered Threshold for Baseline Trauma (IES > 33)

Table 7: Main Outcomes With Trauma Interaction (Trauma defined as Pre-Registered Index: IES > 33)

	Mental Experiencing (business)	Earnings pre-COVID	Earnings during COVID	Business pre-COVID	Business during COVID
	(1)	(2)	(3)	(4)	(5)
Panel A: Traditional vs Control					
Traditional Training	-0.200*	-0.209*	-0.100	-0.000	-0.027
	(0.116)	(0.121)	(0.089)	(0.040)	(0.039)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
High Trauma (IES > 33)	0.137	0.129	0.023	0.022	0.003
	(0.141)	(0.114)	(0.106)	(0.060)	(0.059)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Traditional Training × High Trauma	-0.049	-0.228	-0.153	-0.028	-0.018
	(0.203)	(0.231)	(0.158)	(0.079)	(0.077)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Trad + Trad × High Trauma	-0.249	-0.437	-0.253	-0.029	-0.045
p-value: Trad + Trad × High Trauma = 0	0.147	0.031	0.057	0.683	0.506
Panel B: GME vs Traditional					
GME training=1	0.154*	0.199*	0.109	0.012	0.064**
	(0.092)	(0.105)	(0.075)	(0.032)	(0.031)
	[0.777]	[0.777]	[0.892]	[1.000]	[0.777]
High Trauma (IES > 33)	0.066	-0.154	-0.150	-0.012	-0.021
	(0.152)	(0.199)	(0.121)	(0.053)	(0.052)
	[1.000]	[1.000]	[0.892]	[1.000]	[1.000]
GME Treatment × High Trauma	0.207	0.160	0.085	0.076	0.032
	(0.185)	(0.229)	(0.147)	(0.066)	(0.064)
	[0.892]	[1.000]	[1.000]	[0.892]	[1.000]
GME + GME × High trauma	0.361	0.359	0.194	0.088	0.095
p-value: GME + GME × High trauma = 0	0.025	0.076	0.124	0.127	0.091
Panel C: GME vs Control					
GME Treatment	0.005	0.013	0.027	0.027	0.047
	(0.095)	(0.084)	(0.075)	(0.036)	(0.035)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
High Trauma (IES > 33)	0.157	0.098	-0.057	0.022	0.005
	(0.142)	(0.102)	(0.106)	(0.058)	(0.059)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
GME Treatment × High Trauma	0.120	-0.076	-0.022	0.042	-0.001
	(0.171)	(0.144)	(0.135)	(0.070)	(0.069)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
GME + GME × High trauma	0.125	-0.063	0.006	0.069	0.046
p-value: GME + GME × High trauma = 0	0.382	0.584	0.961	0.260	0.446
Mean DV in Control	0.00	0.00	0.00	0.70	0.70
Mean DV in Traditional	-0.10	-0.20	-0.10	0.70	0.60
N in Control	390	323	539	333	552
N in Traditional	456	380	642	392	659
N in GME	839	665	1115	679	1145

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the main outcomes of the experiment, interacted with an index for "High Trauma". We categorize respondents as "High Trauma" if they score 33 or above on the Impact of Event Score (IES-R) at baseline. Panel A shows the effect of the traditional training 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. Regressions models and controls are defined as in the main tables of the paper. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 7 replicates the results of Table C.18 of the paper, but using the pre-registered index of "high trauma". Results are qualitatively similar between the two tables. The treatment effects in the "high trauma" sample are also very similar in magnitudes in the two cases.

4.2 Baseline Trauma and Well-Being Outcomes

Table 8: Well-being Outcomes With Trauma Interaction (Trauma defined as pre-registered: IES > 33)

	Overall Index	Sub-indices	
	Wellbeing	Kessler	Resilience and self-efficacy
	(1)	(2)	(3)
Panel A: Traditional vs Control			
Traditional Training	-0.056	-0.063	-0.023
	(0.081)	(0.081)	(0.089)
		[1.000]	[1.000]
High Trauma (IES> 33)	-0.608***	-0.647***	0.025
	(0.135)	(0.132)	(0.133)
		[0.001]	[1.000]
Traditional Training × High Trauma	0.146	0.166	-0.018
	(0.176)	(0.173)	(0.182)
		[1.000]	[1.000]
Trad + Trad × High Trauma	0.090	0.103	-0.040
p-value: Trad + Trad × High Trauma = 0	0.570	0.503	0.802
Panel B: GME vs Traditional			
GME Treatment	-0.002	0.008	-0.026
	(0.067)	(0.066)	(0.076)
		[1.000]	[1.000]
High Trauma (IES> 33)	-0.490***	-0.504***	-0.032
	(0.117)	(0.116)	(0.127)
		[0.001]	[1.000]
GME Treatment × High Trauma	0.085	0.093	-0.039
	(0.146)	(0.144)	(0.161)
		[1.000]	[1.000]
GME + GME × High Trauma	0.084	0.101	-0.065
p-value: GME + GME × High Trauma	0.518	0.433	0.647
Panel C: GME vs Control			
GME Treatment	-0.052	-0.057	-0.005
	(0.077)	(0.076)	(0.083)
		[1.000]	[1.000]
High Trauma (IES> 33)	-0.662***	-0.698***	-0.013
	(0.135)	(0.132)	(0.136)
		[0.001]	[1.000]
GME Treatment × High Trauma	0.242	0.270*	-0.062
	(0.160)	(0.157)	(0.166)
		[0.274]	[1.000]
GME + GME × High Trauma	0.190	0.213	-0.067
p-value: GME + GME × High Trauma = 0	0.181	0.125	0.647
Mean DV in Control	0.00	0.00	0.00
Mean DV in Traditional	0.00	0.00	0.00
N in Control	545	545	545
N in Traditional	649	647	649
N in GME	1134	1133	1134

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the well-being outcomes, interacted with an index for "High Trauma". We categorize respondents as "High Trauma" if they score 33 or above on the Impact of Event Score (IES-R) at baseline. The first sub-index captures psychological resilience (Sinclair and Wallston, 2004; Smith et al., 2008; Chen et al., 2001). The second sub-index reflects psychological distress, as measured by the Kessler K6 non-specific distress scale (Kessler et al., 2002). Panel A shows the effect of the traditional training 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. Regressions models and controls are defined as in the main tables of the paper. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 9: Well-being Outcomes With Trauma Interaction (Trauma defined as in the paper: IES > 33 & Dummy for Being Victim/Venuezuela Migrant)

	Overall Index	Sub-indices	
	Wellbeing	Kessler	Resilience and self-efficacy
	(1)	(2)	(3)
Panel A: Traditional vs Control			
Traditional Training	0.041	0.030	0.049
	(0.090)	(0.089)	(0.097)
		[0.582]	[0.582]
High Trauma	-0.207*	-0.251**	0.161
	(0.107)	(0.106)	(0.105)
		[0.126]	[0.286]
Traditional Training × High Trauma	-0.154	-0.125	-0.214
	(0.137)	(0.136)	(0.143)
		[0.368]	[0.286]
Trad + Trad × High Trauma	-0.113	-0.095	-0.166
p-value: Trad + Trad × High Trauma = 0	0.286	0.366	0.118
Panel B: GME vs Traditional			
GME Treatment	-0.059	-0.042	-0.085
	(0.075)	(0.073)	(0.083)
		[0.888]	[0.811]
High Trauma	-0.396***	-0.412***	-0.060
	(0.094)	(0.093)	(0.103)
		[0.001]	[0.888]
GME Treatment × High Trauma	0.162	0.155	0.118
	(0.118)	(0.116)	(0.129)
		[0.811]	[0.811]
GME + GME × High Trauma	0.103	0.113	0.033
p-value: GME + GME × High Trauma	0.259	0.208	0.738
Panel C: GME vs Control			
GME Treatment	-0.000	-0.002	-0.006
	(0.082)	(0.081)	(0.090)
		[1.000]	[1.000]
High Trauma	-0.237**	-0.276**	0.130
	(0.108)	(0.108)	(0.106)
		[0.067]	[1.000]
GME Treatment × High Trauma	-0.001	0.019	-0.092
	(0.127)	(0.127)	(0.130)
		[1.000]	[1.000]
GME + GME × High Trauma	-0.002	0.017	-0.098
p-value: GME + GME × High Trauma = 0	0.985	0.862	0.302
Mean DV in Control	0.00	0.00	0.00
Mean DV in Traditional	0.00	0.00	0.00
N in Control	545	545	545
N in Traditional	649	647	649
N in GME	1134	1133	1134

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the well-being outcomes, interacted with an index for "High Trauma". We categorize respondents as "High Trauma" if they score 33 or above on the Impact of Event Score (IES-R) at baseline, or they report being a victim of the internal conflict and/or Venezuelan migrant. The first sub-index captures psychological resilience (Sinclair and Wallston, 2004; Smith et al., 2008; Chen et al., 2001). The second sub-index reflects psychological distress, as measured by the Kessler K6 non-specific distress scale (Kessler et al., 2002). Panel A shows the effect of the traditional training 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. Regressions models and controls are defined as in the main tables of the paper. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Tables 8 and 9 broadly show that there are no effects of the treatments on well-being outcomes, either in terms of distress (measured using the Kessler-6 scale) or resilience and self-efficacy. They also highlight how participants with high trauma at baseline have worse well-being outcomes. While the effect of the treatments on wellbeing within the high-trauma sample is generally positive, coefficients are estimated with noise.

4.3 Alternative Threshold for Baseline Trauma (IES > 24)

Table 10: Main Outcomes with Trauma Interaction Using Alternative Trauma Indicator (IES > 24)

	Mental Experiencing (business)	Earnings pre-COVID	Earnings during COVID	Business pre-COVID	Business during COVID
	(1)	(2)	(3)	(4)	(5)
Panel A: Traditional vs Control					
Traditional Training	-0.198	-0.268**	-0.138	-0.013	-0.045
	(0.124)	(0.131)	(0.094)	(0.043)	(0.041)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
High Trauma (IES > 24)	0.166	0.081	-0.063	0.059	0.030
	(0.134)	(0.120)	(0.109)	(0.055)	(0.054)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Traditional Training × High Trauma	-0.054	0.002	0.003	0.016	0.042
	(0.184)	(0.221)	(0.154)	(0.073)	(0.071)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Trad + Trad × High Trauma	-0.253	-0.266	-0.135	0.003	-0.004
p-value: Trad + Trad × High Trauma = 0	0.078	0.133	0.276	0.955	0.952
Panel B: GME vs Traditional					
GME Treatment	0.176*	0.227*	0.126	0.036	0.090***
	(0.100)	(0.116)	(0.080)	(0.035)	(0.033)
	[0.586]	[0.552]	[0.615]	[0.781]	[0.095]
High Trauma (IES > 24)	0.073	-0.000	-0.084	0.057	0.058
	(0.135)	(0.173)	(0.113)	(0.047)	(0.047)
	[1.000]	[1.000]	[1.000]	[0.781]	[0.781]
GME Treatment × High Trauma	0.112	0.037	0.004	-0.010	-0.061
	(0.165)	(0.200)	(0.138)	(0.060)	(0.058)
	[1.000]	[1.000]	[1.000]	[1.000]	[0.781]
GME + GME × High Trauma	0.287	0.264	0.130	0.025	0.029
p-value: GME + GME × High Trauma = 0	0.031	0.099	0.241	0.606	0.549
Panel C: GME vs Control					
GME Treatment	0.040	0.000	0.010	0.041	0.058
	(0.101)	(0.088)	(0.076)	(0.037)	(0.036)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
High Trauma (IES > 24)	0.198	0.060	-0.129	0.052	0.024
	(0.133)	(0.109)	(0.110)	(0.054)	(0.054)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
GME Treatment × High Trauma	-0.001	-0.015	0.033	-0.006	-0.035
	(0.160)	(0.149)	(0.135)	(0.066)	(0.064)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
GME + GME × High Trauma	0.038	-0.014	0.043	0.035	0.023
p-value: GME + GME × High Trauma = 0	0.761	0.900	0.699	0.519	0.672
Mean DV in Control	0.00	0.00	0.00	0.70	0.70
Mean DV in Traditional	-0.10	-0.20	-0.10	0.70	0.60
N in Control	390	323	539	333	552
N in Traditional	456	380	642	392	659
N in GME	839	665	1115	679	1145

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the main outcomes of the experiment, interacted with an index for "High Trauma". We categorize respondents as "High Trauma" if they score 24 or above on the Impact of Event Score (IES-R) at baseline. Panel A shows the effect of the traditional training 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. Regressions models and controls are defined as in the main tables of the paper. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 11: Well-Being Outcomes with Trauma Interaction Using Alternative Trauma Indicator (IES > 24)

	Overall Index	Sub-indices	
	Wellbeing	Kessler	Resilience and self-efficacy
	(1)	(2)	(3)
Panel A: Traditional vs Control			
Traditional Training	-0.020	-0.021	-0.050
	(0.084)	(0.084)	(0.093)
		[1.000]	[1.000]
High Trauma (IES > 24)	-0.539***	-0.564***	-0.031
	(0.123)	(0.122)	(0.126)
		[0.001]	[1.000]
Traditional Training × High Trauma	0.012	0.008	0.072
	(0.158)	(0.156)	(0.167)
		[1.000]	[1.000]
Trad + Trad × High Trauma	-0.009	-0.013	0.022
p-value: Trad + Trad × High Trauma = 0	0.949	0.922	0.875
Panel B: GME vs Traditional			
GME Treatment	-0.027	-0.020	-0.017
	(0.070)	(0.068)	(0.080)
		[1.000]	[1.000]
High Trauma (IES > 24)	-0.536***	-0.561***	0.008
	(0.101)	(0.100)	(0.115)
		[0.001]	[1.000]
GME Treatment × High Trauma	0.107	0.120	-0.064
	(0.130)	(0.128)	(0.146)
		[1.000]	[1.000]
GME + GME × High Trauma	0.079	0.100	-0.081
p-value: GME + GME × High Trauma = 0	0.468	0.356	0.512
Panel C: GME vs Control			
GME Treatment	-0.048	-0.051	-0.019
	(0.079)	(0.078)	(0.085)
		[1.000]	[1.000]
High Trauma (IES > 24)	-0.577***	-0.604***	-0.044
	(0.124)	(0.122)	(0.127)
		[0.001]	[1.000]
GME Treatment × High Trauma	0.136	0.149	-0.006
	(0.149)	(0.147)	(0.155)
		[1.000]	[1.000]
GME + GME × High trauma	0.088	0.098	-0.025
p-value: GME + GME × High Trauma = 0	0.492	0.434	0.849
Mean DV in Control	0.00	0.00	0.00
Mean DV in Traditional	0.00	0.00	0.00
N in Control	545	545	545
N in Traditional	649	647	649
N in GME	1134	1133	1134

This table shows the impact of each treatment in the Colombia experiment (GME treatment and traditional training) on the well-being outcomes, interacted with an index for "High Trauma". We categorize respondents as "High Trauma" if they score 24 or above on the Impact of Event Score (IES-R) at baseline. The first sub-index captures psychological resilience (Sinclair and Wallston, 2004; Smith et al., 2008; Chen et al., 2001). The second sub-index reflects psychological distress, as measured by the Kessler K6 non-specific distress scale (Kessler et al., 2002). Regressions models and controls are defined as in the main tables of the paper. Panel A shows the effect of the traditional training 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.

Tables 10 and 11 show that the heterogeneity of treatment effects by trauma presented in the paper is robust to the precise definition of trauma used. In particular, using as definition of "High Trauma" being above 24 in the Impact of Event Score (IES) gives qualitatively similar results to the alternative definitions of "High Trauma" used in the paper and in this report, for both economic and wellbeing outcomes.

5. Robustness of Main Results Removing Winsorisation

Table 12: Main Economic Outcomes with Non-Winsorised Economic Indicators

	Overall Indices		Pre-covid sub-indices		Covid sub-indices	
	Pre-covid	Covid	Earnings	Earnings	Safety nets	Covid actions
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Traditional vs Control						
Traditional training	-0.157**	-0.090	-0.223**	-0.145**	0.013	-0.072
	(0.078)	(0.069)	(0.095)	(0.071)	(0.068)	(0.067)
			[0.082]	[0.082]	[0.602]	[0.232]
Panel B: GME vs Traditional						
GME training	0.130**	0.138**	0.184**	0.140**	0.025	0.081
	(0.066)	(0.059)	(0.082)	(0.060)	(0.057)	(0.058)
			[0.053]	[0.053]	[0.271]	[0.120]
Panel C: GME vs Control						
GME training	0.008	0.073	-0.014	0.038	0.046	-0.004
	(0.063)	(0.060)	(0.065)	(0.058)	(0.061)	(0.063)
			[1.000]	[1.000]	[1.000]	[1.000]
Mean DV in Control	0.00	0.00	0.00	0.00	0.00	0.00
Mean DV in Traditional	-0.10	-0.10	-0.20	-0.10	0.00	-0.10
N in Control	346	552	323	539	546	551
N in Traditional	407	660	380	642	650	657
N in GME	704	1147	665	1115	1134	1142

This table shows results on the main economic indices used in the paper, but without winsorization at the 99^th percentile. Regressions models and controls are defined as in the main tables of the paper. Clustered standard errors in parentheses, FDR sharpened q-values in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 12 shows that the main results of the paper on economic outcomes are robust when we consider non-winsorized variables in our economic indices.

6. Dealing with Attrition: Lee Bounds (extra, not included in PAP)

Table 13: Lee Bounds

	GME-Traditional		GME-Control		Traditional-Control	
	Mean Treatment Effect [N]	Lower/Upper Bounds [N]	Mean Treatment Effect [N]	Lower/Upper Bounds [N]	Mean Treatment Effect [N]	Lower/Upper Bounds [N]
	Max Diff Attrition 0.9%		Max Diff Attrition 6.7%		Max Diff Attrition 7.6%	
Mental Experiencing (overall)	0.026	0.011/0.049	0.014	-0.106*/0.156**	-0.069	-0.159**/0.100
	[1796]	[1787]	[1690]	[1620]	[1206]	[1165]
Mental Experiencing (business)	0.178**	0.164**/0.201***	0.046	-0.061/0.204***	-0.157*	-0.252***/0.028
	[1295]	[1288]	[1229]	[1178]	[846]	[819/819]
Mental Experiencing (non-business)	-0.006	-0.020/0.020	0.004	-0.105*/0.146**	-0.049	-0.149**/0.120*
	[1796]	[1787]	[1690]	[1620]	[1206]	[1165]
Earnings pre-COVID	0.191**	0.177**/0.289***	-0.007	-0.061/0.135**	-0.222**	-0.297***/0.019
	[1045]	[934]	[988]	[868]	[703]	[620]
Business pre-COVID	0.020	0.026/0.039	0.031	0.025/0.063**	0.001	-0.020/0.018
	[1071]	[956]	[1012]	[890]	[725]	[637]
Earnings during COVID	0.141**	0.135**/0.161***	0.040	-0.008/0.201***	-0.144**	-0.199***/0.058
	[1757]	[1748]	[1654]	[1585]	[1181]	[1142]
Business during COVID	0.048*	0.045*/0.053**	0.033	0.014/0.075***	-0.022	-0.044/0.013
	[1804]	[1794]	[1697]	[1626]	[1211]	[1170/1170]
Safety Index	0.024	0.013/0.045	0.047	-0.045/0.195***	0.014	-0.081/0.142**
	[1784]	[1775]	[1680]	[1610]	[1196]	[1155]
Business Response	0.084	0.042/0.119**	-0.003	-0.172***/0.164***	-0.074	-0.222***/0.100
	[1799]	[1790]	[1693]	[1622]	[1208]	[1167]
Investment	0.066	0.046/0.074	0.062	-0.094*/0.120**	-0.048	-0.200***/-0.000
	[1826]	[1816]	[1723]	[1652]	[1231]	[1190/1190]
Kessler	0.017	0.006/0.041	0.011	-0.076/0.153**	-0.012	-0.096/0.125*
	[1780]	[1771]	[1678]	[1608]	[1192]	[1152]
Psych. Resilience	-0.035	-0.048/-0.001	-0.049	-0.138**/0.118*	-0.052	-0.140*/0.105
	[1783]	[1774]	[1679]	[1609]	[1194]	[1153/1153]

This table shows Lee bounds for estimated treatment effects are shown in Columns (2), (4) and (6). The number of observations included when calculating the upper and lower bounds are shown in square brackets. *, **, *** denote significance at the 10, 5, and 1 percent levels respectively.

Table 13 shows Lee Bounds for the main outcomes of interest in the experiment. It shows that the treatment effects of the GME training compared to the traditional training are bounded away from zero, confirming that there is a positive impact of our GME training compared to a traditional business training. Regarding the comparison between each training arm and the no-intervention group, the Lee bounds on our main economic outcomes of interest (i.e. earnings, business status and mental experiencing) are also directionally aligned with the results presented in the main tables.

References

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies 81*(2), 608–650.

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.

Chen, G., S. M. Gully, and D. Eden (2001). Validation of a new general self-efficacy scale. *Organizational research methods 4*(1), 62–83.

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.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies 76*(3), 1071–1102.

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.

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.

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics 134*(2), 557–598.