

# Managing Emotions: The Effects of Online Mindfulness Meditation on Mental Health and Economic Behavior

ADVIK SHREEKUMAR  
University of California, Berkeley

PIERRE-LUC VAUTREY\*  
MIT

May 7, 2026

## Abstract

Mindfulness meditation has gained popularity, fueled by accessible smartphone apps and rising concerns about mental health. While such apps are claimed to affect mental well-being, productivity, and decision making, existing evidence is inconclusive due to limited sample sizes and high attrition. We address these concerns by conducting a large-scale, low-attrition experiment with 2,384 US adults, randomizing access and usage incentives for a popular mindfulness app. Access to the app improves an index of anxiety, depression, and stress by 0.43 standard deviations (SDs) at two weeks and 0.49 SDs at four weeks, with persistent effects three months later. It also improves earnings on a focused proofreading task by 1.4 percent. However, we find near-zero effects on a standard cognitive test (a Stroop task), and on decisions over risk and information acquisition where past economics research has indicated that emotions affect choice. This study provides evidence that digital mindfulness improves mental health and can raise productivity, but suggests that these effects do not stem from traditional measures of cognitive skills nor do they accompany more primitive changes in the information and risk preferences we measure.

---

\*Corresponding author: Advik Shreekumar ([adviks@berkeley.edu](mailto:adviks@berkeley.edu)). We thank Stacy Wang for excellent research assistance. We are indebted to Abhijit Banerjee, Esther Duflo, Ben Olken, and particularly to Frank Schilbach for their guidance. We thank Charlie Rafkin, Deivy Houeix, Clemence Idoux, Justine Knebelmann, Antoine Levy, Ro'ee Levy, Mathilde Munoz, Oliver Kim, Lucy Page, Matthew Ridley, Garima Sharma, Lena Song, and many seminar and conference participants for helpful comments. We thank all our study participants for their time and patience. We received IRB approval from the MIT Committee for the Use of Humans as Experimental Subjects (COUHES), protocol #2008000210. The experiment was pre-registered on the AEA registry, number AEARCTR-0007876. We thank J-PAL North America and the US Health Care Delivery Initiative for its support of this project. This material is based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374, and the George and Obie Shultz Fund at MIT. It was also supported by the Administration for Children and Families (ACF) of the United States (U.S.) Department of Health and Human Services (HHS) as part of a financial assistance award (Grant #: 90PD0310-01-00) totaling \$25,000 with 100 percent funded by ACF/HHS. For more information, please visit the ACF website, Administrative and National Policy Requirements. Headspace provided premium subscriptions to its app and the associated usage data at no cost but did not influence the design, conduct, analysis, or interpretation of the experiment. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement, by ACF/HHS, the U.S. Government, or Headspace.

# 1 Introduction

Poor mental health casts a long shadow on well-being, with depression and anxiety disorders among the top contributors to disability-adjusted life years for people aged 10 to 49 (Vos et al., 2020). These mental states affect not just our experience of the world, but also how we navigate it. Negative emotions and worries interfere with judgment, tax our limited attention, and reduce productivity (Johnson and Tversky, 1983; Loewenstein, 2000; Banerjee and Mullainathan, 2008; Kaur et al., 2021; Duquenois, 2021). As awareness of these problems has grown, so too has the demand for tools to improve well-being, attention, and productivity. Mindfulness meditation apps are one such popular tool, with hundreds of millions of downloads and billions of dollars in valuation.

Psychologists describe mindfulness meditation as a set of techniques that train users to “pay attention in a certain way, on purpose and non-judgmentally, to the present experience” (Kabat-Zinn, 2003). Medical professionals incorporate it into therapies for depression and anxiety, where its efficacy is well-established (Goldberg et al., 2021). Independently, many meditators also intend to improve their energy, memory, and concentration (Cramer et al., 2016). Large firms, including Google, Ford, and McKinsey, promote mindfulness among their employees for similar reasons.<sup>1</sup> A recent spate of research has also attempted to improve high-stakes decision making by improving reflection and self-regulation (Heller et al., 2017; Blattman et al., 2017; Dube et al., 2025). Underlying all these trends is the idea that self-observation is a learnable skill that can help one manage negative emotions and interrupting thoughts. This resonates with the recent focus in behavioral economics on attention as a primitive underlying behavior (Gabaix, 2019; Bordalo et al., 2020), animating the hypothesis that mindfulness meditation may have economically relevant effects.

However, high-quality evidence on the efficacy of widely-used meditation apps and their influence on economic behavior is scarce. Compared to personalized, professionally-administered therapy, meditation apps offer a more generic, scalable, and unstigmatized introduction to the practice. Do such technologies retain any benefits of individually-tailored, in-person interventions? Past studies of online meditation offer limited answers due to attrition and smaller samples, with meta-analyses raising concerns about methodological quality (Flett et al., 2019; Vonderlin et al., 2020; Sommers-Spijkerman et al., 2021; Whitfield et al., 2022). A rigorous, large-scale experiment testing a digital mindfulness intervention would begin to resolve these open questions about an increasingly common practice.

This paper studies the effects of a leading mindfulness meditation app, Headspace, on mental health and economic behavior by conducting a pre-registered randomized control trial (RCT) with 2,384 US adults interested in meditation, recruited via social media ads. In this three-month experiment, the first group receives free access to a premium version of the app (worth \$39 at the time); a second group receives, in addition, a \$10 incentive to use the app at least four or ten separate days during the first two weeks; and a third group serves as a waitlist control. We assess impacts on anxiety,

---

<sup>1</sup>See “Talking Mindfulness on the C.E.O. Beat”, published in the New York Times on Nov. 28, 2018.

depression, and stress with validated questionnaires up to three months after randomization. After two weeks participants complete an experimental session that includes a paid proofreading task, a cognitive test, and a set of incentivized choices between risky outcomes.

Detailed administrative usage data demonstrate high engagement with the app for the first two weeks after people receive access to it. Of the participants who receive access without additional incentives, 80.5 percent meditate with it at least once and use it on average of 5.3 days for a total of 95 minutes during the first two weeks. Incentives increase usage while they last: they make participants 8.9 percentage points (pp) more likely to try the app and use it on 2 more days and 48 more minutes during the same period. During the first two weeks, 49.3 percent use the app at least once every three days, but usage markedly decreases over time: between the fourth and the eighth week after receiving the app, only 9.5 percent use it every three days, and incentives have no lasting effect in contrast with other studies of wellness habits, such as hand washing or gym attendance (Hussam et al., 2019; Charness and Gneezy, 2009; Acland and Levy, 2015).<sup>2</sup>

Our first main finding is that offering access to the app improves mental health, as measured by a 0.43 standard deviation (SD) reduction in an index of depression, anxiety, and stress compared to the control group after two weeks and 0.49 SDs after four weeks. The effects are statistically significant after adjusting for multiple hypothesis testing (MHT) by a procedure we describe below ( $p < 0.01$ ). The effect at four weeks corresponds to a 14 percentage point reduction in the rate of moderate anxiety and depression symptoms, from a range of 26 to 29 percent in the control group. Effects persist well after Waitlist participants received app access. Three to four months after intervention, the initial treatment group still reports lower anxiety and depression than the initial Waitlist (0.22 SDs). The reduced treatment effect reflects lower uptake of the app by waitlisted participants, suggesting that the main force is the Waitlist group “catching up” rather than the treatment group “backsliding.” These effect sizes are smaller than the estimated effects of conventional therapies from meta-analyses (roughly 0.5 SDs, per Cuijpers et al., 2013; Goldberg et al., 2021). However, they are large given the low cost of the app and suggest that it is highly cost-effective for non-clinical populations. Further, analyzing a measure of social desirability (Marlow-Crowne) as well as sub-randomized interventions suggests that experimenter demand effects are unlikely to explain improvements in mental health.

Our second finding is that participants who received access demonstrate improved focus, as measured by a 1.2 percentage point (1.4%) increase in earnings on a proofreading task (0.09 SDs, MHT-adjusted  $p < 0.05$ ). In this task we pay participants to identify typographical errors in paragraphs of text, measuring their attention to detail. A 1.2pp improvement is modest in absolute terms, but meaningful in context. It is approximately 40% of the 3.2pp performance gap between control participants with and without a Bachelor’s degree. The effect size is also not far from the estimated effects of cash transfers that relieve financial worry among piece-rate workers in India (0.12 SDs; Kaur et al. (2025)). Notably, participants in online experiments tend to show low elasticities of effort

---

<sup>2</sup>App-based meditation differs from other health behaviors in that one can practice the skill without using the app. For example, 46% of the treatment arm reports meditating without the app three months after randomization.

(DellaVigna and Pope, 2018), implying that even modest effects represent true changes in behavior.

We further find that this improved focus is not due to changes in basic cognition such as how the brain processes visual cues and inhibits reflexive judgment. Psychological research posits that mindfulness training directly affect these capabilities (Jha et al., 2007). However, this evidence is based on studies with limited sample size and potential for publication bias: in Whitfield et al. (2022)’s meta-analysis of 45 studies, the mean sample size is 50 and there is “significant variability between studies in the types of scores reported.” We resolve this question in our sample with a precisely estimated null effect on a Stroop task that challenges participants to react quickly to words that appear on screen (95% confidence interval: [-1.1, 0.9] percent change in earnings). This null finding is informative: our 95% confidence interval includes zero and is 45% tighter than the leading meta-analysis on the subject, ruling out large effects that were previously plausible (Whitfield et al., 2022).

Our third main finding is that mindfulness training has no effect preferences, measured by choice under risk in our experimental tasks. A line of work in behavioral economics argues that choice is influenced by the feelings associated with certainty, gain, and loss (Loewenstein and Lerner, 2003); and that aversion to unpleasant feelings affects information acquisition as well as small probability risk taking (Golman et al., 2017; Bénabou and Tirole, 2016; Loewenstein, Weber, et al., 2001). Complementary work in neuroscience conceptualizes mindfulness as improving the ability to prevent emotions from capturing one’s attention (Guendelman et al., 2017). We draw these literatures together to test whether mindfulness blunts the effects of emotions in decision making. Our main investigation elicits the *certainty premium*, defined as the additional utility a participant assigns to receiving \$10 from a guaranteed payment versus as part of a risky prospect. Whereas recent work finds that emotions influence the certainty premium (Callen et al., 2014), we find that two weeks of app access does not measurably alter it: the estimated effect is small and statistically indistinguishable from zero (MHT-adjusted  $p = 0.905$ ), and our 95% confidence interval rules out effects of the magnitude found by Callen et al. (2014).

We report additional findings based on pre-registered cross-randomization and secondary outcomes that examine the interplay between mindfulness practice, a recent meditation session, and emotions. Unpacking these outcomes bolsters the null finding above, and raises hypotheses for future work to explore. First, we cross-randomize half of the App Access group to receive a small cash incentive to meditate immediately before the experimental tasks; the other half proceed directly to the tasks without any meditation prompt. Comparing the unprompted App Access participants to the Waitlist isolates the effect of two weeks’ practice, and we again find no effect on secondary outcomes: demand for pleasant information, demand for unpleasant information, and aversion to small-probability losses. A joint test of the primary and secondary choice-based outcomes reveals no evidence that two weeks of app access affects a set of economic decisions closely tied to a leading conceptualization of mindfulness (joint test:  $p = 0.555$ ).

By contrast, App Access participants prompted to meditate immediately before tasks appear *more* reactive to emotions. During the survey we experimentally induce anxiety and stress before the

proofreading task and certainty premium. These inductions have no effect on App Access participants who were not prompted to meditate. However, negative emotions reduce proofreading performance by 2.3pp and raise the certainty premium by 4.9pp among App Access participants who were prompted to meditate. Even before the emotion induction, the meditation-prompted group shows a preference for pleasant information and aversion to small-probability loss. Aggregating across all tasks suggests that the meditation prompt produces a distinct pattern of behavior from the unprompted App Access group ( $p = 0.03$ ) and Waitlist ( $p < 0.01$ ). Because these results draw on secondary interventions and outcomes, we view the conclusions as suggestive and cautionary for researchers and individuals. Scientifically, one-shot experimental designs that randomize meditation immediately before lab tasks (e.g., Iwamoto et al. (2020)) may identify a different effect than designs that encourage regular practice (e.g., Ash et al. (2021) and Alem et al. (2021)). Practically, while developing mindfulness can improve mental health and attention, a pre-task meditation session does not seem to guarantee better performance or less reactivity to emotions.

This paper investigates several hypotheses, and we take steps to mitigate concerns about spurious findings from multiple testing. First, we report and interpret findings for all pre-specified primary outcomes, including cases where effects are statistically insignificant. Second, we correct for multiple comparisons, applying the Romano and Wolf (2016) method to control the familywise error rate across a single family encompassing all primary outcomes: habit formation, mental health, proofreading performance, Stroop performance, and the certainty premium. In regression tables,  $p$ -values in square brackets are the result of this correction. Each test pools the relevant treatment arms (e.g., all App Access arms versus Waitlist); for transparency we report an unpooled, pre-registered specification immediately below the pooled results so readers can verify that pooling does not conceal meaningful heterogeneity. Finally, we follow Banerjee, Duflo, et al. (2020)'s recommendation and post a "populated pre-analysis plan" alongside the original document, which reports all feasible pre-specified analyses.

Our work contributes to several literatures. First, we contribute to a growing body of knowledge that evaluates therapy-inspired interventions through an economic lens (Blattman et al., 2017; Heller et al., 2017; John and Orkin, 2019; Lund et al., 2021; Bhat et al., 2021; Barker et al., 2022; Roth et al., 2024). We provide the first large-sample experimental evidence on mindfulness meditation's effects on incentivized decision-making, with a sample larger by an order of magnitude than comparable studies (Shapiro et al., 2012; Hafenbrack et al., 2014; Iwamoto et al., 2020; Ash et al., 2021; Alem et al., 2021; Cassar et al., 2023). Similar to our paper in measuring the economic effects of mental health interventions are Baranov et al. (2020) and Angelucci and Bennett (2024), who study therapies that reduce mental illness, but diverge on whether these interventions improve economic outcomes (Baranov et al., 2020) or have no effect (Angelucci and Bennett, 2024). Compared to these studies, which recruit depressed individuals and provide access to either psychiatry or pharmacotherapy, we evaluate a self-directed, light-touch digital intervention in a non-clinical population. Further, by recruiting participants via social media ads, we target the treatment effect on the population that meditation apps and similar wellness products are likely to reach in practice: individuals who respond

to routine advertising, rather than who present clinically or enter a study through institutional channels.

Second, our paper provides experimental evidence that a simple intervention to train attention can improve work productivity, and is to our knowledge the first to study the impacts of meditation on incentivized measures of performance in an adult population. Our findings build on evidence that worries and distractions impair work and test performance more generally (Banerjee and Mullainathan, 2008; Duquenois, 2021), and complement recent findings that training attention can improve school performance among children and college students (Brown et al., 2022; Cassar et al., 2023). Most directly, Kaur et al. (2025) show that cash transfers that relieve financial worry among piece-rate workers in India raise productivity by 0.12 standard deviations. Our experiment operates in the reverse direction: rather than removing worries, our intervention trains participants to manage attention in their presence, showing that training attention is a viable substitute for removing its disruptors.

Third, we contribute to the literature on the role of emotions in economic decision-making, which shows that experimentally induced emotions can distort behavior (Johnson and Tversky, 1983; Loewenstein, 2000; Loewenstein and Lerner, 2003; Slovic and Peters, 2006; Lerner et al., 2015; Callen et al., 2014; Engelmann et al., 2024). Under the assumption that irrelevant emotions bias an individual's evaluation of their decisions (Bernheim and Rangel, 2009; Bernheim and Taubinsky, 2018), reducing their interference with decision-making is welfare-improving. Our findings add two nuanced results to this literature. Sustained mindfulness practice does not measurably alter choices on tasks where emotions are thought to matter, suggesting that the welfare-relevant interference may be harder to modify than a reactivity-based account of mindfulness would predict. However, a recent meditation session may heighten rather than dampen emotional reactivity in subsequent choice tasks, complicating the view that meditating before a decision always improves it.

Finally, we contribute to a growing body of empirical findings on the effects of digital technologies on mental well-being. We study a widely-used digital technology that has positive effects, in contrast with recent work that examines the harmful effects of technologies on well-being (Twenge, 2020; Twenge et al., 2020; Allcott, Braghieri, et al., 2020; Allcott, Gentzkow, et al., 2021; Braghieri et al., 2021). We also conduct the largest-to-date RCT evaluating the mental wellness effects of mindfulness meditation training delivered digitally, contributing significantly to a booming impact evaluation literature reviewed recently by the meta-analysis of Sommers-Spijkerman et al. (2021). Relative to this literature, we also successfully design our online RCT to minimize attrition, a key limitation in many existing experiments, and we collect incentivized measures of productivity and decision-making in addition to mental well-being. Our sample is not selected based on initial symptoms of anxiety or depression nor is it restricted to a student or employee population from any given organization, which contributes to building externally valid evidence that affordable online mindfulness meditation training has large potential for the general population.

The rest of this paper proceeds as follows. [Section 2](#) describes our experimental design, including

details about the mindfulness intervention. In [Section 3](#) we describe takeup of our intervention using administrative data on meditation sessions. [Section 4](#) presents effects on mental health outcomes, and [Section 5](#) continues with effects on economic behavior. Finally, [Section 6](#) concludes with a discussion of our findings and open questions.

## 2 Background and Study Design

### 2.1 Mindfulness Meditation and the Headspace App

Mindfulness was first introduced to clinical psychology to treat chronic pain by incorporating meditation techniques from the Buddhist tradition into secular clinical therapies (Kabat-Zinn, 1982). The techniques typically involve focusing on a sensation, such as the breath, noticing how other thoughts and sensations capture attention, then refocusing on initial sensation. Following large-scale RCTs that extended these interventions to improve mental health, the American Psychological Association Society of Clinical Psychology has listed Mindfulness-Based Cognitive Therapy as an evidence-based treatment for depression with strong research support (American Society of Clinical Psychology, 2019).

Clinical psychologists who became interested in the mechanisms underlying mindfulness-based therapy have produced several characterizations and measurement scales for it, reflecting that mindfulness meditation may affect many mental processes (Davidson and Kaszniak, 2015). The proposed definitions agree on a core idea captured concisely by Kabat-Zinn (2003): mindfulness is the act of “*paying attention in a certain way, on purpose and non-judgmentally, to the present experience.*” First, *on purpose* touches on the idea of controlling attention, rather than letting it be captured by distractions. It also suggests that mindfulness meditation involves meta-cognition, or knowledge about how attention responds to distracting thoughts and stimuli. Second, *non-judgmentally* conveys the idea that attention should not linger on or avoid elements based on emotional reactions to them. Third, *present experience* is what attention should target: current sensations, emotions, and stimuli as opposed to thoughts and emotions referring to the past (including ruminations) or the future (including anticipations).

Headspace is a leading meditation app, ranking among the top five health and wellness smartphone apps in Google and Apple app store revenues in 2020. The free version of the app includes limited content, with full access requiring a subscription priced at \$12.99 per month or \$69.99 per year during our experiment. The design and delivery of this content distinguishes Headspace from most previously studied mindfulness interventions. The app provides a variety of audio and video recordings, often grouped into series or themes. Its core offering—the one we encourage participants to use—is a 10-day introductory mindfulness course. Other popular recordings help users fall asleep or engage in short deep breathing exercises.<sup>3</sup> Crucially, users choose which sessions to engage in and when, in contrast

---

<sup>3</sup>See [Figure A.4](#) for details.

with traditional therapies that require scheduled meetings with a professional.

## **2.2 Recruitment of Participants**

The practice of mindfulness meditation takes effort, time, and regularity. Measuring its effects requires recruiting and retaining participants who will take up the intervention for several weeks. Our recruitment strategy therefore favors individuals who are genuinely interested in trying meditation to maximize take-up. Participants were recruited between July 7 and August 24, 2021 using Facebook and Instagram ads. The ads specified that the study was about meditation, and that it included a Headspace subscription and compensation in the form of gift cards (see [Figure A.1](#) for examples of such ads).

The screening survey presented details about the study and verified that participants were over 18 and lived in the US. To remove inattentive participants, we excluded those who spent less than 20 seconds on the study details page. As our ads might draw active meditators who want a subsidized Headspace subscription, we further exclude participants who report practicing meditation or having previously tried Headspace. As our experiment is funded by agencies that focus on the well-being of low- to middle-income Americans, we also exclude participants reporting an income per household member above the US mean (\$54,000 in 2019; see US Census Bureau, 2020).

We invite the screened-in participants to complete a second baseline survey by SMS message with two main objectives: verifying that participants would remain responsive and have a valid US mobile phone number, and collecting baseline measures of mental health and secondary outcomes. On the following weekend, we emailed participants to complete a third and final baseline survey, which reviewed study structure and collected baseline measures of the proofreading task, Stroop test, and secondary outcomes. A total of 2,384 participants completed the last baseline survey and were randomized into a treatment arm.

## **2.3 Random Assignment to Free App Access and Incentives**

At the end of the last baseline survey, we randomize participants to receive a free three-month Headspace subscription via a voucher code (“App Access”) or be placed on a four-week waitlist (“Waitlist”). Randomization is stratified within eight strata based on age, baseline anxiety score, and baseline willingness to pay for an extension of the Headspace license.

The App Access group is randomized across three subgroups with equal probability: (i) App No Incentive, (ii) App Plus Short Incentive, or (iii) App Plus Long Incentive. The No Incentive group simply receives a voucher code. The incentive groups are additionally told that they will earn a \$10 bonus at the end of the first two weeks if they meditate using the app for at least 10 minutes on at least 4 days (Short Incentive) or at least 10 days (Long Incentive). These incentive treatments are meant to generate further variation in the app usage and to test for the formation of habits. The Short Incentive increases the value of trying the app for a few days and is designed to encourage

experimentation among a broad set of participants. The Long Incentive may increase usage among fewer participants but may create longer-lasting habits through sustained use.

The Waitlist group is randomized evenly across two subgroups: (iv) Pure Waitlist and a (v) Waitlist Cash Transfer. The Pure Waitlist receives no subsequent intervention. To test whether any effects from offering app access may be generated by reciprocity or wealth effects due to the value of the app, the Waitlist Cash Transfer group receives a \$15 cash transfer as a highly fungible online gift card. This amounts to a little more than the cost of a Headspace license for four weeks. We request that waitlisted participants refrain from creating a Headspace account or starting a similar practice while waiting for their voucher.<sup>4</sup> However, it is possible that waitlisted participants privately engage in activities that substitute for Headspace meditation. We therefore pre-register and focus on intent-to-treat estimates, which we expect would underestimate the effects of app access if the waitlist group engages in other wellness behaviors.

## 2.4 Further Randomization

Two weeks after randomization, participants complete an online experimental session containing our main economic outcomes. This module includes two cross-randomized treatments, which we mention now for completeness. We discuss them in [Section 5](#) where they are immediately relevant. The first intervention encourages a subset of the App Access group to meditate immediately before engaging in the economic tasks, and the second is designed to induce stress and anxiety in a subset of participants. We assign these treatments using the same strata as in the assignment among the app versus waitlist groups.

## 2.5 Sample Composition, Balance, and Attrition

Unsurprisingly, our recruitment strategy produces a sample that differs from the general population in a few notable ways. [Table 1](#) summarizes the treatment and control arms, and compares them to the U.S. adult population. The sample is majority female (86%), below 60 years old (85%), and tends to have more formal education (79% have at least a Bachelor’s degree). Few participants identify as Black (2%) or Hispanic (6.3%). Thirty-three percent of participants report symptoms of at least moderate anxiety at baseline, and the answers of 26 percent of the sample correspond to at least moderate depression, while 58 percent begin our study with less than moderate anxiety or depression. Sample characteristics aside, randomization balances the treatment and control arms on observable covariates ( $F$ -test  $p$ -value: 0.285).

[Figure 2](#) presents the sample size and retention rate of our followup surveys, as compared to other recent studies of mindfulness in adults (Galante et al., 2018; Sommers-Spijkerman et al., 2021). We recruit 2,384 participants, retaining 97.7% at two weeks and 94.8% at four weeks. Experiments with

---

<sup>4</sup>Three waitlisted participants obtained voucher codes prematurely by re-taking the enrollment survey; we use their original treatment assignment in all analyses.

comparable retention rates have fewer than 500 respondents, while those with over 1,000 participants have retention rates below 50 percent. The size and stability of our sample enable us to measure the effects mindfulness with minimal concern about attrition polluting our estimates.

### 3 App Usage and Effects of Incentives

We receive administrative data describing the title, start time, and duration of each app-guided meditation session users engage in during their 90-day voucher period. [Figure 3](#) shows the proportion of participants who use the app over time, within a sliding three-day window. Most treatment group participants use the app frequently at the beginning of the study period even without incentives. Eighty percent of participants in the App No Incentive group record at least one session in the first two weeks, using the app for an average of 5.3 days in this period. Usage steadily declines in the following weeks, but 46.7 percent of participants still log at least one session more than four weeks after randomization. This gives us the opportunity to study the effects of offering the mindfulness meditation app in a scenario where individuals engage with it at a high rate.

We test whether incentives encourage habit formation by estimating

$$Y_i = \delta_{s(i)} + \beta_1 \text{AppAccess} + \beta_2 \text{AnyIncentive} + \epsilon_i \quad (3.1)$$

where  $Y_i$  is the outcome for individual  $i$ ,  $\delta_{s(i)}$  is a randomization stratum fixed effect,  $\epsilon_i$  is an error term, and the remaining terms are indicators for treatment assignment. `AnyIncentive` pools both incentive arms for a joint test of whether incentives promote habit formation. For transparency, we also report effects that separate each arm with our pre-specified regression

$$Y_i = \delta_{s(i)} + \beta_1 \text{AppAccess} + \beta_2 \text{ShortIncentive} + \beta_3 \text{LongIncentive} + \epsilon_i, \quad (3.2)$$

where  $\beta_2$  and  $\beta_3$  measure the marginal effects of incentives to meditate on 4 and 10 days in the first two weeks, respectively, relative to the App No Incentive group.

Incentives increase initial usage of the app, especially when they require more usage to qualify for payment. The App No Incentive group meditates for an average of 5.3 days in the first two weeks. Long Incentives increase the number of days meditated by about 50 percent (2.7 days) during this period, compared to a 25 percent increase (1.3 days) from Short Incentives. Both effects are statistically significant.

However, incentives do not promote app use as a habit. In the cumulative post-incentives period, the average App No Incentives participant completes a meditation session on 8.7 days, roughly once in ten days. The pooled incentive arms meditate for 0.6 fewer days during this period, which is indistinguishable from zero ( $p = 0.491$ ; MHT-adjusted:  $p = 0.587$ ). This contrasts with previous findings that incentivizing health behaviors, such as going to the gym or washing hands, generates lasting habits (Charness and Gneezy, 2009; Acland and Levy, 2015; Hussam et al., 2019). One

interpretation is that unlike exercise or hand washing, which require continued practice to reap rewards, meditating *via an app* may be a naturally transient behavior. Once individuals learn the basics of mindfulness from the app, they can apply these skills by meditating independently or exerting attention control and emotional regulation in daily life. The data suggest that this is the case: Table A.15 shows that in our final followup survey three to four months after randomization, roughly 46 percent of treatment participants and 36 percent of waitlist participants report meditating without the app. A more apt comparison, then, might be to structured cognitive behavioral therapy interventions that are designed to teach trainees lasting skills (Blattman et al., 2017; Heller et al., 2017; Dube et al., 2025).

## 4 Effects on Mental Health

We use standard screening questionnaires to measure symptoms of mental distress: the seven-item GAD-7 scale for anxiety (Spitzer et al., 2006), the two- and eight-item PHQ-2 and PHQ-8 scales for depression (Kroenke and Spitzer, 2002), and the ten-item PSS-10 scale for stress (S. Cohen et al., 1994). Both the GAD and PHQ scales are workhorses in psychology, with higher scores being highly predictive of formal anxiety and depression diagnoses from more thorough interviews with medical professionals (Spitzer et al., 2006; Plummer et al., 2016; Kroenke and Spitzer, 2002; Y. Wu et al., 2020).<sup>5</sup> Each scale item asks the respondent how often they have felt an aspect of anxiety in the last two weeks, such as being “nervous, anxious, or on edge”. Responses to these items receive a score of 0–3, and the sum across all items generates a score. The PSS-10 stress scale is not a common clinical tool, but has been extensively used in non-clinical research on mindfulness meditation (e.g., Krusche et al., 2012; Spadaro and Hunker, 2016). It complements clinical measures among individuals who may not have symptoms of mental illness but still experience stress in their everyday lives.

We combine these measurements into a single index of mental health symptoms to improve power. In each followup survey, we standardize each elicited scale by subtracting the Pure Waitlist mean and dividing by the Pure Waitlist standard deviation. We then take the simple average of the standardized scales, so that lower values of the index indicate fewer reported symptoms of mental distress. In line with our pre-registration, the indexes at two and four weeks combine the GAD-7 anxiety scale, the PHQ-8 depression scale, and the PSS-10 stress scale. After fielding the last four-week endline survey, we had enough funding for a short followup three–four months after randomization. To keep the survey brief, we administered only the GAD-7 anxiety scale and PHQ-8 depression scale.

Self-reported psychometric scales are often used in the economics of mental health (e.g., Apouey and Clark (2015), Lindqvist et al. (2020), Liu and Netzer (2023), and Angelucci and Bennett (2024)). However, an important concern is that participants may misrepresent their mental states, perhaps saying what they think we’d like to hear, or what they prefer to think about themselves. We assess

---

<sup>5</sup>For example, cutoffs of 10–13 out of 24 on the PHQ-8 scale have over 90% sensitivity and specificity for predicting formal depression diagnoses.

this possibility in [Section 4.4](#) after presenting main results.

We estimate the intent-to-treat effects of app access with the regression

$$Y_i = \delta_{s(i)} + \beta \text{AppAccess} + \gamma Y_i^{\text{pre}} + \epsilon_i \quad (4.1)$$

where  $Y_i$  is the outcome for individual  $i$ ,  $\delta_{s(i)}$  is a randomization stratum fixed effect,  $\epsilon_i$  is an error term, and the remaining terms are indicators for treatment assignment. The “AppAccess” term compares the App Access group to the Waitlist group. For completeness we also report estimates that separate the incentive arms with [Equation \(3.2\)](#), modified to include baseline covariates.

#### 4.1 Effect of App Access on Mental Health

Providing access to the app leads to marked improvements in mental health. Two weeks after randomization, the App Access group reports an improvement of 0.43 standard deviations (SDs) on the index of symptoms. Effects grow slightly at four weeks, to 0.49 SDs. Effects are large even for the App No Incentive group, at 0.38 SDs at two weeks and 0.46 SDs at four weeks.

Once the Waitlist group gains access to the app, the App-Waitlist gap shrinks to 0.22 SDs at the three to four month mark. The primary explanation for this appears to be take-up among the Waitlist, 44 percent of whom use the app once they have access. Inspecting the individual mental health scales, their average score on the GAD-7 scale drops from 7.25 at four weeks to 6.30 at three months. A secondary explanation is a slight reduction in mental health among the treatment group: their average score on the GAD-7 scale increases slightly from 4.98 at four weeks to 5.21 at three months, which may reflect reduced meditation activity from the height of the experiment. In sum, we find that offering app-based mindfulness meditation improves self-reported mental health for at least three months after intervention.

The effects we measure are not due to improvements among only certain subgroups, nor are they artifacts of treating mental health measurements as continuous variables. [Figure 4](#) presents the cumulative distributions of the anxiety, depression and stress scores, and shows that improvements are distributed across the range of symptom severity. [Tables A.4](#) and [A.5](#) show treatment effects on the proportion of participants who report at least mild, moderate or severe symptoms of anxiety and depression. Effects at all levels are statistically significant. Nor are effects driven by attrition; App Access still exhibits a treatment effect under D. Lee (2009)’s worst-case model of selective attrition ([Table A.2](#)). Results are also robust to removing the baseline mental health index or flexibly controlling for all baseline covariates with debiased machine learning ([Figure A.2](#)).

The standardized effect sizes we measure are similar to those estimated in a recent meta-analysis of digital mindfulness interventions (Sommers-Spijkerman et al., 2021) for depression (0.34 SDs) and stress (0.44 SDs) and are substantially larger than these previous findings (0.26 SDs) for anxiety. However, as Sommers-Spijkerman et al. (2021) emphasize and [Figure 2](#) shows, past studies of digital mindfulness interventions suffer from smaller sample sizes and high attrition rates. In addition,

fewer studies evaluate self-directed programs such as Headspace. Our estimates are thus a significant contribution to the evidence base for widely used mindfulness interventions.

Unsurprisingly, app access has a larger effect on mental health than economic policies that affect well-being indirectly. The effects we find are substantially larger than previously estimated effects from receiving health insurance on depression in the Oregon Medicaid experiment (0.05 SDs in Finkelstein et al. (2012)). They are also about as large as the mental health improvements that a 0.2 SD increase in household income would generate, according to structural estimations that use cross-sectional variations (Alloush and S. Wu, 2023). Of course, health insurance and increased income confer many other benefits and operate at a longer time horizon than our experiment; we only provide these comparisons as a further benchmark.

Finally, effects are smaller than the effects of in-person therapy, estimated at 0.53 SDs by Cuijpers et al. (2013). The comparison requires nuance, as studies of CBT and pharmacotherapy target participants with diagnosed mental illness. While 32% of our sample would screen positive for anxiety and 26% for depression using baseline survey responses, we do not explicitly recruit participants with pre-existing conditions.<sup>6</sup> We expect app-based mindfulness would have smaller effects in a clinical population, owing to the difficulty of treating more severe mental illness.

## 4.2 Effect of Usage Incentives on Mental Health

Incentives deliver small marginal improvements in mental health, if at all. Short and Long Incentives reduce the index of mental health symptoms by 0.07–0.08 standard deviations at two weeks, and 0.05 standard deviations at four weeks. The marginal effect of incentives diminishes to less than 0.04 standard deviations by three months, and are not statistically significant at any time point. To contextualize these null findings, we pool the incentive arms and retrospectively compute power. The minimum detectable effect sizes at 80% power are 0.110 standard deviations at four weeks and 0.127 SDs at three months. This suggests we can rule out moderately sized marginal effects.

The fact that incentives don't have a large effect on mental health despite inducing 25 to 50 percent increases in days meditated may be surprising. One simple interpretation is that the App No Incentive group is well-equipped to choose how and when to use the app, and there is little benefit to additional sessions. This could be the case if there are diminishing returns to more than a few meditation sessions each week. Alternatively, incentivizing extra meditation sessions may result in low-effort or low-quality meditation. This would set meditation apart from more easily measured health behaviors, such as taking steps, exercising, or hand washing (Charness and Gneezy, 2009; Acland and Levy, 2015; Hussam et al., 2019; Aggarwal et al., 2020). Distinguishing these possibilities with more nuanced measures of compliance than app usage than ours remains of interest for future work.

---

<sup>6</sup>These are not formal diagnoses, but higher scores on GAD-7 and PHQ-2 are highly predictive of diagnoses based on longer interviews with a medical professional (Plummer et al., 2016; Arroll et al., 2010).

### 4.3 Heterogeneous Effects on Mental Health

Participants in our experiment are free to choose when and how they use the app, creating natural opportunities for heterogeneous treatment effects. We now turn to estimating heterogeneity along baseline covariates that enrich our understanding of average effects. To do so, we fully interact Equation (4.1) with dummies for whether each respondent is strictly below or weakly above the median value of the baseline covariate. Figure 5 presents estimates, and Table A.6 formally tests for differences in the below- and above-median subgroups.

We first present heterogeneity along four axes: baseline anxiety, baseline mindfulness, prior beliefs about the effect of meditation on anxiety, and a measure of social desirability. Our measure of baseline mindfulness is the Five Facet Mindfulness Questionnaire, a 15-item self-reported scale capturing various aspects of emotional awareness and mental habits.<sup>7</sup> We elicit prior beliefs about the effects of meditation by asking participants to forecast anxiety rates among other subjects who are randomized to treatment and control, with incentives for accurate predictions.<sup>8</sup> We defer discussion of social desirability to the next section, where we address possible demand effects.

Improvements in mental health are largest for individuals who enter the study with more symptoms of anxiety, lower levels of mindfulness, or more optimistic beliefs about treatment effects (Figure 5, Panel A). That said, participants with milder symptoms or more skeptical beliefs still experience large and statistically significant improvements. These groups meditate at similar frequencies, so take-up is unlikely to explain this heterogeneity. Instead, heterogeneity may reflect a combination of ceiling effects and unobserved effort. Individuals with lower levels of anxiety or higher levels of mindfulness mechanically have less room for improvement. In addition, optimists may dedicate more effort to their sessions in the app or supplement these sessions with other complementary activities, reaping larger benefits. Future work can probe these explanations, as well as other possibilities.

### 4.4 Experimenter Demand Effects

An important concern with self-reported outcomes is that they are susceptible to experimenter demand effects: participants may misreport their mental states in response to their beliefs about the study’s purpose (Quidt et al., 2019). This is worrisome because participants can reasonably infer that the study measures the mental health effects of a mindfulness app. We now assess whether experimenter demands can plausibly account for our findings using three pieces of evidence. In doing so, we emphasize that demand effects are a positive theory of behavior with testable implications; if

---

<sup>7</sup>The scale includes positive and negative items, such as “Even when I’m feeling terribly upset I can find a way to put it into words,” and “I believe some of my thoughts are abnormal or bad and I shouldn’t think that way.” Although we cannot verify these statements, the scale is commonly used in research on mindfulness. Taking the scale purely at face value, we can interpret responses as separating participants who represent themselves as more and less mindful.

<sup>8</sup>Specifically, we ask each participant to consider a hypothetical random sample of 10 other subjects with high anxiety, and predict the number who would still report high anxiety in 3 weeks if they did versus did not receive app access. Their expected treatment effect is the difference between these, signed so that larger numbers correspond to greater reductions in anxiety.

they explain large treatment effects, we would expect them to leave discernible signs elsewhere in the data.

First, we exploit heterogeneity in social desirability to determine whether treatment effects are due to participants who tend to misrepresent themselves. Following Dhar et al. (2022), we use the Marlow-Crowne social desirability scale to measure participants' tendencies to exaggerate positive traits and hide negative ones. If participants are adjusting their responses to please experimenters, those with higher propensities to misrepresent such traits may exhibit larger treatment effects. We find no evidence of this: treatment effects are similar for participants with both below- and above-median social desirability scores (Table A.6 Panel D). While social desirability does not perfectly capture responsiveness to experimenter demand, it does rule out the possibility that results are driven by participants prone to misrepresenting themselves.

Second, we benchmark effect sizes against studies that measure demand effects explicitly by revealing the research hypothesis before measuring outcomes. de Quidt et al. (2018) find demand effects of 0.13 SDs in a sample from Amazon's Mechanical Turk, and Mummolo and Peterson (2019) find effects indistinguishable from zero (95% CI: [-0.07, 0.06]).<sup>9</sup> The 0.43–0.49 SD effects of App Access are over three times larger than these effects. Further, demand effects are typically strongest when experimenters and participants are familiar and in proximity with each other (Quidt et al., 2019), and should be severely attenuated by the three-month mark. The App Access-Waitlist gap remains at 0.22 SDs at three months, measured in a survey two months after our last prior contact with participants. Explaining the size and persistence of our estimates would require demand effects that are unusually larger and more durable than those measured in other online surveys.

Finally, we can examine sub-randomization within the App Access and Waitlist groups that should induce experimenter demand and look for effects on outcomes. Within the App Access group, the \$10 usage bonus offered to the Short and Long Incentive arms may imply more strongly to participants that we expect app usage to affect their mental health. Similarly, the Waitlist Cash Transfer group receives a \$15 bonus, also emphasizing the value of the app. Any of these treatments could make our hypotheses more salient, as Mummolo and Peterson (2019) do when paying respondents \$0.25 to behave a particular way. Yet, the Incentive arms report only a 0.03–0.07 SD reduction in symptoms relative to No Incentives arm (Table 3), which are indistinguishable from zero at four weeks and three months. The Pure Waitlist and Cash Transfer arms also report similar mental health at two weeks, four weeks, and three months; the difference is at most 0.05 SDs and is never statistically significant (Table A.7). The absence of large effects in both of these contrasts, with cash incentives much larger than in Mummolo and Peterson (2019), suggests that demand effects are limited in our setting.

Of course, we cannot anticipate every source of demand effects, and a skeptical reader may posit a mechanism we have not discussed. Such a theory would have to generate demand effects in a way that is uncorrelated with social desirability, stronger and more persistent than explicitly induced

---

<sup>9</sup>Mummolo and Peterson (2019) report effects on raw outcomes; we compute standardized effects and confidence intervals from their replication data and include code in our replication files.

demand from other surveys, and not amplified by cash incentives that emphasize the value of the app. While such a theory may exist, these empirical patterns give us some confidence that our results do not purely reflect experimenter demand.

## 5 Effects on Economic Behavior

Immediately after the two-week mental health survey, participants enter an online experimental session designed to measure productivity, cognition, and preferences. The pre-registered primary tasks are a proofreading task measuring sustained attention, a Stroop test measuring executive function, and a pair of incentivized lotteries eliciting preferences over certain and uncertain outcomes. The session also includes two cross-randomized interventions: a prompt incentivizing half of the App Access group to meditate immediately before the rest of the survey, and an emotion induction for all respondents to manipulate stress and anxiety before key tasks. These are pre-registered as secondary interventions, so we begin with [Sections 5.1](#) and [5.2](#) that aggregate across them. [Section 5.3](#) then discusses the cross-randomized interventions as well as secondary outcomes. Because the emotion induction assigns participants to negative or neutral emotions with equal probability, and because roughly 30 percent of the treatment group meditates before the survey, estimates in [Sections 5.1](#) and [5.2](#) capture the effects of mindfulness practice across a range of emotional states and recency of meditation.

### 5.1 Attention

We measure sustained attention with an incentivized proofreading task. Participants highlight spelling and punctuation errors in three paragraphs of text with no time constraint, earning five cents per correctly identified mistake and losing five cents per false positive. They begin with an endowment of 20 cents so that initial mistakes are costly. A practice paragraph familiarizes participants with these incentives. The median participant spends two minutes on the task and earns 90 cents out of a possible \$1.05, for an effective wage of roughly \$27 per hour.

We estimate the intent to treat effect of App Access as before with [Equation \(4.1\)](#) and present results in [Table 4](#). App Access increases performance on the proofreading task by 1.2 percentage points relative to the maximum ( $p = 0.019$ ; MHT-adjusted  $p = 0.034$ ). This effect comes largely from spotting more errors in the paragraphs, not from making fewer mistakes or spending longer on task. The effect is robust to removing the baseline proofreading control, to flexibly adjusting for all baseline covariates with debiased machine learning ([Figure A.3](#)), and to worst-case selective attrition bounds (D. Lee (2009); [Table A.2](#)).

A 1.2 percentage point improvement is modest but meaningful, compared to cross-sectional variation within the control group. The effect is roughly half the gap between participants in the third and fourth quintiles of performance on the baseline proofreading survey ([Table A.14](#)), and about 40

percent of the 3.2 percentage point gap between participants with versus without a Bachelor's degree. Of course, these comparisons do not imply that meditation apps are direct substitutes for education; rather, they indicate that App Access produces a moderate effect relative to the distribution of task performance.

Exploratory analyses to distinguish mechanisms are more consistent with mindfulness improving effort quality, rather than increasing effort quantity or operating indirectly through improved mental health. [Figure A.5](#) shows that the largest benefits accrue to participants with the lowest baseline earnings. One candidate explanation is that these participants exerted little effort at baseline, and that practicing mindfulness induced them to spend longer on task. However, the distributions of time taken on the endline proofreading task are nearly identical in the App Access and Waitlist groups ([Figure A.6](#)) and there is no heterogeneity in treatment effects by time taken on the baseline task ([Figure A.7](#)). A second possibility is that reduced depression or anxiety may indirectly improve productivity (Ridley et al., 2020). However, performance on the proofreading task is similar for participants who report better and worse mental health in the preceding two weeks ([Figure A.8](#)). The absence of a cross-sectional correlation between mental health and proofreading performance makes it less likely that mental health gains are responsible for this effect. A remaining possibility is that participants with low baseline scores were simply less attentive to the text during the time they spend proofreading, and that mindfulness improved the quality of attention participants bring to the task. This interpretation aligns the proofreading effect with interventions that build cognitive endurance (Brown et al., 2022).

The psychology literature also considers the possibility that mindfulness improves more basic cognitive abilities such as executive function (Jha et al., 2007; Whitfield et al., 2022). We test this with a Stroop task. In each of 40 items, a color word (e.g., “yellow”) appears in a potentially incongruent font color (e.g., blue). Participants must select the font color from a multiple choice list within three seconds, earning a bonus for both speed and accuracy. Each correct answer adds 3 cents, minus 1 cent for each second taken; the maximum bonus is \$1.50. The median participant spends 66 seconds on the task and earns \$1.14 out of a possible \$1.50, for an effective wage of roughly \$61 per hour.

App Access has no appreciable effect on Stroop performance: treated participants earn 0.09 percentage points more than control participants, which is indistinguishable from zero. This null is informative. In standard deviation units, our estimate is 0.007 SDs with a 95% confidence interval of [-0.062, 0.076] SDs. This is roughly 45% tighter than the interval from a recent meta-analysis of 45 studies, which finds an effect of 0.15 SDs and 95% confidence interval of [0.02, 0.27] SDs on executive function tasks (Whitfield et al., 2022). A simple Bayesian update treating the meta-analysis as a  $N(0.15, 0.06)$  prior and our study as a  $N(0.007, 0.035)$  likelihood yields a posterior credible interval of [-0.04, 0.12]. This update shifts the interval to include zero and more than halves its upper bound, ruling out larger effects that the prior literature had left plausible.

Taken together, these results suggest that mindfulness practice improves performance on tasks requiring sustained attention, without sharpening executive function as measured by the Stroop test.

One explanation for this divergence is the difference in the tasks: the median participant spends about 40 seconds per proofreading paragraph, but only 1.65 seconds per Stroop item. Mindfulness may matter more for tasks that require maintaining focus over longer stretches of time, which is more consistent with improvements in cognitive endurance than executive function.

## 5.2 Choice

Mindfulness meditation trains practitioners to notice emotions as they arise without immediately acting on them. Because a body of work in behavioral economics finds that emotions associated with certainty, gain, and loss influence risk preferences (Loewenstein and Lerner, 2003; Callen et al., 2014), an intervention targeting emotional reactivity may affect choice under risk. We test this by eliciting a certainty premium: the additional utility a participant assigns to receiving \$10 as a guaranteed payment versus as part of a risky prospect. The measure is constructed from two incentivized lottery choices that separately identify the utility of \$10 under certainty and under uncertainty.

The elicitation involves two choices. First, we elicit the probability  $P$  at which participants are indifferent between a \$10 payment and  $(P, 1 - P)$  gamble paying \$30 or \$0. Second we elicit the probability  $Q$  at which participants are indifferent between a safer 50-50 gamble paying \$30 or \$10, and a  $(Q, 1 - Q)$  gamble over \$30 or \$0. Under expected utility maximization these choices imply utilities for receiving \$10:  $v(10)_c$  in the first choice and  $v(10)_u$  in the second. The certainty premium is  $CP = v(10)_c - v(10)_u$ , so that positive values indicate a preference for guaranteed outcomes. We normalize units so that the certainty premium is interpretable as a percentage of the utility of receiving \$30. Table 4 presents results.

The average Waitlist participant has a certainty premium of 14.3, indicating that a guaranteed \$10 is worth an added 14.3 percent chance of receiving \$30, relative to \$10 received from a gamble. This is consistent with Callen et al. (2014), who also find positive certainty premia, and similar in magnitude to the certainty premium from Andreoni and Sprenger (2011). App Access reduces the certainty premium by 0.15 points (i.e., 0.15% the implied utility of \$30), which is both small in magnitude and indistinguishable from zero ( $p = 0.905$ ; MHT-adjusted:  $p = 0.905$ ). This null is informative, as Callen et al. (2014) find that promoting the recall of frightening memories shifts the certainty premium by 3.4 to 4.3 percentage points; in Section 5.3 below we find a similar effect from inducing stress and anxiety in the Waitlist group. Our 95% confidence interval of  $[-2.67, 2.37]$  excludes these values and our minimum detectable effect at 80% power is 3.6 percentage points, allowing us to rule out effects of a magnitude similar to theirs. Two weeks of mindfulness practice thus does not measurably alter how participants value certainty in such lottery choices. That said, we do observe nuanced variation in behavior induced by our cross-randomized treatments; we turn to those now.

### 5.3 Secondary Outcomes and Cross-randomization

The experimental session includes additional outcomes and cross-randomized interventions to further study the relationship between emotion, meditation, and behavior. We pre-register three secondary outcomes and two auxiliary cross-randomized interventions. We highlight relevant findings here and defer details to [Section A.1](#).

The first cross-randomized intervention prompts half of the App Access group to meditate immediately before the experimental tasks, separating them into an “App + Prompt” and “App No Prompt” group. The distinction matters because habitual practices like meditation may have differing short-run and long-run effects. Distinguishing these can inform practitioners about the value of meditation sessions as decision aids, and help researchers assess whether one-shot meditation experiments provide valid estimates of the effects of sustained practice. Compliance is strong, with roughly 60 percent of prompted participants completing a meditation session before the experimental tasks versus less than 1 percent of unprompted participants. The second cross-randomized intervention induces negative or neutral emotions prior to the proofreading task and certainty premium lotteries, intended to measure how mindfulness affects performance across different emotional states. Negative emotions reduce the certainty premium by 3.8pp in the Waitlist group ( $p = 0.049$ ) but have a negligible effect on proofreading performance (-0.3pp;  $p = 0.710$ )

Our additional outcomes cover three decisions where emotions may play a role. The first captures avoidance to unpleasant but potentially useful information, which may reflect aversion to negative emotion. Participants are offered four hyperlinks to information on dementia risk factors, the risk of job loss to automation, a retirement financial risk calculator, and a life expectancy calculator; the outcome is the share of links declined. The second captures demand for pleasant but non-actionable information, which may reflect a preference for positive affect in the present. Participants are offered a gamble whose payoff depends on three simulated coin tosses; before resolution, they choose between payoff-relevant information (the outcome of one toss) and irrelevant but potentially pleasing information (whether they received a past bonus, or the age of the world’s oldest tree). The outcome is an indicator for choosing either piece of irrelevant information. The third captures aversion to small-probability salient losses, which may reflect emotions associated with contemplating loss (Loewenstein, Weber, et al., 2001). Participants are offered a gamble with a 99% chance of gaining \$1 and a 1% chance of losing \$10 (equal to their base payment); the outcome is an indicator for rejecting the gamble.

Secondary outcomes support the results above: practicing mindfulness improves attention but does not alter preferences. [Table A.12](#) shows that the App No Prompt group outperforms the Waitlist in the proofreading task, regardless of whether we induce Negative or Neutral emotions beforehand ( $p < 0.05$ , each comparison). Turning to the choice outcomes, we find no effects on aversion to unpleasant information, demand for pleasant information, or loss avoidance, and the certainty premium in both the Negative and Neutral conditions. A joint test of the choice outcomes finds

no difference between the No Prompt and Waitlist groups ( $p = 0.555$ ), bolstering the finding that mindfulness practice does not affect choice in situations where emotions may play a role.

However, the short-term effects of a meditation session appear distinct from the longer-term effects of practice. In the Prompt group, the Negative treatment reduces proofreading earnings by 2.3 percentage points; the prompt also affects the certainty premium in both the Negative and Neutral conditions ( $p < 0.05$ , each comparison), and the Negative treatment reduces performance by 2.3 percentage points ( $p = 0.010$ ). Even before the emotion induction, the Prompt group is 5.3 percentage points more likely to seek pleasant information and 6.0 percentage points more likely to avoid a gamble with a salient probability of loss; both effects are consistent with emotions having a stronger effect on action. Aggregating across all comparisons allowed by the meditation prompt and emotion induction, the Prompt group's behavior differs from both the Waitlist ( $p = 0.006$ ) and No Prompt ( $p = 0.032$ ) groups.

These results are consistent with either the view that the Prompt group is more sensitive to irrelevant emotions (and therefore worse off), or simply more attuned to their emotional state (with ambiguous effects on welfare). In either case, we read these results as generating rather than confirming hypotheses: they come from auxiliary treatments and outcomes, and are not corrected for multiple comparisons. However these findings complicate the view that pre-task meditation is a surefire shortcut to improved performance. They also suggest that research designs that implement one-shot meditation sessions may identify different treatment effects than those that encourage longer-term practice.

## 6 Discussion

Our results demonstrate the potential of inexpensive mindfulness meditation apps to improve mental health when used consistently over a period of weeks. The app we evaluate is less expensive and more accessible than in-person psychotherapy, but still delivers sizeable improvements in mental health. While we do not show that app-based meditation is a proper substitute for clinical therapy — particularly for people with serious mental illness — we do establish it as a bulwark against stress and anxiety in a more general population.

In addition, we provide the first large-sample evidence that mindfulness affects economic behavior on an online attention task. If these effects translate into productivity on the job, the benefits of the app easily pay for the costs. Such effects may be larger if mindfulness counteracts larger worries than our experiment induces, or improves focus on tasks that require more than a few minutes to complete. Resolving the effects of mindfulness in response to natural shocks to mental health and measuring its effects on workplace productivity are open areas for future research.

Usage of the app declines over time, and incentives have nearly no long-term effect on meditation behavior. This may reflect a combination of our participants' high baseline interest making the incentives unnecessary, long-term effects of mindfulness rendering future sessions less useful, and

the difficulties of adopting a new habit. Still, incentives may prove effective for encouraging adoption of mindfulness in a less engaged sample. Future work can study methods of increasing interest in mindfulness in a general population. In addition, the growing popularity of wearable health devices, like exercise trackers and smart watches, presents an opportunity to design mindful nudges and develop a richer understanding of treatment compliance and self-directed meditation.

Comparing participants with app access to those on the waitlist, we do not find evidence that practicing mindfulness affects decisions where emotions may play a role. Our results are based on three tasks where the behavioral economics literature has argued that mental states matter: two choices over risky prospects, and an information acquisition decision. Participants engage in these tasks after the treatment group has had two weeks of access to the app. Effects may take longer to materialize, and may be more pronounced on other tasks or in a population with less initial familiarity with mindfulness. But our finding that these choices are affected shortly after a meditation session suggests that the effects of mindfulness may reflect short-term results of directing attention inwards rather than lasting changes to fundamental preferences.

Our study design targets the effects of mindfulness practice and meditation sessions *before* productive tasks and choices over lotteries and information. Our analyses thus investigate whether mindfulness has instrumental value as a pre-task skill that improves attention or aids decision-making. An alternative view is that mindfulness may be a post-task rest activity, and may renew attention or productivity after periods of work or stress; future empirical work can identify such effects. A third, altogether different perspective is that of mindfulness as a consumption good, which practitioners may engage in simply because they prefer the mental states it confers. Economic theory may offer interesting insights into such choices, especially if they involve longer-term shifts in beliefs or outlook (Bernheim, Braghieri, et al., 2021).

Overall, our findings support the idea that policymakers and organizations should consider subsidizing inexpensive digital mindfulness interventions. More broadly, they suggest that such programs might be a way to invest in preventive mental health with better returns than physical wellness programs (Jones et al., 2019). These investments may use a combination of information campaigns, direct cash incentives, and other levers to increase awareness of the effects of practicing mindfulness to manage emotions.

## References

- Acland, Dan and Matthew R. Levy (Jan. 2015). “Naiveté, Projection Bias, and Habit Formation in Gym Attendance”. In: *Management Science* 61.1, pp. 146–160.
- Aggarwal, Shilpa, Rebecca Dizon-Ross, and Ariel D. Zucker (May 2020). *Incentivizing Behavioral Change: The Role of Time Preferences*. Tech. rep. w27079. National Bureau of Economic Research.
- Alem, Yonas et al. (2021). “Mind, Behaviour and Health: A Randomised Experiment”. In: *IZA Discussion Paper*, p. 70.
- Allcott, Hunt, Luca Braghieri, et al. (Mar. 2020). “The Welfare Effects of Social Media”. In: *American Economic Review* 110.3, pp. 629–676.
- Allcott, Hunt, Matthew Gentzkow, and Lena Song (June 2021). *Digital Addiction*. Working Paper 28936. National Bureau of Economic Research.
- Alloush, Mo and Stephen Wu (2023). “Income, Psychological Well-Being, and the Dynamics of Poverty”. In: *Economic Development and Cultural Change*, p. 76.
- American Society of Clinical Psychology (Apr. 2019). *Mindfulness-Based Cognitive Therapy — Society of Clinical Psychology*. <https://div12.org/treatment/mindfulness-based-cognitive-therapy/>.
- Andreoni, James and Charles Sprenger (2011). *Uncertainty equivalents: Testing the limits of the independence axiom*. Tech. rep. National Bureau of Economic Research.
- Angelucci, Manuela and Daniel Bennett (2024). “The economic impact of depression treatment in India: Evidence from community-based provision of pharmacotherapy”. In: *American economic review* 114.1, pp. 169–198.
- Apouey, Benedicte and Andrew E Clark (2015). “Winning big but feeling no better? The effect of lottery prizes on physical and mental health”. In: *Health economics* 24.5, pp. 516–538.
- Arroll, Bruce et al. (July 2010). “Validation of PHQ-2 and PHQ-9 to Screen for Major Depression in the Primary Care Population”. In: *The Annals of Family Medicine* 8.4, pp. 348–353.
- Ash, Elliott et al. (Aug. 2021). “Mindfulness Reduces Information Avoidance”. In: *Center for Law & Economics Working Paper Series* 2021.13.
- Banerjee, Abhijit, Esther Duflo, et al. (2020). *In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics*. Tech. rep. National Bureau of Economic Research.
- Banerjee, Abhijit and Sendhil Mullainathan (May 2008). “Limited Attention and Income Distribution”. In: *American Economic Review* 98.2, pp. 489–493.
- Baranov, Victoria et al. (Mar. 2020). “Maternal Depression, Women’s Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial”. In: *American Economic Review* 110.3, pp. 824–859.
- Barker, Nathan et al. (Dec. 2022). “Cognitive Behavioral Therapy among Ghana’s Rural Poor Is Effective Regardless of Baseline Mental Distress”. In: *American Economic Review: Insights* 4.4, pp. 527–45.

- Bénabou, Roland and Jean Tirole (2016). “Mindful Economics: The Production, Consumption, and Value of Beliefs”. In: *Journal of Economic Perspectives* 30.3, pp. 141–64.
- Bernheim, B Douglas, Luca Braghieri, et al. (2021). “A theory of chosen preferences”. In: *American Economic Review* 111.2, pp. 720–54.
- Bernheim, B Douglas and Antonio Rangel (Feb. 2009). “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics”. In: *The Quarterly Journal of Economics* 124.1, pp. 51–104.
- Bernheim, B Douglas and Dmitry Taubinsky (2018). “Behavioral Public Economics”. In: *Handbook of Behavioral Economics: Applications and Foundations* 1 1, pp. 381–516.
- Bhat, Bhargav et al. (2021). “The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Preferences”. In: *Working Paper*.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan (Apr. 2017). “Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia”. In: *American Economic Review* 107.4, pp. 1165–1206.
- Bolte, Lukas and Collin Raymond (Oct. 2023). “Emotional Inattention”. In: *Working Paper*.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2020). “Memory, attention, and choice”. In: *The Quarterly Journal of Economics* 135.3, pp. 1399–1442.
- Braghieri, Luca, Roe Levy, and Alexey Makarin (Aug. 2021). *Social Media and Mental Health*. SSRN Scholarly Paper ID 3919760. Rochester, NY: Social Science Research Network.
- Brown, Christina L et al. (2022). *Cognitive endurance as human capital*. Tech. rep. National Bureau of Economic Research.
- Callen, Michael et al. (Jan. 2014). “Violence and Risk Preference: Experimental Evidence from Afghanistan”. In: *American Economic Review* 104.1, pp. 123–148.
- Cassar, Lea, Mira Fischer, and Vanessa Valero (2023). “Keep Calm and Carry On: The Short vs. Long Run Effects of Mindfulness Meditation on (Academic) Performance”. In: *Working Paper*, p. 75.
- Cattaneo, Matias D et al. (2024a). “Nonlinear binscatter methods”. In: *arXiv preprint arXiv:2407.15276*. — (2024b). “On binscatter”. In: *American Economic Review* 114.5, pp. 1488–1514.
- Charness, Gary and Uri Gneezy (2009). “Incentives to Exercise”. In: *Econometrica* 77.3, pp. 909–931.
- Chernozhukov, Victor et al. (Feb. 2018). “Double/Debiased Machine Learning for Treatment and Structural Parameters”. In: *The Econometrics Journal* 21.1, pp. C1–C68.
- Cohen, Sheldon, Tom Kamarck, and Robin Mermelstein (1994). “Perceived Stress Scale”. In: *Measuring stress: A guide for health and social scientists* 10.2, pp. 1–2.
- Cramer, Holger et al. (Nov. 2016). “Prevalence, Patterns, and Predictors of Meditation Use among US Adults: A Nationally Representative Survey”. In: *Scientific Reports* 6.1, p. 36760.
- Cuijpers, Pim et al. (July 2013). “A Meta-Analysis of Cognitive-Behavioural Therapy for Adult Depression, Alone and in Comparison with Other Treatments”. In: *The Canadian Journal of Psychiatry* 58.7, pp. 376–385.

- Davidson, Richard J. and Alfred W. Kaszniak (2015). “Conceptual and Methodological Issues in Research on Mindfulness and Meditation”. In: *American Psychologist* 70.7, pp. 581–592.
- de Quidt, Jonathan, Johannes Haushofer, and Christopher Roth (Nov. 2018). “Measuring and Bounding Experimenter Demand”. In: *American Economic Review* 108.11, pp. 3266–3302.
- DellaVigna, Stefano and Devin Pope (Apr. 2018). “What Motivates Effort? Evidence and Expert Forecasts”. In: *The Review of Economic Studies* 85.2, pp. 1029–1069.
- Dhar, Diva, Tarun Jain, and Seema Jayachandran (2022). “Reshaping adolescents’ gender attitudes: Evidence from a school-based experiment in India”. In: *American economic review* 112.3, pp. 899–927.
- Dube, Oeindrila, Sandy Jo MacArthur, and Anuj K Shah (2025). “A cognitive view of policing”. In: *The Quarterly Journal of Economics* 140.1, pp. 745–791.
- Duquenois, Claire (2021). “Fictional Money, Real Costs: Impacts of Financial Salience on Disadvantaged Students”. In: *American Economic Review*, p. 60.
- Engelmann, Jan B et al. (2024). “Anticipatory anxiety and wishful thinking”. In: *American Economic Review* 114.4, pp. 926–960.
- Finkelstein, Amy et al. (Aug. 2012). “The Oregon Health Insurance Experiment: Evidence from the First Year\*”. In: *The Quarterly Journal of Economics* 127.3, pp. 1057–1106.
- Flett, Jayde A. M. et al. (May 2019). “Mobile Mindfulness Meditation: A Randomised Controlled Trial of the Effect of Two Popular Apps on Mental Health”. In: *Mindfulness* 10.5, pp. 863–876.
- Gabaix, Xavier (2019). “Behavioral Inattention”. In: *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2. Elsevier, pp. 261–343.
- Galante, Julieta et al. (Feb. 2018). “A Mindfulness-Based Intervention to Increase Resilience to Stress in University Students (the Mindful Student Study): A Pragmatic Randomised Controlled Trial”. In: *The Lancet Public Health* 3.2, e72–e81.
- Goldberg, Simon B. et al. (Feb. 2021). “The Empirical Status of Mindfulness-Based Interventions: A Systematic Review of 44 Meta-Analyses of Randomized Controlled Trials”. In: *Perspectives on Psychological Science*, p. 1745691620968771.
- Golman, Russell, David Hagmann, and George Loewenstein (Mar. 2017). “Information Avoidance”. In: *Journal of Economic Literature* 55.1, pp. 96–135.
- Guendelman, Simón, Sebastián Medeiros, and Hagen Rampes (2017). “Mindfulness and Emotion Regulation: Insights from Neurobiological, Psychological, and Clinical Studies”. In: *Frontiers in Psychology* 8, p. 220.
- Hafenbrack, Andrew C., Zoe Kinias, and Sigal G. Barsade (Feb. 2014). “Debiasing the Mind Through Meditation: Mindfulness and the Sunk-Cost Bias”. In: *Psychological Science* 25.2, pp. 369–376.
- Heller, Sara B. et al. (Feb. 2017). “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago\*”. In: *The Quarterly Journal of Economics* 132.1, pp. 1–54.

- Hussam, Reshmaan et al. (2019). “Rational Habit Formation: Experimental Evidence from Handwashing in India”. In: *Harvard Business School BGIE Unit Working Paper* 18-030, pp. 18–030.
- Imbens, Guido W. and Charles F. Manski (Nov. 2004). “Confidence Intervals for Partially Identified Parameters”. In: *Econometrica* 72.6, pp. 1845–1857.
- Iwamoto, Sage K. et al. (Apr. 2020). “Mindfulness Meditation Activates Altruism”. In: *Scientific Reports* 10.1, p. 6511.
- Jha, Amishi P., Jason Krompinger, and Michael J. Baime (June 2007). “Mindfulness Training Modifies Subsystems of Attention”. In: *Cognitive, Affective, & Behavioral Neuroscience* 7.2, pp. 109–119.
- John, Anett and Kate Orkin (Apr. 2019). *Can Simple Psychological Interventions Increase Preventive Health Investment?* Tech. rep. w25731. Cambridge, MA: National Bureau of Economic Research, w25731.
- Johnson, Eric J. and Amos Tversky (1983). “Affect, Generalization, and the Perception of Risk”. In: *Journal of Personality and Social Psychology* 45.1, pp. 20–31.
- Jones, Damon, David Molitor, and Julian Reif (Nov. 2019). “What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study\*”. In: *The Quarterly Journal of Economics* 134.4, pp. 1747–1791.
- Kabat-Zinn, Jon (Apr. 1982). “An Outpatient Program in Behavioral Medicine for Chronic Pain Patients Based on the Practice of Mindfulness Meditation: Theoretical Considerations and Preliminary Results”. In: *General Hospital Psychiatry* 4.1, pp. 33–47.
- (2003). “Mindfulness-Based Interventions in Context: Past, Present, and Future”. In: *Clinical Psychology: Science and Practice* 10.2, pp. 144–156.
- Karlsson, Niklas, George Loewenstein, and Duane Seppi (2009). “The ostrich effect: Selective attention to information”. In: *Journal of Risk and uncertainty* 38, pp. 95–115.
- Kaur, Supreet et al. (Jan. 2021). *Do Financial Concerns Make Workers Less Productive?* SSRN Scholarly Paper ID 3770928. Rochester, NY: Social Science Research Network.
- (2025). “Do financial concerns make workers less productive?” In: *The Quarterly Journal of Economics* 140.1, pp. 635–689.
- Kroenke, Kurt and Robert L. Spitzer (Sept. 2002). “The PHQ-9: A New Depression Diagnostic and Severity Measure”. In: *Psychiatric Annals* 32.9, pp. 509–515.
- Krusche, Adele et al. (2012). “Mindfulness Online: A Preliminary Evaluation of the Feasibility of a Web-Based Mindfulness Course and the Impact on Stress”. In: *BMJ open* 2.3, e000803.
- Lee, David (2009). “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects”. In: *REVIEW OF ECONOMIC STUDIES*, p. 32.
- Lerner, Jennifer S. et al. (2015). “Emotion and Decision Making”. In: *Annual Review of Psychology* 66.1, pp. 799–823.
- Lindqvist, Erik, Robert Östling, and David Cesarini (2020). “Long-run effects of lottery wealth on psychological well-being”. In: *The Review of Economic Studies* 87.6, pp. 2703–2726.

- Liu, Shuo and Nick Netzer (Dec. 2023). "Happy Times: Measuring Happiness Using Response Times". In: *American Economic Review* 113.12, pp. 3289–3322.
- Loewenstein, George (May 2000). "Emotions in Economic Theory and Economic Behavior". In: *The American Economic Review* 90.2, pp. 426–432.
- Loewenstein, George and Jennifer S. Lerner (2003). "The Role of Affect in Decision Making". In: *Handbook of Affective Sciences*. Series in Affective Science. New York, NY, US: Oxford University Press, pp. 619–642.
- Loewenstein, George, Elke U. Weber, et al. (Mar. 2001). "Risk as Feelings". In: *Psychological Bulletin* 127.2, pp. 267–286.
- Lund, Crick et al. (2021). *Economic impacts of Mental Health Interventions in Low and middle-Income countries: A Systematic review and meta-Analysis*.
- Mummolo, Jonathan and Erik Peterson (2019). "Demand effects in survey experiments: An empirical assessment". In: *American Political Science Review* 113.2, pp. 517–529.
- Oster, Emily, Ira Shoulson, and E Ray Dorsey (2013). "Optimal expectations and limited medical testing: Evidence from Huntington disease". In: *American Economic Review* 103.2, pp. 804–830.
- Plummer, Faye et al. (Mar. 2016). "Screening for Anxiety Disorders with the GAD-7 and GAD-2: A Systematic Review and Diagnostic Metaanalysis". In: *General Hospital Psychiatry* 39, pp. 24–31.
- Quidt, Jonathan de, Lise Vesterlund, and Alistair J. Wilson (July 2019). "Experimenter Demand Effects". In: *Handbook of Research Methods and Applications in Experimental Economics*.
- Ridley, Matthew et al. (Dec. 2020). "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms". In: *Science* 370.6522.
- Romano, Joseph P and Michael Wolf (2016). "Efficient Computation of Adjusted P-Values for Resampling-Based Stepdown Multiple Testing". In: *Statistics & Probability Letters*, p. 3.
- Roth, Christopher, Peter Schwardmann, and Egon Tripodi (2024). "Misperceived effectiveness and the demand for psychotherapy". In: *Journal of Public Economics* 240, p. 105254.
- Shapiro, Shauna L., Hooria Jazaieri, and Philippe R. Goldin (Nov. 2012). "Mindfulness-Based Stress Reduction Effects on Moral Reasoning and Decision Making". In: *The Journal of Positive Psychology* 7.6, pp. 504–515.
- Slovic, Paul and Ellen Peters (Dec. 2006). "Risk Perception and Affect". In: *Current Directions in Psychological Science* 15.6, pp. 322–325.
- Sommers-Spijkerman, Marion et al. (July 2021). "New Evidence in the Booming Field of Online Mindfulness: An Updated Meta-Analysis of Randomized Controlled Trials". In: *JMIR Mental Health* 8.7, e28168.
- Spadaro, Kathleen C and Diane F Hunker (2016). "Exploring the Effects of an Online Asynchronous Mindfulness Meditation Intervention with Nursing Students on Stress, Mood, and Cognition: A Descriptive Study". In: *Nurse education today* 39, pp. 163–169.

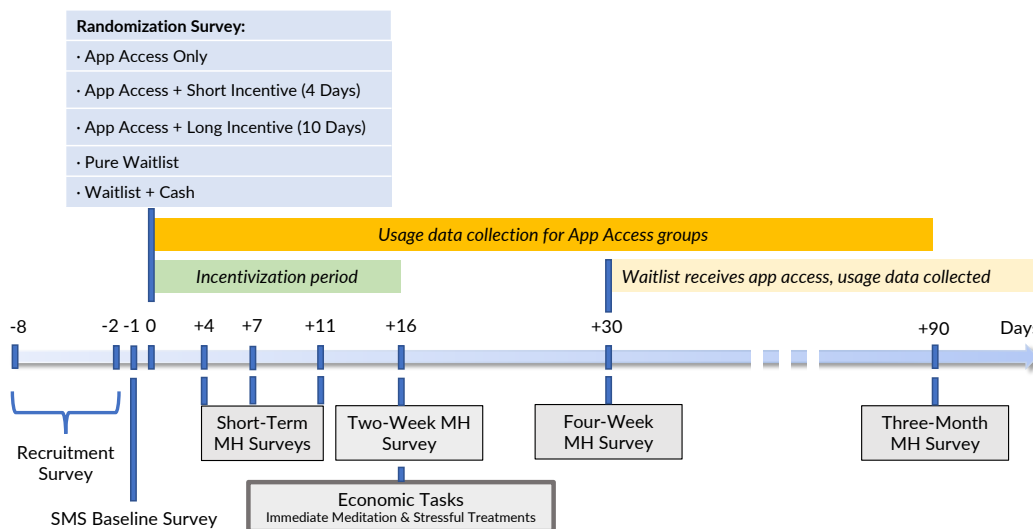
- Spitzer, Robert L. et al. (May 2006). “A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7”. In: *Archives of Internal Medicine* 166.10, pp. 1092–1097.
- Sydnor, Justin (Oct. 2010). “(Over)Insuring Modest Risks”. In: *American Economic Journal: Applied Economics* 2.4, pp. 177–199.
- Twenge, Jean M. (2020). “Increases in Depression, Self-Harm, and Suicide Among U.S. Adolescents After 2012 and Links to Technology Use: Possible Mechanisms”. In: *Psychiatric Research and Clinical Practice* 2.1, pp. 19–25.
- Twenge, Jean M. et al. (Apr. 2020). “Underestimating Digital Media Harm”. In: *Nature Human Behaviour* 4.4, pp. 346–348.
- US Census Bureau (2020). *Real Mean Personal Income in the United States*. <https://fred.stlouisfed.org/series/MAPAINUSA672>
- Vonderlin, Ruben et al. (July 2020). “Mindfulness-Based Programs in the Workplace: A Meta-Analysis of Randomized Controlled Trials”. In: *Mindfulness* 11.7, pp. 1579–1598.
- Vos, Theo et al. (2020). “Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019”. In: *The lancet* 396.10258, pp. 1204–1222.
- Whitfield, Tim et al. (2022). “The effect of mindfulness-based programs on cognitive function in adults: A systematic review and meta-analysis”. In: *Neuropsychology Review*, pp. 1–26.
- Wu, Yin et al. (June 2020). “Equivalency of the Diagnostic Accuracy of the PHQ-8 and PHQ-9: A Systematic Review and Individual Participant Data Meta-Analysis”. In: *Psychological Medicine* 50.8, pp. 1368–1380.



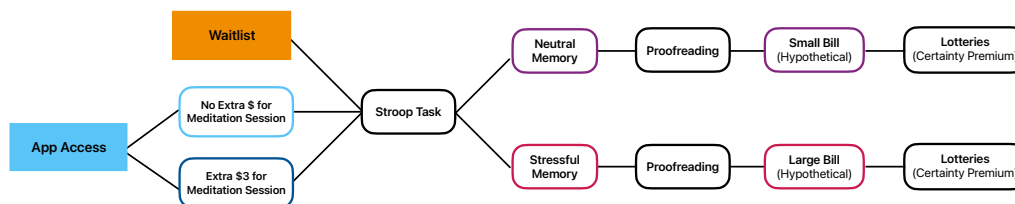
## 7 Figures and Tables

Figure 1: Experiment Overview

(a) Timeline



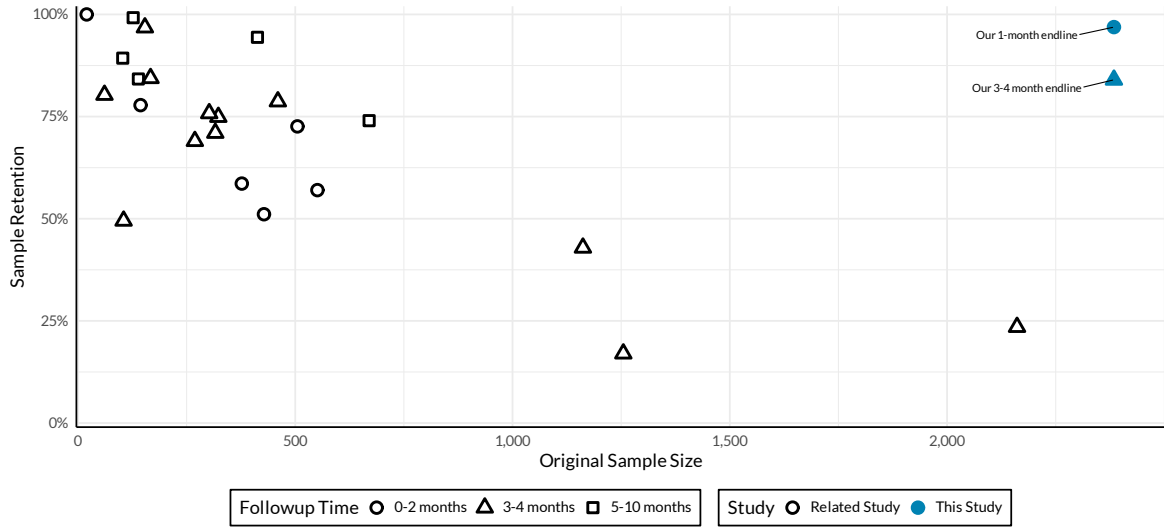
(b) Economic Tasks



*Notes:* This figure depicts the structure of our experiment. Panel A presents the timeline of the entire experiment. We recruited participants in July and August 2021 and conduct three baseline surveys. Randomization occurs weekly in the last baseline across one of five arms: (i) App Access Only, immediately receiving a 90-day license for the Headspace app; (ii + iii) App Access + Short (Long) Incentives, additionally being offered a \$10 bonus for using the app on at least 4 (at least 10) days in the first two weeks; (iv) Pure Waitlist, receiving the license after 30 days; and (v) Waitlist + Cash, additionally receiving a \$15 multi-use gift card. We conduct three short surveys during the first two weeks to track mental health, beliefs about the effects of the app, and willingness to pay for an extension of the license. The main followup survey occurs after 2 weeks, starting with an assessment of mental health and finishing with effort and decision-making tasks. Participants also complete an assessment of their mental health at four weeks, after which the waitlist receives their licenses. We obtain administrative data on the usage of the app for 90 days after license activation. Lower takeup among the waitlist group spurred us to conduct a mental health follow-up survey approximately 3 months after randomization, which was not pre-registered.

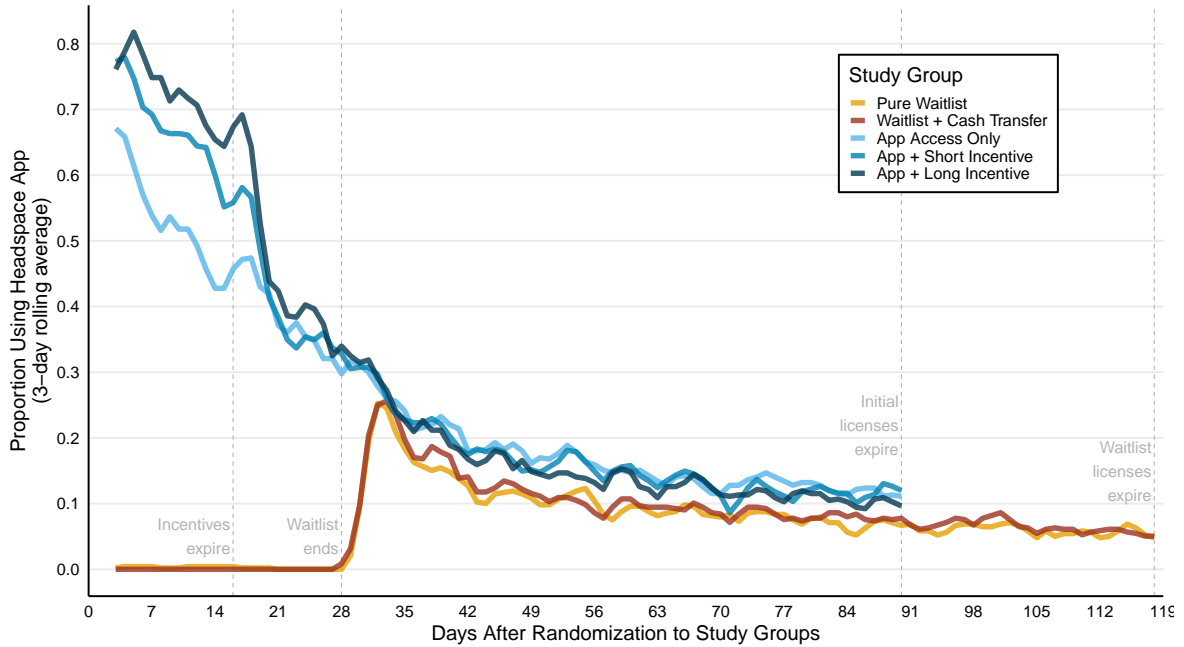
Panel B summarizes the economic tasks completed at two weeks. Those with app access are randomly assigned to receive incentives to meditate using the app right before continuing into the survey. All participants then complete tasks that have built-in distracting or emotion-inducing elements: a Stroop test of cognitive ability to control attention, decisions to avoid useful but potentially unpleasant information, and a decision to accept a risky prospect with high expected value but a low-probability salient loss. All participants are then randomized into one of two conditions: the Neutral or Stressful tasks, which ask participants to think about neutral or worrying memories and situations. They then complete an incentivized proofreading task and risk-taking choices.

Figure 2: Recent Studies of Mindfulness Interventions



Notes: This figure presents the original sample size, retention rate, and time to followup for recent randomized control trials of mindfulness meditation. Solid blue shapes describe our study, and hollow shapes describe related studies described in recent meta-analyses by Galante et al. (2018) and Sommers-Spijkerman et al. (2021).

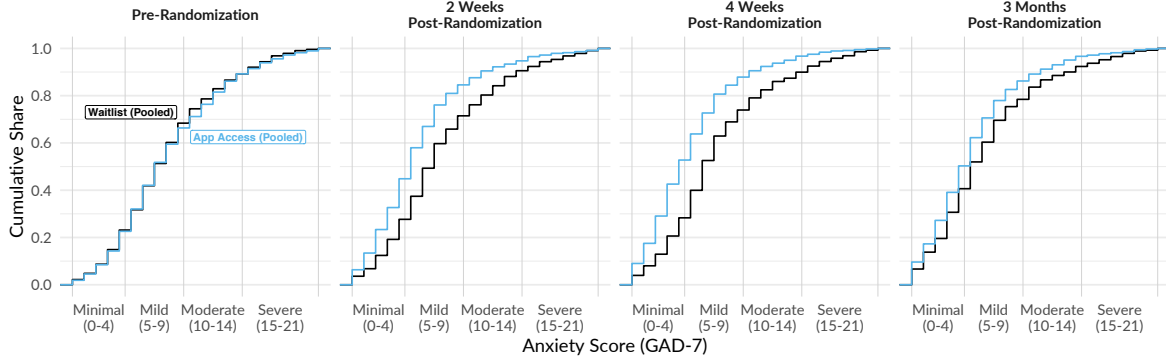
Figure 3: Proportion of Participants Using Headspace in a 3-Days Rolling Window



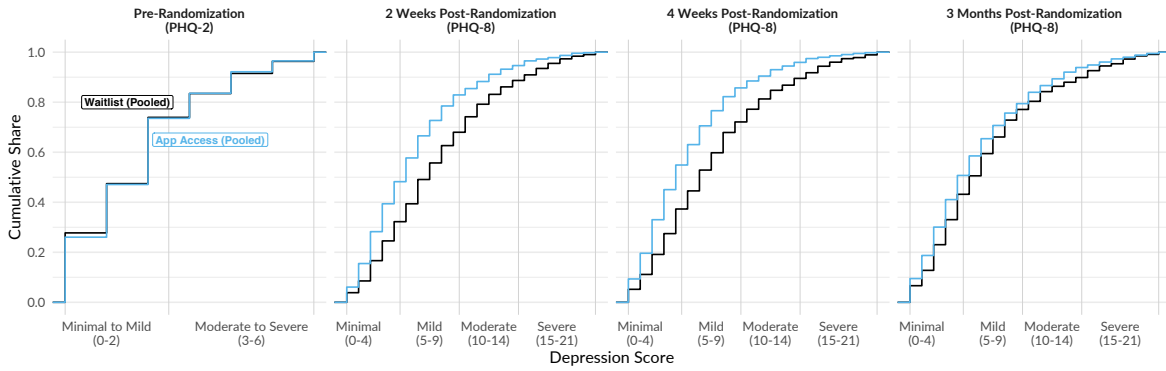
*Notes:* This figure shows the proportion, over time, of participants who recorded at least one session on the Headspace app within the last 3 days. Participants are randomized within one of five arms: (i) App Access Only, receiving free access to the Headspace app; (ii) App + Short Incentive, additionally being offered a \$10 bonus for using the app on at least 4 days during the first two weeks; (iii) App + Long Incentive, being offered a \$10 bonus for using the app on at least 10 days during the first two weeks; (iv) Pure Waitlist, receiving free access to the Headspace app after 30 days; and (v) Waitlist + Cash Transfer, receiving a \$15 multi-use gift card in addition to being placed on the waitlist. Usage data is observed in the 90 days after a participant activates the license we provide.

Figure 4: Long-Term Distribution of Mental Health Scores

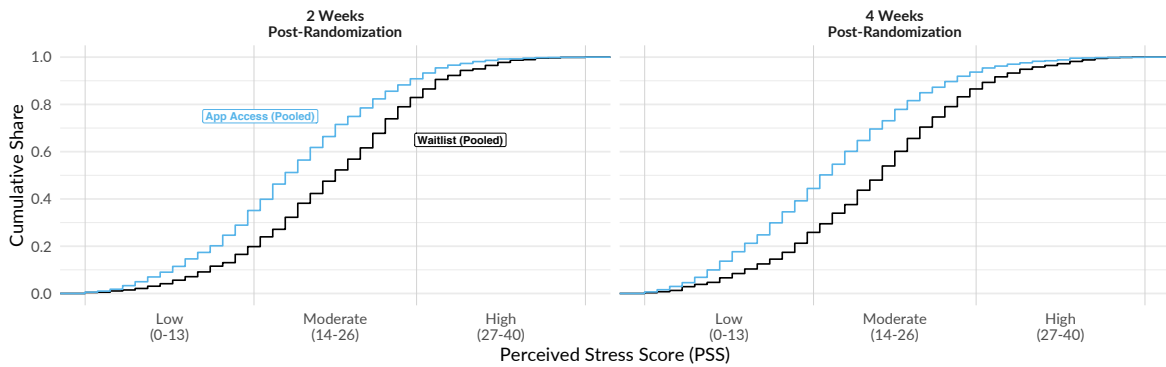
(a) Anxiety



(b) Depression

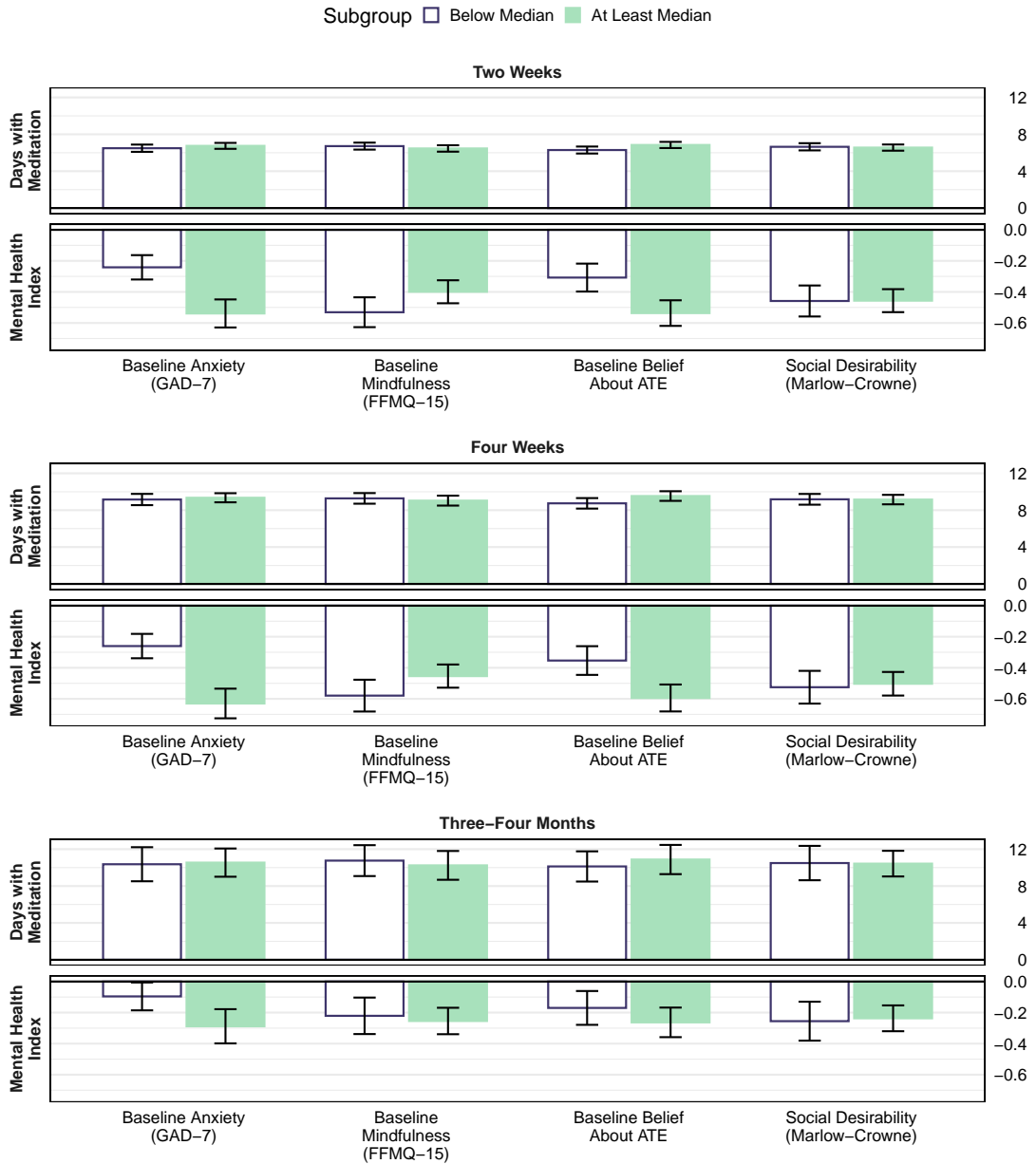


(c) Stress



Notes: This figure presents the empirical cumulative distribution functions for anxiety (GAD-7, panel A), depression (PHQ-2 and PHQ-8, panel B), and stress (PSS-10, panel C). In each figure, the black line represents the Waitlist group, and the blue line represents the App Access groups. To reduce survey length, we measure depression using the shortened PHQ-2 at baseline, and do not measure stress at baseline or at 3 months. For all three scales, lower scores indicate fewer symptoms or signs of the condition.

Figure 5: Heterogeneous Effects of License Provision and Incentives



Notes: This figure presents total effect of App Access on take-up and mental health in subgroups defined by pre-treatment covariates. The estimating equation generalizes Equation (4.1), separating all parameters for observations strictly below the median value of the covariate from those weakly above it:  $Y_i = \delta_{g(i),s(i)} + \beta_{g(i)} \text{AppAccess} + \gamma_{g(i)} Y_i^{\text{pre}} + \epsilon_i$ , where  $g(i) = 0$  for the below-median group and 1 for the above-median group. Days with Meditation counts the number of distinct days a users completed a meditation session on the Headspace app. Social Desirability refers to a 13-item version of the Marlow-Crowne Social Desirability Scale, which measures the tendency to portray oneself favorably in a survey. Table A.6 presents an complementary regression specification to test the equality of effects between the below and above median groups.

Table 1: Sample Characteristics and Balance

	US Adults	Study Sample			
	Mean	Waitlist Group Mean	(St. Dev.)	App Access Group Diff.	(St. Err)
<b>Age Group</b>					
18-29	0.213	0.202	(0.402)	-0.010	(0.011)
30-39	0.172	0.204	(0.403)	0.014	(0.016)
40-49	0.158	0.250	(0.433)	-0.010	(0.017)
50-59	0.163	0.209	(0.407)	-0.016	(0.016)
60-69	0.150	0.109	(0.312)	0.016	(0.013)
70+	0.143	0.025	(0.157)	0.006	(0.007)
<b>Female</b>	0.492	0.848	(0.359)	0.019	(0.015)
<b>Education</b>					
No Bachelor's degree	0.640	0.166	(0.373)	-0.023	(0.015)
Bachelor's degree	0.225	0.345	(0.475)	0.020	(0.020)
Graduate or professional degree	0.135	0.441	(0.497)	0.003	(0.020)
<b>Household Size</b>	2.261	2.874	(1.328)	-0.087	(0.055)
<b>Household Income</b>					
\$34,999 or less	0.265	0.231	(0.422)	-0.022	(0.017)
\$35,000-\$74,999	0.293	0.321	(0.467)	0.040**	(0.020)
\$75,000-\$149,000	0.285	0.332	(0.471)	-0.034*	(0.019)
\$150,000 or more	0.157	0.042	(0.200)	0.004	(0.008)
Prefer not to answer		0.073	(0.261)	0.013	(0.011)
<b>Race &amp; Ethnicity</b>					
White	0.600	0.834	(0.373)	0.015	(0.015)
Black	0.124	0.023	(0.150)	0.000	(0.006)
Hispanic	0.184	0.063	(0.243)	-0.007	(0.010)
Asian	0.056	0.087	(0.282)	-0.001	(0.011)
Other race	0.036	0.057	(0.231)	-0.013	(0.009)
<b>Political Party</b>					
Democrat		0.624	(0.485)	-0.003	(0.020)
Republican		0.027	(0.163)	0.013*	(0.007)
Other		0.349	(0.477)	-0.010	(0.020)
<b>Mental Health at Baseline</b>					
Anxiety Score (GAD-7)		8.012	(4.505)	0.168	(0.123)
Depression Score (PHQ-2)		1.796	(1.645)	0.018	(0.061)
<b>Sample Size</b>					
Observations	—	955		1,429	
<b>Omnibus F-test</b>			<i>F</i> -statistic	DF	<i>p</i> -value
			1.144	2359	0.285

*Notes:* This table presents demographic characteristics of our sample, compares them to the US adult population, and reports differences between our Waitlist control group and the License treatment group at randomization. Demographics for the US adult population come from the 2019 American Community Survey, which does not collect information about partisanship or mental health.

Table 2: Effects of Usage Incentives on Days with Any Meditation Session

	Incentives Active (Weeks 1–2)	After Incentives (Week 3+)	After Waitlist (Week 5+)	Total (Week 1+)
<b>A. Pooled Incentives</b>				
App Access	5.291***	4.279***	1.720**	9.526***
(S.E.)	(0.215)	(0.826)	(0.721)	(0.955)
Any Incentive	2.025***	−0.600	−0.756	1.414
(S.E.)	(0.263)	(0.875)	(0.735)	(1.043)
[MHT $p$ -value]		[0.587]		
Waitlist Mean	0.015	4.504	4.497	4.518
Sample Size	2384	2384	2384	2384
<b>B. Separating Arms</b>				
No Incentive	5.291***	4.279***	1.720**	9.526***
(S.E.)	(0.215)	(0.826)	(0.721)	(0.955)
Short Incentive	6.610***	3.832***	1.245*	10.389***
(S.E.)	(0.197)	(0.771)	(0.679)	(0.885)
Long Incentive	8.019***	3.527***	0.683	11.489***
(S.E.)	(0.227)	(0.745)	(0.656)	(0.861)
Waitlist Mean	0.015	4.504	4.497	4.518
Sample Size	2384	2384	2384	2384

*Notes:* This table presents the average treatment effects of app access and usage incentives on app usage, based on administrative data associated with each participant’s unique voucher code. The outcome is the number of days on which participants completed any meditation session using the app. We randomize some treated participants to receive a \$10 incentive to complete at least 10 minutes of meditation on at least 4 days in the first 2 weeks (“Short Incentive”) or at least 10 days in the first 2 weeks (“Long Incentive”). Panel A estimates the average effect of the two incentive arms relative to unincentivized treatment group participants, and Panel B separates the incentive arms. Usage incentives are active in the first two weeks (column 1), and the waitlist remains in effect for 2 more weeks after incentives expire (column 2). After the waitlist ends, all participants have full app access (column 3). Our administrative data ends 90 days after participants activate their vouchers, and we report cumulative usage in this period in column 4. The estimating equations are Equation (3.1) (panel A) and Equation (3.2) (panel B), which include stratum fixed effects. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . We apply a multiple hypothesis testing (MHT) correction to control the familywise error rate across all primary outcomes in the paper using the Romano-Wolf procedure with 9,999 bootstrap iterations (Romano and Wolf, 2016). Table A.3 presents an analogous table, where the outcome is takeup on the extensive margin (i.e., a binary indicator for any app-based meditation).

Table 3: Effects of App Access and Usage Incentives on Mental Health

	Time from Randomization		
	Two Weeks	Four Weeks	Three–Four Months
<b>A. Pooling Treatment</b>			
App Access	−0.431***	−0.487***	−0.216***
(S.E.)	(0.026)	(0.028)	(0.034)
[MHT $p$ value]	[< 0.001]	[< 0.001]	[< 0.001]
Baseline Outcome	0.626***	0.578***	0.510***
(S.E.)	(0.021)	(0.023)	(0.029)
Waitlist Mean	0.035	0.056	0.046
Sample Size	2330	2311	2004
<b>B. Separating Arms</b>			
No Incentive	−0.381***	−0.456***	−0.198***
(S.E.)	(0.036)	(0.039)	(0.046)
Short Incentive	−0.462***	−0.503***	−0.215***
(S.E.)	(0.036)	(0.038)	(0.045)
Long Incentive	−0.451***	−0.501***	−0.236***
(S.E.)	(0.035)	(0.038)	(0.045)
Baseline Outcome	0.626***	0.579***	0.510***
(S.E.)	(0.021)	(0.023)	(0.029)
Waitlist Mean	0.035	0.056	0.046
Sample Size	2330	2311	2004
<b>C. Index Components</b>			
Anxiety	GAD-7	GAD-7	GAD-7
Depression	PHQ-8	PHQ-8	PHQ-8
Stress	PSS-10	PSS-10	

*Notes:* This table presents average treatment effects of app access on reported symptoms of mental distress over time. We measure symptoms of anxiety using the seven-item Generalized Anxiety Disorder scale (GAD-7); symptoms of depression using the eight-item Patient Health Questionnaire (PHQ-8); and stress using the ten-item Perceived Stress Scale (PSS-10). The outcome at each timepoint is a standardized index that combines the mental health scales measured at that time. We first standardize each scale in each time period by subtracting the Pure Waitlist mean and dividing by the Pure Waitlist standard deviation. The index is the average of these standardized scales. Lower scores indicate lower reported levels of distress. Panel A uses Equation (4.1) to estimate the pooled effect of App Access arms relative to the Waitlist. Panel B uses our pre-registered specification, Equation (3.2), to separate the No Incentive, Short Incentive, and Long Incentive arms within the App Access group. In both panels, the reference group combines the Pure Waitlist and Waitlist Cash Transfer arms. All regressions include stratum fixed effects. Standard errors are from the heteroskedasticity-robust HC3 estimator. Figure A.2 presents a version of this analysis that flexibly adjusts for baseline covariates using debiased machine learning. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . We apply a multiple hypothesis testing (MHT) correction to control the familywise error rate across all primary outcomes in the paper using the Romano-Wolf procedure with 9,999 bootstrap iterations (Romano and Wolf, 2016).

Table 4: Effect of App Access on Economic Behavior

	Proofreading Earnings	Stroop Earnings	Certainty Premium
<b>A. Pooling Treatment</b>			
App Access	1.223**	0.089	-0.149
(S.E.)	(0.524)	(0.438)	(1.286)
[MHT $p$ -value]	[0.034]	[0.831]	[0.905]
Baseline Outcome	0.135***	0.444***	
(S.E.)	(0.014)	(0.029)	
Waitlist Mean	87.928	74.120	14.337
Sample Size	2257	2256	2247
<b>B. Separating Arms</b>			
No Incentive	1.282*	0.270	0.614
(S.E.)	(0.673)	(0.536)	(1.753)
Short Incentive	1.203*	-0.174	-0.779
(S.E.)	(0.699)	(0.626)	(1.765)
Long Incentive	1.185*	0.167	-0.296
(S.E.)	(0.685)	(0.596)	(1.726)
Baseline Outcome	0.135***	0.443***	
(S.E.)	(0.014)	(0.029)	
Waitlist Mean	87.928	74.120	14.337
Sample Size	2257	2256	2247

*Notes:* This table presents average treatment effects of app access on the proofreading task, Stroop test, and certainty premium. The proofreading task involves scanning three paragraphs of text for typographical errors, with a bonus for each error found, a penalty for false positives, and no time limit. The outcome is earnings relative to the maximum possible amount in percentage points; for example, the average Waitlist participant earned 87.9% of the maximum possible bonus. The Stroop test involves rapidly reacting to words that appear on-screen, with a penalty for time taken and incorrect answers. The outcome is earnings relative to the maximum possible amount. The certainty premium, derived from choices on two incentivized lottery tasks, measures the difference between the utility a participant assigns to receiving \$10 with versus from a lottery. We normalize units by the implied utility of receiving \$30 from a lottery; for example, the average Waitlist participant responds as if preferring a guaranteed \$10 payout over a risky one, valuing the added certainty as worth 14% of \$30 received from a risky prospect. Panel A uses Equation (4.1) to estimate the average effect of App Access arms relative to the Waitlist. Panel B uses our pre-registered specification, Equation (3.2), to separate the No Incentive, Short Incentive, and Long Incentive arms within the App Access group. In both panels, the reference group combines the Pure Waitlist and Waitlist Cash Transfer arms. All regressions include stratum fixed effects. Standard errors are from the heteroskedasticity-robust HC3 estimator. Figure A.3 presents a version of this analysis that flexibly adjusts for baseline covariates using debiased machine learning. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . We apply a multiple hypothesis testing (MHT) correction to control the familywise error rate across all primary outcomes in the paper using the Romano-Wolf procedure with 9,999 bootstrap iterations (Romano and Wolf, 2016).

## A Appendix

### A.1 Secondary Randomization and Outcomes

This appendix describes two pre-registered auxiliary treatments and three pre-registered auxiliary outcomes from the experimental session. The treatments are a meditation prompt offered to half of the App Access group at the start of the experimental session (after the mental health module), and a negative / neutral emotion induction administered to all participants partway through the session. The outcomes cover a choice over distressing information, a choice over pleasant information, and a gamble to measure loss aversion. As these are pre-registered as secondary, we do not correct the resulting tests for multiple comparisons, and treat the analyses as hypothesis-generating rather than confirmatory.

#### A.1.1 Secondary Treatments

At the start of the experimental session we offer half of the App Access group a \$3 bonus to complete a 10-minute meditation session before proceeding to the survey. We refer to this subgroup as the Prompt group and the remainder as the No Prompt group. Compliance is high: roughly 60 percent of the Prompt group meditated at this time, compared to 3 percent of the No Prompt group.

During the survey, all participants are randomized with equal probability to a Negative or Neutral emotion condition prior to the proofreading and certainty premium tasks. The Negative condition asks participants to describe an unresolved source of anxiety before the proofreading task, and how they would pay for a large unexpected medical bill (\$8,900) before the lottery tasks. The Neutral condition asks them to describe part of their daily routine and how they would pay a small medical bill (\$100). To incentivize engagement we tell participants that randomly selected answers will be evaluated by an independent reader, with thoughtful answers earning a \$5 bonus. The Negative treatment has mixed success. While inducing worries makes Waitlist participants reduces the certainty premium by 3.8 percent ( $p = 0.049$ ), it does not appreciably reduce proofreading performance.<sup>10</sup>

#### A.1.2 Secondary Outcomes

The three secondary outcomes are administered at the start of the survey, before the negative/neutral randomization (see [Figure 1](#)). Each is designed to isolate a distinct channel through which mindfulness might operate.

---

<sup>10</sup>The direction of the effect in the Waitlist group differs from that in Callen et al. (2014), where recalling violent memories increases the certainty premium. These opposing signs may reflect the stakes in the lotteries, differing risk preferences, or behavior in lab tasks in our study sample and theirs. The difference may be due to differing stakes, as the relationship between risk and emotion may reasonable depend on how much participants stand to gain. In addition, while we view our lottery outcomes, ranging from 0 to 30, as reasonable, they are lower than the stakes in Callen et al. (2014)'s experiment where participants stand to gain between one and three days' wages. Alternatively, our study samples may simply behave differently in lab-style lottery tasks. Callen et al. (2014) study a population with an average of 10 years' formal education. Most of our participants have completed at least a college degree, which may lead them to evaluate the tasks differently.

First, we offer participants four optional hyperlinks: (i) a life expectancy calculator; (ii) risk factors for developing dementia; (iii) the probability of one’s job being replaced by automation; and (iv) a calculator of financial risk in retirement. Selected links are presented at the end of the survey. Our measure is the share of links accepted, which captures the tendency to avoid useful but unpleasant information (Bénabou and Tirole, 2016; Golman et al., 2017).

Second, we offer participants a gamble where they can choose between instrumental information and pleasing information. Participants begin with a digital wallet containing \$0 or \$1 with equal probability and decide whether to acquire an imaginary stock whose value is determined by three coin tosses. Two or more heads add \$1 to their bonus, and two or more tails subtract the same. Before the decision, participants choose one of three pieces of information they would most like to know: the outcome of one coin toss (relevant and emotionally neutral), the starting value of their wallet, or the age of the world’s oldest tree (irrelevant).<sup>11</sup> By default we present all three pieces, but with small probability we present only the chosen one, making the choice incentive-compatible. Choosing an irrelevant piece suggests the participant derives less utility from future payoffs than from immediate ones, such as a warm glow from learning the wallet’s value, or intrigue at a novel fact.

Third, we offer a gamble with a 99% probability of a \$1 gain and a salient 1% probability of losing the survey’s \$10 base payment. We increase the salience of the loss by bolding and coloring the potential loss. This is a stylized version of real-world choices to purchase high-premium insurance against low-probability risks (Sydnor, 2010). Rejecting the gamble suggests overweighting of unlikely losses, which may “stem from disproportionate fear and pleasurable anticipation evoked by such prospects” (Loewenstein, Weber, et al., 2001).

### A.1.3 Results

For outcomes measured before the Negative / Neutral treatment, we estimate

$$Y_i = \alpha + \beta_1 \text{Prompt} + \beta_2 \text{NoPrompt} + \gamma Y_i^{\text{pre}} + \epsilon, \quad (\text{A.1})$$

with the Waitlist group as the omitted category. For outcomes measured after the Negative / Neutral treatment we estimate

$$Y_i = \alpha + \beta_1 \text{Prompt} + \beta_2 \text{NoPrompt} + \beta_3 \text{Negative} \times \text{Prompt} \\ + \beta_4 \text{Negative} \times \text{NoPrompt} + \beta_5 \text{Negative} \times \text{Waitlist} + \gamma Y_i^{\text{pre}} + \epsilon. \quad (\text{A.2})$$

Here, the Waitlist Neutral group is the omitted category. There are three relevant contrasts. The No Prompt–Waitlist comparison isolates the effect of two weeks of practice, the Prompt–No Prompt comparison isolates the short-term effect of a meditation session among practitioners, and the Prompt–Waitlist comparison combines both. [Table A.12](#) presents results, restricting to participants

---

<sup>11</sup>At the time of the survey, the oldest known tree was 5,071 years old.

who completed all items to allow for joint hypothesis tests with Seemingly Unrelated Regressions.

App Access without a recent meditation session improves attention without affecting preferences. On the proofreading task, the No Prompt group outperforms the Waitlist by 1.8 percentage points in the Neutral condition ( $p = 0.032$ ) and 1.7 percentage points in the Negative condition ( $p = 0.040$ ). We find a negligible 0.3 percentage point effect on the Stroop test. A joint test that uses seemingly unrelated regression to pool these three contrasts registers a significant effect ( $p = 0.044$ ). Turning to choice-based outcomes, we find statistically insignificant differences between the Prompt and the Waitlist. There are small and insignificant differences on the certainty premium in both the Negative and Neutral conditions, as well as on all three secondary outcomes. A joint test pooling both certainty premium contrasts with the secondary choice outcomes indicates no difference in choice behavior between the No Prompt and Waitlist groups ( $p = 0.555$ ).

By contrast, prompting participants to meditate appears to make them more sensitive to emotion. On proofreading, while the Prompt group outperforms the Waitlist by 1.6 percentage points in the Neutral condition ( $p = 0.041$ ), they *underperform* the Waitlist by 0.3 percentage points in the Negative condition. That is, the Negative treatment reduces their performance by 2.3 percentage points ( $p = 0.011$ ). A joint test that pools these contrasts with the Stroop test finds no difference between the Prompt and Waitlist groups ( $p = 0.261$ ). The choice-based outcomes also show a pattern consistent with sensitivity to emotion. The Prompt group's certainty premium is 4.3 percentage points lower than the Waitlist's in the Neutral condition ( $p = 0.047$ ) and 4.4 percentage points higher in the Negative condition ( $p = 0.045$ ). Even before the emotion induction, the Prompt group behaves distinctly: they are 5.3 percentage points more likely to seek pleasant information and 6.0 percentage points more likely to avoid a gamble with a salient probability of loss. Aggregating across both attention and choice, the Prompt group's behavior differs from the Waitlist ( $p = 0.006$ ) and perhaps from the No Prompt ( $p = 0.032$ ) group.

#### A.1.4 Discussion

One concern is that the meditation prompt may affect behavior through channels unrelated to emotional states. Compliers spent an additional 10 minutes meditating and received a \$3 bonus, either of which could affect the marginal utility of time or effort spent on subsequent tasks. However, the distribution of time spent on the experimental tasks is nearly identical across the prompt and waitlist groups below median, with statistically insignificant differences of less than 5% above the median (Figure A.9). In short, it does not appear as if the Prompt group is rushing through the survey.

Two mechanisms for this finding remain. First, by putting participants in a more relaxed state, meditation sessions may heighten the contrast between the Negative and Neutral conditions. Second, meditation sessions may focus attention to the emotions, so that participants are more attuned to the effects of recounting an anxious memory or emotions in the secondary tasks. If so, they may focus less on the survey and more on calming themselves, in a form of labor-leisure tradeoff. These

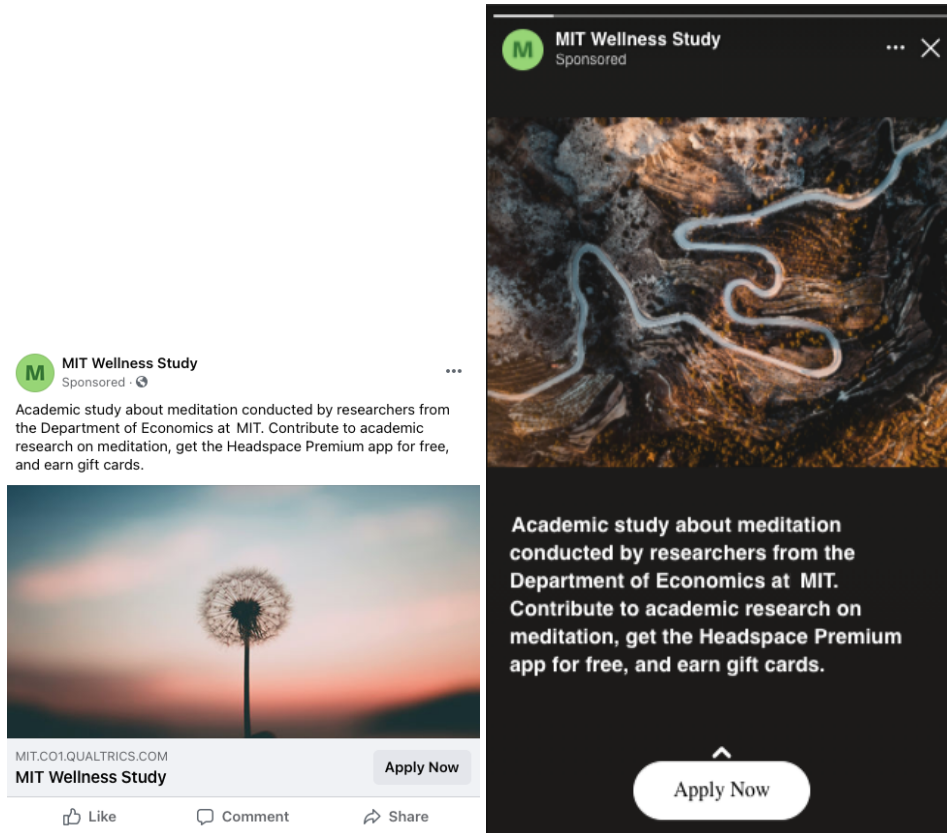
mechanisms are not mutually exclusive, and we cannot further distinguish between them with our data.

The pattern of results does not imply that meditation sessions degrade decision-making in a welfare sense. For example, information avoidance can stem from rational anticipation of emotional costs, such as when investors postpone checking a portfolio or patients delay medical tests (Karlsson et al., 2009; Oster et al., 2013; Bolte and Raymond, 2023). The Prompt group's behavior is also consistent with sharper awareness of one's own mental state. That said, the fact that the Negative treatment reduces the Prompt group's proofreading performance suggests skepticism for the view that pre-task meditation always improves performance.

These results also caution against treating the long-run effects of sustained mindfulness practice and the short-run effects of a single meditation session as interchangeable. Sustained practice over two weeks improves attention without measurably altering preferences. A single meditation session layered on top of sustained practice appears to heighten rather than dampen emotional reactivity and shifts choice in ways that look more responsive to immediate states. One-shot designs that randomize meditation immediately before lab tasks may therefore target a fundamentally different effect than designs that encourage regular practice. Researchers using meditation as an experimental treatment should report both whether participants have prior practice and the time elapsed between any prompted session and outcome measurement.

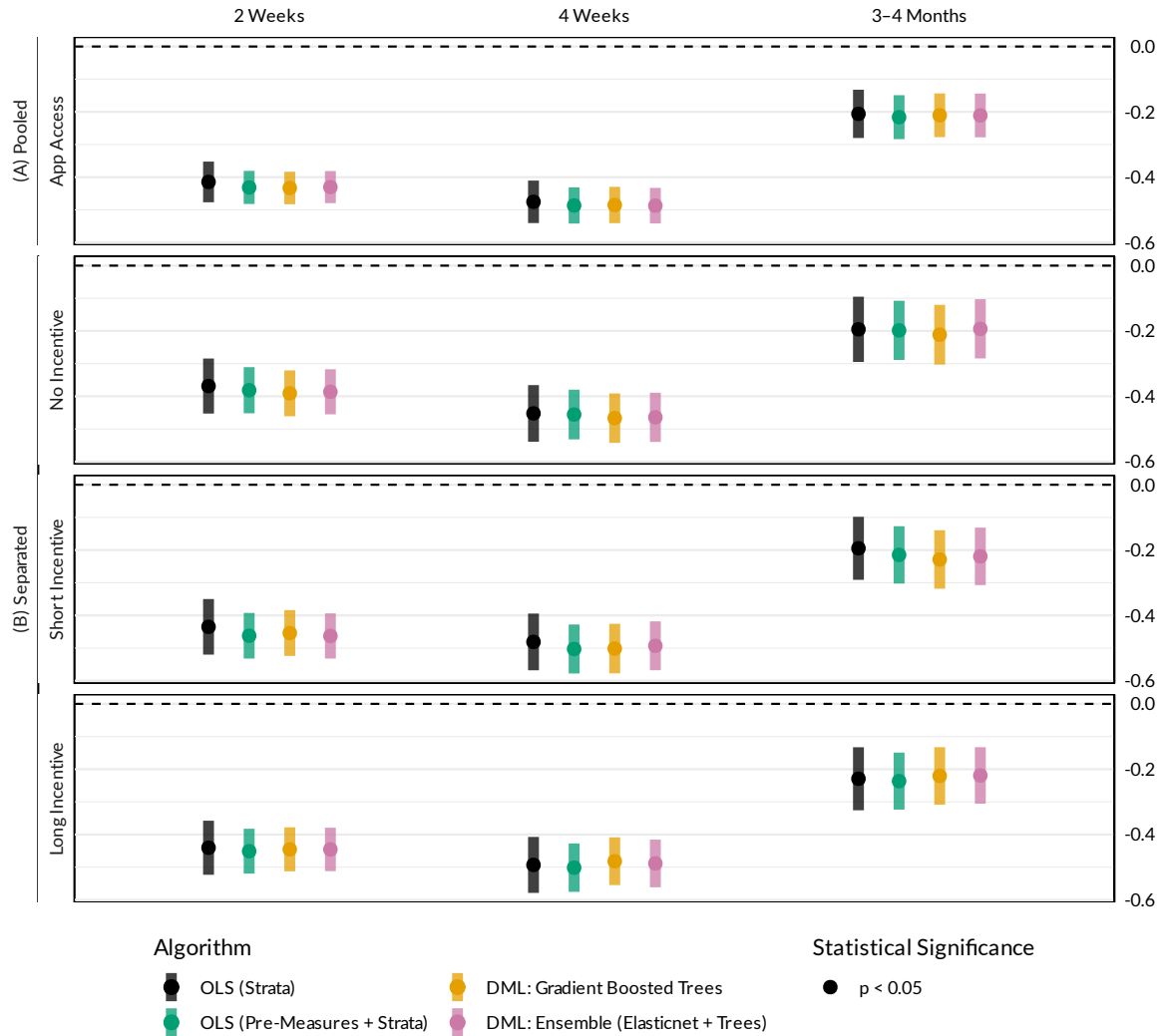
## A.2 Appendix Figures

Figure A.1: Example of Facebook and Instagram Ads Used for Recruitment



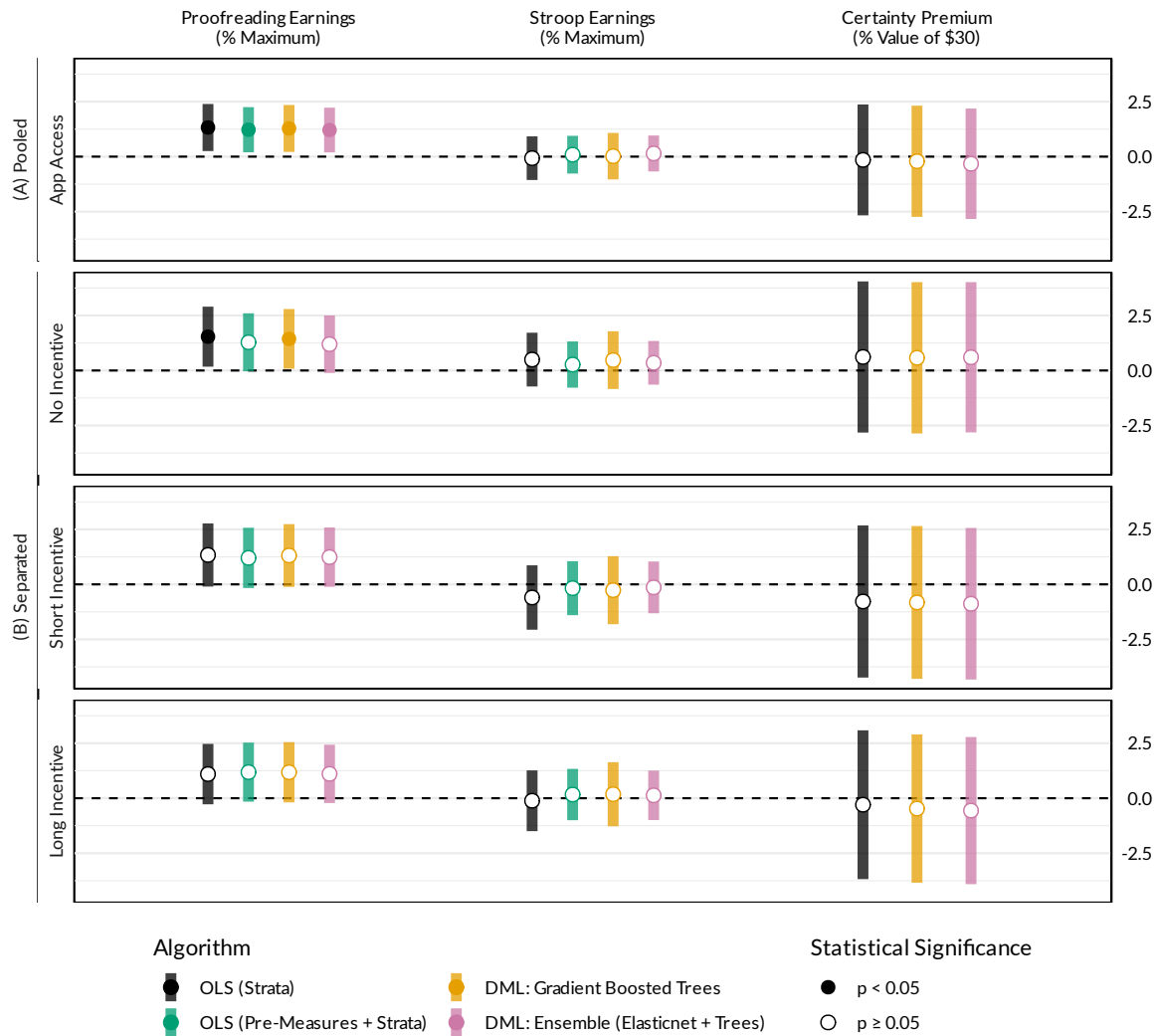
*Notes:* This figure shows examples of ads that were used to recruit participants for the study. A variety of images were used, with the most effective being automatically selected to be distributed more widely. The ad text was always the same.

Figure A.2: Effects of App Access and Usage Incentives on Mental Health, Adjusting for Covariates



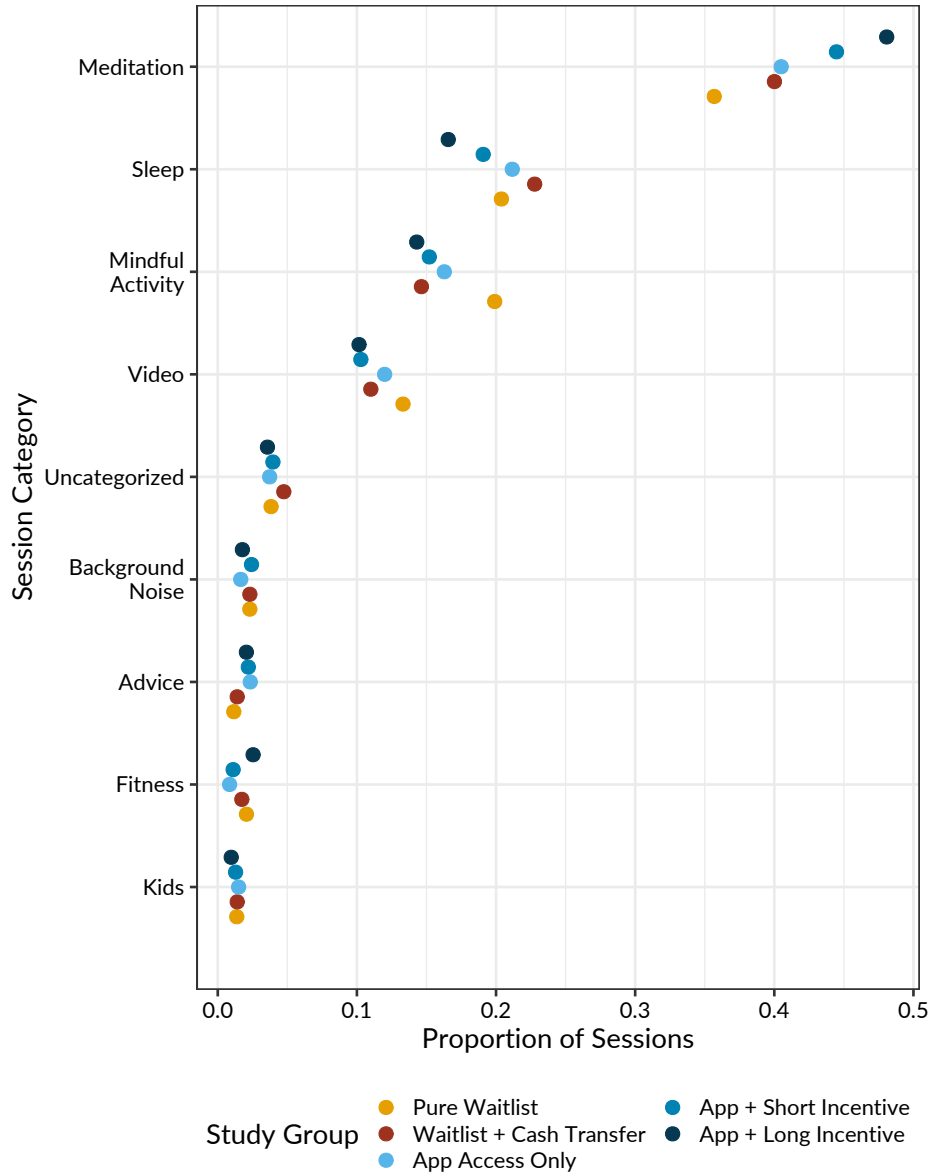
*Notes:* This figure presents average treatment effects of app access and usage incentives on reported symptoms of mental distress over time, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The layout mirrors Table 3. In each column, the green point corresponds to the point estimate reported in Table 3, and the bar is a 95% confidence interval. The black point and bar remove baseline covariates. The orange and pink points and bars present the debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of two algorithms: gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from elasticnet. Our debiased machine learning approach generalizes Equations (3.2) and (4.1) by including a broader set of baseline covariates and allowing them to enter the regression nonlinearly and interactively.

Figure A.3: Effects of App Access on Earnings in Proofreading Tasks, Adjusting for Covariates



Notes: This figure presents average treatment effects of app access on on earnings in the proofreading task, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The layout mirrors Table 4. In each column, the green point corresponds to the point estimated treatment effect after controlling for a baseline measure of the outcome; the black point does not control for a baseline measure (none are available for the certainty premium). The orange and pink points and bars present the debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of two algorithms: gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from elasticnet. Bars are 95% confidence intervals. Our debiased machine learning approach generalizes Equations (3.2) and (4.1) by including a broader set of baseline covariates and allowing them to enter the regression nonlinearly and interactively.

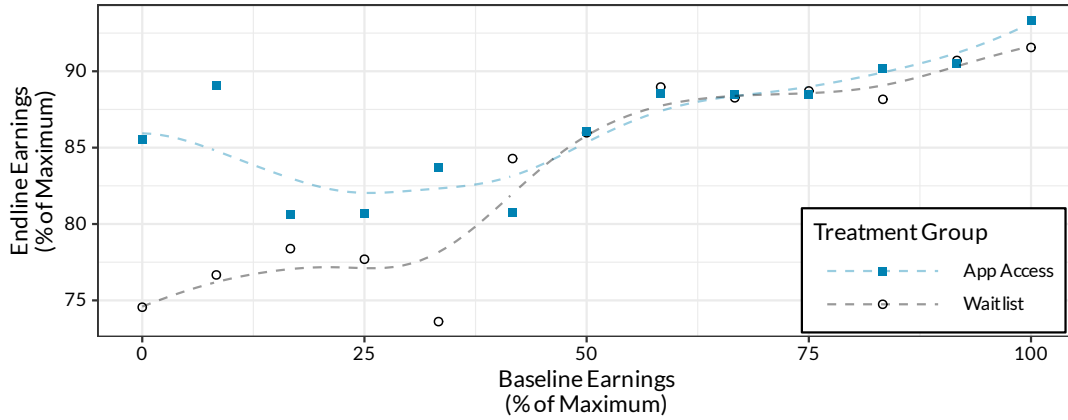
Figure A.4: Session Types



*Notes:* This figure presents the composition of Headspace app sessions each treatment group engaged in during their three months of app access. We derive the categories from the classification of these sessions in the app. Meditation sessions are guided or unguided recordings that teach mindfulness, either in general or in a specific context (e.g., handling stress at work). Sleep sessions are meant to help users fall asleep through a combination of guided meditation, background noise, or a combination of these. Mindful activities include deep breathing and walking. Videos contain a variety of content, ranging from natural environments (e.g., a river or savanna) to inspirational material. Uncategorized sessions were those with missing metadata, making them impossible to place into a category. Background Noise includes music and ambient noise, typically to accompany work or other focused tasks. Advice sessions are short recordings that relate to common problems (e.g., obsession, procrastination). Fitness sessions involve an activity like walking, running, or dance, but are not tagged as a “Mindful Activity” by Headspace. Kids sessions are targeted at teaching mindfulness to children.

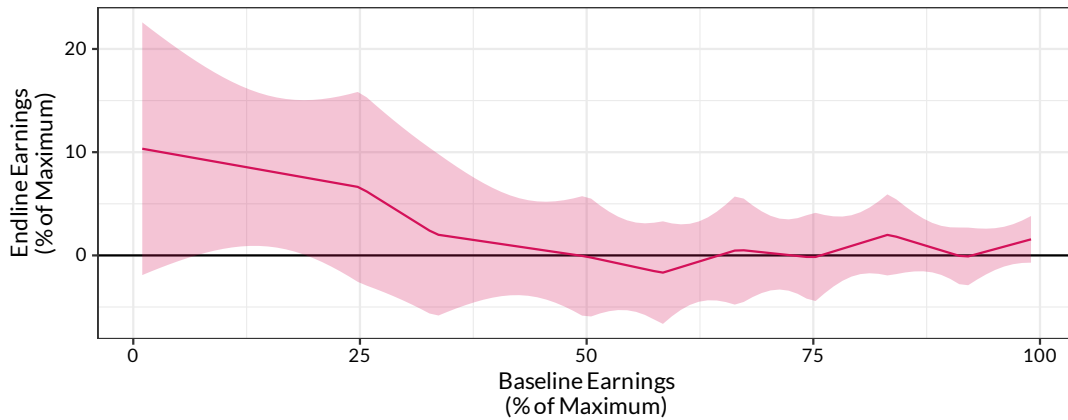
Figure A.5: Endline Proofreading Earnings, by Baseline Proofreading Earnings

(A) Binned Means and Local Linear Regression



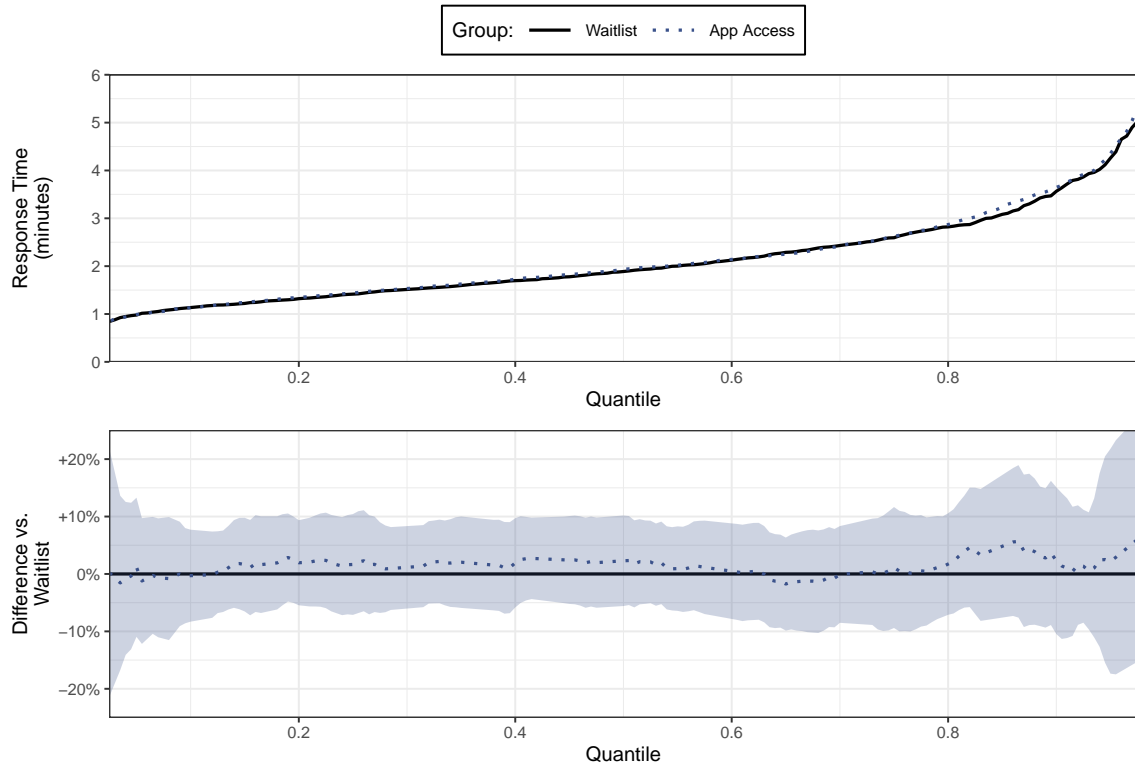
(B) Conditional Average Treatment Effect

Uniform 95% Confidence Band.  $H_0: \text{CATE} = 0; p = 0.016$ .



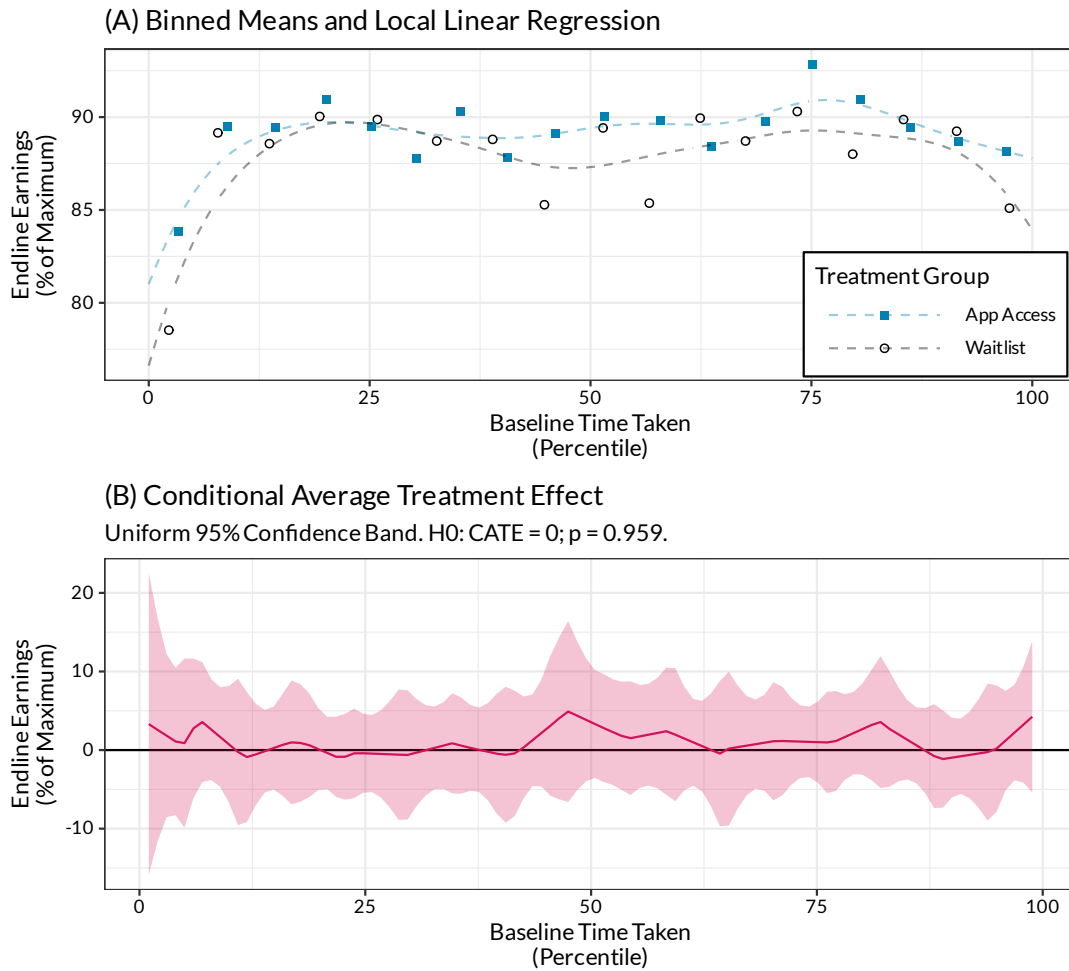
Notes: This figure presents the relationship between baseline and endline performance on the proofreading task. Panel A is presents a binned scatterplot and local linear regression, estimated separately in the Waitlist and App Access groups. The bins correspond to the 13 distinct values of baseline earnings, and the bandwidth for the local linear regression minimizes the integrated mean-square error. Panel B presents presents the conditional average treatment effect (solid line) and uniform 95% confidence band (shaded region), estimated by the method from Cattaneo et al. (2024a). The  $p$ -value is a joint test for equality of means in all bins; rejecting it implies that the conditional average treatment effect is not zero everywhere.

Figure A.6: Time Taken on Proofreading Task



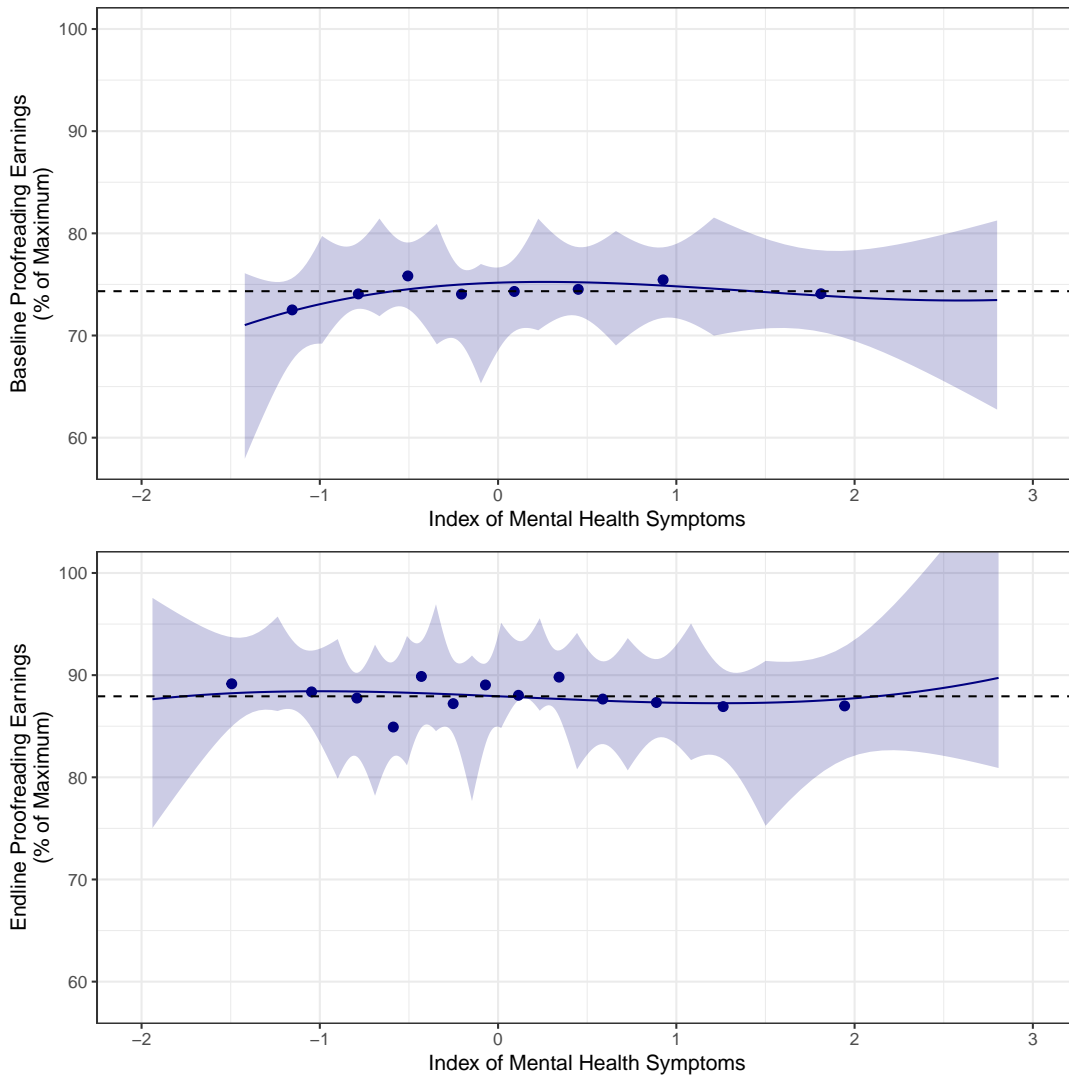
*Notes:* This figure presents the total time taken on the proofreading task. The top panel presents the empirical quantile functions for time taken in the Waitlist (solid black), App Access (dotted blue) groups. The horizontal axis starts at the 2.5% quantile and ends at the 97.5% quantile. The bottom panel presents the difference in quantiles for each App Access groups from the Waitlist group. The shaded area is a 95% uniform confidence band, computed via bootstrap with 9,999 iterations.

Figure A.7: Endline Proofreading Earnings, by Time Spent on Baseline Proofreading Task



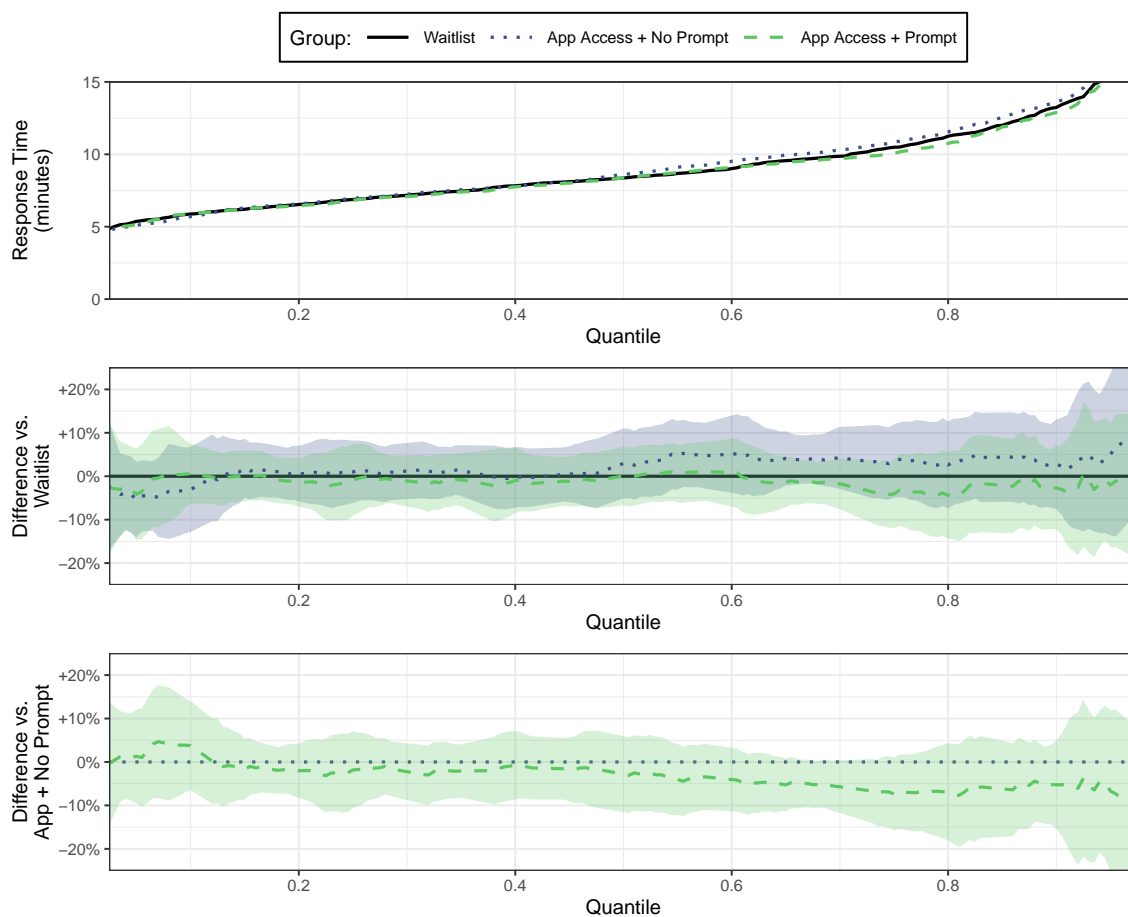
Notes: This figure presents the relationship between endline performance on the proofreading task and time taken on the baseline proofreading task. Panel A presents a binned scatterplot and local linear regression, estimated separately in the Waitlist and App Access groups. The number of bins for the binned scatterplot and the bandwidth for the local linear regression are chosen to minimize the integrated mean-square error. Panel B presents the conditional average treatment effect (solid line) and uniform 95% confidence band (shaded region), estimated by the method from Cattaneo et al. (2024a). The  $p$ -value is a joint test for equality of means in all bins; rejecting it implies that the conditional average treatment effect is not zero everywhere.

Figure A.8: Endline Proofreading Earnings by Mental Health, Among Waitlist



*Notes:* This figure presents the relationship between self-reported mental health symptoms and performance on the proofreading task pre-randomization (top panel) and two weeks after randomization (bottom panel). The points are means from a binned scatterplot, with bins chosen according to Cattaneo et al. (2024b). The dashed black line is the mean, the solid blue line is a global cubic polynomial fit for the conditional mean, and the shaded area is a 95% confidence band for the conditional mean. The pre-randomization mental health index combines the GAD-7 anxiety and PHQ-2 depression scales; we plot the relationship for all participants. The two-week index combines the GAD-7 anxiety, PHQ-8 depression, and PSS-10 stress scales; we plot the relationship for the waitlist group (control). Higher values indicate that respondents report more frequent and severe symptoms.

Figure A.9: Time Taken on Experimental Session Tasks



*Notes:* This figure presents the total time taken on tasks in the experimental session. The top panel presents the empirical quantile functions for time taken in the Waitlist (solid black), App Access No Prompt (dotted blue), and App Access Prompt (dashed green) groups. The horizontal axis starts at the 2.5% quantile and ends at the 97.5% quantile. The middle panel presents the difference in time taken for the treatment arms versus the waitlist, and the bottom panel presents the difference between the Prompt and No Prompt arms. The shaded area is a 95% uniform confidence band, computed via bootstrap with 9,999 iterations.

### A.3 Appendix Tables

Table A.1: Sample Size and Attrition

	Baseline	2 Weeks (Mental Health)	2 Weeks (Decision Making)	4 Weeks	3-4 Months
<b>Waitlist Group</b>					
N	955	943	925	936	817
Attrition Rate		0.013	0.031	0.020	0.145
<b>App Access Group</b>					
N	1429	1388	1336	1377	1212
Attrition Rate		0.029	0.065	0.036	0.152
Difference in Rates		0.016	0.034	0.016	0.007
<i>p</i> -value		0.013	<0.001	0.028	0.663

*Notes:* This table presents sample sizes and attrition rates for the pooled Waitlist and App Access groups. Sample sizes refer to the number of participants who begin each survey. The *p*-value for the difference in rates comes from Pearson's  $\chi^2$  test for a difference between proportions. We observe small but statistically significant differential attrition, with more dropout in the Waitlist group. [Table A.2](#) presents treatment effects on main outcomes under a worst-case model of selective attrition.

Table A.2: Bounds on Population Intent to Treat Effects, Given Attrition

<b>Outcome</b>	<b>ATE for Always Responders</b>	
	Estimate	<i>p</i> -value
Mental Health Index (2 weeks)	[-0.44, -0.44]	<0.001
Mental Health Index (4 weeks)	[-0.52, -0.48]	<0.001
Mental Health Index (3-4 months)	[-0.29, -0.18]	<0.001
Stroop Earnings	[0.06, 0.23]	0.760
Proofreading Earnings	[1.15, 1.35]	0.025

*Notes:* this table presents worst-case bounds on the main study’s treatment effects using the procedure from D. Lee (2009), which involves manually equalizing the attrition rates in the treatment and control groups by eliminating the most extreme values from the more responsive group. Eliminating extreme responses from the top of the distribution provides one bound, and eliminating from the bottom of the distribution provides the other. It is possible to tighten these bounds by conditioning on discrete covariates to estimate a set of conditional bounds, and then aggregating these conditional bounds. Under the assumption that propensity to respond is monotonic in treatment given the conditioning variables, the resulting estimates are worst-case bounds on the treatment effect for the “always responder” population (i. e., respondents who would complete followup surveys whether they received app access or were placed on the waitlist). We condition on quintiles of the baseline value of each outcome when estimating bounds. The “Estimate” column presents the identified set: treatment effects that are consistent with the data, under the worst-case selection model. We also report a *p*-value, formed by inverting Imbens and Manski (2004)’s confidence intervals for partially identified parameters.

Table A.3: Effects of Usage Incentives on App Take-up (Extensive Margin)

	Incentives Active (Weeks 1–2)	After Incentives (Week 3+)	After Waitlist (Week 5+)	Total (Week 1+)
<b>A. Pooling Incentives</b>				
App Access (S.E.)	0.805*** (0.018)	0.210*** (0.027)	0.031 (0.028)	0.408*** (0.023)
Any Incentive (S.E.)	0.089*** (0.021)	0.057** (0.026)	0.037 (0.028)	0.059*** (0.019)
Waitlist Mean	0.002	0.438	0.437	0.439
Sample Size	2384	2384	2384	2384
<b>B. Separating Arms</b>				
No Incentive (S.E.)	0.805*** (0.018)	0.210*** (0.027)	0.031 (0.028)	0.408*** (0.023)
Short Incentive (S.E.)	0.888*** (0.014)	0.270*** (0.026)	0.090*** (0.028)	0.465*** (0.021)
Long Incentive (S.E.)	0.899*** (0.014)	0.265*** (0.026)	0.045 (0.028)	0.469*** (0.021)
Waitlist Mean	0.002	0.438	0.437	0.439
Sample Size	2384	2384	2384	2384

*Notes:* This table presents the average treatment effects of app access and usage incentives on app usage, based on administrative data associated with each participant’s unique voucher code. The outcome is a binary variable equal to one if participants completed any meditation session using the app during the specified period. We randomize some treated participants to receive a \$10 incentive to complete at least 10 minutes of meditation on at least 4 out of first 2 weeks (“Short Incentive”) or at least 10 out of the first 2 weeks (“Long Incentive”). Panel A estimates the average effect of the two incentive arms relative to unincentivized treatment group participants, and Panel B separates the incentive arms. Usage incentives are active in the first two weeks (column 1), and the waitlist remains in effect for 2 more weeks after incentives expire (column 2). After the waitlist ends, all participants have full app access (column 3). Our administrative data ends 90 days after participants activate their vouchers, and we report cumulative usage in this period in column 4. The estimating equations are [Equation \(3.1\)](#) (panel A) and [Equation \(3.2\)](#) (panel B). Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.4: Effects of App Access and Usage Incentives on Anxiety Symptoms Using Score Cutoffs

	Two Weeks	Four Weeks	Three–Four Months
<b>A. Mild or Worse (Score <math>\geq 5</math>)</b>			
App Access	-0.172***	-0.243***	-0.099***
(S.E.)	(0.017)	(0.017)	(0.020)
Baseline Outcome	0.301***	0.270***	0.246***
(S.E.)	(0.029)	(0.029)	(0.031)
Waitlist Mean	0.723	0.717	0.594
Sample Size	2330	2312	2009
<b>B. Moderate or Worse (Score <math>\geq 10</math>)</b>			
App Access	-0.137***	-0.143***	-0.080***
(S.E.)	(0.015)	(0.015)	(0.016)
Baseline Outcome	0.319***	0.278***	0.218***
(S.E.)	(0.023)	(0.022)	(0.024)
Waitlist Mean	0.285	0.261	0.216
Sample Size	2330	2312	2009
<b>C. Severe (Score <math>\geq 15</math>)</b>			
App Access	-0.042***	-0.067***	-0.041***
(S.E.)	(0.010)	(0.010)	(0.010)
Baseline Outcome	0.365***	0.280***	0.166***
(S.E.)	(0.032)	(0.030)	(0.029)
Waitlist Mean	0.094	0.100	0.076
Sample Size	2330	2312	2009

Notes: This table presents average treatment effects of app access and usage incentives on reported symptoms of depression. It supplements Table 3. Anxiety is measured using the GAD-7 scale, where scores range from 0 (no symptoms) to 21 (every symptom is severe). This table discretizes these scales into indicators for obtaining a score above commonly used thresholds. The estimating equation is Equation 4.1. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.5: Effects of App Access and Usage Incentives on Depression Symptoms Using Score Cutoffs

	Two Weeks	Four Weeks	Three–Four Months
<b>A. Mild or Worse (Score <math>\geq 5</math>)</b>			
App Access	-0.160***	-0.178***	-0.080***
(S.E.)	(0.017)	(0.018)	(0.020)
Baseline Outcome	0.168***	0.180***	0.150***
(S.E.)	(0.009)	(0.009)	(0.010)
Waitlist Mean	0.680	0.631	0.573
Sample Size	2330	2311	2004
<b>B. Moderate or Worse (Score <math>\geq 10</math>)</b>			
App Access	-0.150***	-0.142***	-0.031*
(S.E.)	(0.016)	(0.015)	(0.017)
Baseline Outcome	0.175***	0.144***	0.134***
(S.E.)	(0.009)	(0.009)	(0.010)
Waitlist Mean	0.324	0.287	0.238
Sample Size	2330	2311	2004
<b>C. Severe (Score <math>\geq 15</math>)</b>			
App Access	-0.062***	-0.072***	-0.048***
(S.E.)	(0.011)	(0.011)	(0.012)
Baseline Outcome	0.108***	0.089***	0.086***
(S.E.)	(0.008)	(0.007)	(0.009)
Waitlist Mean	0.119	0.115	0.111
Sample Size	2330	2311	2004

Notes: This table presents average treatment effects of app access and usage incentives on reported symptoms of anxiety. It supplements Table 3. Depression is measured using the PHQ-8 scale, where scores range from 0 (no symptoms) to 24 (every symptom is severe). This table discretizes these scales into indicators for obtaining a score above commonly used thresholds. The estimating equation is Equation 4.1. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.6: Heterogeneous Effects of App Access and Usage Incentives

	2 Weeks		4 Weeks	
	Days Used	Mental Health	Days Used	Mental Health
<b>A. Baseline Anxiety</b>				
App Access (S.E.)	6.492*** (0.201)	-0.256*** (0.034)	9.156*** (0.311)	-0.273*** (0.036)
App Access × At Least Median (S.E.)	0.255 (0.260)	-0.302*** (0.050)	0.191 (0.400)	-0.368*** (0.054)
<b>B. Baseline Mindfulness</b>				
App Access (S.E.)	6.724*** (0.194)	-0.536*** (0.039)	9.286*** (0.291)	-0.576*** (0.043)
App Access × At Least Median (S.E.)	-0.149 (0.263)	0.201*** (0.049)	-0.028 (0.394)	0.171*** (0.054)
<b>C. Belief about ATE</b>				
App Access (S.E.)	6.292*** (0.195)	-0.349*** (0.035)	8.759*** (0.290)	-0.392*** (0.039)
App Access × At Least Median (S.E.)	0.610** (0.261)	-0.145*** (0.049)	0.897** (0.393)	-0.164*** (0.053)
<b>D. Social Desirability</b>				
App Access (S.E.)	6.558*** (0.202)	-0.463*** (0.040)	9.060*** (0.298)	-0.511*** (0.044)
App Access × At Least Median (S.E.)	0.099 (0.266)	0.055 (0.050)	0.288 (0.398)	0.047 (0.054)

*Notes:* This table presents the heterogeneous treatment effects of app access and usage incentives on app usage and mental health, along several baseline covariates. Each panel contrasts participants at or above the median baseline value of the covariate with those who are below the baseline. Panel A separates participants by their baseline anxiety score on the GAD-7 scale, and Panel B does the same for baseline mindfulness on the FFMQ-15 scale. Panel C splits participants based on their beliefs about the average treatment effect of app access on anxiety. Above-median beliefs indicate that the participant believes app access will cause a larger decline in anxiety. Panel D presents heterogeneity by social desirability, measured on the Marlow-Crowne scale. In all cases, the estimating equation is:  $Y_i = \delta_{s(i)} + \beta \text{AppAccess} + \gamma Y_i^{\text{pre}} + H_i \cdot (\delta_{h,s(i)} + \beta_h \text{AppAccess} + \gamma_h Y_i^{\text{pre}}) + \epsilon_i$ , where  $H_i$  is an indicator for the covariate in question being weakly above the median. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.7: Effects of App Access and Waitlist Cash Transfer on Mental Health

	Two Weeks	Four Weeks	Three–Four Months
Waitlist Cash Transfer	0.011	0.052	0.031
(S.E.)	(0.039)	(0.044)	(0.054)
App No Incentive	−0.376***	−0.430***	−0.183***
(S.E.)	(0.040)	(0.044)	(0.054)
Short Incentive	−0.457***	−0.477***	−0.199***
(S.E.)	(0.040)	(0.043)	(0.052)
Long Incentive	−0.445***	−0.476***	−0.221***
(S.E.)	(0.039)	(0.043)	(0.052)
Baseline Outcome	0.626***	0.578***	0.510***
(S.E.)	(0.021)	(0.023)	(0.029)
Pure Waitlist Mean	0.000	0.000	0.000
Sample Size	2330	2311	2004
Index Components:			
Anxiety	GAD-7	GAD-7	GAD-7
Depression	PHQ-8	PHQ-8	PHQ-8
Stress	PSS-10	PSS-10	

*Notes:* This table presents average treatment effects of app access and a \$15 cash transfer on reported symptoms of mental distress over time. We measure symptoms of anxiety using the seven-item Generalized Anxiety Disorder scales (GAD-7); symptoms of depression using the eight-item Patient Health Questionnaire (PHQ-8); and stress using the ten-item Perceived Stress Scale (PSS-10). The outcome at each timepoint is a standardized index that combines the mental health scales measured at that time. We first standardize each scale in each time period by subtracting the Pure Waitlist mean and dividing by the Pure Waitlist standard deviation. The index is the average of these standardized scales. Lower scores indicate lower reported levels of distress. The estimating equation adds a term for the Waitlist Cash Transfer group to Equation (3.2). Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.8: Effect of App Access on Proofreading Task (Other Measures)

	Typos Found	False Positives	Seconds Taken (log)
<b>A. Pooling Treatment</b>			
App Access	0.237**	-0.016	0.008
(S.E.)	(0.103)	(0.033)	(0.017)
Baseline Outcome	0.318***	-0.027***	0.548***
(S.E.)	(0.046)	(0.010)	(0.024)
Waitlist Mean	14.756	0.291	4.756
Sample Size	2257	2257	2257
<b>B. Separating Arms</b>			
No Incentive	0.240*	-0.055	-0.008
(S.E.)	(0.135)	(0.041)	(0.023)
Short Incentive	0.242*	0.019	0.033
(S.E.)	(0.135)	(0.048)	(0.023)
Long Incentive	0.229*	-0.013	-0.002
(S.E.)	(0.136)	(0.041)	(0.024)
Baseline Outcome	0.318***	-0.027***	0.549***
(S.E.)	(0.046)	(0.010)	(0.024)
Waitlist Mean	14.756	0.291	4.756
Sample Size	2257	2257	2257

*Notes:* This table presents average treatment effects of app access and usage incentives the proofreading task. The primary outcome is earnings relative to the maximum possible amount in percentage points. For example, the average waitlist participant earned 87.9% of the maximum possible bonus. Participants receive a bonus for each typo they correctly identify during the task (column 1), but suffer a penalty for false positives (column 2). There is no time limit on the task, leading to a skewed distribution of time taken; column 3 reports time taken in log seconds. Panel A uses Equation (4.1) to estimate the average effect of App Access arms relative to the Waitlist. Panel B uses our pre-registered specification, Equation (3.2), to separate the No Incentive, Short Incentive, and Long Incentive arms within the App Access group. In both panels, the reference group combines the Pure Waitlist and Waitlist Cash Transfer arms. All regressions include stratum fixed effects and a pre-randomization measure of the outcome. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.9: Effect of App Access on Stroop Task (Other Measures)

	Errors Made	Seconds Taken
<b>A. Pooling Treatment</b>		
App Access	-0.044	0.043
(S.E.)	(0.123)	(0.307)
Baseline Outcome	0.296***	0.209
(S.E.)	(0.050)	(0.439)
Waitlist Mean	1.439	67.817
Sample Size	2256	2254
<b>B. Separating Arms</b>		
No Incentive	-0.132	-0.029
(S.E.)	(0.169)	(0.394)
Short Incentive	0.064	0.017
(S.E.)	(0.170)	(0.499)
Long Incentive	-0.064	0.141
(S.E.)	(0.152)	(0.431)
Baseline Outcome	0.295***	0.209
(S.E.)	(0.050)	(0.439)
Waitlist Mean	1.439	67.817
Sample Size	2256	2254

*Notes:* This table presents average treatment effects of app access on earnings on an incentivized Stroop test. Participants start with 30 cents and complete a 40-item test. Panel A uses Equation (4.1) to estimate the average effect of App Access arms relative to the Waitlist. Panel B uses our pre-registered specification, Equation (3.2), to separate the No Incentive, Short Incentive, and Long Incentive arms within the App Access group. In both panels, the reference group combines the Pure Waitlist and Waitlist Cash Transfer arms. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.10: Effect of App Access on Certainty Premium Components

	Switch Point (Certainty)	Switch Point (Uncertainty)	$v(10)_c$	$v(10)_u$
<b>A. Pooling Treatment</b>				
App Access (S.E.)	0.774 (0.862)	0.462 (0.700)	0.774 (0.862)	0.923 (1.400)
Waitlist Mean	58.380	72.022	58.380	44.043
Sample Size	2247	2247	2247	2247
<b>B. Separating Arms</b>				
No Incentive (S.E.)	-0.402 (1.179)	-0.508 (0.959)	-0.402 (1.179)	-1.016 (1.917)
Short Incentive (S.E.)	0.963 (1.175)	0.871 (0.974)	0.963 (1.175)	1.742 (1.948)
Long Incentive (S.E.)	1.777 (1.162)	1.037 (0.946)	1.777 (1.162)	2.074 (1.893)
Waitlist Mean	58.380	72.022	58.380	44.043
Sample Size	2247	2247	2247	2247

*Notes:* This table presents average treatment effects of app access on choices over risky prospects. The certainty premium measures the additional utility a participant assigns to receiving \$10 as a guaranteed payment versus as the outcome of a risky lottery. We elicit  $P$ , the probability at which a participant is indifferent between a guaranteed \$10 payment and a (\$30,  $P$ ; \$0  $1 - P$ ) lottery; and  $Q$ , the probability at which a participant is indifferent between a (\$30, 0.5; \$10 0.5) lottery and a (\$30,  $Q$ ; \$0  $1 - Q$ ) lottery. Under expected utility maximization, these indifference probabilities imply  $v(10)_c = Pv(30)_u - (1 - P)v(0)_u$  and  $0.5v(10)_u + 0.5v(30)_u = Qv(30)_u - (1 - Q)v(0)_u$ . Normalizing the value of \$0 and \$30 to  $v(0)_u = 0$  and  $v(30)_u = 100$ , these indifference probabilities imply that an expected utility maximizer values \$10 at  $v(10)_c = 30P$  in the first task (certainty) and  $v(10)_u = 100 \frac{Q-0.5}{0.5}$  in the second (uncertainty). The certainty premium is  $v(10)_c - v(10)_u$ . Columns 1 and 2 are the probabilities,  $P$  and  $Q$ , expressed as percentages. Higher values indicate greater risk aversion. Columns 3 and 4 are the implied valuations of \$10 under certainty and uncertainty, respectively. Panel A estimates the average effect of App Access relative to the Waitlist using Equation (4.1). Panel B uses our pre-registered specification, Equation (3.2), to separate the No Incentive, Short Incentive, and Long Incentive arms within the App Access group. In both panels the reference group combines Pure Waitlist and Waitlist Cash Transfer. All regressions include stratum fixed effects but no baseline measures, as none are available. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.11: Effect of App Access on Economic Behavior (Secondary)

	Avoid Unpleasant Info	Seek Pleasant Info	Avoid Salient Loss
<b>A. Pooling Treatment</b>			
App Access	-0.001	0.038*	0.036*
(S.E.)	(0.014)	(0.020)	(0.020)
Waitlist Mean	0.454	0.294	0.319
Sample Size	2257	2257	2257
<b>B. Separating Arms</b>			
No Incentive	0.006	0.034	0.021
(S.E.)	(0.019)	(0.027)	(0.027)
Short Incentive	-0.009	0.033	0.018
(S.E.)	(0.020)	(0.027)	(0.027)
Long Incentive	-0.002	0.046*	0.068**
(S.E.)	(0.020)	(0.027)	(0.028)
Waitlist Mean	0.454	0.294	0.319
Sample Size	2257	2257	2257

*Notes:* This table presents average treatment effects of app access and usage incentives on three secondary choice outcomes measured in the experimental session, each designed to capture a distinct channel through which emotion may influence decisions. The first outcome (column 1) is an indicator for declining information: participants are offered four hyperlinks to potentially unpleasant but useful information (dementia risk factors, risk of job loss to automation, a retirement financial risk calculator, and a life expectancy calculator), and the outcome is the share of links declined. The second outcome (column 2) is an indicator for choosing pleasant but non-instrumental information: before a gamble resolved by three coin tosses, participants choose between learning the outcome of one toss (payoff-relevant) or one of two irrelevant pieces of information (the starting value of a digital wallet, or the age of the world's oldest tree); the outcome equals one if the participant selects irrelevant information. The third outcome (column 3) is an indicator for rejecting a gamble with a 99% chance of gaining \$1 and a salient 1% chance of losing the survey's \$10 base payment, capturing aversion to small-probability salient losses. Panel A uses Equation [Equation \(4.1\)](#) to estimate the average effect of App Access arms relative to the Waitlist. Panel B uses our pre-registered specification, [Equation \(3.2\)](#), to separate the No Incentive, Short Incentive, and Long Incentive arms within the App Access group. In both panels, the reference group combines the Pure Waitlist and Waitlist Cash Transfer arms. All regressions include stratum fixed effects. No baseline measures are available for these outcomes, as they were administered only at endline. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted p-values: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. These are pre-registered secondary outcomes and are not included in the Romano-Wolf multiple hypothesis testing correction applied to primary outcomes.

Table A.12: Effect of Pre-Task Meditation Prompt and Emotion Induction on Behavior

	Pre-Survey			Pre-Negative/Neutral Treatment				Post-Negative/Neutral		
	Meditation Session (1)	Avoid Unpleasant Info (2)	Stroop Earnings (3)	Seek Pleasant Info (4)	Avoid Salient Loss (5)	Proofreading Earnings (6)	Certainty Premium (7)			
Prompt (S.E.)	0.619*** (0.019)	0.018 (0.017)	-0.147 (0.523)	0.054** (0.024)	0.061** (0.024)	1.641** (0.807)	-4.343** (2.183)			
Negative $\times$ Prompt (S.E.)						-2.294** (0.909)	4.859** (2.377)			
No Prompt (S.E.)	0.008** (0.003)	-0.020 (0.017)	0.324 (0.512)	0.021 (0.023)	0.010 (0.024)	1.839** (0.860)	-1.912 (2.143)			
Negative $\times$ No Prompt (S.E.)						-0.432 (0.852)	-0.521 (2.354)			
Negative $\times$ Waitlist (S.E.)						-0.318 (0.853)	-3.841** (1.953)			
Baseline Outcome (S.E.)			0.444*** (0.029)			0.135*** (0.014)				
Waitlist Mean	0.000	0.454	74.120	0.294	0.319	88.362	16.255			
Sample Size	2261	2257	2256	2257	2257	2257	2247			
Pre-registered Primary Task		✓			✓	✓				
<b>Joint Tests</b>	All Outcomes (2-7)			Attention (3,6)		Choices (2,4,5,7)				
Prompt = Waitlist	$p = 0.006$	$p = 0.261$	$p = 0.005$							
Prompt = No Prompt	$p = 0.032$	$p = 0.092$	$p = 0.064$							
No Prompt = Waitlist	$p = 0.141$	$p = 0.044$	$p = 0.555$							

Notes: This table presents the average treatment effects of app access and a prompt to meditate on choice and effort, based on two cross-randomized interventions. Outcomes are presented in survey order. Before the experimental tasks, half of the App Access group receives a prompt offering a \$3 incentive to complete a meditation session on the app immediately before the survey; 60.3% comply. Participants then complete an unpleasant information task, Stroop test, a pleasant information task, and a loss avoidance task. They then are randomized to a negative or neutral emotional prime: recalling a stressful or neutral memory before the proofreading task, and to describing how they would handle a large or small medical bill before the certainty premium. Figure 1b depicts the survey flow. For outcomes prior to the Negative / Neutral treatment (columns 1-5), the regression specification is  $Y_i = \delta_s + \beta_1 \text{Prompt} + \beta_2 \text{NoPrompt} + \gamma Y_i^{\text{pre}} + \epsilon_i$ ; afterwards it is  $Y_i = \alpha + \beta_1 \text{Waitlist} \times \text{Negative} + \beta_2 \text{Prompt} + \beta_3 \text{Prompt} \times \text{Negative} + \beta_4 \text{NoPrompt} + \beta_5 \text{NoPrompt} \times \text{Negative} + \beta_6 Y_i^{\text{pre}} + \epsilon$ . Baseline outcomes are only available for the Stroop and proofreading tasks; other tasks were only conducted at endpoint. Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted regression  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . The heteroskedasticity-robust  $p$ -value for the joint test is computed from the Wald statistic for seemingly unrelated regressions, using a bootstrap stratified by treatment status with 9,999 iterations. Joint tests are Wald tests for equality of all coefficients associated with each arm, including interactions with the Negative treatment.

Table A.13: Effect of Pre-Task Meditation Prompt and Emotion Induction on Certainty Premium Components

	Switch Point (Certainty)	Switch Point (Uncertainty)	$v(10)_c$	$v(10)_u$
Prompt	1.100	2.721**	1.100	5.443**
(S.E.)	(1.456)	(1.197)	(1.456)	(2.394)
Negative $\times$ Prompt	-0.977	-2.918**	-0.977	-5.837**
(S.E.)	(1.575)	(1.338)	(1.575)	(2.676)
No Prompt	1.697	1.804	1.697	3.608
(S.E.)	(1.466)	(1.153)	(1.466)	(2.305)
Negative $\times$ No Prompt	-0.140	0.191	-0.140	0.382
(S.E.)	(1.602)	(1.286)	(1.602)	(2.571)
Negative $\times$ Waitlist	0.704	2.272**	0.704	4.545**
(S.E.)	(1.310)	(1.048)	(1.310)	(2.096)
Baseline Outcome				
(S.E.) Waitlist Mean	58.030	70.887	58.030	41.775
Sample Size	2247	2247	2247	2247

*Notes:* This table presents the average treatment effects of app access, a meditation prompt, and an induction of negative emotions on components of the certainty premium. The certainty premium measures the additional utility a participant assigns to receiving \$10 as a guaranteed payment versus as the outcome of a risky lottery. We elicit  $P$ , the probability at which a participant is indifferent between a guaranteed \$10 payment and a (\$30,  $P$ ; \$0  $1 - P$ ) lottery; and  $Q$ , the probability at which a participant is indifferent between a (\$30, 0.5; \$10 0.5) lottery and a (\$30,  $Q$ ; \$0  $1 - Q$ ) lottery. Under expected utility maximization, these indifference probabilities imply  $v(10)_c = Pv(30)_u - (1 - P)v(0)_u$  and  $0.5v(10)_u + 0.5v(30)_u = Qv(30)_u - (1 - Q)v(0)_u$ . Normalizing the value of \$0 and \$30 to  $v(0)_u = 0$  and  $v(30)_u = 100$ , these indifference probabilities imply that an expected utility maximizer values \$10 at  $v(10)_c = 100P$  in the first task (certainty) and  $v(10)_u = 100\frac{Q-0.5}{0.5}$  in the second (uncertainty). The certainty premium is  $v(10)_c - v(10)_u$ . Columns 1 and 2 are the probabilities,  $P$  and  $Q$ , expressed as percentages. Higher values indicate greater risk aversion. Columns 3 and 4 are the implied valuations of \$10 under certainty and uncertainty, respectively. The regression specification is  $Y_i = \alpha + \beta_1 \text{Waitlist} \times \text{Negative} + \beta_2 \text{Prompt} + \beta_3 \text{Prompt} \times \text{Negative} + \beta_4 \text{NoPrompt} + \beta_5 \text{NoPrompt} \times \text{Negative} + \epsilon$ . Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted regression  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Table A.14: Performance on Baseline and Endline Proofreading Tasks

Baseline Quintile	Observations	Endline Earnings (SE)
1	183	80.171 (1.313)
2	131	88.586 (1.086)
3	229	88.375 (0.845)
4	184	90.710 (0.766)
5	198	91.558 (0.696)

*Notes:* This table presents the performance of Waitlist group participants in the endline proofreading task, based on their performance on baseline survey's proofreading task. Endline earnings are reported as a percent of the maximum possible earnings (\$1.50).

Table A.15: Independent Meditation Activities

	Any Non-Headspace Meditation	With Another App	Without an App
App Access	0.097***	0.054**	0.071***
(S.E.)	(0.030)	(0.025)	(0.027)
Short Incentive	0.136***	0.030	0.113***
(S.E.)	(0.030)	(0.024)	(0.027)
Long Incentive	0.119***	0.051**	0.093***
(S.E.)	(0.030)	(0.025)	(0.027)
Waitlist Mean	0.341	0.171	0.218
N	1994	1994	1994

*Notes:* This table presents average treatment effects of app access and usage incentives on self-reported meditation activities outside the mindfulness app. In the three–four month endline survey, we ask participants to indicate whether they have used Headspace, another app, or meditated without an app in the previous month. The estimating equation is [Equation \(3.2\)](#). Standard errors are from the heteroskedasticity-robust HC3 estimator. Unadjusted  $p$ -values: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .