

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

ADVIK SHREEKUMAR

MIT

PIERRE-LUC VAUTREY*

MIT

January 24, 2023

Abstract

Emotions and worries can reduce individuals' available attention and affect economic decisions. In a four-week experiment with 2,384 US adults, offering free access to a popular mindfulness meditation app that costs \$13 per month improves mental health, productivity, and decision-making. First, it causes a 0.44 standard deviation reduction in symptoms of stress, anxiety, and depression, comparable to the impacts of more expensive in-person therapy, with improvements even among participants with minimal or mild symptoms at baseline. Second, it increases earnings on a proofreading task by 1.9 percent. Third, it modifies the effects of emotion on decision making, and may reduce the interference of personal worries on choices over risky prospects.

*Corresponding author: Advik Shreekumar (adviks@mit.edu). We thank Stacy Wang for excellent research assistance. Vautrey is indebted to Abhijit Banerjee, Esther Duflo and particularly to Frank Schilbach for their continued guidance and support, and Shreekumar to Ben Olken and Frank Schilbach for the same. We thank seminar and conference participants at MIT, Cornell, UCSD, Stanford (SITE), brig, the University of Chicago (AFE) and Deivy Houeix, Clemence Idoux, Justine Knebelmann, Antoine Levy, Ro'ee Levy, Sendhil Mullainathan, Mathilde Munoz, Lucy Page, Matthew Ridley, Garima Sharma, Lena Song and particularly Charlie Rafkin for helpful comments and suggestions. We thank all our study participants for their time and patience. Experimental instructions are available at <https://adviksh.com/stable/00001.pdf>. 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 project 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. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement, by ACF/HHS, or the U.S. Government. For more information, please visit the ACF website, Administrative and National Policy Requirements. This material is also based on work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. This research received support from the George and Obie Shultz Fund at MIT. 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 in any way. The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement by, Headspace.

1 Introduction

Research in behavioral economics has shown that individuals regularly make mistakes and inconsistent choices that may reduce welfare. Such departures from rationality often arise from emotions interfering with their judgement and to competing demands on their limited attention (Johnson and Tversky, 1983; Loewenstein, 2000; Loewenstein et al., 2001; Lerner et al., 2004; Callen et al., 2014; Gabaix, 2019). Distractions and worries can also affect performance on attention-demanding tasks and reduce productivity (Banerjee and Mullainathan, 2008; Kaur et al., 2021; Duquenois, 2021). As stress and poor mental health are rising, with about one in two American adults reporting being very stressed in any two weeks (American Psychological Association, 2020), so is the demand from the general public, firms, and organizations for tools to reduce the burden of worries on well-being, attention, productivity, and decision-making.

Mindfulness meditation apps are one such popular tool, with hundreds of millions of downloads and billions of dollars in valuation. They are based on a secular 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). When professionally administered in clinical settings, these techniques are effective treatments for depression and anxiety (Goldberg et al., 2021). Evidence from neuroimaging and self-reports also suggest that mindfulness improves the regulation of emotions (Guendelman et al., 2017). Many meditators cite general wellness as a reason for their practice, but also often intend to improve their energy, memory, and concentration (Cramer et al., 2016), and several large firms promote mindfulness among their employees (e.g., Google, Nike, McKinsey).

However, high-quality evidence on the longer-term impacts of these apps on mental health, productivity, and decision-making with the general public remains scarce. The existing literature provides promising evidence from a combination of correlations, small-scale experiments, and analyses of immediate effects from one-time mindfulness training.¹ This paper studies the effects of a leading mindfulness meditation app (Headspace) on mental health, productivity, and decision-making by conducting a pre-registered RCT with 2,384 US adults interested in meditation, recruited via social media ads. In the four-week experiment, the first group receives free access to a premium version of the app (worth \$13); 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 group.² We assess impacts on stress, anxiety, and depression over time, with validated questionnaires at

¹A recent meta-analysis notes that the majority of existing randomized control trials (RCTs) of online mindfulness interventions suffer from attrition issues and small sample sizes (Sommers-Spijkerman et al., 2021). The RCTs with larger samples (e.g., Mak et al., 2018) typically evaluate “mobile health” versions of structured therapies, an approach that is less flexible than the unstructured meditation app we evaluate. Experiments evaluating popular apps typically have limited statistical power (e.g., Flett et al., 2019). A separate meta-analysis of workplace mindfulness programs concludes that these programs improve employee mental health and have ambiguous effects on productivity (Vonderlin et al., 2020). Some experiments test the effects of mindfulness meditation on decision-making outcomes such as information avoidance (Ash et al., 2021), altruism (Iwamoto et al., 2020), intertemporal choice and choice under risk (Alem et al., 2021), or sunk-cost bias (Hafenbrack et al., 2014). These studies report suggestive effects but, with the exception of Iwamoto et al. (2020), have limited statistical power to detect moderately sized effects.

²Half of the control is further randomized to receive an unconditional cash transfer of \$15, that is, more than the price of one month of access to the app. This transfer has no effect on mental health or other outcomes.

six different stages of the experiment. After two weeks, participants complete a set of incentivized economic decisions and effort tasks where attention and emotional regulation play a central role.

Detailed administrative usage data demonstrate high engagement for the first two weeks after people receive access to the app. Of the participants who receive access without additional incentives, 80.3 percent try meditating with it at least once and use it an average of 11 times for a total of 95 minutes during the first 16 days.³ Incentives successfully increase usage while they last: they make participants 8.8 percentage points more likely to try the app and use it 4 more times 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 as they appear to do in other studies of wellness habits such as hand washing or gym attendance (Hussam et al., 2019; Charness and Gneezy, 2009; Acland and Levy, 2015).⁴

Our first main finding is that offering access to the app without incentives leads to meaningful improvements in mental health, as measured by a 0.37 standard deviation (SD) reduction in an index of depression, anxiety, and stress compared to the control group after two weeks ($p < 0.001$).⁵ App usage declines after three weeks, but the treatment effects persist at four weeks (0.44 SDs), corresponding to a 11 to 13 percentage point reduction in the fraction of participants with moderate symptoms of anxiety and depression relative to the control group rate of 26 to 29 percent. Effects are smaller but still significant (0.31 SDs) even among participants with only mild or minimal symptoms at baseline. They are also persistent. Three to four months after intervention, well after Waitlist participants received app access, the initial treatment group still reports lower anxiety than the initial Waitlist (0.22 SDs). The reduction in treatment effect is commensurate with uptake of the app by waitlisted participants, suggesting that the main force is the control group “catching up” rather than the treatment group “backsliding”. While these effect sizes are smaller than the estimated effects of cognitive behavioral therapy, in-person mindfulness therapies, and pharmacotherapy on depression in meta-analyses that correct for publication bias (roughly 0.5 SDs, per Cuijpers et al., 2013; Goldberg et al., 2021), they are remarkably large given the low cost of the app and suggest that it is highly cost-effective.⁶

Second, participants who received access to the meditation app earn an average of 1.9 percent more in a proofreading task, a 0.13 SD increase ($p < 0.01$). In this task we pay participants to

³Most participants receive the license on a Saturday and incentives for usage during the next two weeks.

⁴Mindfulness meditation is different from other health behaviors in that one can learn meditation skills quickly and apply them without keeping using an app, and they can, more generally, learn skills of emotional awareness and regulation that can be impactful beyond the time spent learning them through meditation.

⁵This improvement occurs incrementally over the course of two weeks, suggesting the overall effects on mental health are unlikely to be driven by reciprocity toward the experimenter or experimenter demand effects. We discuss demand effects in Section 4.5.

⁶A limiting factor in interpreting such comparisons is that the RCTs of in-person psychotherapy often include a passive placebo intervention, which was not easily implementable in our remote field setting. Additionally, clinical effects are often measured in the longer term than in our study. Finally, our study population is neither clinical nor a patient population, and meta-analysis suggests that treatment effects are likely to be smaller for such groups with more severe mental illness (Cuijpers et al., 2013).

identify simple spelling and grammar errors in paragraphs of text with no time pressure,⁷ measuring a dimension of productivity that is essential in many workplaces. However, we find no effect on performance in a time-limited Stroop test measuring the cognitive ability to ignore visual distractions (95% confidence interval: $[-0.7, 1.3]$ percent change in earnings). One interpretation is that access to the app specifically improves abilities to sustain effortful attention required by proofreading, but not the more basic cognitive control tested by the Stroop task. A 1.9% improvement in this kind of attention may seem economically modest; however, the intervention’s costs were also quite modest.⁸ In addition, participants in online experiments often show low elasticities of effort (DellaVigna and Pope, 2018), limiting the effect size we would expect to observe. A low-cost modest improvement in a widely applied skill like attention control may produce large cumulative effects across tasks and time.

Our third main finding is that mindfulness meditation mitigates the interference of emotions with decision making. An important cost of worries and emotions lies in their tendency to make choices inconsistent (Johnson and Tversky, 1983; Loewenstein and Lerner, 2003; Raghunathan et al., 2006; Slovic and Peters, 2006; Callen et al., 2014). We investigate the relationship between mindfulness, emotion, and choice by randomizing some participants to engage with stressful thoughts and then presenting all participants with incentivized decisions. One decision elicits an indifference point between two lotteries, and the other elicits an indifference point between a lottery and a sure payment. Comparing these indifference points allows us to calculate the *certainty premium*, which quantifies how an individual’s risk aversion varies when choosing between stochastic versus certain prospects. Past work shows that the certainty premium responds to stress-like emotion (Callen et al., 2014), making it a useful metric for studying emotions and decision making.

In the Waitlist group, inducing stress reduces the certainty premium by 0.13 SD ($p < 0.05$),⁹ contrasting with Callen et al. (2014)’s finding that recalling violence increases the certainty premium. This may reflect natural heterogeneity in risk preferences between our study populations and differing lottery stakes between our experiments.¹⁰ However, our focus is not to document *how* risk-taking reacts to emotions but rather test *if* it does and whether mindfulness mitigates such interference. While stressful thoughts reduce the certainty premium in the Waitlist, they have a weak, positive effect on the certainty premium in the App Access group (0.07 SD increase) that is distinct from its effect on the Waitlist group ($p < 0.05$ for the difference). This comparison averages across additional cross-randomization that Section 2.5 discusses in more detail, but put simply, gives evidence that mindfulness reduces and potentially reverses the effects of stressful emotions on choice. To the extent

⁷Pay increases linearly in correctly identified errors and decreases by the same amount in words incorrectly flagged.

⁸A one-month subscription for the app costs \$13, and the median participant in the treatment group spent 74 minutes using the app in the two weeks before the proofreading task.

⁹This is driven by an increase in risk aversion in choices between lotteries (0.15 SDs, $p < 0.05$). We find no effect on risk aversion in choices between a lottery and a sure payment (0.03 SD increase, $p > 0.5$), which is qualitatively consistent with Callen et al. (2014)’s finding that emotions affect choice under uncertainty but not certainty.

¹⁰In particular, Callen et al. (2014) study Afghan civilians. They note that certainty premia vary by educational attainment in their sample, where the average participant completed roughly 10 years of school. By contrast, roughly 80% of our sample of U.S. adults have earned a Bachelor’s degree. In addition, our lottery stakes range from \$0 to \$30, while theirs are higher, ranging from one to three days’ wages.

that these emotions are not a part of true or long-run preferences, reducing emotional interference is welfare enhancing (Bernheim and Rangel, 2009). Regardless of the normative implications, we believe this is the first piece of revealed-behavior evidence that meditation affects the interference of emotions with decisions.

In contrast with the positive medium-term impacts of access to the app, we find evidence that a mindfulness meditation session can have short-lived adverse effects in decisions and productivity when negative emotions are involved. To study the *immediate* effects of a meditation session, we incentivize a subset of the participants who received access to the app to complete a meditation session immediately before the effort and decision-making tasks.¹¹ This prompt to meditate reduces earnings on the proofreading task when participants are first asked to describe worrying memories, but not otherwise. It also increases participants’ unpleasant information avoidance and small probability loss aversion. None of these effects are present in the group that had access to the app but no incentive to meditate right before the tasks, demonstrating that experiments studying only the immediate effects of mindfulness meditation are insufficient to conclude about the longer-term, sustained effects of the practice. These findings may also caution organizations against encouraging meditation sessions as a default before important decision-making events.

We leverage this rich set of results to propose a simple framework for modeling mindfulness meditation with economic primitives in the style of Gabaix (2019). An agent must rationally allocate her attention between a payoff-relevant variable and an irrelevant emotion.¹² The emotion distorts the agent’s utility function and shifts her rational attention allocation away from the relevant dimensions of the task at hand. We model mindfulness meditation as temporarily overriding this process, allocating all of the agent’s attention to her mental state, which helps her realize that emotions do not bear consequences on the outcomes of many situations. In the long run, this sophistication reduces the interference of unrelated emotions with decisions and increases attention available for the task at hand, in line with our main findings. During or immediately after meditating, however, the agent is more attentive to her emotions and is therefore distracted from the task at hand. Consistent with our results for the short-term effects of a meditation session, the agent is also more likely to avoid making decisions that could result in negative emotions, such as seeking potentially unpleasant information.

In studying the economic implications of online mindfulness meditation, this paper investigates

¹¹One concern is that this treatment causes other differences in the survey-taking experience, such as participants postponing the survey when they learn that they should complete a meditation session before it. To limit such effects, we told all participants to plan for up to an hour of time (much longer than actually needed) to be able to take this survey in one sitting in a quiet environment. When we announced incentives for a meditation session, we made it clear that the bonus would only be granted if the session was completed within 30 minutes of seeing the announcement.

¹²We do not assert that all emotions should be ignored, as they are often meaningful signals (e.g. a “gut feeling” as a summary of information). Our framework specifically considers emotions that alter a decision-maker’s understanding of the consequences of their decision (a “characterization failure” in the words of Bernheim and Taubinsky (2018)). We find it reasonable to assert that some agents may “prefer” to make some choices without the influence of certain emotions, especially when the emotions are disconnected from the choice at hand. For instance, a CEO may want to make her management decisions independently of potentially stressful family issues. If this assumption seems strong, our framework can also be interpreted descriptively as capturing agents who “choose” to exclude emotions from their preferences (Bernheim et al., 2021), with no normative consequence.

a number of relevant hypotheses. We take steps to mitigate concerns about spurious findings arising from multiple testing. First, we adhere closely to our pre-analysis plan and describe minor deviations from it in Section B.1. This includes reporting results of our main analyses, regardless of whether treatment effects are large (e.g., on mental health) or small (e.g., of incentives on long-run usage), and statistically significant (e.g., on proofreading) or insignificant (e.g., on the Stroop test). Second, we explicitly correct p -values for multiple testing in Section B.6 and find that our main claims stand up to stricter statistical scrutiny.

Our work contribute to several literatures. First, we evaluate the effects of mindfulness meditation on economic behavior and provide experimental evidence on both the immediate and sustained effects of meditation. Our experiment is larger by an order of magnitude than recent studies with incentivized decision-making and is the first to evaluate both the immediate and sustained effects of mindfulness meditation (Shapiro et al., 2012; Hafenbrack et al., 2014; Iwamoto et al., 2020; Ash et al., 2021; Alem et al., 2021; Cassar et al., 2022). More broadly, we contribute to a growing body of knowledge on the effects of therapy-inspired psychological interventions on economic outcomes (Blattman et al., 2017; Heller et al., 2017; John and Orkin, 2019; Baranov et al., 2020; Lund et al., 2021; Bhat et al., 2021). Relative to this economic literature, we study a light-touch, self-directed intervention based on another approach to clinical psychology—mindfulness—and focus on a more general population.

Second, our paper shows how a simple intervention to train attention and emotional control can help improve work productivity, and it is the first to study the impacts of meditation on incentivized measures of performance in an adult population. Previous work argues that distractions and worries can reduce work or test performance (Banerjee and Mullainathan, 2008; Kaur et al., 2021; Duquenois, 2021). Our paper shows that inexpensive interventions can actually boost productivity by helping adults sustain their attention to the task at hand. We complement recent findings that practicing attention-based tasks can improve school performance among children (Brown et al., 2022) and college students (Cassar et al., 2022).

Third, we contribute to the literature on the role of emotions in economic decision-making, which shows that experimentally induced irrelevant emotions can distort decision-making (Johnson and Tversky, 1983; Loewenstein, 2000; Loewenstein et al., 2001; Loewenstein and Lerner, 2003; Lerner et al., 2004, 2015; Slovic and Peters, 2006; Callen et al., 2014). With the assumption that emotions are skewing individuals’ evaluation of the consequences of their decisions (Bernheim and Rangel, 2009; Bernheim and Taubinsky, 2018; Bolte and Raymond, 2022), reducing their interference with decision-making is welfare-improving. We provide the first experimental evidence on a viable way to protect from such interference. This result also complements a vast psychology literature articulating the mechanisms of meditation in terms of improved emotion regulation, relying on self-reports and neuroimaging (Lindsay and Creswell, 2017).

Fourth, we contribute to a growing body of empirical findings on the effects of digital technologies on mental well-being. We study an increasingly popular digital technology that has positive effects, in contrast with recent work examines the harmful effects of technologies on well-being (Twenge,

2020; Twenge et al., 2020; Allcott et al., 2020, 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 most 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 of the existing experiments, and we collect incentivized measures of productivity and decision-making in addition to mental well-being. Finally, our sample is not selected based on high initial symptoms of anxiety or depression nor is it restricted to a student or employee population from any given specific 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 gives an account of our experimental design, including details about the mindfulness intervention and definitions of key outcomes. 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 productivity and decision making. We then present a model that motivated our design and helps interpret our findings in Section 6. While we place this material after our empirical findings, the curious reader can also peruse it before Section 2 for a mental model of mindfulness. Finally, Section 7 concludes with a discussion of our findings and open questions.

2 Experiment

2.1 Recruitment of Participants

The practice of mindfulness meditation takes effort, time, and regularity. Measuring its effects requires recruiting participants who will take up the intervention and retaining them for several weeks. After all, with an encouragement design we can only estimate local average treatment effects for people whom we induce to meditate (Angrist and Imbens, 1995). Our recruitment strategy therefore favors individuals who are genuinely interested in trying meditation to maximize take-up. To minimize post-randomization attrition, we assign participants to an intervention arm after they have completed three distinct surveys on separate days, demonstrating the willingness and ability to participate in a longer study.

Participants are recruited using Facebook and Instagram ads that specify that the study is about meditation, and it offers the opportunity to obtain a free subscription to the paid version of the Headspace app as well as compensation in the form of gift cards (see Figure B.1 for examples of such ads). Recruitment occurred between July 7 and August 24, 2021. Ads were shown 2,158,678 times to 1,131,495 unique US adults and were clicked on by 32,432 individuals.¹³ All told, 10,615 respondents completed our screening survey.¹⁴

¹³These clicks include Facebook reactions (e.g., “liking” the ads), comments, visits to the associated Facebook page, and opening the link to the screening survey.

¹⁴Seventy-six percent of the screening survey responses resulted from directly opening one of our Facebook links. The remainder followed the survey link in another manner, such as copying and pasting it into a web browser.

The screening survey included a consent form to verify that participants were over 18 and lived in the US, a demographic questionnaire, and other baseline information. In addition, it presents a detailed layout of the study, including the information that they would either receive free access to Headspace at the beginning of the study or after a follow-up survey four weeks later. We verified comprehension and agreement to the basic requirements of the study, such as completing seven surveys over the course of four weeks, owning a smartphone and being able to install an app on it, having time to use the app on a regular basis, and having an email and a phone number to receive surveys, messages, and compensation.

There were 7,033 participants who consented, finished the screening survey, and provided unique contact information. To remove inattentive participants, we exclude those who spent less than 20 seconds on the study presentation page. Our study ads might draw people who already consistently meditate and simply want a subsidized Headspace subscription. Therefore, we further exclude participants who report already practicing meditation consistently or having previously tried Headspace.¹⁵ As our experiment is funded by agencies that focus on improving the well-being of low- to middle-income Americans, we exclude participants reporting an income per household member above the US mean (\$54,000 in 2019; see US Census Bureau, 2020).¹⁶

This screening procedure resulted in a sample of 3,356 participants. We invite these participants to complete a second baseline survey by SMS message with two main objectives: screening out unresponsive participants or those who had entered an incorrect phone number, and collecting baseline measures of several outcomes. This includes measures of mental health, subjective well-being, willingness to pay for an extension of the Headspace license beyond the three months offered in the study, and incentivized beliefs about the effects of meditation on other participants. A total of 2,768 participants completed this second survey.

On the following weekend, we email participants to complete the third and final baseline survey. This survey reminds participants of the study structure and collects additional baseline variables to improve the precision of treatment effect estimates: incentivized baseline measures of their performance on a proofreading task and a Stroop test (a standard cognitive test of selective attention capability; see Jensen, 1965), a psychometric scale of mindfulness (FFMQ-15; see Baer et al., 2008), and self-reported risk, social, and time preferences (Falk et al., 2018). A total of 2,384 participants completed the last baseline survey and were randomized into a treatment arm.

2.2 Random Assignment to Free App Access and Incentives

At the end of the last baseline survey, we randomize participants into one of five groups with equal probability: (i) Headspace No Incentive, (ii) Headspace Short Incentives, intended to encourage at

Proceeding through the study requires a functioning email and receiving SMS messages at a US phone number, giving us confidence that our participants are unique and live in the US.

¹⁵Participants might anticipate this and avoid disclosing their meditation habits, but this should, if anything, reduce measured treatment effects.

¹⁶We chose income per household member to be more inclusive of parents with children at home, who have been particularly hit by COVID-19-related stay-at-home orders (Cuadrado et al., 2021). The income screening criterion was not enforced at the beginning of the recruitment period (first 107 randomized participants).

least some experimentation with the app among most users, (iii) Headspace Long Incentives, intended to generate lasting habits through frequent initial usage, (iv) Pure Waitlist Control, or (v) Waitlist Control Cash Transfer, to compare effects to that of offering a cash value equivalent to the cost of Headspace. Randomization is stratified within eight strata based on age, baseline anxiety score, and baseline willingness to pay for an extension of the Headspace license.

2.2.1 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 instruct practitioners to direct their focus on a sensation, such as the breath, to notice how other thoughts and sensations capture attention and to refocus back to the initial object. The approach was quickly extended to treating a variety of mental health issues and led to the creation of a set of mindfulness-based interventions. Two leading examples are Mindfulness-Based Stress Reduction (Kabat-Zinn and Hanh, 2009) to tackle chronic stress and Mindfulness-Based Cognitive Therapy (Segal et al., 2018) to treat depression. These interventions have been extensively evaluated in research designs of increasing quality, including large-scale RCTs (Kuyken et al., 2015; Segal et al., 2020), which has led the American Psychological Association Society of Clinical Psychology to list Mindfulness-Based Cognitive Therapy as an evidence-based treatment for depression with strong research support (American Society of Clinical Psychology, 2019).

As the evidence for efficacy of these therapies grew, clinical psychologists became interested in understanding the mechanisms through which mindfulness-based therapies operate. This interest spawned a variety of characterizations and measurement scales exist, (Davidson and Kaszniak, 2015), reflecting that mindfulness meditation may affect many mental processes.¹⁷

The proposed definitions do 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 attention control, that is, directing one’s attention to certain objects by choice rather than letting it be captured by distractions. It also suggests that mindfulness meditation increases meta-cognition, or one’s knowledge about how attention works and is captured by 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.¹⁸ Third, *present experience* is what attention should targeted: 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, focused on training its users in the skill of mindfulness. It ranked among the top five health and wellness smartphone apps in Google and Apple app store

¹⁷Notable articles from the clinical psychology literature attempting to unify theoretical foundations and produce a definition of mindfulness include Kabat-Zinn (2003), Baer (2003), Bishop et al. (2004), and Brown et al. (2007), which have collectively been cited over 20,000 times as of October 2022

¹⁸This capacity to limit focusing or avoiding attention to certain aspects of the experience based on emotions is often labeled as *acceptance* or *non-reactivity* (Baer, 2003).

revenues in 2020, and reports over 70 million users worldwide. 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 recordings and short videos, often grouped into series or themes. Its core offering—and 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.¹⁹ Crucially, users choose which sessions to engage in and when, in contrast with traditional therapies that require scheduled meetings with a healthcare professional or trainer.

2.2.2 App Treatment and Usage Incentives

We randomly assign three-fifths of participants to receive a free three-month subscription to Headspace in the form of a voucher code immediately at the end of the third baseline survey. Participants are advised to start with a specific series called “Basics” that trains novice users in the technique of mindfulness meditation.

Participants who receive Headspace right away are further randomized into one of three groups with equal probability: (i) Headspace No Incentive, (ii) Headspace Short Incentive, or (iii) Headspace Long Incentive. Those in the first group receive their voucher code and instructions to use it. Participants assigned to the second or third groups are further told that they will earn an additional \$10 bonus if they meditate using the app for at least 10 minutes on at least 4 separate days (Short Incentives) or at least 10 separate days (Long Incentives) over the first two weeks of the study. These incentive treatments are meant to generate further variation in the app usage and to test for the formation of habits. The Headspace Short Incentive steeply increases the value of trying the app for a few days and is designed to encourage experimentation among a broad set of participants, including skeptical ones who believe the app will deliver fewer benefits. The Headspace Long Incentive may increase usage among fewer participants but encourages frequent usage of the app over the first two weeks and may create longer-lasting habits.

2.2.3 Waitlist Control Group and Cash Transfer

As we promise a Headspace license to all participants during recruitment, we assign the remaining two-fifths of participants to a waitlist control group and tell them that they will receive a voucher for Headspace Plus at the end of the final survey, sent four weeks after randomization. We ask these participants to refrain from creating a Headspace account in the meantime and to avoid starting a new meditation practice or taking up new habits on their own while waiting for their Headspace license.²⁰

¹⁹See Appendix B.5 and particularly Figure B.8 for details

²⁰We cannot be sure about what waitlist participants do during this period. We do ask them to prioritize their well-being over the requirements of the study and that they should start any practice that they feel a need to start. They may take up mindfulness meditation or other wellness activities outside the app, which would reduce observed treatment effects and is thus not a cause of concern for the robustness of our findings.

We assign half of the waitlist participants to Pure Waitlist Control, with no subsequent intervention. To test whether any effects from offering the app access may be generated by reciprocity or any kind of wealth effect due to the value of the app, the other half of the control group, the Waitlist Control Cash Transfer, additionally receives a \$15 cash transfer in the form of a highly fungible gift card (valid at Walmart, Amazon.com, and over a hundred other retailers). This amounts to a little more than the cost of a Headspace license for four weeks.

2.3 Sample Composition, Balance, and Attrition

Table 1 presents demographic characteristics of the randomized sample of participants and shows comparable statistics in the general US adult population when available. Our sample differs from the general population in a few important ways. Not unexpectedly given our social media recruitment, only 2 percent of our participants are above 70 years old, but our sample includes meaningful numbers of participants from all other age groups. Our screening criteria also produce a sample with very few participants having household incomes over \$150,000, or with more children in the household than average.

By selection into our ads, our sample is composed of a majority of females, is more educated than the general population, and is more likely to politically identify as a Democrat or independent. Individuals identifying as Black or Hispanic are also under-represented in our sample relative to the general population. Thirty-three percent of our randomized 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 of the sample begins our study with less than moderate anxiety or depression.²¹

Table 1 reports estimated differences between the Waitlist and App Access groups obtained by regressing each variable on a treatment dummy. Out of 26 baseline covariates, three have small but statistically nonzero imbalance at conventional significance levels. We view these imbalances as small in absolute terms and our preferred regression specifications will not adjust for them. This is consistent with our pre-analysis plan and avoids complicating inference with *post hoc* model selection. That said, our findings are robust to adjusting for observed imbalances. In Appendix B.3 we employ the double/debiased machine learning approach of Chernozhukov et al. (2018) to flexibly adjust for potential imbalance, finding that point estimates and standard errors are relatively unchanged.

The vast majority of participants responded to our followup surveys. Table B.2 presents sample size for each of our surveys. At the main two- and four-week endlines, the attrition rates in each arm are 6.5 percent or less. That said, individuals the App Access group are slightly less likely to participate in the main endline surveys (1-3 percentage points, $p < 0.05$). We explore attrition more thoroughly in Appendix B.4, finding that the attrition does not result in problematic differences between treatment and control respondents. This implies that differential attrition is unlikely to introduce large biases in our main analyses.

²¹We describe our elicitation of these symptoms in Section 2.4

2.4 Measures of Mental Health and Subjective Well-Being

Throughout the study, we use standard, validated psychometric questionnaires to measure symptoms of generalized anxiety disorder (“anxiety”), major depressive disorder (“depression”), and stress. Our main measures of mental health are collected two and four weeks after randomization, but participants also complete shorter questionnaires at baseline and shortly after randomization.

Anxiety. We measure anxiety using the Generalized Anxiety Disorder 7-item scale (GAD-7; see Spitzer et al., 2006) as well as the shortened 2-item version (GAD-2; see Plummer et al., 2016). 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 between 0 and 21 for GAD-7. These scales are workhorses in psychology, demonstrating high test-retest reliability (Spitzer et al., 2006) and the ability to predict diagnoses from more thorough interviews with mental health professionals (Plummer et al., 2016).

We collect the full GAD-7 score at baseline, after informing participants that all screening based on survey responses has already occurred, and again at our main followups at 2 weeks, 4 weeks, and 3–4 months. To sketch the trajectory of anxiety in the short term, we administer the GAD-2 scale in brief surveys 4, 7, and 11 days after randomization. Using the shorter scale avoids exhausting participants with long set of questions too many times and because these intermediate measures are primarily intended to capture the rate of improvement rather than to precisely assess treatment effects.

Depression. We measure depression with the Patient Health Questionnaire, which is similar in spirit to GAD-7. Depression is less prevalent in the general population than anxiety, so we limit our measurement of it to reduce the burden on participants. At baseline, we administer the two-item PHQ-2, and at 2 and 4 weeks we administer the longer PHQ-8 (see Arroll et al. (2010) and Wu et al. (2020) for validation of these scales).

Stress. We measure stress—the degree to which participants find their lives to be unpredictable, uncontrollable, and overloaded—using the Perceived Stress Scale (PSS-10; see Cohen et al., 1994). Unlike GAD-7 and PHQ-8, this scale is not systematically used to screen for mental health issues in clinical settings, but it 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.

2.5 Experimental Session with Cross-Randomized Treatments: Immediate Meditation Incentives and Stressful Tasks

Two weeks after randomization, participants complete an online experimental session that contains our main economic outcomes. Figure 1b presents the outline of experimental tasks. The session starts by incentivizing a random subset of participants with Headspace access to meditate using the app. It then continues with an incentivized Stroop cognitive test that measures attention control, decisions to acquire or avoid information, and a decision to take a risky gamble with salient

low probability losses. Participants are then randomly assigned to a task that induces neutral or stressful thoughts, before completing an incentivized proofreading task and making a set of decisions between lotteries.

We designed the sequence of tasks to start with outcomes that have built-in distracting or emotion-inducing elements and to finish with outcomes where emotional interference is induced by an external experimental treatment. We pre-specified the Stroop test, the proofreading task, and the final set of risk-taking choices as our primary outcomes.

2.5.1 Immediate Meditation Incentives Treatment

To disentangle the immediate effects of meditation from longer-term, sustained effects of regular practice, the survey starts by randomizing participants who already received access to the app into being offered a bonus to pause the survey and complete a meditation session right away. We do this for two main reasons. First, findings from this treatment are policy relevant for organizations considering implementing or encouraging mindfulness meditation session during the workday, before meetings or decision-making events. Second, at a methodological level, it is easier to conduct experiments in one day, experimentally assigning a meditation session at the beginning of a lab session where outcomes are observed. However, no evidence exists to assess the validity of extrapolating such short-term findings to hypothesized longer-term effects of sustained meditation practice.

One concern is that this treatment causes other differences in the survey-taking experience, such as participants postponing the survey to a later, more quiet time when they learn that they should complete a meditation session. To limit such effects, we told all participants to plan for up to an hour of time (much longer than the time actually needed for the survey itself) to be able to take this survey in one sitting, in a quiet environment. When we announced incentives for a meditation session, we made it clear that the bonus would only be granted if the session was completed within 30 minutes of seeing the announcement.

2.5.2 Stressful Treatment: Induction of Unrelated Emotions

In addition to affecting how individuals react to emotions *associated with* decisions such as acquiring information or accepting small probability losses, mindfulness meditation may, first and foremost, affect the management of emotions and worries *unrelated to* the task at hand. To investigate this, before our main productivity and decision-making outcomes, we experimentally assign all participants to one of two conditions: Neutral or Stressful. The Neutral acts as a control condition: participants are asked to describe something that they do to stay grounded, for example, a routine. The Stressful condition asks participants to remember vividly and describe a memory that makes them feel anxious and is unresolved. In both cases, to ensure high engagement with this treatment, answers are incentivized for thoughtfulness: participants are told that some answers will be randomly selected and evaluated by an independent reader. If the answer is deemed thoughtful and personal, an additional bonus will be awarded.

Later in the experiment, participants receive a “second dose” of the Stressful treatment: They are asked to describe how they would handle an unexpected medical bill. In the Neutral condition, the ambulance bill is only \$100, while in the Stressful condition it is \$8,900. As a manipulation check, and to better understand how mindfulness meditation impacts affective responses, we collect a measure of self-assessed mood at the beginning of the survey (after the Immediate Meditation Treatment) and after each dose of the Neutral or Stressful treatment.

2.6 Measures of Productivity and Cognitive Function

Our main real effort outcome is a proofreading task, which participants complete after the first dose of the Stressful or Neutral treatment. Participants proofread three paragraphs of text with a total of 17 spelling or punctuation errors. There is no time constraint, and they earn five cents per correctly highlighted error but lose five cents if they highlight correct words. Participants begin with an endowment of 20 cents, so that initial mistakes are costly. The median participant spends two minutes on the task and earns 90 cents from it. A practice paragraph presents these incentives clearly to the participants. To better capture effects from the Stressful condition, the instructions and practice paragraph are completed before the Stressful or Neutral task so that participants immediately start the incentivized portion of the proofreading task after they complete the elicitation of potentially worrying thoughts.

Estimating effects on productivity with high external validity is challenging in our online setting. We recruit participants with a variety of skills and backgrounds, and survey them with a general-purpose survey software. This proofreading task allows us to capture differences in attention to detail, a dimension of productivity involved in many real jobs. Its goal is simple to communicate and requires no complicated training, making it a natural part of a longer survey where we also explain and collect other measures.

One salient mechanism that could be responsible for improvements in work performance is cognitive ability. Because mindfulness meditation is hypothesized to train cognitive functions related to attention control (Jha et al., 2007), participants complete an incentivized cognitive test designed to measure them Dean et al., 2019: the Stroop test (Jensen, 1965). Words are displayed on the screen for three seconds, and the participant must select among several options the name of the color that the word is displayed in. The words themselves are color names that must be ignored even though they capture attention. We incentivize the task by paying participants a bonus that increases with how much time is left for each time they click the right color to reward both speed and accuracy.

2.7 Interference of Emotions with Risk Decisions

After the second dose of the Neutral or Stressful treatment, participants complete a series of binary choices between lotteries. This task’s main objective is to study whether mindfulness meditation affects the previously documented propensity of emotions to interfere with decision-making, in particular in choices between risky prospects (Loewenstein and Lerner, 2003; Lerner et al., 2004; Raghunathan et al., 2006; Callen et al., 2014). This literature finds that experimentally induced

emotions affect risk choices, which we attempt to replicate in the control group. We then study whether the effects are different among participants who received access to the app.

We therefore follow Callen et al. (2014) and elicit two menus of binary choices between lotteries. The first one is designed to estimate the probability P_{certain} that a participant is indifferent between a lottery that pays \$30 with P_{certain} and \$0 with $(1 - P_{\text{certain}})$, against a guaranteed \$10. The second menu estimates $P_{\text{uncertain}}$ such that the participant is indifferent between a lottery that pays \$30 with $P_{\text{uncertain}}$ and \$0 with $(1 - P_{\text{uncertain}})$ or a lottery that pays \$30 or \$10 with equal probabilities. The survey questionnaire prevents participants from submitting a set of answers such that the lottery is accepted for a given P and then rejected for a larger P . Participants also must choose the lottery with $P = 1$ and the alternative with $P = 0$. Doing so allows us to keep all participants in the analysis with well-defined switching points on each menu, and gives participants who might be confused another chance to understand the instructions. P_{certain} and $P_{\text{uncertain}}$ are computed as the midpoints between values of P where the participants switch from the alternative to the lottery.

These two outcomes are not meant to be compared directly. Rather, as in Callen et al. (2014), we derive the implied certainty premium. Assuming expected utility, and assuming that the utility values of \$30 and \$0 are constant across choices, we can derive the utility values of \$10 implied by P_{certain} and $P_{\text{uncertain}}$, $u_{\text{certain}}(10)$ and $u_{\text{uncertain}}(10)$. We can then define the certainty premium as $CP = u_{\text{certain}}(10) - u_{\text{uncertain}}(10)$. Under expected utility, $CP = 0$, but Callen et al. (2014) show that $CP > 0$ for most individuals in their sample. They also find that experimentally induced recollection of violent events decreases $P_{\text{uncertain}}$, marginally decreases P_{certain} , and increases CP . We test for similar effects by experimentally inducing stressful thoughts in our waitlist control group and to study whether the effects are any different among participants who received access to the app.

2.8 Sensitivity of Decisions to Emotional Consequences

Prior to receiving the Neutral or Stressful treatment, participants complete two sets of decisions which may generate negative emotions that weigh in the chosen option: decisions to receive or avoid potentially unpleasant information, and to accept a high expected-value lottery with low-probability, salient losses. Both of these outcomes were pre-registered as secondary, and are described along with corresponding results in Appendix B.2. We also document how these outcomes help us narrow down the likely mechanisms in Section 6.

3 App Usage and Effects of Incentives

Headspace collects and shares with us detailed administrative usage data from each account associated with a code given in the study.²² The data contain information on the name, starting time, and duration of each recording users listen to while their voucher was active. In this section, we

²²This was explained to participants in the consent form and subsequently. This is the only data that Headspace collected about our study participants.

describe usage statistics of participants over time. We find that usage is high even among participants assigned to the Headspace No Incentive group, perhaps unsurprisingly given our recruitment of interested participants. This gives us the opportunity to study the effects of offering the mindfulness meditation app in a best-case scenario where individuals do engage with it at a high rate. We then show that the Headspace Short Incentives and Headspace Long Incentives treatments generate additional usage in the period of time when they are active, without any lasting effects on usage once they expire.

App Usage without Incentives. Figure ?? displays the proportion of participants in each treatment arm who use the app at least once within a sliding three-day window, and Table 2 presents usage statistics. Overall, most participants use the app frequently at the beginning of the study period even without incentives, with 80.3 percent of participants in the Headspace No Incentive group recording at least one session and spending an average of 95 minutes on the app during the first two weeks. The frequency of usage then declines rapidly, but 56 percent still use the app at least once between the third and the eighth weeks after randomization (Table 2, row 6, column 1).

Short-Term Effects of Incentives. The incentives we offer increase usage of the app, especially so when incentives require more usage before participants qualify for payment. Table 2 reports a series of usage statistics in the different intervention arms. We estimate effects of receiving any usage incentive using the following regression:

$$Y_i = \beta_1 + \beta_2 \text{AnyIncentive}_i + \delta_{\text{stratum}} + \epsilon_i, \quad (1)$$

where Y_i is the usage outcome for individual i , AnyIncentive_i is an indicator for receiving any incentives to meditate in the two weeks after randomization, and δ_{stratum} are fixed effects for randomization strata. We calculate heteroskedasticity-robust standard errors using the HC1 estimator.

To separately estimate the additional effect of the Headspace Long Incentives relative to the Headspace Short Incentives, we extend Equation (3) to

$$Y_i = \beta_1 + \beta_2 \text{AnyIncentive}_i + \beta_3 \text{LongIncentive}_i + \delta_{\text{stratum}} + \epsilon_i, \quad (2)$$

where LongIncentive_i is an indicator for being eligible for the higher powered Headspace Long Incentives, which require 10 days of usage in the first two weeks after randomization.

Incentives do significantly increase usage, by any measure: they increase the proportion of individuals who go on to try the app at least once by about 11 percent and increase the number of sessions and the total time spent on the app in the first two weeks by 36 percent and 50 percent (Table 2, column 5). As expected, most of the increase in the propensity to try the app at all can be achieved with the easier-to-get Headspace Short Incentives, although they do not increase this extensive margin any more than the Headspace Long Incentives, contrary to our hypothesis. On the other hand, as we expected, the Headspace Long Incentives further increases usage on the intensive margin.

No Sustained Effects of Incentives. While incentives significantly increase app usage while

they are active, they do not result in any sustained app usage once they expire. Figure ?? captures this trend. In Table 2, we report statistics of usage beyond the incentive validity period and find no detectable increase in usage from past incentives apart from a small increase in the extensive margin, likely reflecting that more people installed the app in the first place.

This result contrasts with previous findings that temporarily incentivizing certain health behaviors, such as going to the gym or washing hands, generate lasting habits (Charness and Gneezy, 2009; Acland and Levy, 2015; Hussam et al., 2019). Our preferred explanation is that unlike exercise or hand washing, which require continued practice to reap rewards, using a meditation 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 emotion regulation in daily life. The data suggest that this is the case: in a 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.

4 Effects on Mental Health

4.1 Empirical Framework

To estimate the impact of making mindfulness available on mental health outcomes, we focus on intent-to-treat (ITT) estimates. Panel A of Table 3 reports results of a regression of the form

$$Y_i^{post} = \beta_1 \text{App Access}_i + \beta_2 \text{AnyIncentive}_i + \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i. \quad (3)$$

where Y_i^{post} is the outcome for individual i , App Access_i is an indicator for receiving a Headspace license, AnyIncentive_i is an indicator for receiving any incentives to meditate in the two weeks after randomization, δ_{stratum} are fixed effects for randomization strata, and Y_i^{pre} are pre-randomization measures of the outcome, where available. We calculate heteroskedasticity-robust standard errors using the HC1 estimator.

This approach pools our two incentive arms (Headspace Short Incentives and Headspace Long Incentives). As we discuss below, the Long Incentive arm has a minimal marginal effect beyond the Short Incentives. Pooling the arms provides a succinct summary of the effect of incentives and improves the precision of the estimate. For completeness and in line with our pre-registration, we also report estimates of effect from each incentive treatment in Panel B of the table.

4.2 Effects of Offering App Access on Mental Health

Before diving into treatment effect estimates, we first sketch the evolution of anxiety over time. Figure 3 presents average GAD-2 score in each treatment arm as reported in each survey. During the first 30 days after randomization, anxiety declines slightly for participants on the waitlist. The App Access groups report steady improvements in anxiety, starting at 4 days and accumulating gradually for the next 30.

These improvements are captured more precisely by our detailed questionnaires about symptoms of anxiety, depression, and stress. Table 3 reports the treatment effects that we observe on standardized scores. App access alone leads to an improvement of 0.37 SDs in an index of the three measures, with effect in the 0.33 to 0.39 SD range across the outcomes. These effects become larger at four weeks, at 0.44 SDs for the index and 0.38 to 0.47 SDs on each outcome. Once the Waitlist group gains access to the app they report reduced anxiety, and the App-Waitlist gap shrinks to 0.22 SDs at the three to four month mark.²³ The persistent effect of initial app access are consistent with higher uptake among the treatment group. In sum, we find that encouraging app-based mindfulness meditation improves self-reported mental health for at least three months after intervention.

The effects we measure are not artifacts of treating mental health measurements as continuous variables, nor are they due to improvements among only certain subgroups. 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. Table B.5 shows treatment effects on the proportion of participants who report at least mild, moderate or severe symptoms of anxiety and depression, as defined by the 5-points, 10-points and 15-points cutoffs in GAD-7 and PHQ-8. After App access reduces the fractions of participants with at least mild, moderate and severe anxiety by 32 percent, 50 percent and 64 percent, and the fractions with at least mild, moderate and severe depression by 24 percent, 47 percent and 65 percent. Effects at all levels are highly statistically significant.

The standardized effect sizes 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. Given the limitations of many of the existing studies emphasized by Sommers-Spijkerman et al. (2021), and the small number of studies evaluating popular, unstructured app such as Headspace, our estimates are a significant contribution toward building the evidence base for these interventions.

The effects are also smaller but comparable to previously measured effects of expensive in-person therapy. For instance, Cuijpers et al. (2013) conduct a meta-analysis of in-person cognitive behavioral therapy (CBT) interventions and estimate an effect size of 0.53 SDs on depression.²⁴ The effects of pharmacotherapy alone are also estimated as similar to CBT by Cuijpers et al. (2013). The effects we find are also about ten times larger than previously estimated effects (0.05 SDs) from receiving health insurance on depression in the Oregon Medicaid experiment (Finkelstein et al., 2012), although such comparisons are difficult since we measure short-term effects. 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, 2018).

²³We expect the actual effect in our study population to be smaller due to attrition. In our three-month followup, nonrespondents in the app access arm report worse initial mental health than nonrespondents in the Waitlist. It's reasonable to believe that the treatment participants who dropped out were those who experienced the fewest benefits. That said, our response rate at three months remains at 84%, and we do not expect attrition to fully explain lower anxiety in the treatment group

²⁴Their initial estimates of about 0.71 falls to 0.53 after they correct for publication bias or filter out low-quality studies.

4.3 Incentive Effects

The second row in both panels of Table 3 estimates the additional effect of receiving either of the short and long usage incentive schemes. Incentives only appear to moderately increase mental health improvements, if at all: the additional effect of incentives is only significant at the 16-day mark, where we estimate an additional 0.07 SD improvement to the index of mental health scores, or about 19 percent of the effect of app access alone.

This small effect may appear surprising given our finding that incentives increase the number of sessions and time spent on the app by 41 percent and 75 percent during the first 16 days.²⁵ One interpretation is that individuals are well equipped to know how much of the app to use on their own and that meditation is more effective when one *wants* to do it and applies oneself. In other words, it is more difficult to observe a meaningful measure of compliance than in other health contexts, such as taking steps, exercising, or washing hands (Charness and Gneezy, 2009; Acland and Levy, 2015; Hussam et al., 2019; Aggarwal et al., 2020). Testing incentives based on better measures of compliance than app usage only and testing the effects of incentives with participants who might be less interested than ours to begin with remain of interest for future work.

4.4 Heterogeneous Effects

Participants in our experiment are free to choose when and how they use the app, creating natural opportunities for heterogeneous treatment effects. We investigate possible axes of heterogeneity in Figure 6. Panel (A) shows that in the license-only group, improvements in mental health are largest for individuals who have more severe baseline symptoms, hold more optimistic beliefs about treatment effects, and score lower on a baseline measurement of mindfulness.²⁶ That said, participants with milder symptoms or more skeptical beliefs still experience large and statistically significant improvements in mental health. Any differences in these effects may be due to some combination of differential app usage and differential app effectiveness between the two. Finally, we note that although most of our sample is female, men in the license-only group experience comparable improvements in mental health.

Next, we note that participants who entered the experiment with higher levels of mindfulness benefit less from app access. We propose two mechanical reasons this may be the case. First, in exploratory analyses, we find that highly mindful individuals report lower levels of depression and anxiety, leaving less room for improvement in their mental health. Second, highly mindful individuals may have less to learn from the app, limiting the effectiveness of the lessons it provides.

We offer a tentative interpretation of these patterns. Larger treatment effects among optimists suggest that participants generally know whether meditation will be personally helpful. When left

²⁵Table 2 shows that incentives increase the extensive margin of trying the app at least once by approximately 10 percent, more in line with the size of the effect on mental health. Of course, we do not have a disciplined way to attribute the effects of incentives on mental health with either the intensive or extensive margin effect on usage or a combination of both.

²⁶Participants' baseline mental health and beliefs about treatment effects are uncorrelated with each other. Figure B.10 presents the relationship between the two.

to their own devices, more optimistic individuals meditate slightly more than others, and derive larger mental health benefits.

Given that app usage improves mental health, a natural question is whether incentives can induce individuals who would benefit from the app to use it. People who suffer from mental illness or are skeptical of the benefits are two such groups. Panel B of Figure 6 presents the marginal effect of receiving usage incentives compared to receiving only a license. Incentives increase app sessions and cause similar improvements in mental health for those with more and less severe anxiety, as well as those with more and less skeptical beliefs about its efficacy. However, our sample comprises individuals who are already interested in mindfulness meditation. While we present evidence that incentives induce further meditation without incurring negative effects in this group, we advise against naively extrapolating to individuals who are uninterested in meditation.

4.5 Robustness to Experimenter Demand Effects

Experimenter demand effects are an important concern when using self-reported measures (see, e.g., de Quidt et al., 2019), such as the mental health scales in this study. We begin by noting that the scales we use, PHQ-8 and GAD-7, are validated, clinical diagnostic tools used to measure the mental health effects of other interventions (Ruiz et al., 2011; Beard and Björgvinsson, 2014; Beard et al., 2016; Toussaint et al., 2020). However, we wish to take these concerns seriously, and now describe features of the data that suggest demand effects are unlikely to be the primary explanation for the treatment effects we measure.

An immediate possibility is that participants who receive the app may answer mental health questions optimistically. Perhaps participants are simply telling us what they think we expect to hear. With this in mind, we wrote our experimental instructions to mitigate such concerns. Before measuring mental health, we inform participants that our sole intention is to understand how participants were feeling.²⁷ Participants appear to take this seriously, and the right column of Figure 3 shows that many report short-term increases in anxiety. Across the five surveys we administer in the first two weeks, 46.5 percent of participants in the App Access groups and 56.4 percent of participants in the Waitlist report higher anxiety on one survey than on the survey immediately prior. Moreover, Figure 3 reveals that the gap between the App Access and Waitlist groups emerges gradually via the steady accumulation of small improvements on the anxiety scale, marching in tandem with the App Access group’s cumulative app usage. For demand effects to explain this pattern of improvement, participants would have to be sophisticated enough to report progressively improving symptoms over time. A more parsimonious reading is that the treatment group’s declining anxiety reflects their increased meditation on the app.

A similar threat would be participants reporting improved mental health in reciprocity for receiving something from us. If this were the case, we may expect Waitlist participants who received a cash transfer to report improved mental health relative to those who did not. However, in Table 3

²⁷This preface read “There are no right or wrong answers. Some people will feel better, some will not. We are simply interested in how you feel.”

we do not find that the cash transfer improves mental health. If anything, the cash transfer group appears to experience worse mental health than the Pure Waitlist group, which is inconsistent with reciprocity effects.

Another source of demand effect might be that participants believe that they can receive the app sooner by reporting poor mental health. This could produce spurious treatment effects as participants given a license might stop inflating their symptoms, while the Waitlist participants continue to do so. We believe this is unlikely, as our experimental instructions clearly state that all participants would eventually receive a license, and that the timing was decided by a computerized lottery. The data also do not bear out this concern. Waitlist participants report gradually lower levels of anxiety over time, declining from 31.6% with moderate or severe anxiety at baseline to 28.5% at two weeks and 26.1% at four weeks.

Overall, while we cannot entirely rule out that some of our treatment effects are driven by experimenter demand, the time structure and repeated nature of our measurements, the absence of effect from a cash transfer, and the small effects from usage incentives suggest that experimenter demand effects are unlikely to drive a large fraction of the effects that we observe. Additionally, we find significant effects from access to the app on revealed-preference outcomes that we review in Section 5.

5 Effects on Productivity and Decision-Making

5.1 Empirical Framework

As pre-specified, to maximize power when studying effects on economic outcomes, we pool the three initial license treatments (no incentives, short incentives, and long incentives) and consider the license group as a whole. We first estimate regressions of the form:

$$Y_i^{post} = \beta_1 \text{App Access}_i + \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i, \quad (4)$$

where Y_i^{post} is the outcome for individual i , App Access_i is an indicator for receiving a Headspace license, δ_{stratum} are fixed effects for randomization strata, and Y_i^{pre} is a pre-randomization measure of the outcome, where available. All standard errors use an HC1 estimator.

To separate the impacts of offering access to the mindfulness meditation app on economic behavior from the immediate effects of a meditation session, we offered half of participants with app access a bonus to meditate in the app immediately before taking the survey. We estimate effects of offering the app only and additional effects from the immediate meditation incentives using the following regression equation:

$$Y_i^{post} = \beta_1 \text{App Access}_i + \beta_2 \text{App Access}_i \times \text{Immediate Meditation}_i + \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i, \quad (5)$$

where $\text{Immediate Meditation}_i$ is an indicator for receiving an incentive to meditate immediately before the survey where outcomes were measured. Only users who initially received a license were randomized to receive immediate meditation incentives, so β_1 can be interpreted as the sustained ITT effect of offering access to the app (and any medium-term usage incentives) on behavior, while β_2 is the ITT effect of encouraging participants to meditate before making decisions.

We further investigate the effects of worrying thoughts on economic behavior and the potential for mindfulness to reduce the interference of emotions on decision-making. To do so, we cross-randomized all participants between executing a task designed to induce stress and a more neutral task, and we estimate the following regressions:

$$Y_i^{post} = \beta_1 \text{App Access}_i + \beta_2 \text{Stressful Task}_i + \beta_3 \text{Stressful Task}_i \times \text{App Access}_i + \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i, \quad (6)$$

and:

$$\begin{aligned} Y_i^{post} = & \beta_1 \text{App Access}_i + \beta_2 \text{Stressful Task}_i \\ & + \beta_3 \text{Stressful Task}_i \times \text{App Access}_i \\ & + \beta_4 \text{App Access}_i \times \text{Immediate Meditation}_i \\ & + \beta_5 \text{Stressful Task}_i \times \text{App Access}_i \times \text{Immediate Meditation}_i \\ & + \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i, \end{aligned} \quad (7)$$

where Stressful Task_i is an indicator for assignment to the stressful task. The coefficient β_3 estimates the difference in the effect of being exposed to the stressful task on individuals with access to a license (but no incentive to meditate immediately when using Equation 7) and on those on the Waitlist. The coefficient β_5 estimates the difference in the effect on stress within the App Access group, comparing individuals with an incentive to meditate immediately before the survey to those without such an incentive.

Tables 4 and 6 present treatment effects estimated using the above regressions on our main outcomes, with additional measures presented in Tables B.6, 5 and B.7. The Stroop test as well as our secondary outcomes are conducted before administering the Stressful Task, and thus are only analyzed using regressions 4 and 5. Tables B.8 and B.9 report effects of offering the app and incentivizing an immediate mediation on our secondary outcomes which measure avoidance of losses and of unpleasant information.

5.2 Productivity and Attention

We begin by studying the effects of the mindfulness meditation app on productivity in the attention-intensive proofreading task and on our cognitive test of attention control. Our first finding, reported in Table 4, column 2, is that access to the app increases earnings in the proofreading task

by 1.4 percent (1.2 percentage points as a fraction of the maximum possible earnings), corresponding to a 0.09 SD increase. When focusing on the isolated effect of offering the app (excluding individuals who were incentivized to meditate before the session, in column 3), the increase in earnings reaches 2.0 percent, or 0.13 SD.²⁸ Considering that effort in experimental tasks like this one has been documented to be very inelastic to payment (DellaVigna et al., 2016; DellaVigna and Pope, 2018), this represents a sizable increase, obtained from only two weeks of using a very affordable app. In other words, to the extent that these benefits translate into real work environments, offering such apps to employees and giving them some time to use them could be a high-return investment for firms and organizations—without even including possible improvements in retention and reductions in absenteeism from better mental health.

Table 4 columns 3 and 7 present the effects from eliciting stressful thoughts on self-reported feelings of being stressed immediately before the task, and on earnings in the task. The Stressful treatment induces a large effect in self-reported feelings of being stressed immediately before the task (column 3), which worsen by 1 SD. This effect is larger for participants with app access (by 0.21 SD) who otherwise enjoy a more relaxed state (by 0.16 SD). In other words, participants with access to the app are less stressed than Waitlist participants after the neutral task, but they report being about as stressed as them after the stressful task. However, the Stressful treatment does not affect participants’ earnings in the proofreading task, independent of app access.²⁹

This result could be interpreted in two ways. First, it is possible that the stress we induce in this experiment is not enough to impact performance in the proofreading task, as there are no large differences in performance between Waitlist participants who recall neutral versus worrying memories. If so, the relevant treatment effect is the one comparing the App Access and Waitlist groups, averaging across the stressful and neutral memory conditions (Table 4 Column 3, Row 1). Second, recalling worrying memories does cause a larger subjective increase in stress for the App Access group than for the Waitlist. To the extent that this additional stress should degrade performance, this is evidence that the effects of mindfulness are robust to the larger burden stress places on the App Access group. In this case, the relevant treatment effects come from Column 5 of Table 4.

We find no significant effect on earnings per time spent, although the estimated coefficient is positive (Table B.6). Access to the app non-significantly increases the time spent on the task, leading to earnings per time spent being less affected. Notably, the task was not timed, and incentives did not encourage fast work. Presumably, our participants did not have a very high opportunity cost of time since they were likely not “professional paid survey takers” as can be found on research platforms.³⁰ One interpretation is that access to the app makes participants able to sustain their attention to the task longer, for example, by reducing the cost of paying attention. Given that the increase in time spent is not statistically significant, these mechanisms may possibly be at play, but

²⁸We come back to the reduction in earnings from the Immediate Meditation treatment in Section 5.4.

²⁹Except for participants who additionally receive incentive to meditate, see Section 5.4.

³⁰They often took the time to answer non-incentivized feedback questions with lengthy answers, further suggesting they are not trying to optimize their earnings per time spend on the survey.

we do not have strong evidence for any of them.

We detect no effect on performance in a Stroop test (Table 5), our primary measure of cognitive function. We cannot reject a small reduction in errors made by participants, but this result nevertheless suggests that the increase in output in the proofreading task is unlikely to be fully driven by improvements in cognitive functions linked to attention control.

As a result, we are unable to take a definitive stance on the mechanism behind improvements in work output from the mindfulness meditation app. Yet, regardless of the mechanism, we can conclude that the offering the app makes individuals perform better in the task.

Finally, we investigate heterogeneous treatment effects on earnings in the proofreading task in Figure 6. We do not find significant heterogeneity.³¹ Gains in proofreading earnings appear to be smaller for individuals with better baseline mental health, although this heterogeneity is not significant, and even individuals with less symptoms directionally appear to benefit from the app in that regard.

5.3 Interference of Unrelated Emotions with Risk-Taking Decisions

We next study whether access to the app can mitigate the interference of stressful thoughts with people’s risk-taking decision. We report effects on the certainty premium in Table 6, and on its two components (risk aversion under uncertainty and under certainty) in Table B.7.³²

First, we verify that the Stressful tasks, which ask participants to recall worrying memories and think about a hypothetical large health bill, affect risk-taking decisions among Waitlist participants. These tasks reduce the Waitlist’s certainty premium by 23 percent (Table 6, column 4). This is consistent with Callen et al. (2014)’s finding that inducing negative emotions affects risk aversion under uncertainty but not certainty. However, the effects we find are in the opposite direction (reducing, rather than increasing the certainty premium), which may reflect natural differences between risk preferences in our study sample and theirs, or the stakes in our respective lotteries. Callen et al. (2014) work with a population of Afghan citizens. They note that certainty premia vary by educational attainment in their sample, where the average participant completed 10 years of education. By contrast, 36% of our participants hold a Bachelor’s degree and 44% hold an additional graduate degree, typically corresponding to between 16 and 22 years of education. The higher level of education in our sample may translate to differing reactions to emotion and uncertainty, in a manner consistent with Callen et al. (2014)’s findings. In addition, while we view our lottery outcomes, ranging from \$0 to \$30, as reasonable stakes for a mechanisms experiment, they are lower

³¹We only find significant heterogeneity of the marginal effect of incentives, by gender. Notably, males only make up about 15 percent of our sample, and this strong selection by gender into the study suggests that other important unobserved characteristics might be correlated with gender in our sample. We also find that incentives do not increase app usage for males on average, suggesting they may have a backfiring effect in some way.

³²Recall from Section 2.7 that we elicit risk aversion as the value of P such that participants are indifferent between a lottery paying \$30 with probability P and \$0 otherwise, and an outside option. Risk aversion under uncertainty, $P_{\text{uncertainty}}$ corresponds to the case when the outside option is itself a lottery, that pays \$30 or \$10 with equal probability. Risk aversion under certainty, $P_{\text{certainty}}$, is the indifference point when the outside option is a sure \$10. The two values are not compared directly. We construct the certainty premium by fixing the utility of \$0 and \$30 and assuming expected utility in each decisions, as in Callen et al. (2014).

than the stakes in Callen et al. (2014)’s lotteries, where participants stand to gain between one and three days’ wages. The relationship between risk and emotion may reasonably depend on stakes, which could also explain the differing signs on our effects.

Our focus, however, is not to document *how* risk-taking reacts to emotions in general, but *if* it does and specifically whether mindfulness meditation mitigates such interference. Our main result is that access to the app reverses the interference of this Stressful treatment on choice under risk. While the Stressful treatment reduces the certainty premium by 23 percent among Waitlist participants, access to the app undoes this effect, with a difference-in-differences estimate of +37 percent of the control mean (column 4, $p < 0.05$). As shown in Table B.7, this is led by access to the app undoing the effect of the Stressful treatment on risk aversion under uncertainty. On its own, this finding suggests that mindfulness meditation can insulate decision making from the effects of stressful thoughts, at least at the levels of stress and mindfulness we induce in this experiment.

That said, we do not observe strong evidence that the App Access group is entirely insensitive to emotions. The Stressful treatment does increase their certainty premium by 14 percent relative to the control mean. Though not statistically significant at conventional levels ($p \approx 0.2$), this is consistent with stress moderately increasing the certainty premium (95% confidence interval: [-6%, +31%]). Scrutinizing the App Access group more closely in column 5 of Table 6, we find that stress has a larger effect on participants whom we incentivize to meditate immediate before the survey (+29.4%, $p < 0.05$) and a near-zero effect on participants who did not receive this incentive (+3%, $p \approx 0.8$). However, this latter estimate is again too imprecise to entirely rule out moderate sensitivity to emotions. Our preferred interpretation of these findings focuses on the more precisely estimated effects that come from pooling the App Access group, though we detail further evidence that the immediate effects of meditation sessions may differ from the long-term effects of mindfulness in Section 5.4.

Next, considering only the estimates of treatment effects overall in Table 6 columns 2 and 3 (as well as Table B.7 columns 1, 2, 5 and 6), pooling the Neutral and Stressful treatments, access to the app appears to have no significant effect on measures of risk-taking behavior. This result suggests that we should not think of meditation as changing fundamental preference parameters (such as risk aversion) per se. Instead, we find evidence that mindfulness meditation makes people’s choices less susceptible to changes in their emotional states.

These results provide a proof of concept that mindfulness meditation can help prevent unrelated emotions from interfering with decision-making. Stress caused Waitlist participants to choose differently on a set of lottery tasks, and mindfulness meditation appears to diminish or reverse this effect. Crucially, we do not need to take a stance on whether participants should be more or less risk averse in these decisions. The implications of these findings depend on the reader’s philosophical commitments. For those who assume that the stressful thoughts we induce should not affect preferences on this narrowly-framed lottery choice, a natural conclusion is that stress impedes the decision maker’s assessment of outcomes and reduces welfare (Bernheim and Rangel, 2009). Another interpretation without the normative implication is that mindfulness simply alters the practitioner’s preferences to

become less susceptible to the induced emotions. In this case, mindfulness is a consumption good that affects decision making and the practice of meditation is more akin to maintaining regular sleep for clarity of mind, for example.

5.4 Further Immediate Effects of Meditation Sessions

We now turn to understanding the immediate effects of mindfulness meditation sessions. These can have policy relevance for organizations that might consider implementing encouragements to practice meditation in the workplace or before certain decision-making events. Further, it is methodologically valuable to understand whether the short-term effects of sessions resemble the more sustained effects from regular practice. We find that it is not the case. In fact, an incentivized mindfulness meditation session reduces work performance and increases small probability loss aversion and the avoidance of unpleasant information in the short-term. Finally, these short-term effects help us understand likely mechanisms through which mindfulness meditation affects behavior in the longer-term (see Section 6).

Table 4, column 5 shows that the immediate meditation session partially reduces the gain in earnings in the proofreading task due to receiving the app. Column 7 reveals that this reduction is entirely driven by the Stressful treatment, suggesting that participants are more susceptible to attention capture by negative emotions immediately after a meditation session, although we measure this interaction imprecisely. It also suggests that reduced earnings are not merely driven by participants running out of time due to the meditation session, as we would then expect the reduction to be independent of the Stressful treatment. Instead, the role of the Stressful treatment may be understood through the lens of our model (see Section 6) as directing more of the agent’s attention to their emotions, as opposed to the task at hand.

In addition, incentives to meditate at the beginning of the survey have notable effects on our secondary outcomes, described in Appendix B.2 and further interpreted with the lens of our model in Section 6. In the short-term, a meditation session increases the propensity to refuse a small-stakes lottery with high expected value but salient low probability losses, and increases avoidance of emotionally unpleasant information.

Overall, we detect several statistically significant effects from incentivizing a mindfulness meditation session immediately before the effort task and decisions that are absent from the group that only received access to the app. From a standard economic view, these effects are mostly adverse ones since they make participants less productive and less likely to seek useful information or to take a high expected value, small-stakes lottery. Our model (section 6) presents an interpretation for the discrepancy of the short run and long run effects, based on psychological conceptualizations of mindfulness (Lindsay and Creswell, 2017). Meditation sessions make the individual more attentive to their emotions in the short term, which allows them to build emotional sophistication and avoid attentional capture in the medium to long term.

One might alternatively wonder whether these effects are amplifications of insignificant but directionally similar effects when the treatment is access to the app only. Our data do not support

this hypothesis in general: as we saw, app access *increases* earnings in the proofreading task, and, suggestively, it *reduces* unpleasant information avoidance (albeit not significantly). Overall, these results do *not* lend support to the methodological validity of inferring sustained effects of mindfulness meditation from the immediate effects measured right after an experimentally induced session.

In simpler terms, our leading interpretation is that a mindfulness meditation session puts individuals in a relaxed and emotionally sensitive state that reduces their willingness to generally engage with the experimental tasks, especially when the tasks bring unpleasant emotions to the forefront. Either way, these findings may caution organizations against encouraging meditation sessions immediately before important decision-making events or in the middle of a stressful workday.

6 Conceptual Framework: Emotions and Attention Misallocation

We now present a theoretical framework using economic primitives, in line with the key existing ideas from psychology, that unify our empirical findings. As mindfulness encompasses a variety of techniques, our goal is not to provide the definitive theoretical treatment of it. Rather, we intend to provide a guide for interpreting our empirical results by incorporating the core ideas from the psychology literature into a workhorse model of economic behavior.

This framework combines a model of costly attention allocation with a utility function that depends in part on emotions. The agent must direct her attention to a choice problem, then make a decision based on her understanding of the problem. Emotions draw some of the agent’s attention away from aspects of the task at hand, and also directly influence her decision making. An influential hypothesis suggests that mindfulness training increases emotional sensitivity, allowing practitioners to build emotional awareness over time (Lindsay and Creswell, 2017). This is in line with the consensus definition that mindfulness involves non-judgmentally attending to one’s mental state. We capture this by modeling mindfulness sessions as directing the agent’s attention to her thoughts and emotions. This makes the agent more sensitive to emotions in the short run, but gradually improves her ability to avoid attention capture later on.

We begin by providing motivating examples to animate our discussion. We intend these to be illustrative, though not exhaustive, of the kinds of settings we consider.

- **Emotionally Demanding Work.** X’s job requires him to interact with a stream of clients, one after the other. He may be a social worker, public defender, or counselor in a prison. X sympathizes with his clients, often thinking about their situations well after each appointment is done. This poses a challenge: each case requires careful attention to the problem at hand, but X grows increasingly tired and distracted as the day progresses. By his last meeting, X’s abilities to attend to his last client and choose the best course of action are impaired by the emotionally taxing nature of his earlier meetings.
- **Negotiation.** Y is about to begin an important negotiation. Perhaps she is about to ask for a raise, request an individualized learning plan from her daughter’s school, or propose a change in care for a relative in a nursing home. However, Y cares deeply about this issue and

is worried about growing emotional and losing credibility during the meeting. She is convinced that the best way to achieve her desired outcome is to dispassionately state the facts of her case to the relevant decision maker, and would prefer to be able to do this.

We note that emotions do convey important information in these cases, signaling the importance of certain tasks. However, emotions also affect the agent’s decision making in ways that run counter to their preferences. That is, both agents may wish to focus on a particular task and desire emotionally neutral preferences in service of this goal (Bernheim et al., 2021). With these examples in mind, we now sketch our model.

6.1 Framework

Decision Problem with Emotions. An agent finds herself in an particular situation. She must learn the nature of her situation, and then choose an action to maximize her utility. We represent the true state of situation with a two-dimensional vector, $x = (x_1, x_2) \in \mathbb{R}^2$. From the agent’s perspective, these are random variables drawn from a distribution with mean $x^d = (x_1^d, x_2^d)$. Her action is a scalar, $a \in \mathbb{R}$.

In this structure, we intend x_1 to capture a payoff-relevant aspect of the situation, while x_2 models some external or extraneous influence —perhaps a passing thought or a strong but unrelated emotion. To make x_1 as payoff relevant, we model the agent’s action as producing utility given by:

$$u(a, x_1, x_2) = -(a - x_1)^2.$$

That is, the agent obtains the greatest utility from matching a to x_1 regardless of the value of x_2 . However, the agent has emotional sensitivity μ , and acts as if choosing a will yield utility

$$\tilde{u}(a, x_1, x_2) = -(a - x_1)^2 - \mu(a - x_2)^2.$$

That is, the agent acts as if she obtains the most utility from matching a to some combination of x_1 and x_2 .

The influence of x_2 may capture a variety of tendencies. For example, the agent may be mischaracterizing the situation, such as if emotions cause her to overestimate the probability of an undesirable outcome (Bernheim and Taubinsky, 2018). Alternatively, x_2 may reflect a kind of projection bias that causes the agent to choose differently in the short run than they would prefer to in the long run (Loewenstein et al., 2003). In either case, while the agent is aware that x_1 and x_2 are distinct, she behaves as if unaware of x_2 ’s irrelevance to her payoff.

Attention Allocation. The agent does not know the values of x_1 and x_2 , and must exert costly attention to discern them. She begins with default beliefs $x^d = E[x]$, and we represent her attention to dimension i with $m_i \in [0, 1]$. Given an attention vector m , the agent estimates the attributes to be

$$\hat{x}(m) = \left(m_1 x_1 + (1 - m_1) x_1^d, m_2 x_2 + (1 - m_2) x_2^d \right).$$

Paying attention incurs an effort cost, $C(m)$. We consider attention paid to different dimensions to be substitutes, modeling $C(m) = \kappa (\sum_i m_i)^2$, with $\kappa \geq 0$ an attention cost parameter.

The agent's action will depend on her reading of the situation, which in turn depends on her allocation of attention. We model the agent as optimally selecting her attention to solve

$$m^* = \arg \max_m \mathbb{E} [\tilde{u}(a(m), x)] - C(m),$$

where $a(m)$ is the optimal action given the beliefs $\hat{x}(m)$. Solving for m , we note two comparative statics that will be central to our predictions. The amount of attention m_1^* allocated to the relevant dimension:

1. increases when emotional sensitivity μ decreases.
2. increases with the variance of the first dimension ($\mathbb{E}[(x_1 - x_1^d)^2]$) and decreases with the variance of the competing dimension ($\mathbb{E}[(x_2 - x_2^d)^2]$).
3. weakly increase when the cost of paying attention κ decreases.

Sequence of Decisions. We consider an agent who goes through an infinite sequence of static decision problems involving independent realizations of x . We assume that the agent behaves at each period as if she was only solving this static problem. The long-run effects of mindfulness meditation in our model arise from the following assumption:

Assumption 1: Building Emotional Awareness Through Introspection At each period t , agent with positive emotional sensitivity μ may become emotionally neutral and switch to $\mu = 0$ for all periods $t' > t$. This happens with a probability that increases in the amount of attention paid to emotions in period t , $m_{2,t}$.³³

6.2 Main Predictions

Prediction 1: In the long run, mindfulness meditation practice increases the attention available to spend on the task or decision at hand.

Prediction 2: In the long run, mindfulness meditation practice reduces the interference of irrelevant emotions with decisions.

Prediction 3: In the short run, mindfulness meditation sessions reduce the attention available to spend on the task or decision at hand.

³³One could microfound this model through learning or practice. For example, suppose the agent observes her realized utility $u(a, x) = -(a - x_1)^2$ in each period. If the likelihood of obtaining this utility given her reading of the situation $(\hat{x}_1(m_1), \hat{x}_2(m_2))$ is sufficiently small under the misspecified $\tilde{u}(a, x, \mu)$, she wakes up to her misspecification as in Gagnon-Bartsch et al. (2018). Attention paid to x_2 throws a sharper light on the gap between her realized utility and her anticipated utility under the misspecified model. Alternatively, one could consider mindfulness to be a technology that directly reduces emotional sensitivity, by offering the agent practice at the task of noticing and then moving past emotions.

We find evidence in favor of Prediction 1 (Table 4, Column 5, Row 1), suggestive of Prediction 2 (Table 6 Column 5), and in favor of Prediction 3 (Table 4, Column 5, Row 5).

Alternative Mechanisms Prediction 1 follows without assuming any change in the attention cost parameter κ . Regular mindfulness meditation may decrease κ by building cognitive endurance, as in CITE. Our experiment includes a cognitive test (the Stroop task), which can be interpreted as a direct measure of κ to assess the extent of this mechanism. We do not find evidence in favor of this mechanism, since app access does not improve performance on the Stroop test. Therefore, our results demonstrate that mindfulness can affect various aspect of economic behavior without directly improving cognitive function.

6.3 Auxiliary Predictions: Emotions in Utility

We consider additional effects of mindfulness meditation on decisions that may affect emotions, such as decisions to avoid potentially unpleasant information. Intuitively, if the agent derives a concave hedonic value from her perceived emotions, she is more likely to make decisions that avoid creating emotional variance when she is paying attention to her emotions than when she is ignoring them. Such decisions may include avoiding potentially unpleasant information, for example. We extend the utility function as:

$$\tilde{u}(a, x, \mu, \nu) = -(a - x_1)^2 - \mu(a - x_2)^2 + \nu \log(x_2(m)),$$

where ν is a parameter capturing the direct utility from emotions. By Jensen’s inequality, concavity of the log implies that the agent is worse off when the variance of x_2 is larger, especially if m_2 is large. As we have seen, mindfulness meditation practice may lead to a short-run increase in m_2 but a long-run decrease in the same because of emotional awareness. This implies short-run and long-run predictions about the effects of mindfulness meditation on decisions which reduce anticipated emotional variance, such as avoidance of unpleasant information or avoidance of loss-framed gambles:

Auxiliary Prediction A1: In the short run, mindfulness meditation sessions increase the extent of emotionally avoidant decisions.

Auxiliary Prediction A2: In the long run, mindfulness meditation practice reduces the extent of emotionally avoidant decisions.

We test these predictions with secondary outcomes that are extensively described in Appendix B.2. We find evidence in favor of Prediction A1 (Table B.8, row 2, columns 2 and 4), and no strong evidence in either direction for Prediction A2 (Table B.8, row 1, columns 2 and 4).

Effects on Mental Health. We do not attempt to explicitly model mental health. Depression, stress and anxiety are in themselves complex, multidimensional concepts and incorporating them into this model is beyond our scope. We do however note that our framework emphasizes emotional awareness, the ability to avoid focusing attention on thoughts and emotions that are irrelevant to

the task at hand, as the main skill trained by mindfulness. Since rumination on negative thoughts about the past and worries about the future are leading hypothesized mechanisms of depression and anxiety (Gotlib and Joormann, 2010), our framework includes a natural way in which mindfulness meditation may reduce symptoms of mental ill-health.

7 Discussion

Our results demonstrate the potential of inexpensive mindfulness meditation apps to improve mental health, work performance, and consistent decision-making. Even though the app we evaluate is vastly less expensive than in-person psychotherapy, it leads to comparable short-run improvements in mental health. More generally, we provide the first large-sample evidence that mindfulness meditation can affect important dimensions of economic behavior with revealed-preference outcomes. If our estimated effects on work performance translate into real-world workplace productivity improvements, the cost of the app will rapidly pay for itself.³⁴

One related limitation resides in our approach to recruiting subjects. To establish a proof of concept, and because it was a natural and cost-effective first step, we recruited subjects who were interested in trying the mindfulness meditation app. Given our positive results, future research should study the determinants of such interest, and the effects of providing access to similar apps—as well as information about them—to all the individuals in a subject pool rather than the interested ones only.

It is more difficult to estimate the real-world consequences of reducing interference of emotions with economic decisions, such as choices under risk. These types of decisions are at the core of many organizations, and making the best decisions in high-stakes, high-uncertainty environments is a leading part of the wage premium of top executives. Many stressors potentially interfere with individuals emotions, fueled by political polarization, racial tensions or global health crises. As we have argued, decision-making is likely sub-optimal in some instances if it can be swayed by irrelevant emotional states and worries of the decision-maker. Our results therefore provide a proof of concept that these decisions can be made more consistent, but a challenging open question remains to study the magnitude of emotional distortions outside of controlled decision-making experiments.

We also make a methodological contribution that could inform future research on mindfulness and related psychological interventions. Our results show that evaluating behavioral consequences immediately after a meditation session can be a poor yardstick to measure the more sustained effects. This suggests that research should refrain from drawing strong conclusions from one-session experiments in the lab, unless immediate effects are the object of interest.

Overall, our findings support the idea that policymakers and organizations should consider subsidizing inexpensive tools such as this one. More broadly, they suggest that mental wellness programs might be a way to invest in preventive health with better returns than physical wellness programs (Jones et al., 2019). Our heterogeneity results also indicate that when designing such programs,

³⁴The opportunity cost of time spent meditating is less trivial, but should still come below other mental health care options.

organizations should use monetary incentives with caution, as they only appear to help individuals with more symptoms or optimistic beliefs about the app at baseline, and may carry lesser long-term benefits.

References

- Acland, Dan and Matthew R. Levy**, “Naiveté, Projection Bias, and Habit Formation in Gym Attendance,” *Management Science*, January 2015, 61 (1), 146–160.
- Aggarwal, Shilpa, Rebecca Dizon-Ross, and Ariel D. Zucker**, “Incentivizing Behavioral Change: The Role of Time Preferences,” Technical Report w27079, National Bureau of Economic Research May 2020.
- Alem, Yonas, Hannah Behrendt, Michele Belot, and Anikó Bíró**, “Mind, Behaviour and Health: A Randomised Experiment,” 2021, p. 70.
- Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow**, “The Welfare Effects of Social Media,” *American Economic Review*, March 2020, 110 (3), 629–676.
- , **Matthew Gentzkow, and Lena Song**, “Digital Addiction,” Working Paper 28936, National Bureau of Economic Research June 2021.
- Alloush, M**, “Income, Psychological Well-Being, and the Dynamics of Poverty,” 2018, p. 76.
- American Psychological Association**, “Stress in America 2020: A National Mental Health Crisis,” *Diakses dari: apa. org/news/press/releases/stress*, 2020.
- American Society of Clinical Psychology**, “Mindfulness-Based Cognitive Therapy | Society of Clinical Psychology,” <https://div12.org/treatment/mindfulness-based-cognitive-therapy/> April 2019.
- Angrist, Joshua D. and Guido W. Imbens**, “Identification and Estimation of Local Average Treatment Effects,” Working Paper 118, National Bureau of Economic Research February 1995.
- Arroll, Bruce, Felicity Goodyear-Smith, Susan Crengle, Jane Gunn, Ngaire Kerse, Tana Fishman, Karen Falloon, and Simon Hatcher**, “Validation of PHQ-2 and PHQ-9 to Screen for Major Depression in the Primary Care Population,” *The Annals of Family Medicine*, July 2010, 8 (4), 348–353.
- Ash, Elliott, Daniel Sgroi, Anthony Tuckwell, and Shi Zhuo**, “Mindfulness Reduces Information Avoidance,” *Center for Law & Economics Working Paper Series*, August 2021, 2021 (13).
- Baer, Ruth A.**, “Mindfulness Training as a Clinical Intervention: A Conceptual and Empirical Review,” *Clinical Psychology: Science and Practice*, 2003, 10 (2), 125–143.
- , **Gregory T. Smith, Emily Lykins, Daniel Button, Jennifer Krietemeyer, Shannon Sauer, Erin Walsh, Danielle Duggan, and J. Mark G. Williams**, “Construct Validity of the Five Facet Mindfulness Questionnaire in Meditating and Nonmeditating Samples,” *Assessment*, September 2008, 15 (3), 329–342.
- Banerjee, Abhijit V. and Sendhil Mullainathan**, “Limited Attention and Income Distribution,” *American Economic Review*, May 2008, 98 (2), 489–493.
- Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko**, “Maternal Depression, Women’s Empowerment, and Parental Investment: Evidence from a Randomized Controlled Trial,” *American Economic Review*, March 2020, 110 (3), 824–859.

- Beard, C. and T. Björgvinsson**, “Beyond Generalized Anxiety Disorder: Psychometric Properties of the GAD-7 in a Heterogeneous Psychiatric Sample,” *Journal of Anxiety Disorders*, August 2014, 28 (6), 547–552.
- , **K. J. Hsu, L. S. Rifkin, A. B. Busch, and T. Björgvinsson**, “Validation of the PHQ-9 in a Psychiatric Sample,” *Journal of Affective Disorders*, March 2016, 193, 267–273.
- Bernheim, B. Douglas and Antonio Rangel**, “Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics*,” *The Quarterly Journal of Economics*, February 2009, 124 (1), 51–104.
- Bernheim, B Douglas and Dmitry Taubinsky**, “Behavioral Public Economics,” *Handbook of Behavioral Economics: Applications and Foundations 1*, 2018, 1, 381–516.
- , **Luca Braghieri, Alejandro Martínez-Marquina, and David Zuckerman**, “A theory of chosen preferences,” *American Economic Review*, 2021, 111 (2), 720–54.
- Bhat, Bhargav, Jonathan de Quidt, Johannes Haushofer, Vikram Patel, Gautam Rao, Frank Schilbach, and Pierre-Luc Vautrey**, “The Long-Run Effects of Psychotherapy on Depression, Beliefs, and Preferences,” *Working Paper*, 2021.
- Bishop, Scott R., Mark Lau, Shauna Shapiro, Linda Carlson, Nicole D. Anderson, James Carmody, Zindel V. Segal, Susan Abbey, Michael Specia, Drew Velting, and Gerald Devins**, “Mindfulness: A Proposed Operational Definition,” *Clinical Psychology: Science and Practice*, 2004, 11 (3), 230–241.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan**, “Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia,” *American Economic Review*, April 2017, 107 (4), 1165–1206.
- Bolte, Lukas and Collin Raymond**, “Emotional Inattention,” October 2022.
- Braghieri, Luca, Roe Levy, and Alexey Makarin**, “Social Media and Mental Health,” SSRN Scholarly Paper ID 3919760, Social Science Research Network, Rochester, NY August 2021.
- Brown, Christina, Supreet Kaur, Geeta Kingdon, and Heather Schofield**, “COGNITIVE ENDURANCE AS HUMAN CAPITAL,” 2022, p. 86.
- Brown, Kirk Warren, Richard M. Ryan, and J. David Creswell**, “Mindfulness: Theoretical Foundations and Evidence for Its Salutary Effects,” *Psychological Inquiry*, October 2007, 18 (4), 211–237.
- Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger**, “Violence and Risk Preference: Experimental Evidence from Afghanistan,” *American Economic Review*, January 2014, 104 (1), 123–148.
- Cassar, Lea, Mira Fischer, and Vanessa Valero**, “Keep Calm and Carry On: The Short vs. Long Run Effects of Mindfulness Meditation on (Academic) Performance,” 2022, p. 75.
- Charness, Gary and Uri Gneezy**, “Incentives to Exercise,” *Econometrica*, 2009, 77 (3), 909–931.

- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins**, “Double/Debiased Machine Learning for Treatment and Structural Parameters,” *The Econometrics Journal*, February 2018, *21* (1), C1–C68.
- Cohen, Sheldon, Tom Kamarck, and Robin Mermelstein**, “Perceived Stress Scale,” *Measuring stress: A guide for health and social scientists*, 1994, *10* (2), 1–2.
- Cramer, Holger, Helen Hall, Matthew Leach, Jane Frawley, Yan Zhang, Brenda Leung, Jon Adams, and Romy Lauche**, “Prevalence, Patterns, and Predictors of Meditation Use among US Adults: A Nationally Representative Survey,” *Scientific Reports*, November 2016, *6* (1), 36760.
- Cuadrado, Esther, Alicia Arenas, Manuel Moyano, and Carmen Tabernero**, “Differential Impact of Stay-at-Home Orders on Mental Health in Adults Who Are Homeschooling or “Childless at Home” in Time of COVID-19,” *Family Process*, 2021, *n/a* (n/a).
- Cuijpers, Pim, Matthias Berking, Gerhard Andersson, Leanne Quigley, Annet Kleiboer, and Keith S Dobson**, “A Meta-Analysis of Cognitive-Behavioural Therapy for Adult Depression, Alone and in Comparison with Other Treatments,” *The Canadian Journal of Psychiatry*, July 2013, *58* (7), 376–385.
- Davidson, Richard J. and Alfred W. Kaszniak**, “Conceptual and Methodological Issues in Research on Mindfulness and Meditation,” *American Psychologist*, 2015, *70* (7), 581–592.
- de Quidt, Jonathan, Lise Vesterlund, and Alistair J. Wilson**, “Experimenter Demand Effects,” *Handbook of Research Methods and Applications in Experimental Economics*, July 2019.
- Dean, Emma Boswell, Frank Schilbach, and Heather Schofield**, “2. Poverty and Cognitive Function,” in “The Economics of Poverty Traps,” University of Chicago Press, 2019, pp. 57–118.
- DellaVigna, Stefano and Devin Pope**, “What Motivates Effort? Evidence and Expert Forecasts,” *The Review of Economic Studies*, April 2018, *85* (2), 1029–1069.
- , **John A. List, Ulrike Malmendier, and Gautam Rao**, “Estimating Social Preferences and Gift Exchange at Work,” Working Paper 22043, National Bureau of Economic Research February 2016.
- Duquenois, Claire**, “Fictional Money, Real Costs: Impacts of Financial Salience on Disadvantaged Students,” *American Economic Review*, 2021, p. 60.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde**, “Global Evidence on Economic Preferences,” *The Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.
- Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group**, “The Oregon Health Insurance Experiment: Evidence from the First Year*,” *The Quarterly Journal of Economics*, August 2012, *127* (3), 1057–1106.
- Flett, Jayde A. M., Harlene Hayne, Benjamin C. Riordan, Laura M. Thompson, and Tamlin S. Conner**, “Mobile Mindfulness Meditation: A Randomised Controlled Trial of the Effect of Two Popular Apps on Mental Health,” *Mindfulness*, May 2019, *10* (5), 863–876.

- Gabaix, Xavier**, “A Sparsity-Based Model of Bounded Rationality,” *The Quarterly Journal of Economics*, November 2014, 129 (4), 1661–1710.
- , “Behavioral Inattention,” in “Handbook of Behavioral Economics: Applications and Foundations 1,” Vol. 2, Elsevier, 2019, pp. 261–343.
- Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein**, *Channeled Attention and Stable Errors*, Harvard Business School Boston, 2018.
- Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Becerra**, “Testing Attrition Bias in Field Experiments,” 2020.
- Goldberg, Simon B., Kevin M. Riordan, Shufang Sun, and Richard J. Davidson**, “The Empirical Status of Mindfulness-Based Interventions: A Systematic Review of 44 Meta-Analyses of Randomized Controlled Trials,” *Perspectives on Psychological Science*, February 2021, p. 1745691620968771.
- Golman, Russell, David Hagmann, and George Loewenstein**, “Information Avoidance,” *Journal of Economic Literature*, March 2017, 55 (1), 96–135.
- Gotlib, Ian H. and Jutta Joormann**, “Cognition and Depression: Current Status and Future Directions,” *Annual Review of Clinical Psychology*, 2010, 6 (1), 285–312.
- Guendelman, Simón, Sebastián Medeiros, and Hagen Rampes**, “Mindfulness and Emotion Regulation: Insights from Neurobiological, Psychological, and Clinical Studies,” *Frontiers in Psychology*, 2017, 8, 220.
- Hafenbrack, Andrew C., Zoe Kinias, and Sigal G. Barsade**, “Debiasing the Mind Through Meditation: Mindfulness and the Sunk-Cost Bias,” *Psychological Science*, February 2014, 25 (2), 369–376.
- Heller, Sara B., Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack**, “Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago*,” *The Quarterly Journal of Economics*, February 2017, 132 (1), 1–54.
- Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol**, “Rational Habit Formation: Experimental Evidence from Handwashing in India,” *Harvard Business School BGIE Unit Working Paper*, 2019, (18-030), 18–030.
- Imbens, Guido W. and Charles F. Manski**, “Confidence Intervals for Partially Identified Parameters,” *Econometrica*, November 2004, 72 (6), 1845–1857.
- Iwamoto, Sage K., Marcus Alexander, Mark Torres, Michael R. Irwin, Nicholas A. Christakis, and Akihiro Nishi**, “Mindfulness Meditation Activates Altruism,” *Scientific Reports*, April 2020, 10 (1), 6511.
- Jensen, Arthur R.**, “Scoring the Stroop Test,” *Acta psychologica*, 1965, 24 (5), 398–408.
- Jha, Amishi P., Jason Krompinger, and Michael J. Baime**, “Mindfulness Training Modifies Subsystems of Attention,” *Cognitive, Affective, & Behavioral Neuroscience*, June 2007, 7 (2), 109–119.

- John, Anett and Kate Orkin**, “Can Simple Psychological Interventions Increase Preventive Health Investment?,” Technical Report w25731, National Bureau of Economic Research, Cambridge, MA April 2019.
- Johnson, Eric J. and Amos Tversky**, “Affect, Generalization, and the Perception of Risk,” *Journal of Personality and Social Psychology*, 1983, 45 (1), 20–31.
- Jones, Damon, David Molitor, and Julian Reif**, “What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study*,” *The Quarterly Journal of Economics*, November 2019, 134 (4), 1747–1791.
- Kabat-Zinn, Jon**, “An Outpatient Program in Behavioral Medicine for Chronic Pain Patients Based on the Practice of Mindfulness Meditation: Theoretical Considerations and Preliminary Results,” *General Hospital Psychiatry*, April 1982, 4 (1), 33–47.
- Kabat-Zinn, Jon**, “Mindfulness-Based Interventions in Context: Past, Present, and Future,” *Clinical Psychology: Science and Practice*, 2003, 10 (2), 144–156.
- Kabat-Zinn, Jon and Thich Nhat Hanh**, *Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness*, Random House Publishing Group, July 2009.
- Kaur, Supreet, Sendhil Mullainathan, Suanna Oh, and Frank Schilbach**, “Do Financial Concerns Make Workers Less Productive?,” SSRN Scholarly Paper ID 3770928, Social Science Research Network, Rochester, NY January 2021.
- Krusche, Adele, Eva Cyhlarova, Scott King, and J Mark G Williams**, “Mindfulness Online: A Preliminary Evaluation of the Feasibility of a Web-Based Mindfulness Course and the Impact on Stress,” *BMJ open*, 2012, 2 (3), e000803.
- Kuyken, Willem, Rachel Hayes, Barbara Barrett, Richard Byng, Tim Dalgleish, David Kessler, Glyn Lewis, Edward Watkins, Claire Brejcha, Jessica Cardy, Aaron Causley, Suzanne Cowderoy, Alison Evans, Felix Gradinger, Surinder Kaur, Paul Lanham, Nicola Morant, Jonathan Richards, Pooja Shah, Harry Sutton, Rachael Vicary, Alice Weaver, Jenny Wilks, Matthew Williams, Rod S Taylor, and Sarah Byford**, “Effectiveness and Cost-Effectiveness of Mindfulness-Based Cognitive Therapy Compared with Maintenance Antidepressant Treatment in the Prevention of Depressive Relapse or Recurrence (PREVENT): A Randomised Controlled Trial,” *The Lancet*, July 2015, 386 (9988), 63–73.
- Lee, David S**, “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *REVIEW OF ECONOMIC STUDIES*, 2009, p. 32.
- Lerner, Jennifer S., Deborah A. Small, and George Loewenstein**, “Heart Strings and Purse Strings: Carryover Effects of Emotions on Economic Decisions,” *Psychological Science*, May 2004, 15 (5), 337–341.
- , **Ye Li, Piercarlo Valdesolo, and Karim S. Kassam**, “Emotion and Decision Making,” *Annual Review of Psychology*, 2015, 66 (1), 799–823.
- Lindsay, Emily K. and J. David Creswell**, “Mechanisms of Mindfulness Training: Monitor and Acceptance Theory (MAT),” *Clinical Psychology Review*, February 2017, 51, 48–59.
- Loewenstein, George**, “Emotions in Economic Theory and Economic Behavior,” *The American Economic Review*, May 2000, 90 (2), 426–432.

- and **Jennifer S. Lerner**, “The Role of Affect in Decision Making,” in “Handbook of Affective Sciences” Series in Affective Science, New York, NY, US: Oxford University Press, 2003, pp. 619–642.
- Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch**, “Risk as Feelings,” *Psychological Bulletin*, March 2001, *127* (2), 267–286.
- Loewenstein, George, Ted O’Donoghue, and Matthew Rabin**, “Projection Bias in Predicting Future Utility*,” *The Quarterly Journal of Economics*, November 2003, *118* (4), 1209–1248.
- Lund, Crick, Kate Orkin, Witte, Davies, Johannes Haushofer, Bass, Bolton, Murray, Murray, Tol, Graham Thornicroft, and Vikram Patel**, “Economic impacts of Mental Health Interventions in Low and middle-Income countries: A Systematic review and meta-Analysis,” 2021.
- Mak, Winnie WS, Alan CY Tong, Sindy YC Yip, Wacy WS Lui, Floria HN Chio, Amy TY Chan, and Celia CY Wong**, “Efficacy and Moderation of Mobile App-Based Programs for Mindfulness-Based Training, Self-Compassion Training, and Cognitive Behavioral Psychoeducation on Mental Health: Randomized Controlled Noninferiority Trial,” *JMIR Mental Health*, October 2018, *5* (4), e8597.
- Plummer, Faye, Laura Manea, Dominic Trepel, and Dean McMillan**, “Screening for Anxiety Disorders with the GAD-7 and GAD-2: A Systematic Review and Diagnostic Metaanalysis,” *General Hospital Psychiatry*, March 2016, *39*, 24–31.
- Raghunathan, Rajagopal, Michel T. Pham, and Kim P. Corfman**, “Informational Properties of Anxiety and Sadness, and Displaced Coping,” *Journal of Consumer Research*, March 2006, *32* (4), 596–601.
- Romano, Joseph P and Michael Wolf**, “Efficient Computation of Adjusted P-Values for Resampling-Based Stepdown Multiple Testing,” 2016, p. 3.
- Ruiz, Miguel A., Enric Zamorano, Javier García-Campayo, Antonio Pardo, Olga Freire, and Javier Rejas**, “Validity of the GAD-7 Scale as an Outcome Measure of Disability in Patients with Generalized Anxiety Disorders in Primary Care,” *Journal of Affective Disorders*, February 2011, *128* (3), 277–286.
- Segal, Zindel V., Mark Williams, and John Teasdale**, *Mindfulness-Based Cognitive Therapy for Depression, Second Edition*, Guilford Publications, June 2018.
- , **Sona Dimidjian, Arne Beck, Jennifer M. Boggs, Rachel Vanderkruik, Christina A. Metcalf, Robert Gallop, Jennifer N. Felder, and Joseph Levy**, “Outcomes of Online Mindfulness-Based Cognitive Therapy for Patients With Residual Depressive Symptoms: A Randomized Clinical Trial,” *JAMA Psychiatry*, June 2020, *77* (6), 563–573.
- Shapiro, Shauna L., Hooria Jazaieri, and Philippe R. Goldin**, “Mindfulness-Based Stress Reduction Effects on Moral Reasoning and Decision Making,” *The Journal of Positive Psychology*, November 2012, *7* (6), 504–515.
- Slovic, Paul and Ellen Peters**, “Risk Perception and Affect,” *Current Directions in Psychological Science*, December 2006, *15* (6), 322–325.

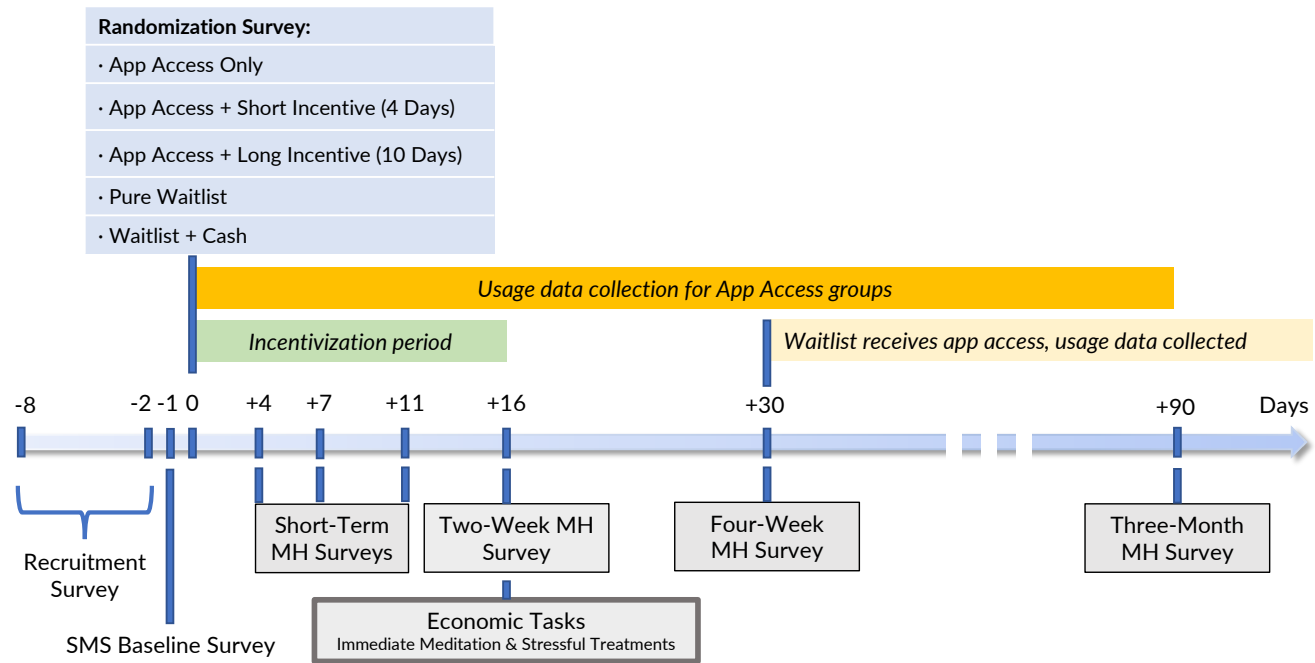
- Sommers-Spijkerman, Marion, Judith Austin, Ernst Bohlmeijer, and Wendy Pots**, “New Evidence in the Booming Field of Online Mindfulness: An Updated Meta-Analysis of Randomized Controlled Trials,” *JMIR Mental Health*, July 2021, 8 (7), e28168.
- Spadaro, Kathleen C and Diane F Hunker**, “Exploring the Effects of an Online Asynchronous Mindfulness Meditation Intervention with Nursing Students on Stress, Mood, and Cognition: A Descriptive Study,” *Nurse education today*, 2016, 39, 163–169.
- Spitzer, Robert L., Kurt Kroenke, Janet B. W. Williams, and Bernd Löwe**, “A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7,” *Archives of Internal Medicine*, May 2006, 166 (10), 1092–1097.
- Sydnor, Justin**, “(Over)Insuring Modest Risks,” *American Economic Journal: Applied Economics*, October 2010, 2 (4), 177–199.
- Toussaint, Anne, Paul Hüsing, Antje Gumz, Katja Wingenfeld, Martin Härter, Elisabeth Schramm, and Bernd Löwe**, “Sensitivity to Change and Minimal Clinically Important Difference of the 7-Item Generalized Anxiety Disorder Questionnaire (GAD-7),” *Journal of Affective Disorders*, March 2020, 265, 395–401.
- Twenge, Jean M.**, “Increases in Depression, Self-Harm, and Suicide Among U.S. Adolescents After 2012 and Links to Technology Use: Possible Mechanisms,” *Psychiatric Research and Clinical Practice*, 2020, 2 (1), 19–25.
- , **Jonathan Haidt, Thomas E. Joiner, and W. Keith Campbell**, “Underestimating Digital Media Harm,” *Nature Human Behaviour*, April 2020, 4 (4), 346–348.
- US Census Bureau**, “Real Mean Personal Income in the United States,” <https://fred.stlouisfed.org/series/MAPAINUSA672N> 2020.
- Vonderlin, Ruben, Miriam Biermann, Martin Bohus, and Lisa Lyssenko**, “Mindfulness-Based Programs in the Workplace: A Meta-Analysis of Randomized Controlled Trials,” *Mindfulness*, July 2020, 11 (7), 1579–1598.
- Wu, Yin, Brooke Levis, Kira E. Riehm, Nazanin Saadat, Alexander W. Levis, Marleine Azar, Danielle B. Rice, Jill Boruff, Pim Cuijpers, Simon Gilbody, John P. A. Ioannidis, Lorie A. Kloda, Dean McMillan, Scott B. Patten, Ian Shrier, Roy C. Ziegelstein, Dickens H. Akena, Bruce Arroll, Liat Ayalon, Hamid R. Baradaran, Murray Baron, Charles H. Bombardier, Peter Butterworth, Gregory Carter, Marcos H. Chagas, Juliana C. N. Chan, Rushina Cholera, Yeates Conwell, Janneke M. de Man-van Ginkel, Jesse R. Fann, Felix H. Fischer, Daniel Fung, Bizu Gelaye, Felicity Goodyear-Smith, Catherine G. Greeno, Brian J. Hall, Patricia A. Harrison, Martin Härter, Ulrich Hegerl, Leanne Hides, Stevan E. Hobfoll, Marie Hudson, Thomas Hyphantis, Masatoshi Inagaki, Nathalie Jetté, Mohammad E. Khamseh, Kim M. Kiely, Yunxin Kwan, Femke Lamers, Shen-Ing Liu, Manote Lotrakul, Sonia R. Loureiro, Bernd Löwe, Anthony McGuire, Sherina Mohd-Sidik, Tiago N. Munhoz, Kumiko Muramatsu, Flávia L. Osório, Vikram Patel, Brian W. Pence, Philippe Persoons, Angelo Picardi, Katrin Reuter, Alasdair G. Rooney, Iná S. Santos, Juwita Shaaban, Abbey Sidebottom, Adam Simning, Lesley Stafford, Sharon Sung, Pei Lin Lynnette Tan, Alyna Turner, Henk C. van Weert, Jennifer White, Mary A. Whooley, Kirsty Winkley, Mitsuhiro Yamada, Andrea Benedetti, and Brett D. Thombs**, “Equivalency of the

Diagnostic Accuracy of the PHQ-8 and PHQ-9: A Systematic Review and Individual Participant Data Meta-Analysis,” *Psychological Medicine*, June 2020, *50* (8), 1368–1380.

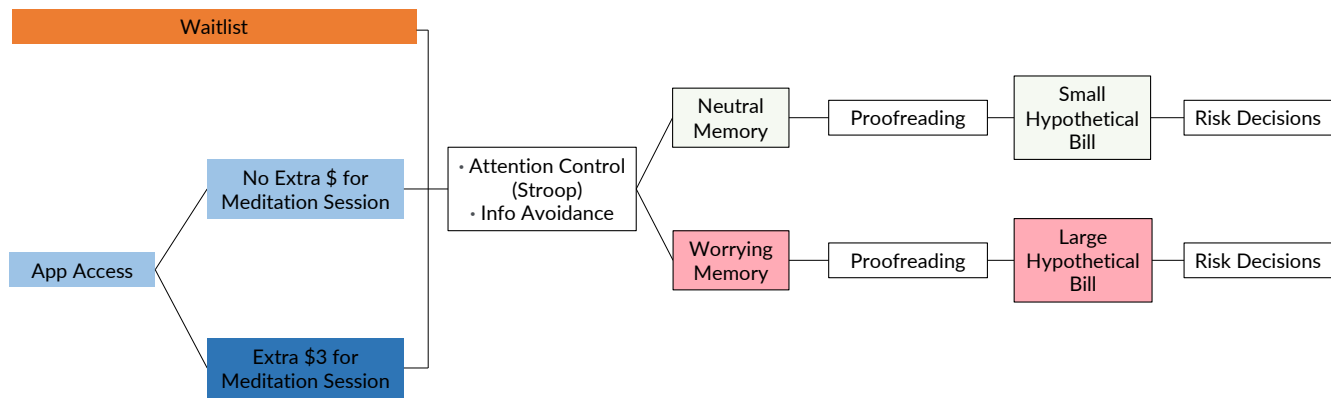
8 Figures and Tables

Figure 1: Experiment Overview

(a) Timeline



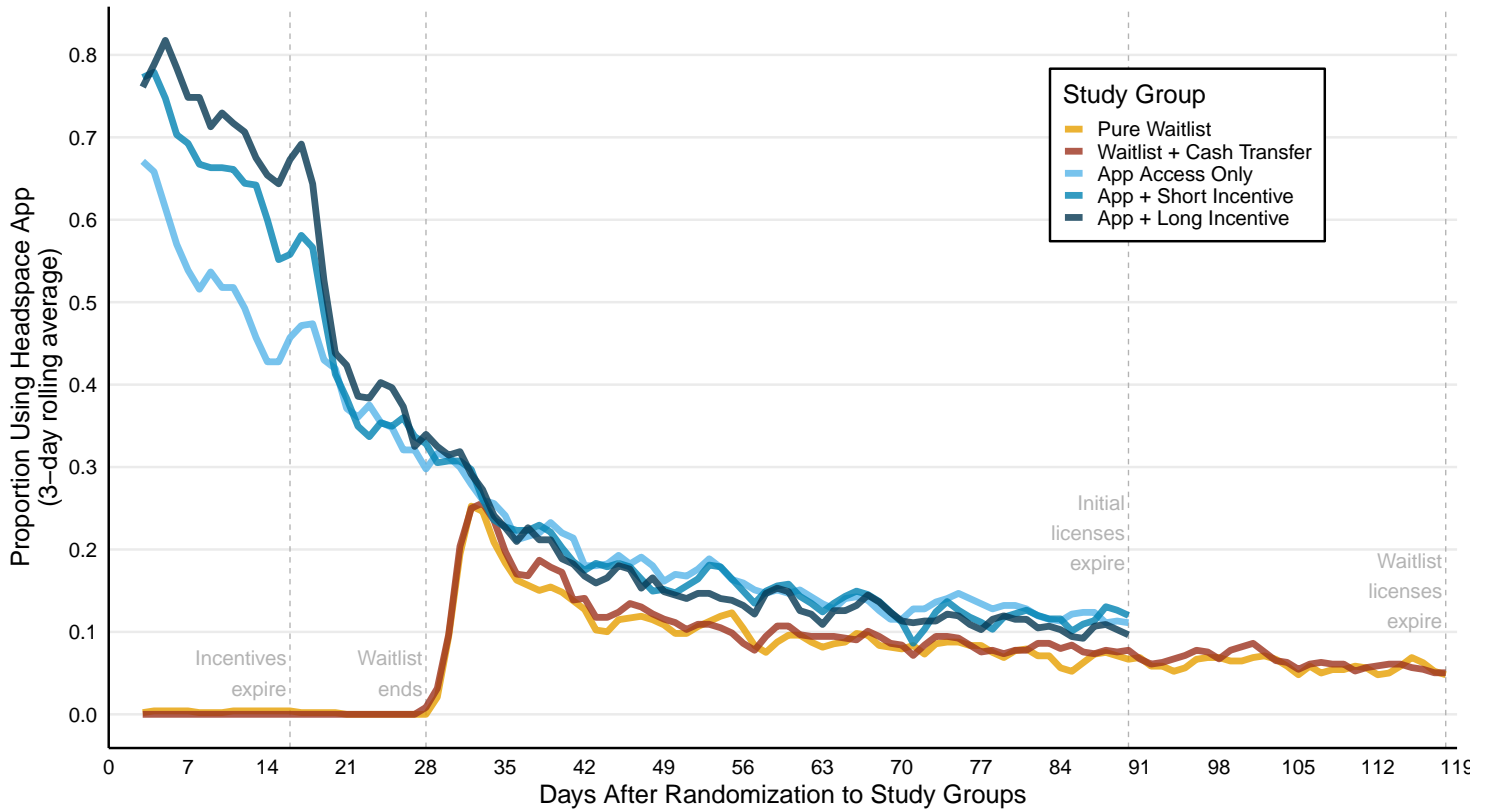
(b) Economic Tasks



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) of the first 16 days; (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.

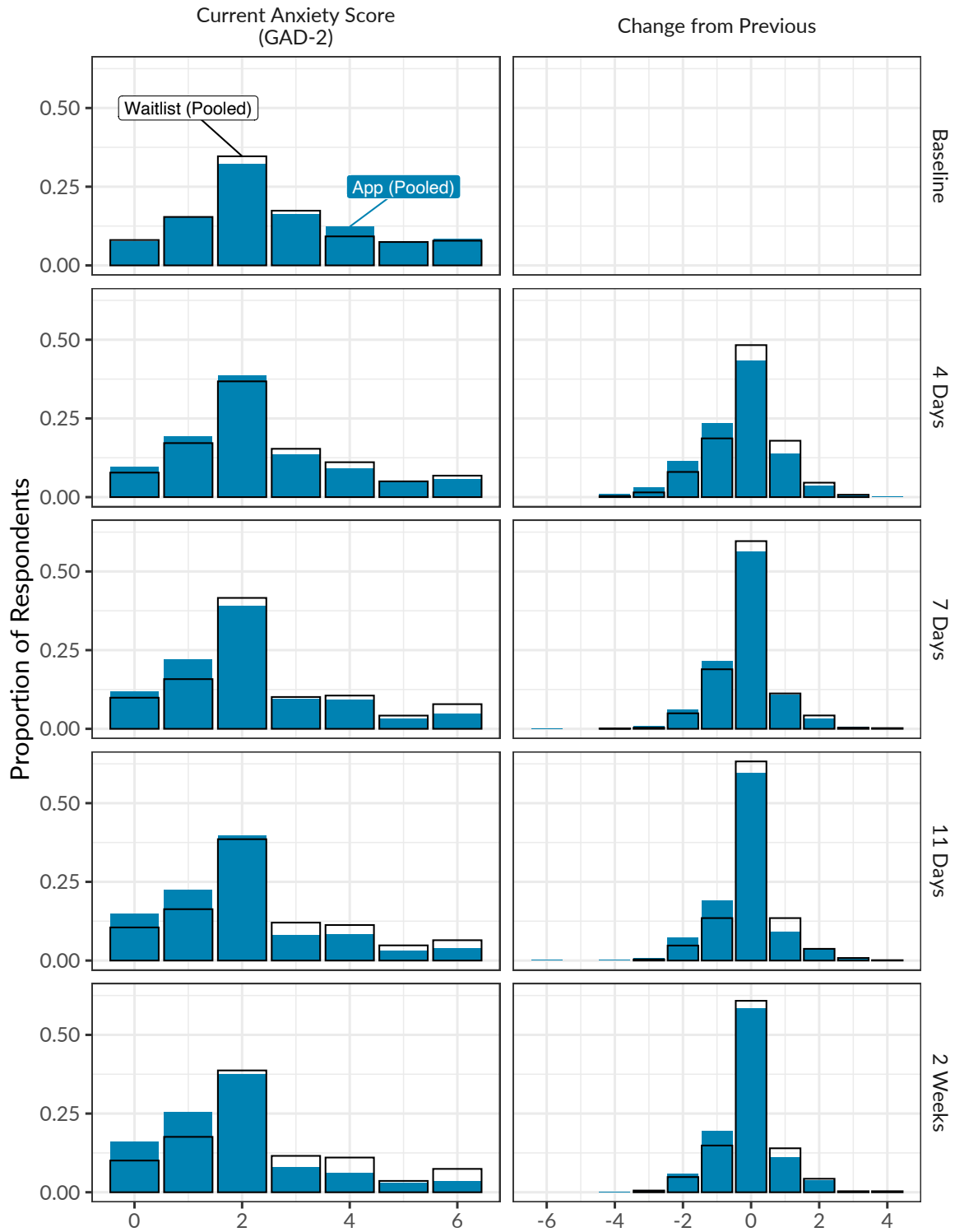
Panel B summarizes the economic tasks completed at two weeks. Those with app access are randomly assigned to receive incentives to meditate using the app right before continuing into the survey. All participants then complete tasks that have built-in distracting or emotion-inducing elements: a Stroop 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: Proportion of Participants Using Headspace in a 3-Days Rolling Window In Each Intervention Arm



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 16 days; (iii) App + Long Incentive, being offered a \$10 bonus for using the app on at least 10 days during the first 16 days; (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.

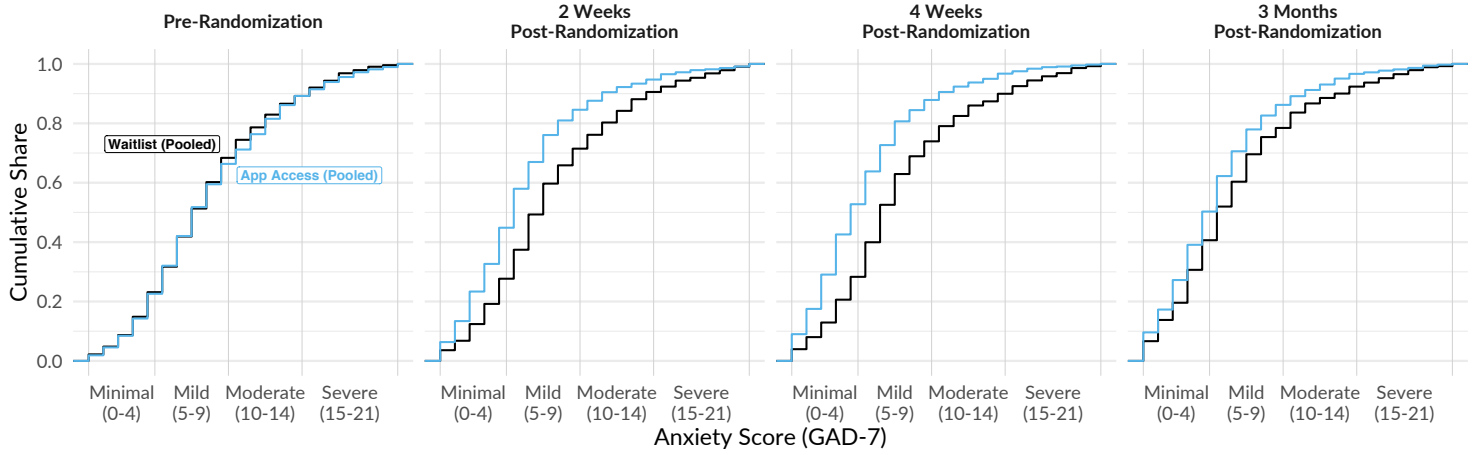
Figure 3: Distribution of Individual Anxiety Scores, and Change in Anxiety Scores, Over Time



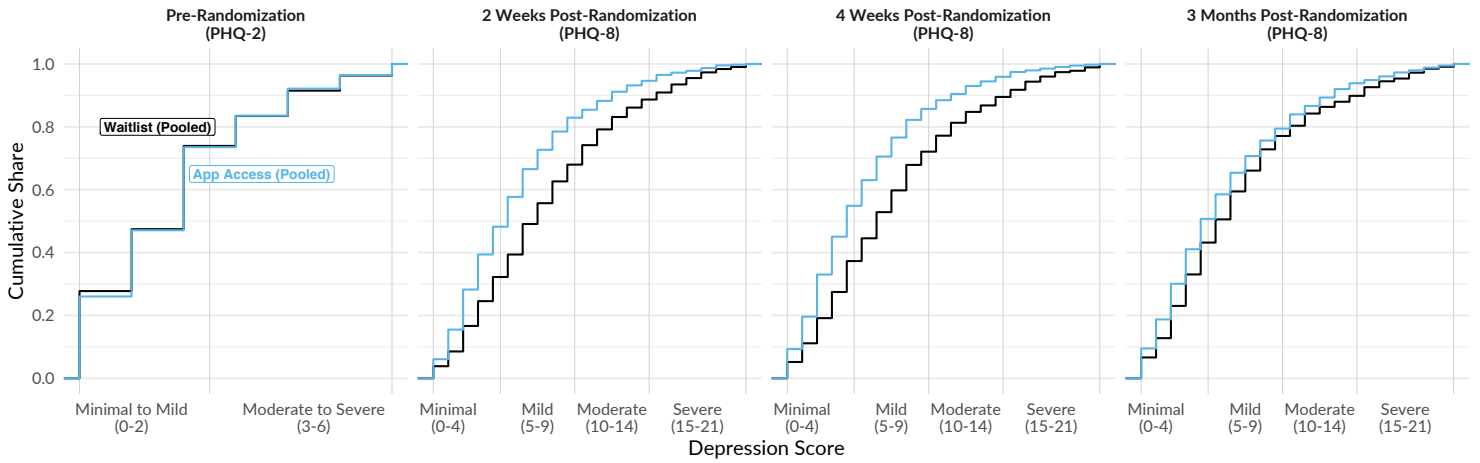
Notes: This figure presents the evolution of the GAD-2 anxiety score in the Waitlist and App Access groups during the first two weeks of the study. The GAD-2 score is calculated from the first two elements of the GAD-7 anxiety scale, and ranges from 0 (no reported symptoms of anxiety) to 6 (maximum reported symptoms of anxiety). Let $Y_{i,t}$ represent respondent i 's anxiety score at time t . The left column of the figure presents the distribution of scores $Y_{i,t}$, while the right column presents the distribution of within-person changes from the previous score, $Y_{i,t} - Y_{i,t-1}$. Each row of the figure presents outcomes at a particular point in time. Hollow bars with black outlines represent the Waitlist group, and solid blue bars represent the App Access group. For precise statistical tests, Table B.4 presents estimates of the average treatment effect at each time point.

Figure 4: Long-Term Distribution of Mental Health Scores

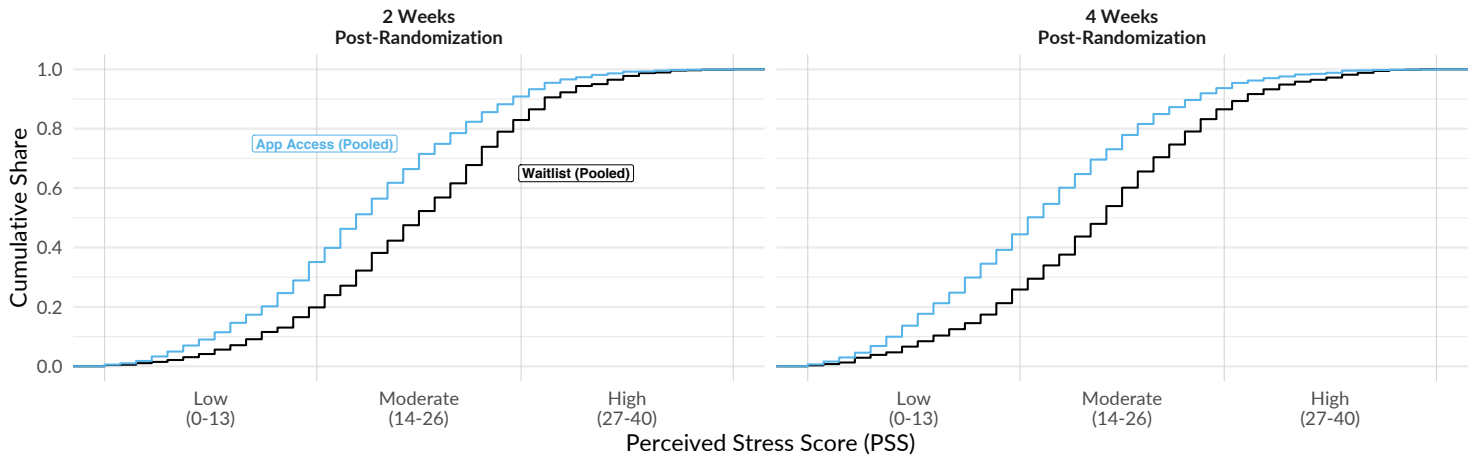
(a) Anxiety



(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.

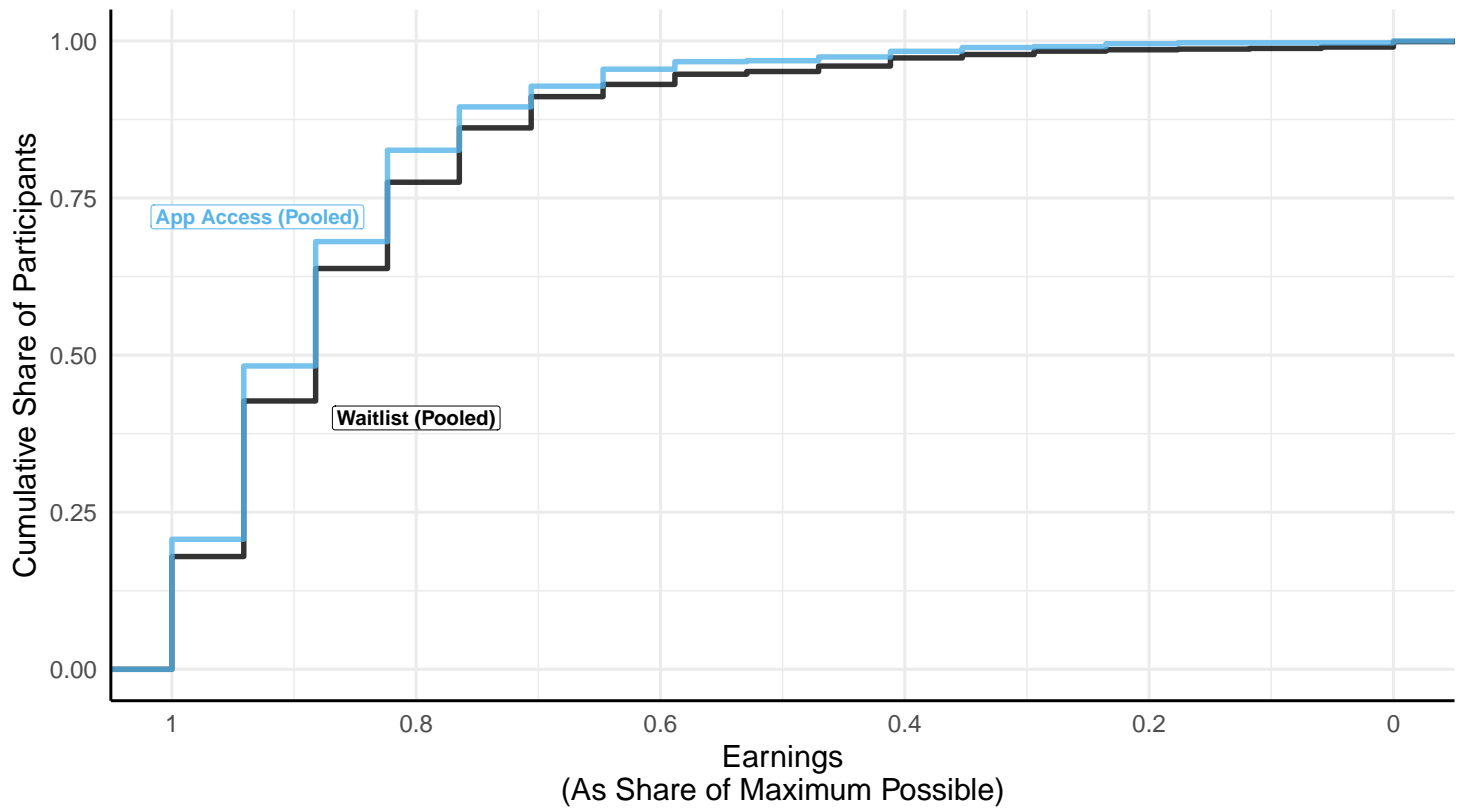


Figure 5: Distribution of Earnings in the Proofreading Task

Notes: This figure presents the empirical cumulative distribution function of earnings in the proofreading task, expressed as the share of the maximum possible earnings for flagging all 17 errors without flagging any non-errors. The App Access group does excludes participants who were offered incentives to meditate right before the session. The x-axis runs in reverse, starting from maximum earnings and diminishing to the right. Table 4 presents formal statistical tests for the average treatment effect of App Access on the proofreading score.

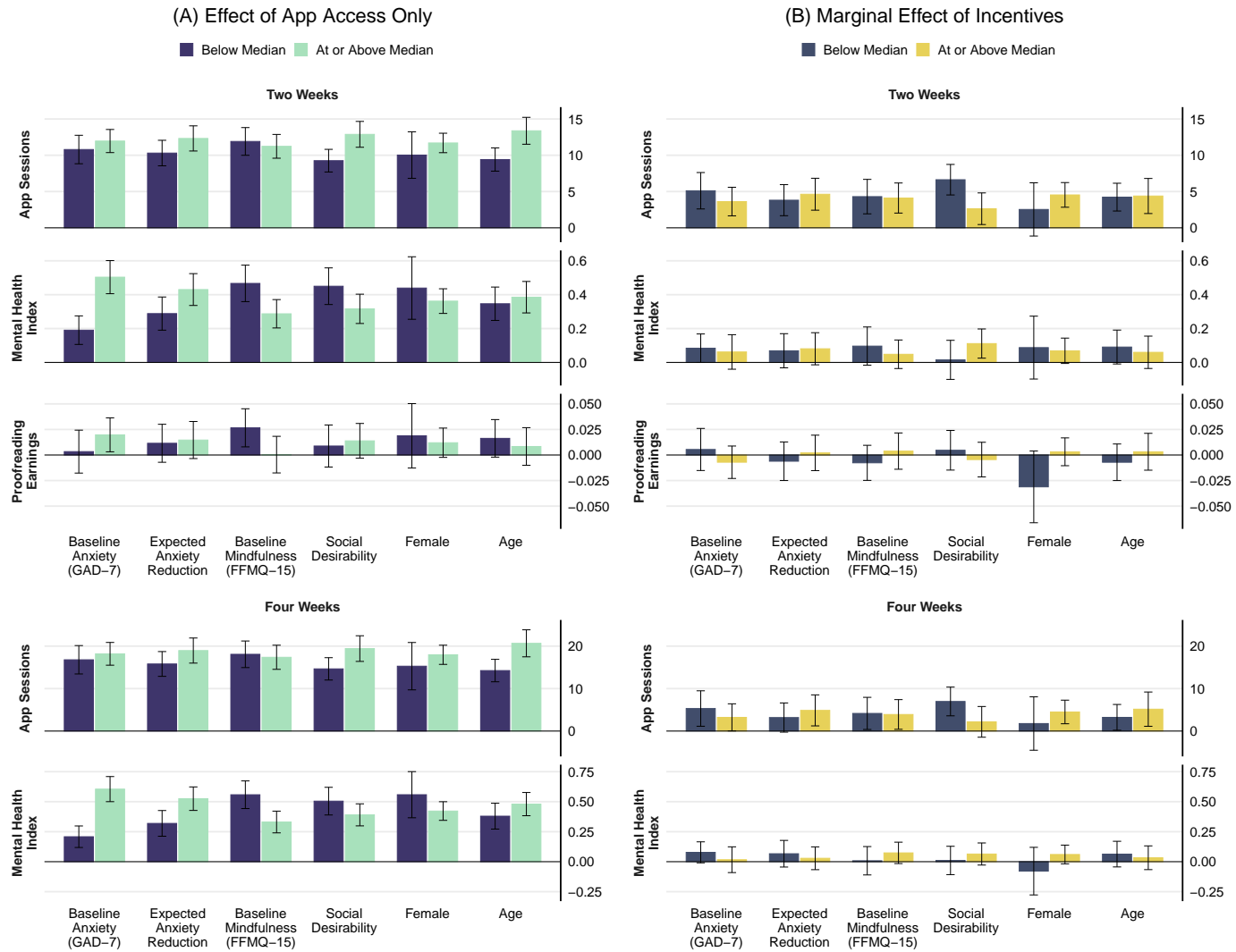


Figure 6: Heterogeneous Effects of License Provision and Incentives

Notes: This figure presents total effect of providing a Headspace license (Panel A) and marginal effect of offering usage incentives over the first 16 days (Panel B), in subgroups defined by several pre-treatment covariates. For each covariate, we divide the sample into two subsamples: those observations below the sample median, and those at or above the sample median. We then estimate equation 3 separately in each subsample. Panel A presents estimates and 95% confidence intervals for the coefficient on App Access, and Panel B presents the same for Any Incentive. All baseline covariates are continuous, except for Female, where "Below Median" refers to males and "At or Above Median" refers to females. Social Desirability refers to a 13-item version of the Marlow-Crowne Social Desirability Scale, which is one way to measure respondents' tendency to portray themselves favorably in a survey. App Sessions counts the number of distinct Headspace sessions users completed on the Headspace app within the first 16 days, when incentives are active; and Proofreading Earnings refers to the percent bonus participants received on the proofreading task, compared to the maximum possible bonus.

Table 1: Sample Characteristics and Balance

	US Adults Mean (1)	Baseline			
		Waitlist Group Mean (2)	(St. Dev.) (3)	App Access Group Diff. (4)	(St. Err) (5)
Age Group					
18-29	0.213	0.202	(0.402)	−0.010	(0.017)
30-39	0.172	0.204	(0.403)	0.014	(0.017)
40-49	0.158	0.250	(0.433)	−0.010	(0.018)
50-59	0.163	0.209	(0.407)	−0.016	(0.017)
60-69	0.150	0.109	(0.312)	0.016	(0.013)
70+	0.143	0.025	(0.157)	0.006	(0.007)
Female	0.492	0.848	(0.359)	0.019	(0.015)
Education					
No Bachelor’s degree	0.640	0.215	(0.411)	−0.024	(0.017)
Bachelor’s degree	0.225	0.345	(0.475)	0.019	(0.020)
Graduate or professional degree	0.135	0.441	(0.497)	0.004	(0.021)
Household Size	2.261	2.874	(1.328)	−0.086	(0.055)
Household Income					
\$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	(0.009)
Prefer not to answer		0.073	(0.261)	0.013	(0.011)
Race & Ethnicity					
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)	−0.002	(0.012)
Other race	0.036	0.057	(0.231)	−0.013	(0.009)
Political Party					
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)
Mental Health at Baseline					
Moderate or severe anxiety		0.316	(0.465)	0.020	(0.020)
Likely major depressive disorder		0.261	(0.439)	0.004	(0.018)
Sample Size					
N	—	955		1,429	

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. The License group contains more individuals in the \$35,000-\$74,999 household income range, and reporting a Republican political affiliation. In Figures B.2, B.3, B.4, and B.5 we present specifications adjusting for potential imbalance using debiased machine learning and show that our main conclusions do not change.

Table 2: App Usage and Effects of Usage Incentives in Short and Medium-Term

	Mean Usage				Effect of Incentives	
	No Incentives	Any Incentives	Short Incentives	Long Incentives	Any Incentive vs. None	Long Incentive vs. Short
	(1)	(2)	(3)	(4)	(5)	(6)
Any Session by Day 16	0.80 (0.40)	0.89 (0.31)	0.89 (0.32)	0.90 (0.31)	0.09 (0.02)	0.01 (0.02)
Sessions by Day 16	11.03 (13.30)	15.05 (14.09)	13.79 (13.60)	16.30 (14.46)	4.02 (0.76)	2.52 (0.91)
Minutes by Day 16	94.99 (175.08)	142.98 (220.19)	130.94 (192.13)	154.96 (244.57)	47.99 (10.73)	24.01 (14.25)
At Least 4 Days by Day 16	0.54 (0.50)	0.75 (0.43)	0.76 (0.43)	0.75 (0.43)	0.21 (0.03)	-0.01 (0.03)
At Least 10 Days by Day 16	0.20 (0.40)	0.33 (0.47)	0.24 (0.43)	0.42 (0.49)	0.12 (0.02)	0.18 (0.03)
Any Session Between Days 21-60	0.56 (0.50)	0.61 (0.49)	0.61 (0.49)	0.61 (0.49)	0.05 (0.03)	0.00 (0.03)
Sessions Between Days 21-60	11.14 (24.81)	9.63 (21.89)	10.23 (24.26)	9.03 (19.26)	-1.52 (1.34)	-1.20 (1.42)
Minutes Between Days 21-60	153.07 (406.23)	148.22 (458.48)	166.18 (559.83)	130.33 (327.46)	-4.85 (23.80)	-35.86 (29.74)
N	2384					

Notes: This table presents usage statistics of the app licenses given to participants in the License Treatment group. The top panels gathers usage statistics related to the period when incentives are in effect, while the bottom panel describes usage past the incentivization period.

- The No Incentives column presents means of various usage statistics and their standard deviations among participants who received access to the app but no further usage incentive. Any Incentives reports the same, pooling the Short Incentives and Long Incentives treatment groups. We presented these groups separately in columns 3 and 4.
- The Effect of Incentives columns report estimates of the effect of receiving either Incentives treatment on usage estimated as β_2 in Equation 1, and the marginal effect or Long Incentives relative to Short Incentives only, estimated as β_3 in Equation 2.
- “At Least N Days by Day 16” is a dummy for having used the app for at least 10 minutes on at least N separate days in the first 16 days after receiving the license. “At Least 4 Days by Day 16” is the criterion to receive a \$15 bonus in the Short Incentives treatment, while “At Least 10 Days by Day 16” is the criterion to receive a \$15 bonus in the Long Incentives treatment.

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

	Two Weeks				Four Weeks				Three-Four Months	
	Index (1)	Anxiety (2)	Depression (3)	Stress (4)	Index (5)	Anxiety (6)	Depression (7)	Stress (8)	Anxiety (9)	Depression (10)
A. Pooling Arms										
Any App Access (S.E.)	-0.366 (0.035)	-0.370 (0.039)	-0.326 (0.041)	-0.391 (0.049)	-0.434 (0.037)	-0.456 (0.040)	-0.377 (0.040)	-0.467 (0.051)	-0.233 (0.048)	-0.157 (0.050)
Any Usage Incentive (S.E.)	-0.073 (0.036)	-0.064 (0.039)	-0.076 (0.041)	-0.072 (0.051)	-0.046 (0.037)	-0.062 (0.040)	-0.045 (0.039)	-0.013 (0.052)	-0.016 (0.046)	-0.032 (0.049)
B. Separating Treatment Arms										
Cash Transfer (S.E.)	0.014 (0.037)	0.038 (0.044)	-0.011 (0.046)	0.082 (0.054)	0.049 (0.041)	0.083 (0.047)	0.029 (0.049)	0.112 (0.056)	0.075 (0.058)	0.001 (0.059)
App Access Only (S.E.)	-0.359 (0.039)	-0.351 (0.045)	-0.332 (0.046)	-0.350 (0.056)	-0.409 (0.042)	-0.415 (0.046)	-0.363 (0.047)	-0.411 (0.057)	-0.195 (0.056)	-0.156 (0.059)
Short Incentive (S.E.)	-0.435 (0.038)	-0.420 (0.045)	-0.380 (0.045)	-0.451 (0.056)	-0.454 (0.041)	-0.493 (0.045)	-0.380 (0.047)	-0.435 (0.056)	-0.207 (0.054)	-0.174 (0.058)
Long Incentive (S.E.)	-0.429 (0.038)	-0.409 (0.043)	-0.436 (0.046)	-0.393 (0.056)	-0.455 (0.041)	-0.461 (0.045)	-0.436 (0.046)	-0.414 (0.056)	-0.215 (0.054)	-0.201 (0.058)
N	2330	2330	2330	2330	2311	2312	2311	2311	2009	2004

Notes: This table presents estimated Intent-to-Treat effects of offering access to the mindfulness meditation app, as well as usage incentives, on reported mental illness over time. The estimating equation is Equation 3. We standardize outcomes by subtracting the Waitlist group mean and dividing by the Waitlist standard deviation, so that effects are in standard deviation units relative to a control mean of zero. We elicit symptoms of mental illness using common psychometric scales. Lower numbers indicate fewer or less severe symptoms. At four days, seven days, and eleven days we report anxiety using the GAD-2 scale. At two and four weeks we report an index that averages anxiety (GAD-7), depression (PHQ-8), and stress (PSS-10), after separately standardizing each score. At three-four months we report anxiety (GAD-7). Panel A is our preferred specification, combining the Long Incentives and Short Incentives treatment arms. The omitted category combines the Pure Waitlist and Waitlist Cash Transfer groups. Panel B separates the two incentive groups, in line with our pre-analysis plan. The omitted category is the Pure Waitlist group.

Table 4: Impacts of App Access, Immediate Meditation and Stressful Treatment on Earnings in the Proofreading Task

	Stressed	Proofreading Earnings			
	(1)	(2)	(3)	(4)	(5)
App Access	−0.171	0.012	0.017	0.016	0.017
(S.E.)	(0.054)	(0.005)	(0.006)	(0.007)	(0.009)
App x Immediate Meditation	0.022		−0.011		−0.002
(S.E.)	(0.055)		(0.006)		(0.008)
Stressful Task	1.094			−0.004	−0.004
(S.E.)	(0.053)			(0.008)	(0.008)
Stressful x App	0.229			−0.009	0.000
(S.E.)	(0.087)			(0.010)	(0.012)
Stressful x App x Immediate	0.222				−0.018
(S.E.)	(0.101)				(0.012)
Control Mean	1.966	0.879	0.879	0.884	0.884
(S.D.)	(1.083)	(0.135)	(0.135)	(0.132)	(0.132)
N	2257	2257	2257	2257	2257

Notes: This table presents the effect of app access, incentives to meditate immediately before the task, and assignment to the Stressful Task on earnings in the proofreading task, expressed as percentage of the maximum possible amount. As a “first stage” of the Stressful task, we present effects on self-reported feelings of being stressed before completing the proofreading task, estimated using Equation 7. We then estimate effects of access to the app, immediate meditation incentives and of the stressful task, as well as its interaction with the former, using regressions presented in Equation 4 (column 2), Equation 5 (column 3), Equation 6 (column 4) and Equation 7 (column 5).

Table 5: Effects of App Access and Immediate Meditation Incentives on Performance in Stroop Test

	Earnings		Errors		Time Spent	
	(1)	(2)	(3)	(4)	(5)	(6)
App Access	0.001	0.003	-0.033	-0.084	0.179	-0.006
(S.E.)	(0.004)	(0.005)	(0.123)	(0.145)	(0.337)	(0.400)
App x Immediate Meditation		-0.004		0.100		0.371
(S.E.)		(0.006)		(0.150)		(0.438)
Control Mean	0.741	0.741	1.439	1.439	67.817	67.817
(S.D.)	(0.125)	(0.125)	(3.156)	(3.156)	(9.051)	(9.051)
N	2256	2256	2256	2256	2256	2256

Notes: This table details the treatment effects of offering access to the app and incentives to use it immediately before the experimental decision-making session, on the number of errors made in the Stroop test as well as the total time spent on it. Earnings is a each participant's total reward, measured as a proportion relative to the largest award earned. For example, the control mean of 0.741 implies that the participants in the control group earned on average 74.1% as much as the highest-earning participant overall. Errors refers to the total number of mistakes a participant made, either by selecting the wrong answer or letting time elapse. Time Spent refers to the total time, in seconds, a participant spent on the test.

Table 6: Impacts of App Access, Immediate Meditation and Stressful Treatment on the Certainty Premium

	Stressed	Certainty Premium			
	(1)	(2)	(3)	(4)	(5)
App Access	−0.070	−0.005	−0.006	−0.034	−0.022
(S.E.)	(0.051)	(0.013)	(0.015)	(0.018)	(0.021)
App x Immediate Meditation	0.062		0.002		−0.025
(S.E.)	(0.052)		(0.017)		(0.023)
Stressful Task	0.352			−0.037	−0.037
(S.E.)	(0.050)			(0.019)	(0.019)
Stressful x App	0.023			0.060	0.032
(S.E.)	(0.077)			(0.026)	(0.030)
Stressful x App x Immediate	0.187				0.056
(S.E.)	(0.088)				(0.033)
Control Mean	2.032	0.143	0.143	0.163	0.163
(S.D.)	(1.040)	(0.298)	(0.298)	(0.308)	(0.308)
N	2257	2247	2247	2247	2247

Notes: This table presents the effect of the Stressful Task on the certainty premium, estimated in the Risk Choices module. As a “first stage” of the Stressful task, we present effects on self-reported feelings of being stressed before completing the risk choices, estimated using Equation 7. We then estimate effects of access to the app, immediate meditation incentives and of the stressful task, as well as its interaction with the former, using regressions presented in Equation 4 (column 2), Equation 5 (column 3), Equation 6 (column 4) and Equation 7 (column 5).

- “Stressed” are the standardized self-reported answers to a Likert scale asking how the participants feel right now, between the second dose of the Stressful task and the Risk Choices module. The lowest number corresponds to the answer “Very calm, very relaxed” and the highest number corresponds to “Very upset, very stressed”.
- The Certainty Premium is constructed using the switching points in two multiple-price lists as in Callen et al. (2014). The first elicits the probability $P_{\text{certainty}}$ that makes a participant indifferent between a lottery that pays \$30 with probability $P_{\text{certainty}}$ and \$0 otherwise, and a certain \$10. The second multiple price list elicits the probability $P_{\text{uncertainty}}$ that makes a participant indifferent between a lottery that pays \$30 with probability $P_{\text{uncertainty}}$ and \$0 otherwise, and a lottery paying \$10 or \$30 with equal probability. Certainty Premium is then obtained by assuming expected utility in each of the two decisions above, assuming $u(30) = 1$ and $u(0) = 0$, deriving the implied value of $u(10)$ for each of the two decisions and taking $CP = u_{\text{certainty}}(10) - u_{\text{uncertainty}}(10)$. We also present effects on the individual components of the Certainty Premium, $P_{\text{certainty}}$ and $P_{\text{uncertainty}}$, in Table B.7.

Appendix

A Model and Proofs

We represent the agent’s attention as a vector m where $m_i \in [0, 1]$ represents the attention paid to the attribute x_i . Given an attention vector, the agent estimates attributes to be

$$\hat{x}(x, m) = \left(m_1 x_1 + (1 - m_1) x_1^d, \dots, m_n x_n + (1 - m_n) x_n^d \right).$$

The agent then takes the action that maximizes her utility, given her reading of the situation $x(m)$.

$$a(x, m) = \arg \max_a u(a, \hat{x}(x, m)).$$

Attention Allocation Problem. Exerting the attention vector m is costly. The agent optimally select her attention vector to solve:

$$m^* = \arg \max_m \mathbb{E} [u(a(x, m), x)] - C(m), \tag{A.1}$$

where the expected value is with respect to x , $C(m)$ is the cost of attention. To model that attention paid to different dimensions are substitutes, we assume $C(m) = \kappa (\sum_i m_i)^2$, with $\kappa \geq 0$ an attention cost parameter.³⁵

True and Distorted Preferences. We now specify a set of utility functions for the agent. We assume that the attributes vary along two dimensions, $x = (x_1, x_2)$, where x_1 are the relevant attributes of the decisions, and x_2 are irrelevant, distracting attributes such as passing thoughts or emotions of the agent at that time. The agent’s welfare-relevant preferences are given by the utility function

$$u(a, x) = -(a - x_1)^2,$$

that is, the agent tries to match the value of the attribute x_1 . Instead, an agent with emotional naïveté μ solves Equation A.1 and selects a using the following distorted utility function:

$$\tilde{u}(a, x, \mu) = -(a - x_1 - \mu x_2)^2,$$

incorrectly taking into account the irrelevant attribute x_2 .

Solving for the Attention Allocation. We can simplify Equation A.1 given the distorted

³⁵Gabaix (2014) aims for tractable solutions for the attention levels, and assumes that attention costs are additively separable in the dimensions of m . By contrast, our focus is on modeling how attention capture by irrelevant emotions can reduce attention to the relevant aspects of the problem, and this requires assuming that attention to different dimensions enter as substitutes in the attention cost function.

utility function:

$$a(x, m) = \hat{x}_1 + \mu\hat{x}_2 = m_1x_1 + (1 - m_1)x_1^d + \mu m_2x_2 + \mu(1 - m_2)x_2^d$$

To find m^* , we begin by adding a constant term to the maximand in Equation A.1:

$$m^* = \arg \max_m \mathbb{E} [u(a(x, m), x) - u(a(x, 1), x)] - C(m),$$

Then, we have that:

$$\begin{aligned} & \mathbb{E} [u(a(x, m), x) - u(a(x, 1), x)] \\ &= \mathbb{E} [(a(x, 1) - x_1 - \mu x_2)^2 - (a(x, m) - x_1 - \mu x_2)^2] \\ &= \mathbb{E} [(a(x, 1) + a(x, m) - 2x_1 - 2\mu x_2)(a(x, 1) - a(x, m))] \\ &= \mathbb{E} [((1 + m_1)x_1 + \mu(1 + m_2)x_2 + (1 - m_1)x_1^d + \mu(1 - m_2)x_2^d - 2x_1 - 2\mu x_2) \\ &\quad * ((1 - m_1)x_1 + \mu(1 - m_2)x_2 - (1 - m_1)x_1^d - \mu(1 - m_2)x_2^d)] \\ &= \mathbb{E} \left[\left((1 - m_1)(x_1 - x_1^d) + \mu(1 - m_2)(x_2 - x_2^d) \right)^2 \right] \\ &= \mathbb{E} \left[((1 - m_1)(x_1 - x_1^d))^2 \right] + \mathbb{E} \left[(\mu(1 - m_2)(x_2 - x_2^d))^2 \right] + 2\mu(1 - m_1)(1 - m_2)\mathbb{E} [(x_1 - x_1^d)] \mathbb{E} [(x_2 - x_2^d)] \\ &= (1 - m_1)^2\sigma_1^2 + \mu^2(1 - m_2)^2\sigma_2^2, \end{aligned}$$

since x_1 and x_2 have mean x_1^d and x_2^d respectively, and defining $\sigma_i^2 = \mathbb{E} [(x_i - x_i^d)^2]$, the variance of attribute i . To keep derivations light, we then define the weighted variances $s_1 = \sigma_1^2$ and $s_2 = \mu^2\sigma_2^2$.

Replacing the utility and the attention cost functions in Equation A.1, we obtain:

$$m^*(\mu, \kappa) = \arg \min_{m \in [0, 1]^2} s_1(1 - m_1)^2 + s_2(1 - m_2)^2 + \kappa(m_1 + m_2)^2,$$

To solve this constrained optimization problem, we introduce Lagrange multipliers associated with each inequality constraint on m_1 and m_2 , defining the following Lagrangean:

$$\mathcal{L} = s_1(1 - m_1)^2 + s_2(1 - m_2)^2 + \kappa(m_1 + m_2)^2 + \lambda_1 m_1 + \lambda_2 m_2 + \nu_1(1 - m_1) + \nu_2(1 - m_2)$$

Taking the first-order conditions w.r.t. m_1 and m_2 , we obtain:

$$2s_1(m_1 - 1) + 2\kappa(m_1 + m_2) + \lambda_1 - \nu_1 = 0 \tag{A.2}$$

$$2s_2(m_2 - 1) + 2\kappa(m_1 + m_2) + \lambda_2 - \nu_2 = 0 \tag{A.3}$$

Inspecting these conditions reveals the relevant comparative statics:

- Increasing s_1 makes the left side of A.2 negative and leaves A.3 unchanged. A corresponding increase in m_1 and decrease in m_2 (if possible) will then satisfy the first-order condition.

- Increasing μ (and therefore s_2) makes the left side of A.3 negative but leaves A.2 unchanged. An increase in m_2 and decrease in m_1 (if possible) will satisfy the first-order condition.
- Increasing κ shifts the left sides of both equations in the positive direction. The sum $m_1 + m_2$ must decrease to satisfy the first-order condition.

Supplementary Materials

Table of Contents

B	Supplementary Materials	58
B.1	PAP Deviations	58
B.2	Information Avoidance and Aversion to Low-Probability Losses	59
B.3	Balance	61
B.4	Attrition	63
B.5	Usage And Beliefs	65
B.6	Multiple Hypothesis Testing	69
B.7	Appendix Figures	70
B.8	Appendix Tables	80

B Supplementary Materials

B.1 PAP Deviations

We now discuss our rationale for modest departures from the pre-registration.

Double/Debiased Machine Learning (DML).

PAP Text: We collect a variety of pre-treatment information from pre-randomization surveys. In addition to the conventional specifications detailed above, we will report an additional set of analysis that adjust for these covariates using a double machine learning approach (Chernozhukov et al., 2018). In particular, because most covariates we observe are categorical, we will report DML results where we perform partialling out using (1) regression tree methods (e.g., random forest or gradient boosted trees), and (2) an ensemble of a penalized linear model and a tree-based method. These exercises will incorporate all available covariates, including strata fixed effects and pre-treatment measures of outcomes where available. For each outcome, our preferred DML specification will be the model that maximizes cross-validated goodness of fit.

Deviation: We pre-registered a double machine learning exercise to assess the sensitivity of our results to potential covariate imbalance. Our PAP outlines a procedure that (1) performs DML with a tree-based algorithm and an ensemble method, (2) reports results from the best-fitting of these. We deviate to present more information in response to common questions: we report results from both the tree-based method and the ensemble (rather than just the best-fitting one), and also report results based on penalized regression.

Mindfulness Mediation Analysis.

PAP Text: If FFMQ-15 score is affected, we will conduct mediation analysis on other affected outcomes.

Deviation: We do not conduct a mediation analysis. In Appendix Table B.10 we find that app access increases scores on every subscale of the FFMQ-15 scale. Performing a mediation analysis would require choosing a functional form to relate each subscale to our main outcomes, which we lack the information to credibly do. While writing our PAP we anticipated effects on some but not all FFMQ subscales, which would have made this analysis more straightforward. For these reasons, we do not perform a mediation analysis.

B.2 Information Avoidance and Aversion to Low-Probability Losses

In this appendix, we describe the design details about how we measure secondary outcomes that help us narrow down mechanisms through which mindfulness meditation affects behavior. We then report results.

B.2.1 Measurement

Information Avoidance Individuals have documented tendencies to avoid useful but potentially unpleasant information (Golman et al., 2017), presumably because they would prefer to avoid facing negative emotions from anticipating and receiving the information. To measure the effects on information avoidance, we allow participants to opt in to receiving as many of the following links (presented in a random order) as they want at the end of the survey: (i) a life expectancy calculator based on simple questions; (ii) risk factors for developing dementia; (iii) the risk of one’s job of being replaced by automation; and (iv) an individual calculator of financial risk in retirement. To shed some light on the mechanism behind any effect on information avoidance, we note that mindfulness meditation may (i) make individuals more or less willing to avoid the possible negative emotions that they anticipate from the information and (ii) affect individuals’ utility from the future instrumental value of the information. To help isolate effects from (ii), we design a short, incentivized information acquisition game where the information has instrumental value but is emotionally neutral. Participants are told that they will be able to choose whether to purchase an imaginary stock whose value is determined by three computerized coin tosses: if at least two coin flips come up as heads, purchasing the stock will add \$2 to their bonus, but if two coins or more come up as tails, the same amount will be subtracted.

The participants have the option to receive one of three pieces of information. The first is the outcome of the first coin toss (which makes the purchase decision an easy one for moderate levels of risk aversion). The second and third are two decision-irrelevant pieces of information: the value of another bonus they have already received or the age of the oldest tree in the world. This decision is binding with a small probability: most participants receive all three pieces of information, and we verify that when they do, there is no difference in the propensity to purchase the stock so that any difference in the choice of information cannot be explained by different plans to purchase the stock *ex ante*.

Small Probability Loss Aversion It has been argued that people sometimes consider a risky prospect based on how it makes them feel rather than rationally evaluating the possible payoffs and their probabilities (Loewenstein et al., 2001). In particular, they may focus on the value of the worst case payoff and decouple outcomes from their associated probabilities, which may correspond behaviorally to over-weighting small probabilities. Because mindfulness meditation is hypothesized to help regulate emotions and reduce reactivity to them, we test for effects on the propensity to refuse a small-stakes bet with excellent expected value but a salient low probability loss: a bet that earns \$1 with a 99 percent probability or loses \$10 otherwise. This decision mirrors documented real-life choices to purchase high-premium insurance against low probability moderate risks (Sydnor,

2010).

This outcome is designed to be a simple decision rather than a more involved menu because we are less interested in measuring effects on small probability loss aversion than in generating a proof of concept that meditation may affect such decisions, and also because the simplicity makes the decision more natural. Given that we measure the uptake of risky prospects with higher probabilities of similar losses and lower expected value later in the survey, we can identify whether any increased propensity to refuse this lottery reflects increased sensitivity to small probability loss-framed outcomes rather than changes in risk preferences.

B.2.2 Results

Effects of Access to the App We first find that app access alone increases the propensity to avoid small probability, salient losses, but the effect is driven by participants who received incentives to meditate immediately before the survey. In the absence of an immediate meditation, app access only does not affect this outcome (column 2, first row). We also find that app access alone has no effect on unpleasant information avoidance behavior. We do not find evidence of detectable effects on an index of unpleasant real information avoidance (columns 3 and 4). However, incentives to meditate immediately before the survey do impact these decisions: we now turn to these immediate effects.

Immediate Effects of an Incentivized Meditation Session Incentives to meditate at the beginning of the survey increase the propensity to refuse a small-stakes lottery with high expected value but salient low probability losses by 5.0 percentage points, from a propensity of 33 percent among participants with app access who are not incentivized to meditate (Table B.8). This is unlikely to be driven by a change in risk preferences, since the immediate meditation treatment does not reduce risk aversion in our other measures that do not make salient the possibility of a loss. Rather, in line with our model, we interpret this effect as evidence of a temporarily increased emotional sensitivity.

Finally, as further evidence of this mechanism, the immediate meditation incentives increase avoidance of emotionally unpleasant information by 3.8 percentage points. The effect is mostly driven by the piece of information about financial risk in retirement, that participants who received the immediate meditation treatment are 8.1 percentage points more likely to decline (Table B.9). However, effects on other pieces of information are directionally similar. As a placebo check, we do not find a significant effect from the immediate meditation session on the propensity to demand decision-relevant information in our neutral information demand game (see Section B.2).

B.3 Balance

A natural concern in any study of causal effects is that estimated treatment effects may reflect idiosyncrasies of the units being studied, rather than a true effect of the intervention. Table 1 compares treatment and control at baseline, and Appendix Table B.1 extends this comparison to our followup surveys. As expected, stratified randomization in a large sample is a powerful safeguard against potential imbalance, and we consider the observed differences between our treatment and control arms to be small.

However, to investigate the sensitivity of our main findings to potential imbalance, we adopt the double/debiased machine learning (DML) procedure of Chernozhukov et al. (2018). We choose this approach to flexibly and systematically account for potential imbalance, rather than taking manually searching across regression specifications that adjust for various combinations of covariates.

At a high level, this procedure involves using covariates to “explain away” variation in the outcome and treatment assignment. Any remaining variation in the outcome is demonstrably not due to variation in observed variables; it must either be due to treatment or to unobserved variables. In our setting, randomization implies balance on the unobserved variables and suggests that the remaining variation corresponds to treatment effects.

Economists conventionally make the above argument via linear regressions of the form

$$Y_i = \beta' D_i + \gamma' X_i + \epsilon_i,$$

where D_i are treatment variables and X_i are observed covariates. Our double/debiased machine learning procedure replaces the linear term $\gamma' X_i$ with a more general function:

$$Y_i = \beta' D_i + f(X_i) + \epsilon_i.$$

Adjusting for a flexible $f(X_i)$ requires estimating the conditional expectations $E[Y|X]$ and $E[D|X]$. We estimate these conditional expectations function three ways:

- Elasticnet: a regularized linear regression, as implemented in the R package *glmnet* version 4.1-3. We tune the regularization (λ), mixing (α), and relaxation (γ) parameters.
- Gradient-Boosted Trees: an ensemble of decision trees, as implemented in the R package *lightgbm* version 3.2.1. We set a learning rate of 0.025 and tune the number of leaves, as well as the minimum number of observations per leaf.
- Ensemble: a combination of elasticnet and gradient-boosted trees. We train a gradient-boosted tree model to predict the residuals from an elasticnet regression, tuning each component as described above.

We present all three methods —particularly the ensemble —to remain relatively agnostic about the functional form of the conditional expectations.

More explicitly, the procedure involves three steps:

1. **Partition observations into folds for cross-fitting.** We divide observations into 10 folds, using stratified random assignment to improve the comparability of observations across folds. For a given regression, we construct strata based on all treatment assignment variables, as well as an indicator for whether the outcome is above its median value. For example, considering the regression

$$\text{Anxiety}_i = \beta_1 \text{License}_i + \beta_2 \text{AnyIncentive}_i + f(X_i) + \epsilon_i,$$

we define strata based on the full crossing of $\text{License} \times \text{AnyIncentive} \times 1(\text{Anxiety} > \text{Median})$, randomly assigning observations within each stratum across the 10 folds.

2. **Obtain cross-fitted residuals for the outcome and treatment assignments.** That is, for all observations in fold k , use observations from all the other folds to estimate the conditional expectation functions for the outcome and for treatment assignment. The cross-fitted residual is the difference between the observed value of a variable and its conditional expectation. In general, for the regression:

$$Y_i = \beta' D_i + f(X_i) + \epsilon_i$$

we compute:

$$\check{Y}_i = Y_i - \hat{E}_{-k[i]}[Y_i|X_i] \quad \check{D}_i = D_i - \hat{E}_{-k[i]}[D_i|X_i]$$

where $\hat{E}_{-k[i]}$ is an estimated conditional expectation, excluding observations in the same cross-validation fold as observation i .

In all cases where algorithms require tuning, we perform nested cross-validation to select the tuning parameters (e.g., $\{\lambda, \alpha, \gamma\}$ for elasticnet). For example, suppose we are generating the cross-fitted residuals for observations in fold 1. We perform a second round of cross-validation within folds 2–9 to choose these parameters, and then use all observations in folds 2–9 to fit the elasticnet that predicts the conditional mean for observations in fold 1.

3. **Regress the residualized outcome on the residualized treatment.** Collecting observation across folds, we estimate the ITT via OLS:

$$\hat{\beta} = \underset{b}{\operatorname{argmin}} \sum_{i=1}^n (\check{Y}_i - b' \check{D}_i)^2$$

We present DML-adjusted treatment effects in Appendix Figures B.2, B.3, B.4, and B.5. In all cases, we find that adjusting flexibly for observed covariates has a negligible effect on point estimates and offers minute improvements in precision. We expect that this is because, in a randomized experiment, treatment is independent of the covariates leaving little room for omitted variables bias.

B.4 Attrition

We attempt to minimize attrition in our study through extensive pre-screening and routine reminders. Still, some of our participants do not respond to each followup survey. In this section, we document the extent of attrition, its implications for the internal validity of this study, and finally present worst-case bounds on treatment effects. We find that participants who don't respond at two weeks tended to perform more poorly on the baseline Stroop and proofreading tasks, and those who don't respond at three months have higher anxiety scores in the waitlist than in the treatment group.

First, we present the sample size and attrition rates in each of our followup surveys. Table B.2 contains the relevant statistics. Overall attrition at our main two and four week endline surveys remains below 5 percent, though it is higher in the App Access Group than in the Waitlist.

Attrition always reduces the information available to us, but is particularly concerning if it biases estimated treatment effects. This may happen if non-respondents differ from respondents overall (so that treatment effect estimates do not generalize to the entire study population), or differ in the treatment and control groups (reducing the comparability of the treatment and control groups). We follow the recommendations in Ghanem et al. (2020) to gauge the possibility of such biases, comparing the baseline responses for treatment and control participants.

Figure B.6 presents these tests. Each panel presents a test of internal validity for an outcome at a particular time horizon. We perform these tests separately because attrition may be problematic for only some outcomes (e.g., if attrition is differential by proofreading performance but not mental health status), or for only some time horizons (e.g., if long-term attrition is differential by mental health status, but short term attrition is not).

First, we test whether treatment and control respondents are comparable on average by comparing the mean baseline outcomes for treatment and control respondents, as well as for attriters. Visually, this is a joint comparison of the orange circle against the blue circle, and the orange square against the blue square. Formally, we fit the regression

$$Y_i = \sum_{s=1}^S \{\delta^s + \beta^s \text{Responded}_{it}\} 1\{S_i = s\} + \pi_{10} \text{AppAccess}_i + \pi_{11} \text{AppAccess}_i \times \text{Responded}_{it} + \epsilon_i \quad (\text{B.1})$$

where Y_i is a baseline measure of the outcome for a given column. The term Responded_{it} is an indicator equal to 1 if respondent i completed the given question at time t . We then test the null hypothesis $\pi_{11} = \pi_{10} = 0$.

Second, we test whether respondents are on average representative of the study population. Visually, this corresponds to asking whether any of the baseline means differ within a panel. Formally

we fit the regression

$$Y_i = \sum_{s=1}^S \delta^s 1\{S_i = s\} + \pi_{01} \text{Responded}_{it} + \pi_{10} \text{AppAccess}_i + \pi_{11} \text{AppAccess}_i \times \text{Responded}_{it} + \epsilon_i. \quad (\text{B.2})$$

We then test the null hypothesis $\pi_{01} = \pi_{10} = \pi_{11} = 0$.

As with many identification tests, reading these results requires an appreciation of nuance. While smaller p -values indicate that attrition is statistically detectable, they do not inform us of the size of this bias. Conversely, larger p -values only imply that there are no detectable differences between average outcomes for the groups being compared, not that the sample guarantees valid estimates of average treatment effects.

We find no strong evidence that attrition should contaminate our study of short-term average treatment effects on mental health. However, by three to four months attrition is higher among treatment group participants who began the study with higher self-reported anxiety. We expect that this makes mindfulness seem more beneficial than it may actually be, on the assumption that the non-respondents are those who experienced the fewest benefits.

We do find evidence that non-respondents tend to have earned less money on our baseline productivity tasks. That said, treatment and control respondents are still comparable to each other. Put another way, attrition induces selection on productivity, but the selection is similar within the treatment and control arms. This limits the interpretation of our analysis of productivity to respondents (roughly 95% of participants); we can't confidently extend these conclusions to the other 5 percent who do not respond.

Finally, we recognize that it is common to bound treatment effects in the presence of attrition. One popular procedure comes from Lee (2009), and involves manually equalizing the attrition rates in treatment and control by eliminating the most extreme values from the more responsive group. Under the assumption that propensity to respond is monotonic in both groups, this allows us to calculate bounds on the treatment effect. We calculate and present such bounds in table B.3, using Imbens and Manski (2004)' approach for estimating confidence intervals for partially identified parameters. Though we present these bounds, we emphasize that our treatment and control respondents appear comparable to each other, and that these bounds are worst-case.

B.5 Usage And Beliefs

B.5.1 Local Average Treatment Effects

Our focus on the Intent to Treat effects in the main text leaves open tempting questions: what are the effects of mindfulness on those who we induce to meditate? And, how much improvement does another day of practicing meditation deliver? Formally the first question asks for the local average treatment effect (LATE) of the app, and the second asks us to draw a dose-response curve relating time spent meditating to our main outcomes.

We can answer the first question by estimating LATE using two-stage least squares. Drawing the dose-response curve is more involved. In principle, we can use comparisons between our incentive arms to sketch the relationship between time meditating and the outcomes. Our short incentives and especially our long incentives increase time spent meditating above the app-only arm, so this exercise effectively involves comparing the average time spent meditating and change in outcomes for each of the treatment groups.

However, there are several reasons that outcomes may differ across our incentive arms:

1. **Complier variation.** Each incentive arm may induce meditation from a different group of individuals, and effects may vary by group. For example, suppose participants know the effect meditation will have on them personally. In this case, those who meditate will be the ones for whom the benefit outweighs the cost. Cash incentives will then induce meditation from participants who expect smaller benefits.
2. **Type of meditation.** Headspace offers a variety of modules, and each incentive arm may induce participants to explore a different set of meditation offerings. For example, offering incentives for 10 days of meditation may encourage participants to choose a 10-day module, if they believe it will be easier to maintain commitment with such a course.
3. **Amount of meditation.** If meditation has cumulative effects, then inducing more (less) meditation should induce larger (smaller) treatment effects.

So, to determine whether we can draw a dose-response curve, we first need to ask whether compliers are comparable across incentive arms, and whether they engage in different kinds of meditation.

We find that compliers for each treatment arm are similar to each other at two and four weeks after intervention, before waitlist participants receive the app. However, cash incentives appear to increase take-up of “Basics” and “Everyday Headspace” modules, and decrease take-up of modules that aid in falling asleep.

Characterizing compliers is straightforward before the waitlist participants receive the app. We assume one-sided noncompliance: that participants use the app only after we grant them access to it. Under this assumption, compliers are simply those who have any recorded meditation sessions in our administrative data. Figure B.7 presents baseline characteristics for compliers. Compliers in each study group are highly similar to each other. This is unsurprising given high take-up in these

arms: with over 80% of participants in each arm using the app in the first four weeks, there is little room for these populations to differ.

Next, we investigate whether the incentive arms induced different types of meditation. To measure this, we divide Headspace sessions into categories based on how they are presented in the app. This corresponds roughly to the menus or search terms a user would use to access the session.

Headspace categorizes its sessions into various groups. For example, the “Basics” group is designed to introduce users to the core tenets of mindfulness. Figure B.8 presents the share of each study arm’s meditation sessions that fall into several prominent groups. We find some heterogeneity: the app access groups (and especially the incentive groups) were more prone to use Basics and Everyday Headspace sessions more than the Waitlist groups, who tended to use sleep-focused sessions and assorted mindful activities, such as deep breathing.

Tables B.14, B.15, and B.16, present the TSLS estimates of LATE considering different definitions of the endogenous variable. Our second-stage regression is of the form:

$$Y_i^{\text{post}} = \beta_1^{(2)} \text{Meditation Amount}_i + \delta_{\text{stratum}}^{(2)} + \gamma^{(2)} Y_i^{\text{pre}} + \epsilon_i^{(2)}. \quad (\text{B.3})$$

The endogenous regressor $\text{Meditation Amount}_i$ measures the amount of meditation respondent i had completed before the survey. We use the superscript (2) to distinguish these coefficients from their analogs in our first-stage equation, which instruments for meditation amount using treatment assignment:

$$\begin{aligned} \text{Meditation Amount}_i = & \beta_1^{(1)} \text{App Access Only}_i + \beta_2^{(1)} \text{Short Incentives}_i + \beta_3^{(1)} \text{Long Incentives}_i \\ & + \delta_{\text{stratum}}^{(1)} + \gamma^{(1)} Y_i^{\text{pre}} + \epsilon_i^{(1)}. \end{aligned} \quad (\text{B.4})$$

Because we assume one-sided compliance during the first four weeks of the study, all estimates in this period involve scaling the ITT estimates by the extent of compliance: the share of people who meditate at all, the number of days per week that they meditate, or the number of minutes per day that they meditate.

We prefer to focus on the first estimate, and to interpret it as the average treatment effect among compliers. It’s tempting to read the estimates in Tables B.15 and B.16 as the marginal effect of another day’s meditation or another 10 minutes of meditation. However, we stress that these estimates are linear approximations to the true relationship between meditation and well-being, which may be highly nonlinear. For example, treatment effects could follow a sigmoid shape where initial sessions have a smaller marginal effect (as a user is learning the skill), as do final sessions (once the user attains mastery). We lack exogenous variation to make plausible claims about time spent meditating.

B.5.2 Beliefs and Valuation

We now briefly discuss how participants’ attitudes towards mindfulness evolve over the course of the study. This subsection contains three secondary analyses from our pre-analysis plan: changes in

willingness to pay for an extension of the meditation license, predictions about the average treatment effect in the study, and subjective ease of practicing meditation. We find that treatment group reports greater ease of meditation and a short-term increase in valuation of the app, but grows less optimistic about the effects of meditation with time.

These outcomes discussed here during our short SMS surveys 4, 7, and 11 days after randomization, and during our endline survey at 30 days after randomization. Specifically:

- We elicit participants’ willingness to pay for a 3-month extension of their license using a probabilistic Becker-DeGroot-Marschak mechanism. Participants can report valuations between \$0 and \$100 on a sliding scale. This question appears on surveys at baseline, as well as post-randomization at 4, 7, 11, and 30 days. We implement the mechanism with 1% probability per participant, selecting one of their responses uniformly at random after excluding any missing values.
- To measure subjective treatment effects, we ask participants to consider a hypothetical scenario. We instruct them to consider 10 other randomly selected participants who report anxiety symptoms at the beginning of the study. Then, we ask them to predict how many of these 10 would report anxiety in 3 weeks if they did not receive a Headspace license (“control”), as well as if they did receive the license and used it for 5 or more days per week (“treatment”). Finally, we calculate participants’ subjective treatment effect as the treatment-minus control improvement, which is an integer ranging from -10 to 10.

These questions appear at baseline, as well as 4, 7, 11, and 30 days post-randomization. In all post-baseline elicitations, we remind participants of their previous responses and present an opportunity to update them.

- We measure ease of meditation along two dimensions: finding time and space to meditate, and focusing on meditation for 10 minutes given the right time and place.

These questions informally separate difficulties with forming habits in general from those with a meditation habit specifically. First we ask participants, “In your experience so far, how easy or difficult is it to find a good time and space to meditate?” They may respond on a scale from 0 (“very difficult”) to 10 (“very easy”). We treat this as an integer between 0 and 10. Next we ask, “If you had the right time and space, how easy or difficult would it be to focus on meditating for 10 minutes without quitting?” Participants may respond with one of five choices: any of the four combinations of “very/somewhat difficult/easy”, or “I don’t know—I have not been meditating”. We focus on the proportion of participants reporting that it is “easy” or “very easy” to find time and space to meditate.

These questions appear 4, 7, 11, and 30 days post-randomization, but only for participants who received immediate access to the meditation app.

To estimate treatment effects on willingness to pay and subjective treatment effects these quan-

tities, we perform regressions of the form

$$Y_{it} = \delta_{\text{stratum}} + \beta_1 \text{AppAccess}_i + \beta_2 \text{AnyIncentive}_i \epsilon_{it} \quad (\text{B.5})$$

where Y_{ist} is the outcome for respondent i at time t , δ_{stratum} are stratum fixed effects, and ϵ_{it} is an error clustered at the respondent level. The remaining terms are indicators for receiving a Headspace license, and for receiving any incentives to meditate in the two weeks after randomization. Because we only elicit ease of meditation for participants who receive immediate app access, when we estimate effects on ease of meditation we perform regressions of the form

$$Y_{it} = \delta_{\text{stratum}} + \beta_1 \text{AnyIncentive}_i + \beta_2 \text{LongIncentive}_i \epsilon_{it} \quad (\text{B.6})$$

where all terms are as before, and LongIncentive_i is an indicator for receiving an incentive to meditate on 10 out of the first 14 post-randomization days.

We find that receiving access to the app increases willingness to pay for a license extension at 4, 7, and 11 days after randomization (Table B.11, row 2, columns 2–4). Effects are largest early on, at roughly 8% of the control mean, and decay to become indistinguishable from zero by 30 days. We find a smaller but similar trend among participants who receive a cash transfer but remain on a waitlist. By contrast, participants on the Pure Waitlist report no average change in WTP while waiting for the license.

Participants in the treatment groups revise their predicted treatment effects modestly downward over the course of the study. At the outset respondents predict that in a group of 10 participants with anxiety, consistent weekly meditation for three weeks would reduce anxiety for 3.5 more people than would naturally remit. This number remains stable in the Pure Waitlist and Cash Transfer groups. By the 30-day mark, each of the treatment groups has grown more pessimistic, revising their treatment effects down by roughly 10-15% of their initial guess (Table ??, rows 2–4, column 5). These updates emerge earliest and are largest for the Long Incentives group, though they are apparent in the App Only and Short Incentive groups as well.

By the end of the study, participants tend to view meditation as an easier activity than when they started. They report similar or higher levels of ease with finding time and space to meditate, as well as focusing on meditation during sessions. The short incentives, which encourage participants to meditate on 5 out of the 14 post-randomization days, increase reported ease along both dimensions. This suggests that while practice increases the ease of meditation, participants' natural choice of meditation frequency leaves room for further low-cost practice. That said, it appears possible to push people too hard. The long incentives group, whom we encourage to meditate on 10 out of the 14 post-randomization days, report that it is harder to find time or space to meditate than does the short incentives group. These additional sessions they perform provide no increased ease of focus.

B.6 Multiple Hypothesis Testing

We test a number of hypotheses in this project, ranging from whether incentives form long-term meditation habits (unlikely), to whether mindfulness improves mental health (likely), and whether mindfulness alters the effects of stress on decision making (potentially). As the number of null hypotheses we test increases, so too does the chance that we spuriously reject one of them. In this subsection we present a multiplicity correction to assess the robustness of our main hypothesis tests to concerns about multiple testing. At conventional significance levels, we find that the ITT effect of the app on mental health and proofreading performance are robust to correcting for multiplicity.

We follow the procedure from Romano and Wolf (2016), which asymptotically controls the probability of falsely rejecting one null hypothesis within a set of tested hypotheses. By using bootstrap resampling to estimate the joint sampling distribution of the test statistics in question, this procedure attains higher power than approaches like the Bonferroni or Bonferroni-Holm corrections. The corrected p -values we present are based on 1,999 resamples, respecting stratified random assignment.

In keeping with our pre-analysis plan, we group outcomes into families based on the Key Contrasts we pre-specify. These are (1) the effects of incentives on habit formation, (2) the effects of mindfulness on mental health, and (3) the effects of mindfulness on attention allocation. Column 4 of Table B.17 contains p -values that are adjusted for multiple testing within these families.

However, the second and third families each contain two distinct questions. For mental health, these are whether there is an ITT effect of offering the app, and whether incentives deliver benefits on top of simply providing the app. For productivity, these are whether there is an ITT effect, and whether app users react differently to stress than the waitlist group. Column 3 of Table B.17 presents a more lenient multiplicity correction, within each of these subfamilies.³⁶

Thus, a reader concerned only with the question, “does app access improve mental health?” can consult Column 2; a companion of theirs interested in both this effect and the effect of incentives on mental health could consult Column 3. Finally, for those who wish to apply multiplicity corrections to all of our main claims, Column 4 presents p -values adjusted for all the tests in the table.

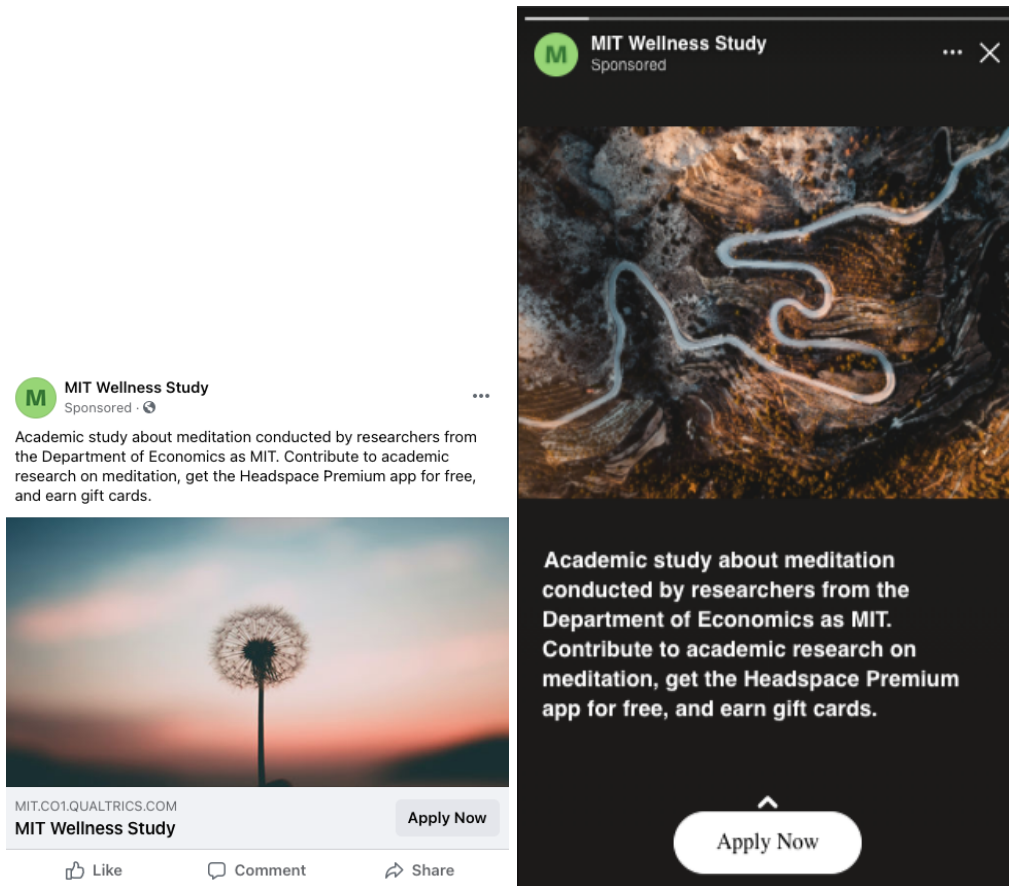
In analyzing the productivity outcomes, we exclude users who received an incentive to meditate immediately. This is not to chase small p -values: Table 6 shows larger differences in emotional interference between immediate meditators and the waitlist control, implying that including them would produce a larger effect and smaller p -value. Instead this is for conceptual coherence: comparing the control group to the treatment group without immediate meditation incentives captures the effect of practicing meditation on the stress response. We believe this is a more relevant hypothesis than that of whether meditating has short-term effects on decision making.

Correcting for multiplicity does not alter our main findings: app access improves mental health from 2 weeks to three months (rows 3-5) as well as proofreading performance (row 9), but not Stroop scores (rows 11-12); incentives cause no long-term change in meditation behavior or mental health (rows 1-2, 6-9); and we lack conclusive evidence to make claims about emotional interference (rows 13-15).

³⁶Note that columns 3 and 4 are identical for the Habit Formation family, since it contains just one subfamily.

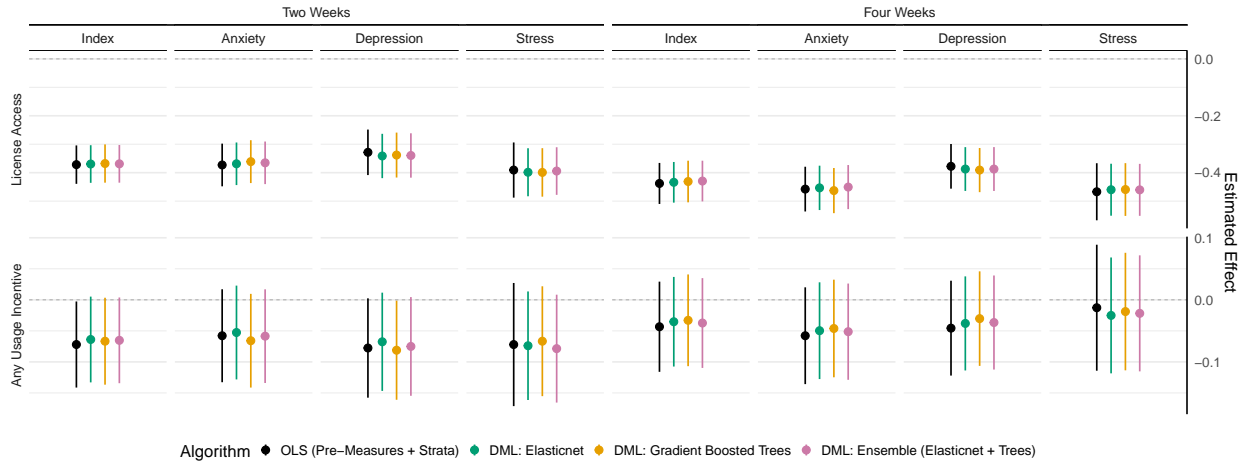
B.7 Appendix Figures

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



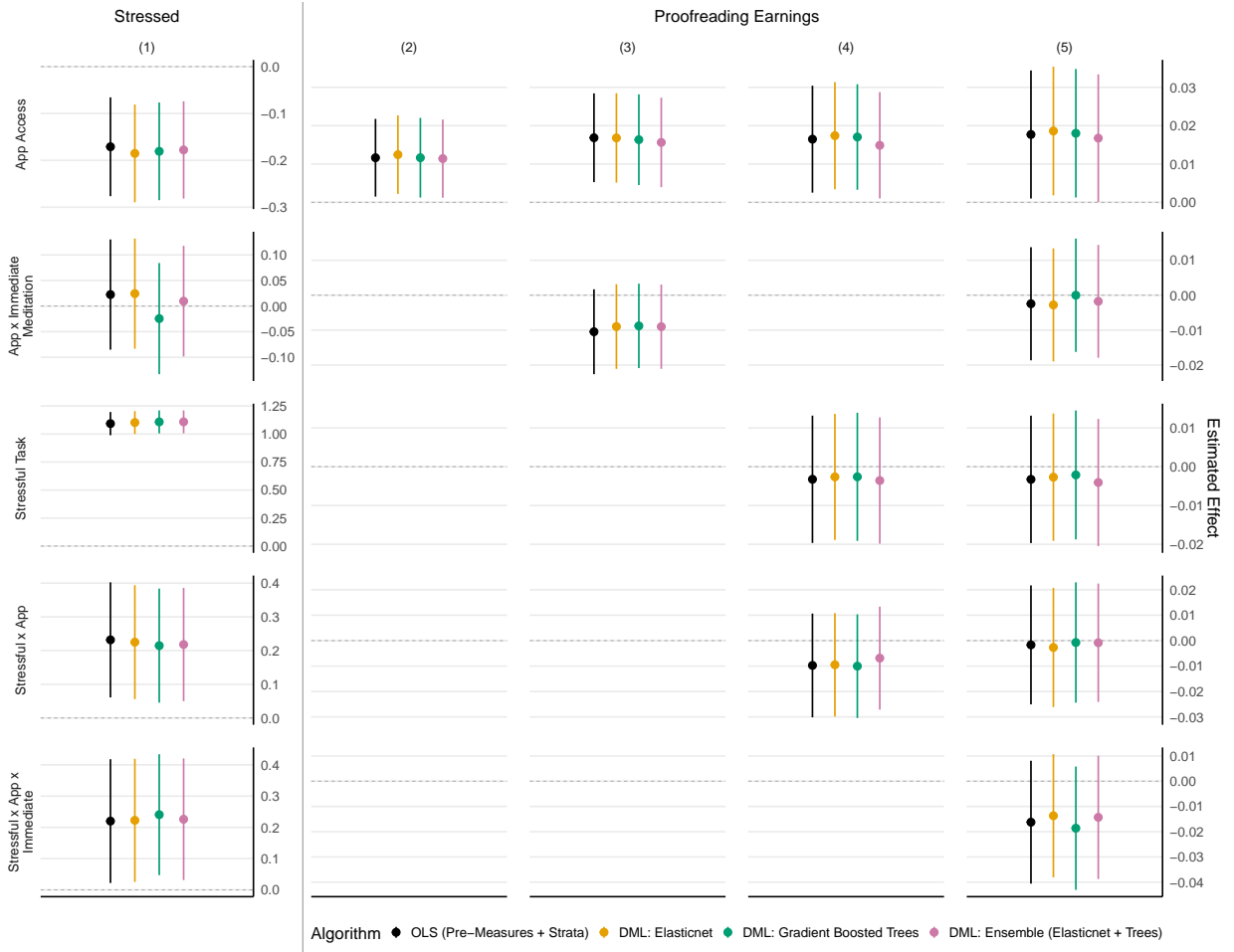
Notes: This figure shows examples of ads that were used to recruit participants for the study. A variety of images were used, with the most effective being automatically selected to be distributed more widely. The ad text was always the same.

Figure B.2: Short-term Effects of App Access and Usage Incentives on Anxiety (Adjusting for Covariates with Double/Debiased Machine Learning)



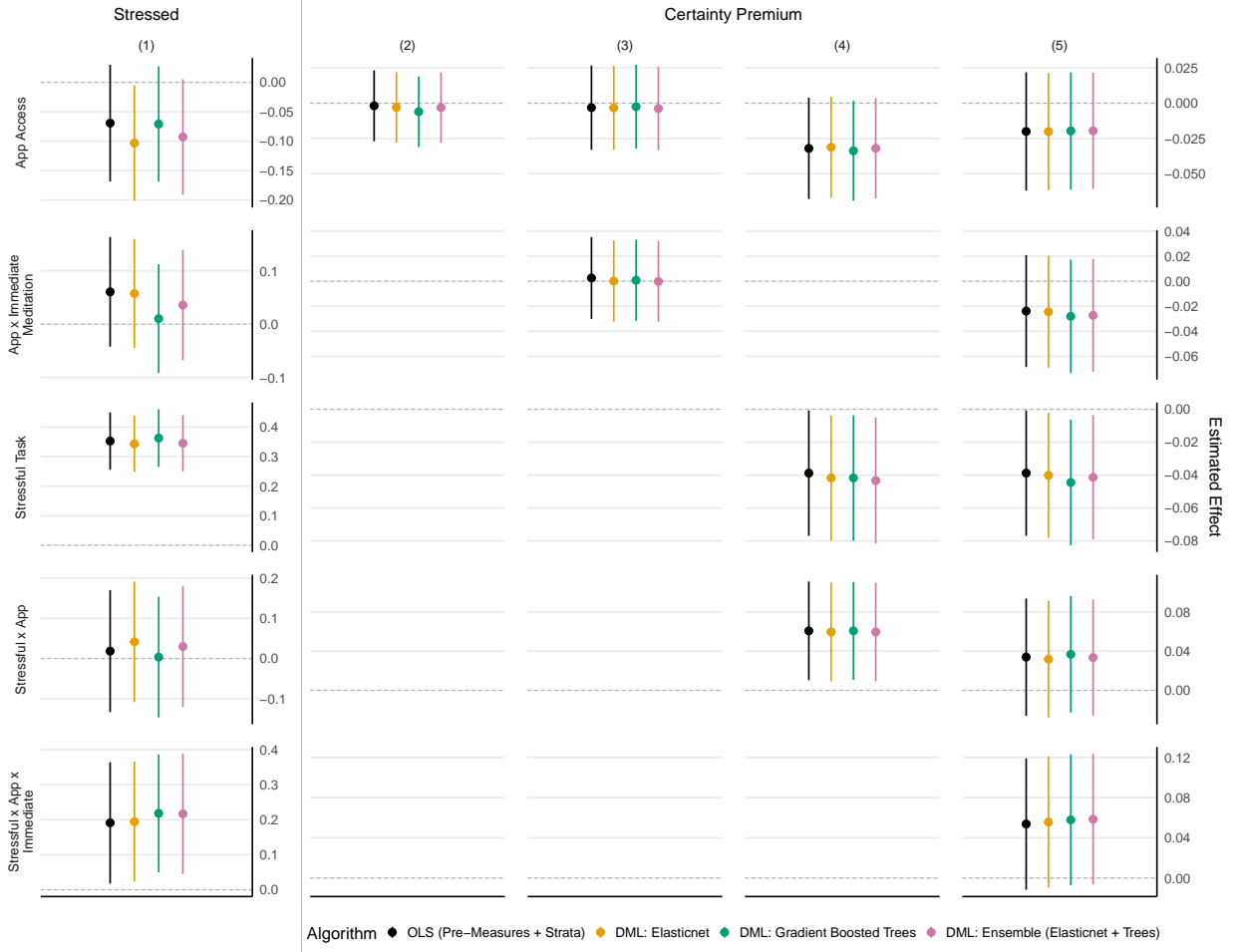
Notes: This figure presents intent-to-treat estimates analogous to those in Panel A of Table 3, adjusting for potential covariate imbalance using double/debiased machine learning. Appendix B.3 describes this procedure in detail. The outcomes are standardized scores of reported anxiety. Anxiety is measured using GAD-2, i.e. the first two items of the GAD-7 scale, a standard psychometric scale commonly used in clinical psychology evaluations. Scores are centered and scaled relative to the control group mean and standard deviation in each period. The regression specification is a generalization of Equation 3 to $Y_i^{post} = \beta_1 \text{AppAccess}_i + \beta_2 \text{AnyIncentive}_i + f(X_i) + \epsilon_i$, where X_i are covariates that include randomization strata, pre-randomization measures of the outcome if available, as well as other pre-randomization covariates. The layout of the figure mirrors the layout of the Table 3, Panel A. In each panel, the black, leftmost point and interval represents an estimate and 95% confidence interval for β_1 (top row) or β_2 (bottom row) from Table 3. The remaining points present the double/debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of three algorithms: elasticnet, gradient boosted trees, and an ensemble of the two. The ensemble uses gradient-boosted trees to predict residuals from elasticnet.

Figure B.3: Impacts of App Access, Immediate Meditation, and Stressful Treatment on Earnings in the Proofreading Task (Adjusting for Covariates with Double/Debiased Machine Learning)



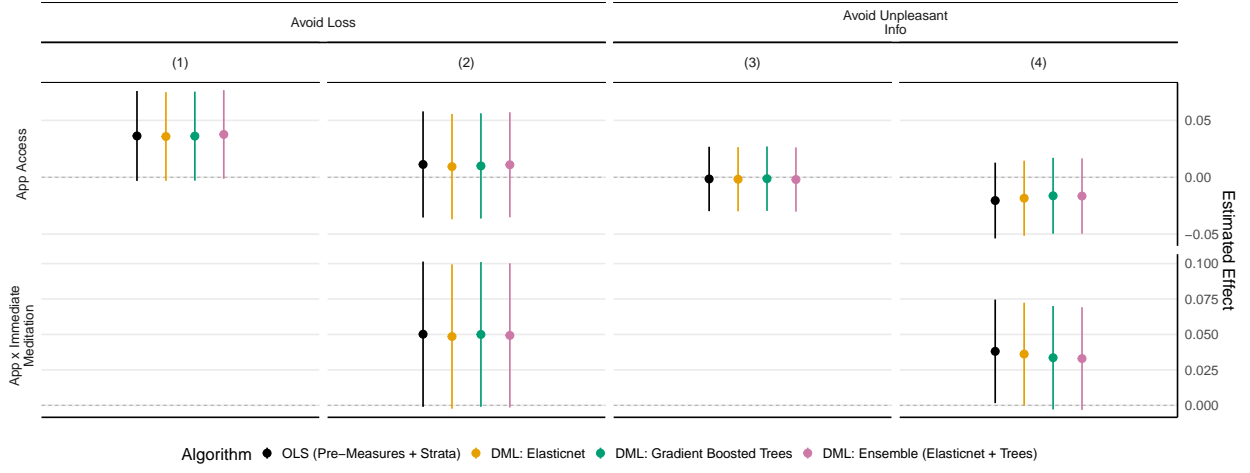
Notes: This figure presents intent-to-treat (ITT) estimates analogous to those in Table 4, adjusting for potential covariate imbalance using double/debiased machine learning. Appendix B.3 describes this procedure in detail. The regression specifications are generalizations of Equations 4, 5, 6, 7, replacing the terms $\delta_{\text{stratum}} + \gamma Y_i^{\text{pre}}$ with $f(X_i)$, a potentially nonlinear function of an extended set of covariates. The layout of the figure mirrors the layout of the Table 4. In each panel, the black, leftmost point and interval represents an ITT estimate and 95% confidence interval after adjusting linearly for randomization strata and the pre-randomization outcome. The remaining points present the double/debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of three algorithms: elasticnet, gradient boosted trees, and an ensemble of the two. The ensemble uses gradient boosted tree model to predict residuals from elasticnet.

Figure B.4: Impacts of App Access, Immediate Meditation, and Stressful Treatment on the Certainty Premium (Adjusting for Covariates with Double/Debiased Machine Learning)



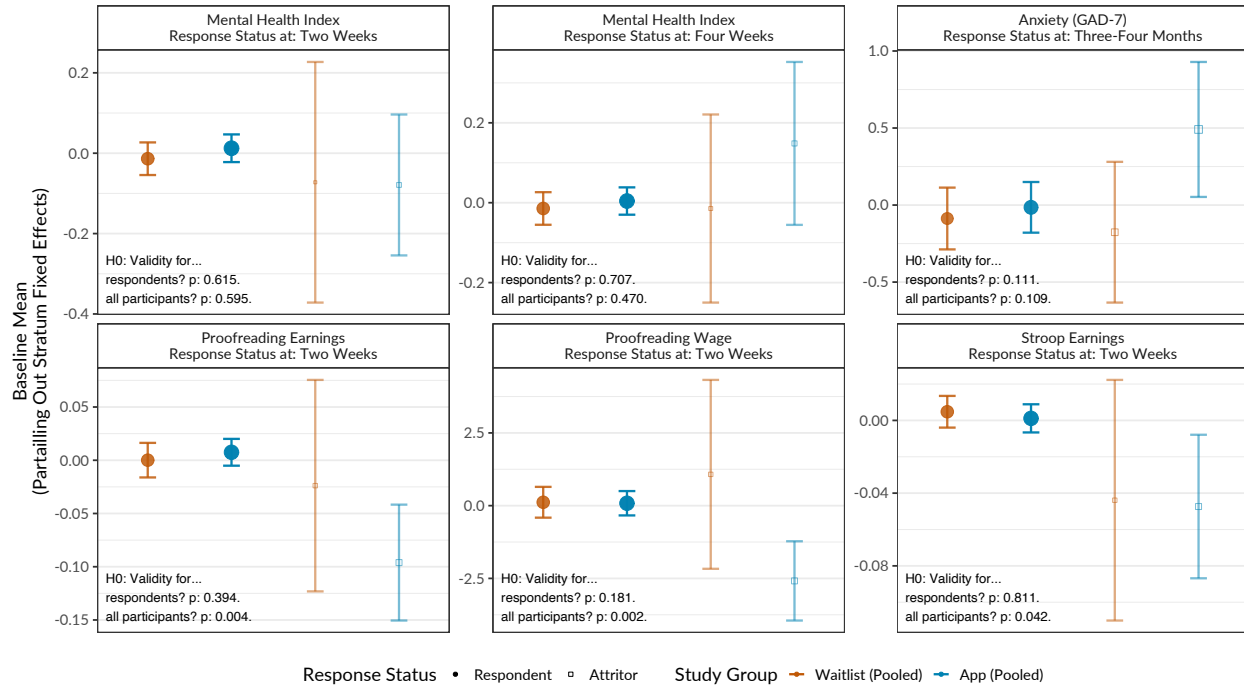
Notes: This figure presents intent-to-treat (ITT) estimates analogous to those in Table 6, adjusting for potential covariate imbalance using double/debiased machine learning. Appendix B.3 describes this procedure in detail. The regression specifications are generalizations of Equations 4, 5, 6, 7, replacing the terms $\delta_{\text{stratum}} + \gamma Y_i^{\text{pre}}$ with $f(X_i)$, a potentially nonlinear function of an extended set of covariates. The layout of the figure mirrors the layout of the Table 6. In each panel, the black, leftmost point and interval represents an ITT estimate after adjusting linearly for randomization strata and the pre-randomization outcome. The remaining points present the double/debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of three algorithms: elasticnet, gradient boosted trees, and an ensemble of the two. The ensemble uses gradient boosted trees to predict residuals from elasticnet.

Figure B.5: Impacts of App Access, Immediate Meditation, and Stressful Treatment on Avoidance of Low Probability Losses and Unpleasant Information (Adjusting for Covariates with Double/Debiased Machine Learning)



Notes: This figure presents intent-to-treat (ITT) estimates analogous to those in Table B.8, adjusting for potential covariate imbalance using double/debiased machine learning. Appendix B.3 describes this procedure in detail. The regression specifications are generalizations of Equations 4 and 5, replacing the terms $\delta_{\text{stratum}} + \gamma Y_i^{\text{pre}}$ with $f(X_i)$, a potentially nonlinear function of an extended set of covariates. The layout of the figure mirrors the layout of the Table B.9. In each panel, the black, leftmost point and interval represents an ITT estimate and 95% confidence interval after adjusting linearly for randomization strata and the pre-randomization outcome. The remaining points present the double/debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of three algorithms: elasticnet, gradient boosted trees, and an ensemble of the two. The ensemble trains a gradient boosted tree model to predict residuals from elasticnet.

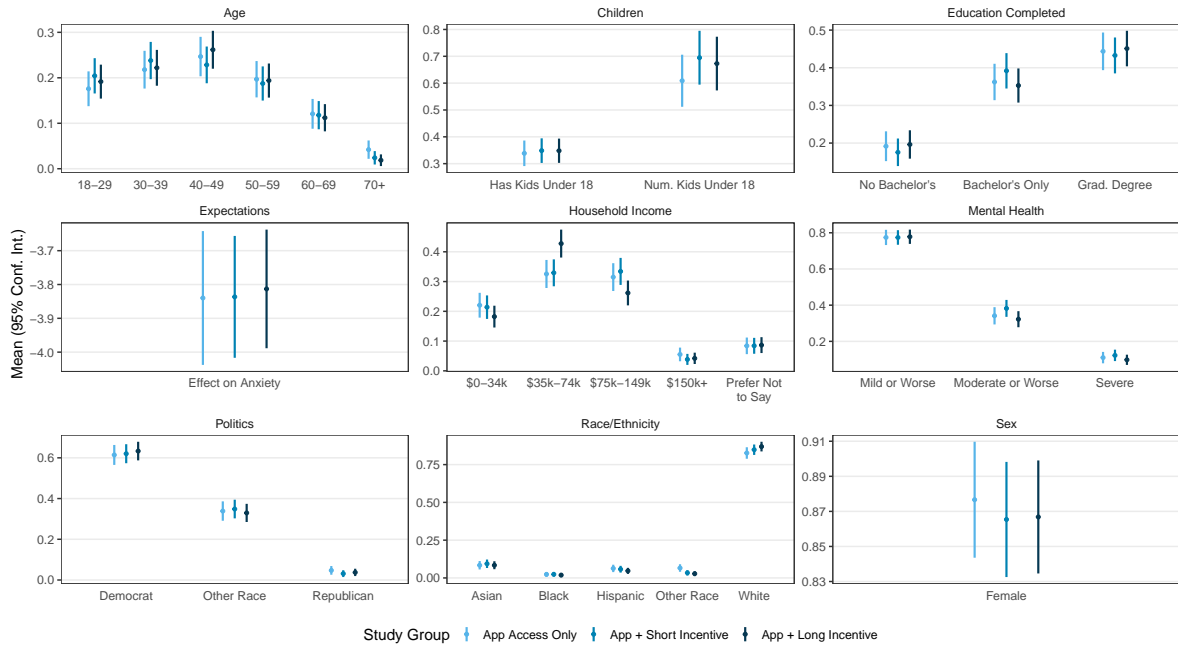
Figure B.6: Tests for Internal Validity Given Attrition



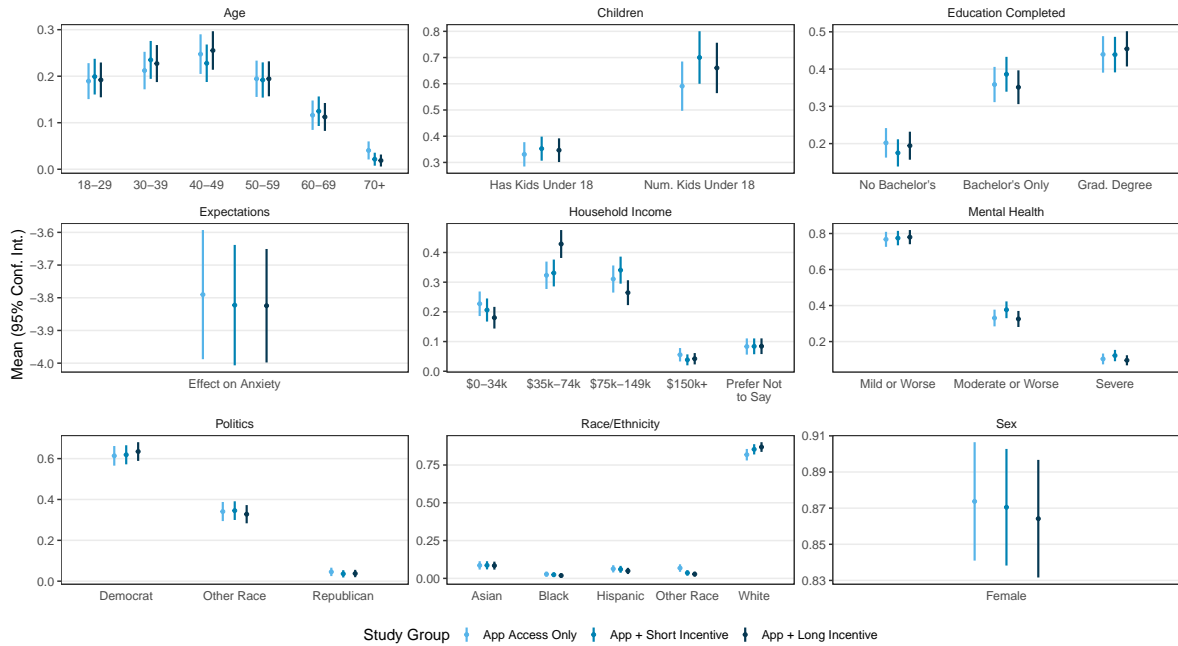
Notes: This figure presents tests of the internal validity of this study, following the recommendations of Ghanem et al. (2020). Each facet investigates the potential role of differential attrition in generating bias in the analysis of a particular outcome at a particular time. Points and interval represent the baseline mean and a 95% confidence interval for the mean for participants in the waitlist (orange) or treatment (blue), who either responded (solid circles, at left) or failed to respond (hollow squares, at right) to an outcome elicitation at the given point in time. A comparison of the two circles roughly corresponds to testing whether survey respondents are comparable to each other; that is, whether attrition biases main analyses of effects on the subpopulation that responds to our survey. A comparison of all four means in each panel roughly corresponds to testing whether attrition biases estimates of treatment effects on the entire study population, including nonrespondents. We provide p -values for formal tests of these hypotheses, based on Equations B.1 (validity for respondents) and B.2 (validity for study population) in each facet. Rejecting a null hypothesis corresponds to finding evidence that attrition induces bias into treatment effect estimates either for the respondent population, or the entire study population.

Figure B.7: Complier Characteristics

(a) Compliers at Two Weeks

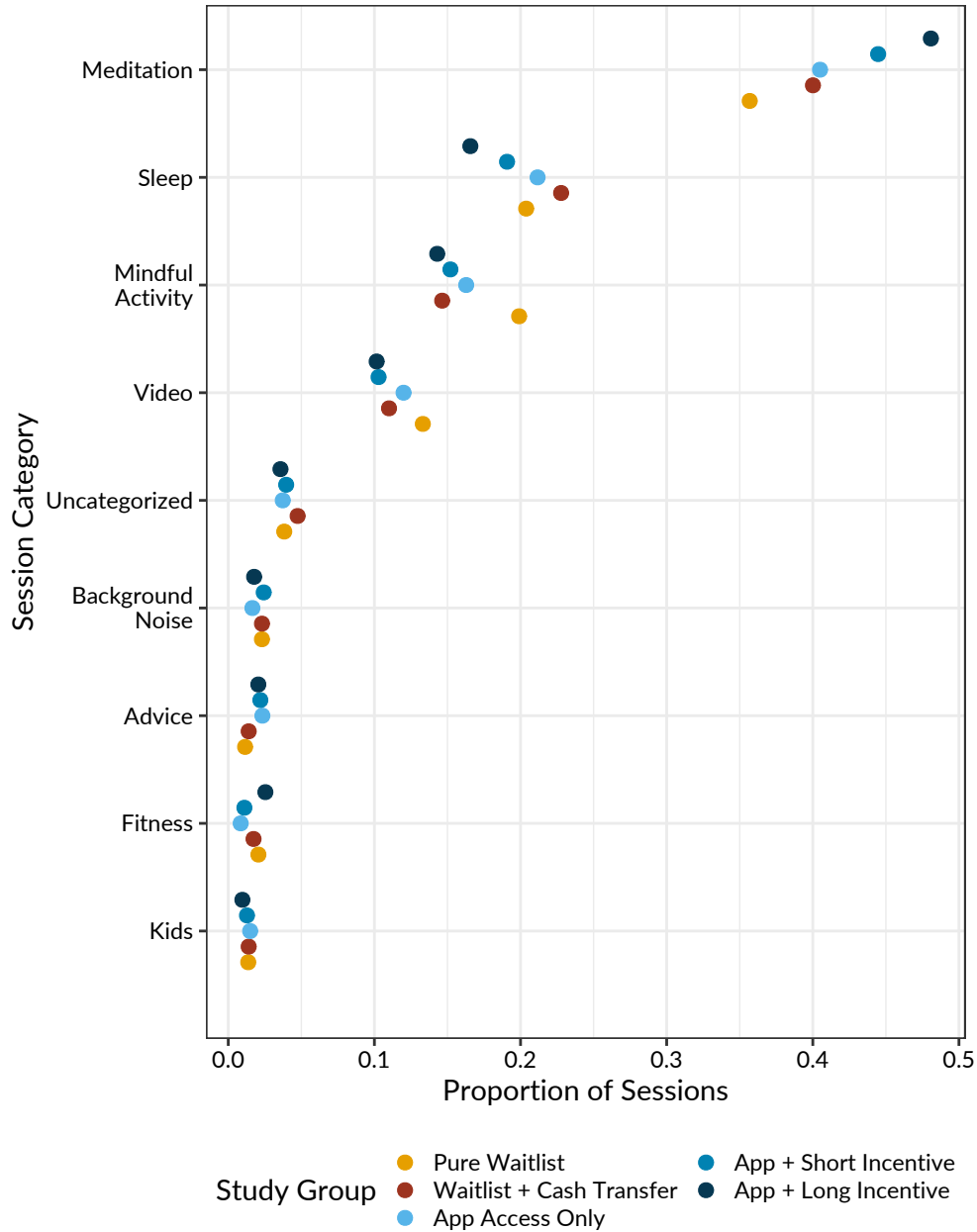


(b) Compliers at Four Weeks



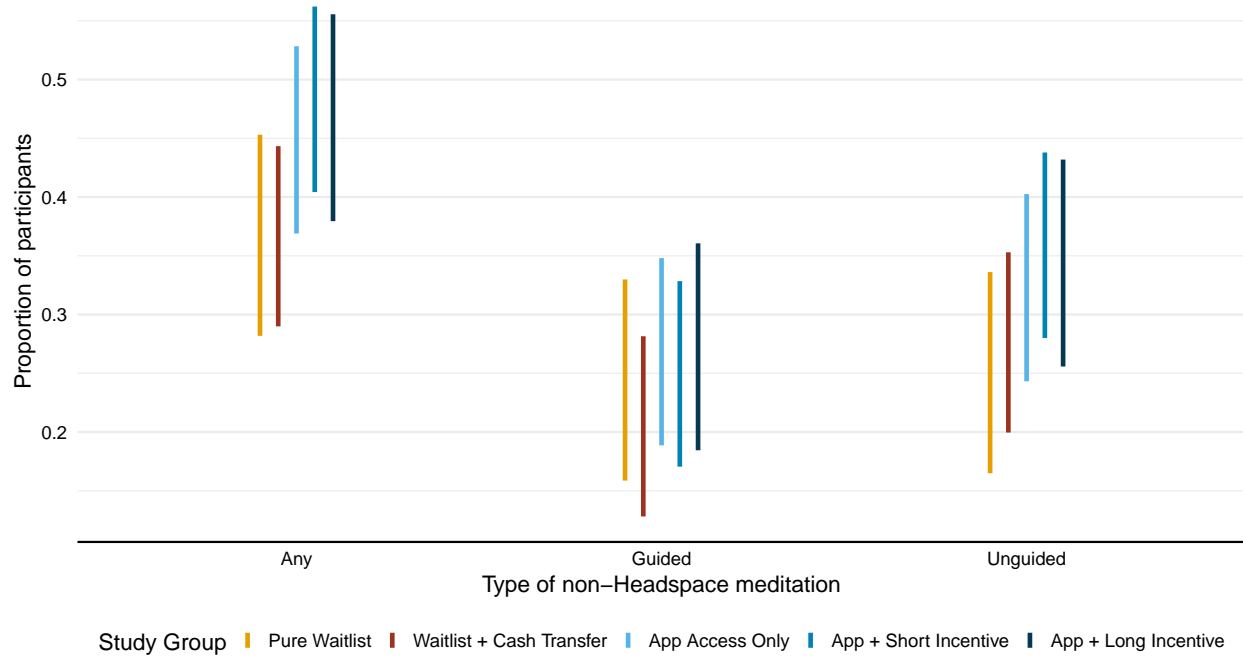
Notes: This figure presents average characteristics for compliers in each treatment arm at two weeks after randomization (Panel A) and four weeks after randomization (Panel B). Within each panel, each facet describes a different baseline characteristic. Points represent the sample mean, and intervals represent 95% confidence intervals. To generate this figure, we assume that participants in the Waitlist group are not using the Headspace app. Compliers are then defined as individuals in the treatment arms who have used the app at least once.

Figure B.8: Session Types



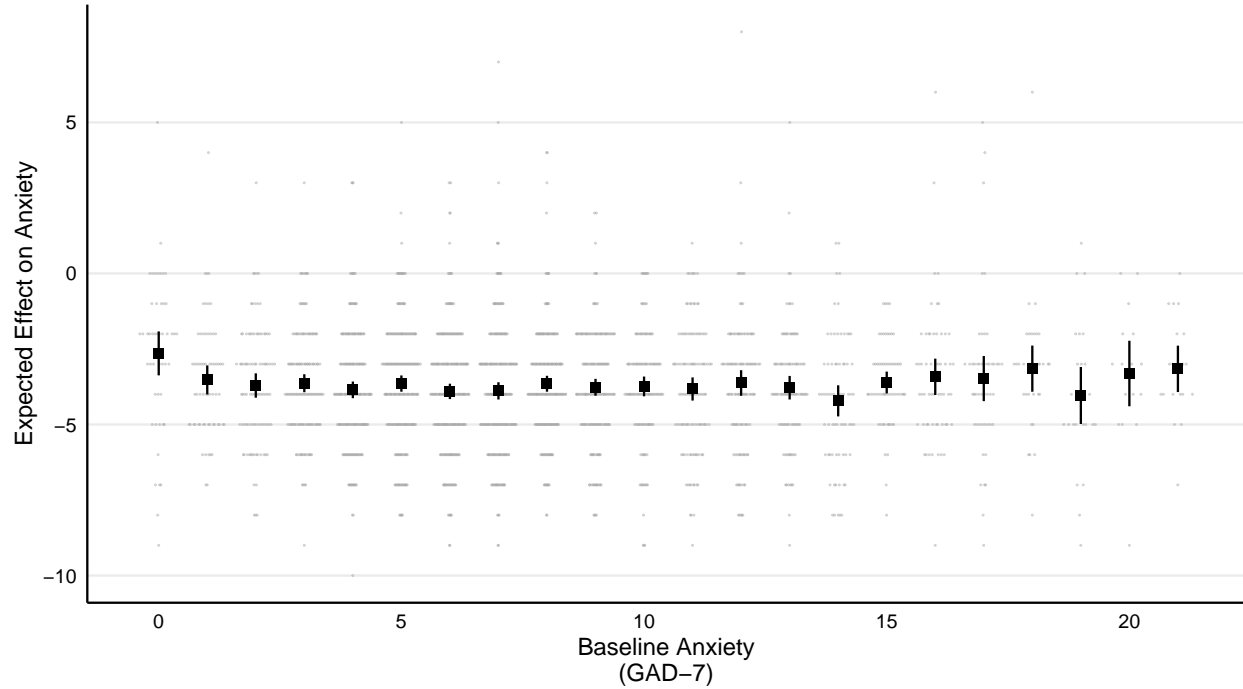
Notes: This figure presents the composition of Headspace app sessions each treatment group engaged in during their three months of app access. We derive the categories from the classification of these sessions in the app. Meditation sessions are guided or unguided recordings that teach mindfulness, either 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 B.9: Self-reported Meditation on Non-Headspace Channels



Notes: This figure presents rates of self-reported meditation outside of the Headspace the joint distribution of baseline anxiety scores (GAD-7) and beliefs about the effectiveness of meditation at reducing anxiety. We elicit beliefs about treatment effects by asking participants to forecast changes in anxiety levels among others, comparing groups who do and do not receive app access. Specifically, we ask each participant to (1) consider a hypothetical random sample of 10 other participants with high anxiety, (2) predict the number who would report high anxiety in 3 weeks if they did versus did not receive app access. The expected treatment effect is the difference in these forecasts. Points in the figure represent individual respondents, with random noise added to x-values to reduce overplotting. Squares present conditional means, and vertical bars give 95% confidence intervals. With the exception of a few participants who report no symptoms of anxiety and more skepticism about treatment effects, there is no systematic relationship between anxiety and beliefs about the efficacy of mindfulness.

Figure B.10: Prior Beliefs about Mindfulness Efficacy, by Baseline Anxiety Score



Notes: This figure presents the joint distribution of baseline anxiety scores (GAD-7) and beliefs about the effectiveness of meditation at reducing anxiety. We elicit beliefs about treatment effects by asking participants to forecast changes in anxiety levels among others, comparing groups who do and do not receive app access. Specifically, we ask each participant to (1) consider a hypothetical random sample of 10 other participants with high anxiety, (2) predict the number who would report high anxiety in 3 weeks if they did versus did not receive app access. The expected treatment effect is the difference in these forecasts. Points in the figure represent individual respondents, with random noise added to x-values to reduce overplotting. Squares present conditional means, and vertical bars give 95% confidence intervals. Except for a few participants who report no symptoms of anxiety and more skepticism about treatment effects, there is no systematic relationship between anxiety and beliefs about the efficacy of mindfulness.

B.8 Appendix Tables

Table B.1: Balance Table

	Baseline																	Two-Week Mental Health				Two-Week Economic Decisions				Four-Week Followup			
	US Adults Mean (1)	Waitlist Group		App Group		Mean (6)	Waitlist Group		App Group		Mean (10)	Waitlist Group		App Group		Mean (14)	Waitlist Group		App Group										
		Mean (2)	(St. Dev.) (3)	Diff. (4)	(St. Err) (5)		(St. Dev.) (7)	Diff. (8)	(St. Err) (9)	(St. Dev.) (11)		Diff. (12)	(St. Err) (13)	(St. Dev.) (15)	Diff. (16)		(St. Err) (17)												
Age Group																													
18-29	0.213	0.202	(0.402)	-0.010	(0.017)	0.201	(0.401)	-0.010	(0.017)	0.203	(0.403)	-0.009	(0.017)	0.202	(0.402)	-0.009	(0.017)												
30-39	0.172	0.204	(0.403)	0.014	(0.017)	0.206	(0.404)	0.015	(0.017)	0.204	(0.403)	0.014	(0.017)	0.207	(0.406)	0.012	(0.017)												
40-49	0.158	0.250	(0.433)	-0.010	(0.018)	0.252	(0.435)	-0.009	(0.018)	0.256	(0.437)	-0.013	(0.019)	0.253	(0.435)	-0.012	(0.018)												
50-59	0.163	0.209	(0.407)	-0.016	(0.017)	0.210	(0.408)	-0.018	(0.017)	0.208	(0.406)	-0.017	(0.017)	0.208	(0.406)	-0.016	(0.017)												
60-69	0.150	0.109	(0.312)	0.016	(0.013)	0.106	(0.308)	0.015	(0.013)	0.106	(0.308)	0.018	(0.014)	0.106	(0.308)	0.019	(0.013)												
70+	0.143	0.025	(0.157)	0.006	(0.007)	0.024	(0.154)	0.006	(0.007)	0.023	(0.149)	0.008	(0.007)	0.024	(0.152)	0.006	(0.007)												
Female	0.492	0.848	(0.359)	0.019	(0.015)	0.848	(0.359)	0.018	(0.015)	0.851	(0.356)	0.014	(0.015)	0.850	(0.357)	0.019	(0.015)												
Education																													
No Bachelor's degree	0.640	0.215	(0.411)	-0.024	(0.017)	0.213	(0.410)	-0.022	(0.017)	0.213	(0.410)	-0.019	(0.017)	0.214	(0.410)	-0.021	(0.017)												
Bachelor's degree	0.225	0.345	(0.475)	0.019	(0.020)	0.347	(0.476)	0.022	(0.020)	0.346	(0.476)	0.017	(0.020)	0.349	(0.477)	0.014	(0.020)												
Graduate or professional degree	0.135	0.441	(0.497)	0.004	(0.021)	0.440	(0.497)	-0.001	(0.021)	0.441	(0.497)	0.002	(0.021)	0.437	(0.496)	0.007	(0.021)												
Household Size	2.261	2.874	(1.328)	-0.086	(0.055)	2.881	(1.329)	-0.093	(0.055)	2.888	(1.324)	-0.102	(0.056)	2.881	(1.328)	-0.090	(0.056)												
Household Income																													
\$34,999 or less	0.265	0.231	(0.422)	-0.022	(0.017)	0.231	(0.422)	-0.022	(0.018)	0.231	(0.422)	-0.023	(0.018)	0.231	(0.422)	-0.024	(0.018)												
\$35,000-\$74,999	0.293	0.321	(0.467)	0.040	(0.020)	0.322	(0.468)	0.039	(0.020)	0.325	(0.469)	0.035	(0.020)	0.322	(0.467)	0.039	(0.020)												
\$75,000-\$149,000	0.285	0.332	(0.471)	-0.034	(0.019)	0.332	(0.471)	-0.033	(0.020)	0.329	(0.470)	-0.030	(0.020)	0.333	(0.472)	-0.031	(0.020)												
\$150,000 or more	0.157	0.042	(0.200)	0.004	(0.009)	0.042	(0.202)	0.002	(0.009)	0.042	(0.201)	0.001	(0.009)	0.042	(0.200)	0.002	(0.009)												
Prefer not to answer		0.073	(0.261)	0.013	(0.011)	0.072	(0.259)	0.014	(0.011)	0.072	(0.259)	0.017	(0.012)	0.073	(0.260)	0.013	(0.011)												
Race & Ethnicity																													
White	0.600	0.834	(0.373)	0.015	(0.015)	0.834	(0.373)	0.013	(0.016)	0.834	(0.373)	0.012	(0.016)	0.835	(0.371)	0.012	(0.016)												
Black	0.124	0.023	(0.150)	0.000	(0.006)	0.023	(0.151)	-0.001	(0.006)	0.023	(0.149)	0.000	(0.006)	0.024	(0.152)	0.000	(0.006)												
Hispanic	0.184	0.063	(0.243)	-0.007	(0.010)	0.063	(0.242)	-0.006	(0.010)	0.063	(0.243)	-0.005	(0.010)	0.064	(0.245)	-0.007	(0.010)												
Asian	0.056	0.087	(0.282)	-0.002	(0.012)	0.088	(0.283)	-0.002	(0.012)	0.089	(0.284)	0.000	(0.012)	0.088	(0.283)	-0.003	(0.012)												
Other race	0.036	0.057	(0.231)	-0.013	(0.009)	0.055	(0.228)	-0.010	(0.009)	0.055	(0.228)	-0.011	(0.009)	0.053	(0.225)	-0.009	(0.009)												
Political Party																													
Democrat		0.624	(0.485)	-0.003	(0.020)	0.624	(0.485)	-0.005	(0.020)	0.626	(0.484)	-0.008	(0.021)	0.625	(0.484)	-0.006	(0.021)												
Republican		0.027	(0.163)	0.013	(0.007)	0.028	(0.164)	0.013	(0.008)	0.027	(0.162)	0.014	(0.008)	0.026	(0.158)	0.016	(0.007)												
Other		0.349	(0.477)	-0.010	(0.020)	0.349	(0.477)	-0.007	(0.020)	0.347	(0.476)	-0.006	(0.020)	0.349	(0.477)	-0.011	(0.020)												
Mental Health at Baseline																													
Moderate or severe anxiety		0.316	(0.465)	0.020	(0.020)	0.317	(0.466)	0.021	(0.020)	0.318	(0.466)	0.012	(0.020)	0.318	(0.466)	0.016	(0.020)												
Likely major depressive disorder		0.261	(0.439)	0.004	(0.018)	0.262	(0.440)	0.003	(0.019)	0.261	(0.439)	-0.005	(0.019)	0.261	(0.439)	0.003	(0.019)												
Sample Size & Attrition																													
N		955		1,429		943		1,388		925		1,336		936		1,377													
Attrition rate	—	—		—		1.26%		2.87%		3.14%		6.51%		1.99%		3.64%													

Notes: This table is an extension of Table 1 that presents demographic characteristics of the sample with differences between the pooled Waitlist and the pooled Headspace groups at different stages of the study. Slight imbalances that exists at the randomization stage persist in the later stages, but attrition is not differential by baseline characteristics.

Table B.2: Sample Size and Attrition

	Baseline	Four Days	Seven Days	Eleven Days	Two Weeks (Mental Health)	Two Weeks (Decision Making)	Four Weeks	Three Months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Waitlist Group								
N	955	938	882	914	943	925	936	817
Attrition Rate		0.018	0.076	0.043	0.013	0.031	0.020	0.145
App Access Group								
N	1429	1370	1269	1280	1388	1336	1377	1212
Attrition Rate		0.041	0.112	0.104	0.029	0.065	0.036	0.152
Difference in Rates		0.023	0.036	0.061	0.016	0.034	0.016	0.007
<i>p</i> -value		(0.002)	(0.005)	(<0.001)	(0.013)	(<0.001)	(0.028)	(0.663)

Notes: This table presents the number of participants who began each survey, by treatment arm. Attrition is lowest for the main surveys at sixteen and thirty days, and highest for the three-month followup. Participants in the License group are detectably less likely to respond to each survey from four days to four weeks, although differential attrition is small in absolute terms at sixteen and thirty days. Although attrition is more severe at three months, we obtain similar response rates from both treatment arms.

Table B.3: Bounds on Intent to Treat, Given Attrition

Outcome	95% Confidence Interval
Mental Health Index (at 2 weeks)	[-0.441, -0.363]
Mental Health Index (at 4 weeks)	[-0.518, -0.409]
Anxiety (GAD-7, at 3-4 months)	[-1.250, -0.808]
Proofreading Earnings (% of maximum)	[-0.008, 0.020]
Proofreading Time Spend (seconds)	[-0.044, 0.156]
Stroop Earnings (% of maximum)	[-0.017, 0.008]

Notes: this table presents worst-case bounds on the intent to treat effect of offering a license. We estimate the identified set using the procedure from Lee (2009), conditioning on randomization strata. This procedure relies on the assumption that receiving a license has a monotonic effect on attrition within a stratum, either making all individuals more likely to respond or less likely to respond. Confidence intervals are based on Imbens and Manski (2004)'s confidence intervals for partially identified parameters.

Table B.4: Short-term Effects of App Access and Usage Incentives on Anxiety

	Four Days (1)	Seven Days (2)	Eleven Days (3)	Two Weeks (4)	Four Weeks (5)
App Access (S.E.)	-0.103 (0.037)	-0.162 (0.041)	-0.239 (0.040)	-0.289 (0.041)	-0.394 (0.044)
Any Usage Incentive (S.E.)	-0.053 (0.038)	-0.029 (0.041)	-0.061 (0.041)	-0.071 (0.041)	-0.060 (0.043)
N	2305	2145	2191	2330	2312

Notes: This table presents estimates of Intent-to-Treat effects from offering access to the mindfulness meditation app, as well as additional usage incentives, on standardized scores of reported anxiety. The estimating equation is Equation 3. Anxiety is measured using GAD-2, the first two items of the GAD-7 scale, a standard psychometric scale commonly used in clinical psychology evaluations. Scores are centered and scaled relative to the pooled Waitlist group mean and standard deviation in each period.

Table B.5: Effects of App Access and Usage Incentives on Symptom Severity Using Score Cutoffs

	Two Weeks			Four Weeks		
	Mild or Worse (Score ≥ 5) (1)	Moderate or Worse (Score ≥ 10) (2)	Severe (Score ≥ 15) (3)	Mild or Worse (Score ≥ 5) (4)	Moderate or Worse (Score ≥ 10) (5)	Severe (Score ≥ 15) (6)
A. Anxiety (GAD-7)						
App Access (S.E.)	-0.151 (0.023)	-0.116 (0.019)	-0.039 (0.012)	-0.228 (0.024)	-0.131 (0.019)	-0.064 (0.012)
Any Usage Incentive (S.E.)	-0.038 (0.024)	-0.034 (0.018)	-0.010 (0.011)	-0.029 (0.025)	-0.020 (0.017)	-0.006 (0.010)
Control Mean (Std. Dev)	0.723 (0.448)	0.285 (0.452)	0.094 (0.293)	0.717 (0.451)	0.261 (0.439)	0.100 (0.301)
B. Depression (PHQ-8)						
App Access (S.E.)	-0.132 (0.024)	-0.139 (0.020)	-0.054 (0.014)	-0.153 (0.025)	-0.136 (0.020)	-0.075 (0.013)
Any Usage Incentive (S.E.)	-0.042 (0.025)	-0.017 (0.020)	-0.011 (0.013)	-0.037 (0.025)	-0.008 (0.019)	0.004 (0.011)
Control Mean (Std. Dev)	0.680 (0.467)	0.324 (0.468)	0.119 (0.324)	0.631 (0.483)	0.287 (0.453)	0.115 (0.320)
N	2,330			2,312		

Notes: This table presents estimates of Intent-to-Treat (ITT) effects from offering access to the mindfulness meditation app, as well as additional usage incentives, on the share of participants reporting various levels of mental illness two and four weeks after randomization, using Equation 3. Anxiety is measured using the GAD-7 scale, where scores range from 0 (no symptoms) to 21 (every symptom is severe). Depression is measured using PHQ-8, where scores range from 0 (no symptoms) to 24 (every symptom is severe). With both metrics, scores above the cutoff of 5, 10 and 15 points are used in clinical practice to screen mild, moderate and severe anxiety and depression.

Table B.6: Effects of App Access, Immediate Meditation Incentives, and Stressful Task on Other Measures of Performance in the Proofreading Task

	True Errors Found					False Positives				Hourly Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
App Access (S.E.)	0.232 (0.103)	0.306 (0.119)	0.328 (0.141)	0.356 (0.171)	-0.012 (0.031)	-0.048 (0.035)	-0.018 (0.042)	-0.016 (0.050)	0.447 (0.475)	0.642 (0.575)	0.297 (0.709)	0.779 (0.888)
App x Immediate Meditation (S.E.)		-0.148 (0.123)		-0.057 (0.166)		0.072 (0.038)		-0.005 (0.048)		-0.390 (0.638)		-0.958 (0.946)
Stressful Task (S.E.)			-0.067 (0.167)	-0.067 (0.167)			0.001 (0.051)	0.001 (0.051)			-0.237 (0.711)	-0.235 (0.712)
Stressful x App (S.E.)			-0.192 (0.207)	-0.100 (0.240)			0.012 (0.063)	-0.065 (0.069)			0.302 (0.953)	-0.273 (1.161)
Stressful x App x Immediate (S.E.)				-0.185 (0.246)				0.156 (0.075)				1.145 (1.291)
Control Mean (S.D.)	14.756 (2.681)	14.756 (2.681)	14.830 (2.576)	14.830 (2.576)	0.291 (0.801)	0.291 (0.801)	0.274 (0.794)	0.274 (0.794)	30.988 (13.449)	30.988 (13.449)	31.760 (14.299)	31.760 (14.299)
N	2257	2257	2257	2257	2257	2257	2257	2257	2257	2257	2257	2257

Notes: This table complements Table 4 by presenting effects on other outcomes from the Proofreading task. True Errors Found is the number of correctly identified mistakes in the proofread excerpts. False Positives is the number of correct words that a participant incorrectly flagged as errors. Hourly earnings divides total earnings by time spent, and is reported in units of dollars per hour. Effects are estimated using Equations 4, 5, 6 and 7.

Table B.7: Effects of App Access, Immediate Meditation Incentives, and Stressful Task on Risk Aversion Under Certainty and Uncertainty

	$P_{\text{uncertainty}}$				$P_{\text{certainty}}$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
App Access (S.E.)	0.490 (0.698)	0.826 (0.827)	2.365 (0.990)	1.938 (1.151)	0.825 (0.858)	1.358 (1.030)	1.483 (1.221)	1.828 (1.459)
App x Immediate Meditation (S.E.)		-0.673 (0.923)		0.851 (1.251)		-1.067 (1.111)		-0.686 (1.567)
Stressful Task (S.E.)			2.380 (1.042)	2.380 (1.043)			0.730 (1.307)	0.730 (1.308)
Stressful x App (S.E.)			-3.767 (1.392)	-2.230 (1.650)			-1.321 (1.716)	-0.942 (2.059)
Stressful x App x Immediate (S.E.)				-3.095 (1.843)				-0.774 (2.223)
Control Mean (S.D.)	72.047 (15.995)	72.047 (15.995)	70.887 (16.531)	70.887 (16.531)	58.409 (19.768)	58.409 (19.768)	58.089 (20.091)	58.089 (20.091)
N	2249	2249	2249	2249	2252	2252	2252	2252

Notes: This table complements Table 6 and presents the effect of the Stressful Task on the two outcomes directly estimated in the Risk Choices module. Effects are estimated using Equations 4, 5, 6 and 7.

- $P_{\text{uncertainty}}$ is the probability that makes a participant indifferent between a lottery that pays \$30 with probability $P_{\text{uncertainty}}$ and \$0 otherwise, and a lottery paying \$10 or \$30 with equal probability.
- $P_{\text{certainty}}$ is the probability that makes a participant indifferent between a lottery that pays \$30 with probability $P_{\text{certainty}}$ and \$0 otherwise.

Table B.8: Impacts of App Access and Immediate Meditation Incentives on Avoidance of Low Probability Losses and Unpleasant Information

	Avoid Loss		Avoid Unpleasant Info	
	(1)	(2)	(3)	(4)
App Access (S.E.)	0.036 (0.020)	0.011 (0.024)	−0.002 (0.014)	−0.021 (0.017)
App x Immediate Meditation (S.E.)		0.050 (0.026)		0.038 (0.019)
Control Mean (S.D.)	0.319 (0.466)	0.319 (0.466)	0.454 (0.335)	0.454 (0.335)
N	2257	2257	2257	2257

Notes: This table presents the effect from offering app access and incentives to meditate immediately before the survey on the propensity to avoid small probability losses and potentially unpleasant information, two decisions where behavior may vary if individuals become more or less emotionally sensitive. Effects are estimated using Equations 4 and 5.

- “Avoid Loss” is a dummy for refusing to take a lottery that pays \$1 with probability 0.99 and loses \$10 with probability 0.01, where the loss is made visually salient.
- “Unpleasant Real Info” is the mean of four dummies for avoiding pieces of potentially unpleasant but useful information, while “Neutral Info” is a dummy for not selecting the decision-relevant, emotionally neutral information in the information selection and stock purchasing game. Both outcomes are described in B.2. We show effects on each individual dummy in Table B.9.

Table B.9: Effects of App Access and Immediate Meditation Incentives on Avoidance of Useful Information

	Avoid Unpleasant Info								Avoid Neutral Info	
	Life Expectancy		Risk Factors of Dementia		Risk of Job Automation		Financial Risk in Retirement		Signal in Lottery	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
App Access	0.018	0.021	−0.019	−0.037	−0.013	−0.032	0.008	−0.033	0.038	0.021
(S.E.)	(0.020)	(0.024)	(0.020)	(0.023)	(0.021)	(0.025)	(0.021)	(0.025)	(0.020)	(0.023)
App x Immediate Meditation		−0.005		0.037		0.039		0.081		0.033
(S.E.)		(0.026)		(0.025)		(0.027)		(0.027)		(0.026)
Control Mean	0.319	0.319	0.328	0.328	0.595	0.595	0.576	0.576	0.294	0.294
(S.D.)	(0.466)	(0.466)	(0.470)	(0.470)	(0.491)	(0.491)	(0.494)	(0.494)	(0.456)	(0.456)
N	2257	2257	2257	2257	2257	2257	2257	2257	2257	2257

Notes: This table details the treatment effects of offering access to the app and incentives to use it immediately before the experimental decision-making session, on the avoidance of various pieces of useful information.

Table B.10: Five Facet Mindfulness Questionnaire

	Total (1)	Observing (2)	Describing (3)	Acting with Awareness (4)	Non-judgment (5)	Non-reactivity (6)
App Access (S.E.)	2.912 (0.440)	0.685 (0.139)	0.451 (0.120)	0.553 (0.117)	0.719 (0.109)	0.515 (0.102)
Any Usage Incentive (S.E.)	0.324 (0.225)	-0.066 (0.082)	0.049 (0.132)	0.128 (0.077)	0.122 (0.088)	0.098 (0.108)
Long Incentive (S.E.)	-0.395 (0.392)	0.013 (0.108)	-0.002 (0.149)	-0.122 (0.092)	-0.155 (0.086)	-0.123 (0.158)
N	2330	2330	2330	2330	2330	2330

Notes: This table presents effects on mindfulness, as measured by the Five Facet Mindfulness Questionnaire. The regressions use a modified version of Equation 3, introducing an additional dummy for receiving Long Incentives.

Table B.11: Willingness to Pay

	Baseline (1)	4 Days (2)	7 Days (3)	11 Days (4)	30 Days (5)
Cash Transfer (S.E.)	−0.53 (1.10)	2.33 (1.25)	2.71 (1.33)	2.38 (1.35)	1.31 (1.55)
App Access (S.E.)	0.08 (1.08)	4.11 (1.32)	3.03 (1.43)	2.46 (1.46)	0.42 (1.70)
Any Usage Incentive (S.E.)	0.27 (0.92)	−0.01 (1.17)	1.51 (1.27)	1.08 (1.34)	0.08 (1.53)
Waitlist Mean (S.D.)	49.06 (29.86)	49.13 (30.42)	49.00 (30.33)	49.20 (30.56)	49.19 (31.02)
Sample Size	2384	2260	2140	2189	2310

Notes: This table presents intent-to-treat effects of offering a cash transfer, access to the mindfulness meditation app, and usage incentives on willingness to pay for a three-month extension of app access. The estimating equation is Equation B.5. Participants report willingness to pay on a scale from 0 to 100 U.S. dollars, as part of a probabilistic Becker-DeGroot-Marschak elicitation.

Table B.12: Subjective ATE

	Baseline Mean (1)	Time After Randomization			
		4 Days (2)	7 Days (3)	11 Days (4)	30 Days (5)
Cash Transfer (S.E.)	−0.02 (0.13)	0.02 (0.13)	−0.06 (0.14)	−0.10 (0.13)	−0.13 (0.13)
App Access (S.E.)	0.09 (0.14)	0.15 (0.13)	0.13 (0.14)	−0.03 (0.13)	−0.28 (0.13)
Any Usage Incentive (S.E.)	0.08 (0.11)	0.01 (0.11)	−0.08 (0.12)	0.04 (0.12)	0.05 (0.12)
Waitlist Mean (S.D.)	3.63 (2.21)	3.58 (2.12)	3.58 (2.24)	3.61 (2.02)	3.66 (2.08)
Sample Size	2384	2289	2131	2185	2309

Notes: This table presents intent-to-treat effects of offering a cash transfer, access to the mindfulness meditation app, and usage incentives on participants' predictions about the average treatment effect in the study. The estimating equation is Equation B.5. See Section B.5.2 for more details on the elicitation.

Table B.13: Ease of Meditating

	4 Days (1)	7 Days (2)	11 Days (3)	30 Days (4)
A. Ease of Finding Time and Space to Meditate				
Any Incentive (S.E.)	0.505 (0.159)	0.469 (0.177)	0.663 (0.177)	0.481 (0.188)
Long Incentive (S.E.)	-0.065 (0.162)	-0.150 (0.170)	-0.550 (0.174)	-0.412 (0.185)
App Access Only Mean (S.D.)	4.70 (2.11)	4.92 (2.07)	4.91 (2.07)	5.01 (2.43)
Sample Size	1098	825	816	956
B. Easy or Very Easy to Focus on Meditation				
Any Incentive (S.E.)	0.045 (0.032)	0.089 (0.033)	0.062 (0.032)	0.053 (0.030)
Long Incentive (S.E.)	0.014 (0.032)	-0.019 (0.032)	0.002 (0.032)	-0.012 (0.030)
App Access Only Mean (S.D.)	0.60 (0.49)	0.62 (0.49)	0.63 (0.48)	0.67 (0.47)
Sample Size	1352	1237	1264	1367

Notes: This table presents intent-to-treat effects of offering usage incentives on self-reported ease of meditating. We attempt to collect responses only from the 1,429 participants in the treatment group. The estimating equation is B.6. In Panel A, the outcome comes from answers to the question “In your experience so far, how easy or difficult is it to find a good time and space to meditate?”, measured on a scale from 0 (“very difficult”) to 10 (“very easy”). We treat this as an integer between 0 and 10. In Panel B, the outcome is the share of participants who responded that it was “very easy” or “somewhat easy” to focus on meditating for 10 minutes without quitting, given the right time and space. The other options were “very difficult” “somewhat difficult”, and “I don’t know—I have not been meditating”. See Section B.5.2 for more details on the elicitation.

Table B.14: Effect of Any App Usage on Key Outcomes

	Two Weeks Index (1)	Four Weeks Index (2)	Proofreading (3)	Three-Four Months Anxiety (4)
<u>A. App Access (Pooled)</u>				
Any App Usage	-0.405	-0.450	0.009	-0.258
(S.E.)	(0.025)	(0.026)	(0.005)	(0.041)
[95% AR Interval]	[-0.554, -0.404]	[-0.604, -0.437]	[-0.003, 0.029]	[-0.748, -0.313]
<u>B. App Access Only</u>				
Any App Usage	-0.391	-0.444	0.013	-0.214
(S.E.)	(0.037)	(0.039)	(0.007)	(0.048)
[95% AR Interval]	[-0.538, -0.377]	[-0.597, -0.430]	[-0.001, 0.032]	[-0.782, -0.307]
<u>C. Short Incentives</u>				
Any App Usage	-0.456	-0.487	0.007	-0.241
(S.E.)	(0.035)	(0.037)	(0.007)	(0.048)
[95% AR Interval]	[-0.576, -0.432]	[-0.610, -0.455]	[-0.004, 0.026]	[-0.718, -0.309]
<u>D. Long Incentives</u>				
Any App Usage	-0.438	-0.486	0.006	-0.260
(S.E.)	(0.035)	(0.036)	(0.007)	(0.048)
[95% AR Interval]	[-0.542, -0.402]	[-0.590, -0.439]	[-0.003, 0.028]	[-0.748, -0.338]
<u>E. Supplementary Information</u>				
<u>Sargan-Hansen Overidentification Test</u>				
J-statistic	0.99	0.18	0.16	0.11
<i>p</i> value	0.609	0.916	0.921	0.944
<u>First-Stage F</u>				
Pooled	2345	2774	2431	227
App Access Only	4014	5209	4225	233
Short Incentives	8783	9568	9022	322
Long Incentives	10220	10923	10389	318
<u>Sample Size</u>				
Pooled	2330	2311	2257	2009
App Access Only	1408	1398	1373	1213
Short Incentives	1400	1389	1366	1213
Long Incentives	1408	1396	1368	1207

Notes: This table presents estimates of the Local Average Treatment Effect (LATE) of using the app at all on key outcomes. The endogenous variable is an indicator for completing at least one meditation session on the app before taking a given followup survey. The estimating equations are B.4 for the first stage and B.3 for the second stage. Panel A presents estimates that instrument for app usage with all three treatment arms. Panels B, C, and D focus on the LATE for each treatment arm separately. Each panel provides a point-estimate, a heteroskedasticity-robust standard error, and a pre-specified 95% confidence interval from the Anderson-Rubin procedure for robustness to weak instruments. Panel E presents additional information useful for interpreting two-stage least squares regressions. First, it provides a Sargan-Hansen overidentification test for Panel A, which roughly corresponds to testing whether every instrument implies the same LATE. Observing a large J-statistic (or small *p*-value) implies that the AR confidence interval in Panel A is likely unreliable. Panel E also provides the first-stage F statistic and number of observations for the regressions in panels A through D.

Table B.15: Effect of Days Meditated on Key Outcomes

	Two Weeks	Four Weeks		Three-Four Months
	Index (1)	Index (2)	Proofreading (3)	Anxiety (4)
<hr/>				
<u>A. App Access (Pooled)</u>				
Days Per Week	-0.295	-0.196	0.008	-0.055
(S.E.)	(0.018)	(0.012)	(0.003)	(0.007)
[95% AR Interval]	∅	[-0.337, -0.304]	[-0.003, 0.024]	[-0.200, -0.083]
<u>B. App Access Only</u>				
Days Per Week	-0.330	-0.210	0.005	-0.051
(S.E.)	(0.031)	(0.020)	(0.006)	(0.012)
[95% AR Interval]	[-0.574, -0.402]	[-0.426, -0.304]	[-0.001, 0.033]	[-0.205, -0.080]
<u>C. Short Incentives</u>				
Days Per Week	-0.359	-0.249	0.010	-0.077
(S.E.)	(0.029)	(0.019)	(0.005)	(0.009)
[95% AR Interval]	[-0.539, -0.404]	[-0.400, -0.297]	[-0.004, 0.024]	[-0.204, -0.088]
<u>D. Long Incentives</u>				
Days Per Week	-0.306	-0.217	0.006	-0.060
(S.E.)	(0.023)	(0.016)	(0.005)	(0.011)
[95% AR Interval]	[-0.430, -0.320]	[-0.332, -0.246]	[-0.002, 0.021]	[-0.187, -0.084]
<hr/>				
<u>E. Supplementary Information</u>				
<u>Sargan-Hansen Overidentification Test</u>				
J-statistic	9.21	7.09	0.55	0.10
<i>p</i> value	0.010	0.029	0.759	0.951
<u>First-Stage F</u>				
Pooled	727	542	721	108
App Access Only	1218	1002	1195	186
Short Incentives	2371	1709	2261	245
Long Incentives	2623	2068	2709	335
<u>Sample Size</u>				
Pooled	2330	2311	2257	2009
App Access Only	1408	1398	1373	1213
Short Incentives	1400	1389	1366	1213
Long Incentives	1408	1396	1368	1207

Notes: This table presents estimates of the Local Average Treatment Effect (LATE) of days using the app on key outcomes. The endogenous variable is the number of days on which a participant completed a meditation session on the app, counting only sessions that took place before the survey in question. The estimating equations are B.4 for the first stage and B.3 for the second stage. Panel A presents estimates that instrument for app usage with all three treatment arms. Panels B, C, and D focus on the LATE for each treatment arm separately. Each panel provides a point-estimate, a heteroskedasticity-robust standard error, and a pre-specified 95% confidence interval from the Anderson-Rubin procedure for robustness to weak instruments. Panel E presents additional information useful for interpreting two-stage least squares regressions. First, it provides a Sargan-Hansen overidentification test for Panel A, which roughly corresponds to testing whether every instrument implies the same LATE. Observing a large J-statistic (or small *p*-value) implies that the AR confidence interval in Panel A is likely unreliable. Panel E also provides the first-stage F statistic and number of observations for the regressions in panels A through D.

Table B.16: Effect of 10 Minutes of Meditation on Key Outcomes

	Two Weeks	Four Weeks		Three-Four Months
	Index (1)	Index (2)	Proofreading (3)	Anxiety (4)
<hr/>				
<u>A. App Access (Pooled)</u>				
10 Minutes Per Day	-0.002	-0.002	0.000	-0.002
(S.E.)	(0.000)	(0.000)	(0.000)	(0.000)
[95% AR Interval]	∅	[-0.009, -0.007]	[0.000, 0.000]	[-0.014, -0.005]
<u>B. App Access Only</u>				
10 Minutes Per Day	-0.002	-0.003	0.000	-0.001
(S.E.)	(0.000)	(0.000)	(0.000)	(0.000)
[95% AR Interval]	[-0.009, -0.006]	[-0.011, -0.008]	[0.000, 0.001]	[-0.013, -0.005]
<u>C. Short Incentives</u>				
10 Minutes Per Day	-0.003	-0.003	0.000	-0.002
(S.E.)	(0.000)	(0.000)	(0.000)	(0.000)
[95% AR Interval]	[-0.008, -0.006]	[-0.010, -0.007]	[0.000, 0.000]	[-0.013, -0.005]
<u>D. Long Incentives</u>				
10 Minutes Per Day	-0.003	-0.004	0.000	-0.002
(S.E.)	(0.000)	(0.000)	(0.000)	(0.000)
[95% AR Interval]	[-0.006, -0.004]	[-0.008, -0.006]	[0.000, 0.000]	[-0.014, -0.006]
<hr/>				
<u>E. Supplementary Information</u>				
<u>Sargan-Hansen Overidentification Test</u>				
J-statistic	8.61	5.73	0.66	0.11
<i>p</i> value	0.014	0.057	0.720	0.945
<u>First-Stage F</u>				
Pooled	213	178	209	37
App Access Only	282	322	276	76
Short Incentives	773	575	749	95
Long Incentives	876	676	873	111
<u>Sample Size</u>				
Pooled	2330	2311	2257	2009
App Access Only	1408	1398	1373	1213
Short Incentives	1400	1389	1366	1213
Long Incentives	1408	1396	1368	1207

Notes: This table presents estimates of the Local Average Treatment Effect (LATE) of minutes using the app on key outcomes. The endogenous variable is the number of minutes meditated using the app, counting only sessions that took place before the survey in question. We divide the number of minutes by 10, which is the length of the typical introductory meditation session on the app. The estimating equations are B.4 for the first stage and B.3 for the second stage. Panel A presents estimates that instrument for app usage with all three treatment arms. Panels B, C, and D focus on the LATE for each treatment arm separately. Each panel provides a point-estimate, a heteroskedasticity-robust standard error, and a pre-specified 95% confidence interval from the Anderson-Rubin procedure for robustness to weak instruments. Panel E presents additional information useful for interpreting two-stage least squares regressions. First, it provides a Sargan-Hansen overidentification test for Panel A, which roughly corresponds to testing whether every instrument implies the same LATE. Observing a large J-statistic (or small *p*-value) implies that the AR confidence interval in Panel A is likely unreliable. Panel E also provides the first-stage F statistic and number of observations for the regressions in panels A through D.

Table B.17: Multiple Hypothesis Testing

Family	Subfamily	Claim	Unadjusted p -value (1)	Adjusted p -value (Subfamily) (2)	Adjusted p -value (Family) (3)	Adjusted p -value (Overall) (4)
Habit formation	Usage after incentives	short incentives	0.611	0.667	0.667	0.984
		long incentives	0.454	0.667	0.667	0.984
Mental health	ITT effect of app	2 week index	<0.001	<0.001	<0.001	<0.001
		4 week index	<0.001	<0.001	<0.001	<0.001
		3-4 month anxiety	<0.001	<0.001	<0.001	<0.001
	Added effect of incentives	2 week index	0.042	0.108	0.108	0.344
		4 week index	0.242	0.400	0.400	0.912
		3-4 month anxiety	0.636	0.640	0.640	0.984
Productivity	ITT effect of app	proofreading score	0.003	0.009	0.021	0.034
		proofreading wage	0.337	0.671	0.859	0.961
		Stroop score	0.592	0.786	0.916	0.984
		Stroop time taken	0.952	0.958	0.958	0.984
	Emotional interference	certainty premium	0.019	0.066	0.120	0.190
		proofreading score	0.537	0.741	0.916	0.984
		proofreading wage	0.190	0.446	0.687	0.859

Notes: Multiple testing corrections using the Romano and Wolf (2016) and 1,999 bootstrap draws to control the familywise error rate (i.e., the probability of falsely rejecting at least one “true” null hypothesis). We perform resampling within strata and treatment groups, so that each resampled dataset contains the same number of respondents in each treatment arm and stratum as in the original dataset.

