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

Advik Shreekumar — MIT

Pierre-Luc Vautrey* — MIT

April 5, 2024

Abstract

Mindfulness meditation has gained popularity, fueled by accessible smartphone apps and rising concerns about mental health. While such apps are claimed to affect mental well-being, productivity, and decision-making, existing experimental evidence is inconclusive due to limited sample sizes and high attrition. We address this gap by conducting a large-scale experiment with 2,384 US adults, randomizing access and usage incentives for a popular mindfulness app. App access improves an index of anxiety, depression, and stress by 0.38 standard deviations (SDs) at two weeks and 0.46 SDs at four weeks, with persistent effects three months later. It also improves earnings on a focused proofreading task by 2 percent. However, we find no effects on a standard cognitive test (a Stroop task), nor on decisions where past economics research has indicated that emotions affect choice. This study alleviates concerns about prior research on mindfulness and mental health, provides evidence for productivity gains, and suggests that these effects do not stem from traditional measures of cognitive improvement or fundamental changes in the preferences we measure.

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

1 Introduction

Poor mental health casts a long shadow on well-being. Depression and anxiety disorders are among the top contributors to disability-adjusted life years for people between 10 and 49, a fact drawn into sharper relief by the uncertainty and isolation of the COVID-19 pandemic (Vos et al., 2020; Santomauro et al., 2021). Such mental states affect not just our experience of the world, but also how we navigate it. Emotions and worries interfere with judgment, tax our limited attention, and reduce productivity (Johnson and Tversky, 1983; Loewenstein, 2000; Banerjee and Mullainathan, 2008; Kaur et al., 2021; Duquennois, 2021). As awareness of these problems has grown, so too has the demand from organizations and the general public for tools to improve well-being, attention, and productivity. Mindfulness meditation apps are one such popular tool, with hundreds of millions of downloads and billions of dollars in valuation.

Psychologists describe mindfulness meditation as a set of secular techniques that train users to "pay attention in a certain way, on purpose and non-judgmentally, to the present experience" (Kabat-Zinn, 2003). Medical professionals currently incorporate it into a variety of therapies for depression and anxiety, where its efficacy is well-established (Goldberg et al., 2021). Independently, many meditators cite general wellness as motivating their practice, but also intend to improve their energy, memory, and concentration (Cramer et al., 2016). Various large firms, including Google, Ford, and McKinsey, promote mindfulness among their employees for similar reasons.[1] Underlying all of these trends is the idea that self-observation is a learnable skill that can help one manage negative emotions and interrupting thoughts. This resonates with the recent focus in behavioral economics on attention as a primitive underlying behavior (Gabaix, 2019; Bordalo et al., 2020), animating the hypothesis that mindfulness meditation may have broader effects.

However, high-quality evidence on the efficacy of widely-used meditation apps and their influence on economic behavior is scarce. Meditation apps offer an introduction to the practice that is more self-directed, designed for a generic audience, and relatively unstigmatized. These factors make app-based meditation scalable, but also raise the question of whether it captures benefits of individually-tailored, in-person therapies. Past studies of online meditation offer limited evidence due to smaller samples or difficulties with attrition, with several meta-analyses raising concerns about methodological quality (Flett et al., 2019; Sommers-Spijkerman et al., 2021).[2] The effect of mindfulness on cognitive function, productivity, and decision-making are similarly ambiguous (Vonderlin et al., 2020; Whitfield et al., 2022).[3] A rigorous, large-scale experiment testing a digital mindfulness intervention would begin to resolve these open questions about an increasingly common practice.

This paper studies the effects of a leading mindfulness meditation app, Headspace, on mental

[1]See "Talking Mindfulness on the C.E.O. Beat", published in the New York Times on Nov. 28, 2018.

[2]Trials with larger samples, like Mak et al. (2018), tend to study more structured "mobile health" interventions that lack the flexibility of common meditation apps.

[3]Several 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 findings but have limited statistical power to detect moderately sized effects, with the exception of Iwamoto et al. (2020).

health and economic behavior by conducting a pre-registered randomized control trial (RCT) with 2,384 US adults interested in meditation, recruited via social media ads. In this three-month experiment, the first group receives free access to a premium version of the app (worth $39); 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.[4] We assess impacts on stress, anxiety, and depression over time, with validated questionnaires at six different stages of the experiment, up to three months after randomization. After two weeks participants complete a cognitive test, a paid proofreading task and a set of incentivized choices between risky outcomes.

Detailed administrative usage data demonstrate high engagement with the app for the first two weeks after people receive access to the app. Of the participants who receive access without additional incentives, 80.3 percent meditate with it at least once and use it an average of 11 times for a total of 95 minutes during the first 16 days.[5] Incentives 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).[6]

Our first main finding is that offering access to the app meaningfully improves mental health, as measured by a 0.38 standard deviation (SD) reduction in an index of depression, anxiety, and stress compared to the control group after two weeks ($p < 0.001$). App usage slowly declines thereafter, 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) 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 waitlist 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.[7]

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

[5]Most participants receive the license on a Saturday and incentives for usage apply during the next two weeks.

[6]Mindfulness meditation differs from other health behaviors in that one can learn meditation skills quickly and apply them without continuing to use the app. For example, three months after the intervention, forty-six percent of participants in the treatment arm report meditating without the app.

[7]A 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

As we are cautious to interpret changes in self-reported measures as real improvements in well-being, we design our experiment and analyses with a critical eye towards demand effects. Several patterns in the data contradict the hypothesis that these effects are driven by experimenter demand. First, we measure participants' tendency to represent themselves favorably with the Marlow-Crowne Social Desirability Scale. Our treatment effects are not driven by participants with a tendency to overstate their own good qualities. Second, we elicit participants' baseline beliefs about the effectiveness of the app, incentivizing them to accurately predict changes in mental health among other participants in the treatment and control conditions. Participants who express that treatment effects will be small are unlikely to be responding to experimenter demand. We find treatment effects even among participants who express the least optimistic beliefs that the app will improve mental health. Third, we examine a subset of the waitlist group who receive an unconditional cash transfer at randomization, equal in value to the license. The cash transfer has no effect on reported mental health, suggesting that reciprocity effects are minimal in our setting. Fourth, simple demand effects explanations would generate a constant treatment effect, not the gradual pattern of improvements we observe. These patterns, along with others that we discuss more fully in the paper, support the claim app-based mindfulness produces real improvements in mental well-being.

Turning towards attention and productivity, we find that 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 identify simple spelling and grammar errors in paragraphs of text with no time pressure,[8] measuring a dimension of productivity that is essential in many workplaces. A 1.9% improvement may seem economically modest; however, the intervention's costs were also quite modest.[9] 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.

By contrast, we find no effect on performance in a time-limited Stroop test, a standard measure of attention control and reaction time in the face of visual distractions (95% confidence interval: $[-0.7, 1.3]$ percent change in earnings). This is striking given an influential psychology literature suggesting that by training one's attention, mindfulness improves cognitive ability as measured with such tasks (Jha et al., 2007). However, a recent meta-analysis concludes that these effects may be driven by studies with smaller samples or greater attrition (Whitfield et al., 2022). We provide an informative null finding, based on a large sample of participants who measurably engage with meditation, suggesting that mindfulness does not generically affect how quickly individuals parse stimuli. It also implies that effects on proofreading task may stem from other mechanisms, such as training the ability to focus (Brown et al., 2022) or the effect of improved mental health on productivity (Mani et al., 2013; Ridley et al., 2020).

We then turn to decision-making, where literatures in economics and psychology intersect to raise hypotheses about the effects of mindfulness on risk-taking and information acquisition. A body of

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).

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

[9]A 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.

work in behavioral economics finds that priming negative emotions affects decisions under uncertainty (Johnson and Tversky, 1983; Loewenstein and Lerner, 2003; Raghunathan et al., 2006; Slovic and Peters, 2006; Callen et al., 2014); and that information acquisition decisions as well as small probability risk taking (Golman et al., 2017; Bénabou and Tirole, 2016; Loewenstein et al., 2001) are influenced by aversion to unpleasant feelings. Complementary work in neuroscience conceptualizes mindfulness as improving the ability to prevent emotions from capturing one's attention, based on neuroimaging and self-report studies (Guendelman et al., 2017). We draw these literatures together to test whether mindfulness blunts the effects of negative emotions in decision making, and affects decisions to acquire unpleasant information or take small-probability risks.

We find no meaningful differences in behavior between participants with access to the app and those on the waitlist. Our main investigation cross-randomizes some participants to engage with stressful thoughts and then presents all participants with incentivized risk decisions. While stressful thoughts do make participants more risk averse, app access does not modulate this effect. In secondary investigations, we elicit participants' willingness to avoid a small probability loss, and to avoid potentially worrying information, such as the the chances that various occupations will be replaced by automation, or risk factors of dementia. Here as well, app access does not shift behavior.[10] While we cannot rule out the possibility that mindfulness training might affect other economic decisions, or that longer-term practice may have some impact, our findings show that two weeks of mindfulness training do not impact choices in a set of economic decisions that closely relate to a leading conceptualization of mindfulness.

Our final contribution is to separately estimate the long-term effects of undergoing mindfulness training from the short-term effects of completing a single meditation session. Such differences arise for many other skills and habits. For example, a student may eke out a higher grade by cramming before an exam even if he has diligently studied in the previous weeks. Or, a brisk jog may tire an athlete in the short term but make her faster if she runs regularly.[11] The effects we have discussed so far are longer-term effects, analogous to regular exercise or studying. Separating these from short-term effects is relevant for both practitioners and researchers of mindfulness. Practically, individuals and organizations who view mindfulness as a decision aid may wonder if they should meditate during the workday or before a big decision. Scientifically, researchers may wish to study mindfulness by conducting a meditation session before administering experimental tasks. Informing both decisions distinguishes our work from existing studies that focus on either short-term or long-term effects, but to our knowledge, never both.

We estimate the short-term effects of a meditation session on decision making by incentivizing half of the group with app access to meditate immediately before the productivity and decision tasks. Sixty percent of those who receive the incentives do so, compared to 3 percent of the unincentivized. Comparing the incentivized participants to participants who have app access but did not receive a

[10]These null findings are unlikely to be due to survey design or lack of participant engagement. As we discuss next, a separate randomized treatment affects information avoidance and salient loss aversion, suggesting that these tasks are capable of capturing treatment effects.

[11]If you prefer: a long runs feel longer in the short run than the long run.

short-term incentive estimates the ITT effect of a single meditation session for practitioners.[12] This is conceptually distinct from the (largely null) long-term effects of mindfulness on decision-making we have previously discussed.[13]

We find some evidence that meditation sessions affects productivity and decision making in the short term. Incentivizing meditation reducing performance on the proofreading task by 1.2 percentage points ($p = 0.084$), especially for participants who engage with stressful thoughts before the task (-2.1 percentage points, $p = 0.028$). The immediate meditation group is also 3.8 percentage points more prone to avoid unpleasant information (8.8%, $p = 0.042$), and 5.1 percentage points more likely to avoid a salient small-probability loss (15.5%, $p = 0.053$). Although a classical perspective would cast changes like reduced productivity or increased information avoidance as mistakes, another view is that they may simply reflect changes in utility based on increased attention to emotions (Bénabou and Tirole, 2016; Bolte and Raymond, 2023). Still, these results caution against making meditation a default practice before important tasks, and suggest that experiments studying the immediate effects of meditation sessions may not inform us about the longer-term effects of the practice. Correcting for multiple testing diminishes the strength of these findings, so we caution against reading them as definitive. Still, we note that despite clear evidence for improvements in mental health, the data are inconsistent with the hypothesis that a meditation session is a surefire shortcut to improved decisions.

This paper investigates a number of relevant hypotheses, and 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 A.1. We report 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 and decision making). Second, we explicitly correct p-values for multiple testing in Section A.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., 2023). 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

[12] One concern is that these incentives 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 (longer than actually needed) to take this survey in one sitting in a quiet environment. We made it clear that the incentives would only be granted if the session was completed within 30 minutes of seeing the announcement.

[13] We estimate longer-term effects by comparing the waitlist group to the app access group that did not receive an incentive to meditate before the making decisions.

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; Duquennois, 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., 2023).

Third, 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 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 initial symptoms of anxiety or depression nor is it restricted to a student or employee population from any given organization, which contributes to building externally valid evidence that affordable online mindfulness meditation training has large potential for the general population.

The rest of this paper proceeds as follows. Section 2 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. Finally, Section 6 concludes with a discussion of our findings and open questions.

2 Background and Study Design

2.1 Mindfulness Meditation and the Headspace App

Mindfulness was first introduced to clinical psychology to treat chronic pain by incorporating meditation techniques from the Buddhist tradition into secular clinical therapies (Kabat-Zinn, 1982). The techniques typically instruct practitioners to direct their focus to a sensation, such as the breath, to notice how other thoughts and sensations capture attention, and to refocus back to the initial sensation. The approach was quickly extended to create a set of interventions to improve mental health. 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 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, (Davidson and Kaszniak, 2015), reflecting that mindfulness meditation may affect many mental processes.[14]

The proposed definitions agree on a core idea captured concisely by Kabat-Zinn (2003): mindfulness is the act of *"paying attention in a certain way, on purpose and non-judgmentally, to the present experience."* First, *on purpose* touches on the idea of 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.[15] Third, *present experience* is what attention should target: current sensations, emotions, and stimuli as opposed to thoughts and emotions referring to the past (including ruminations) or the future (including anticipations).

Headspace is a leading meditation app, 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 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.[16] Crucially, users choose which sessions to engage in and when, in contrast with traditional therapies that require scheduled meetings with a healthcare processional or trainer.

2.2 Recruitment of Participants

The practice of mindfulness meditation takes effort, time, and regularity. Measuring its effects requires recruiting and retaining participants who will take up the intervention for several weeks. Our recruitment strategy therefore favors individuals who are genuinely interested in trying meditation to maximize take-up. To minimize post-randomization attrition, we assign participants to an

[14]Notable 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

[15]This 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).

[16]See Appendix A.5 and particularly Figure B.5 for details

intervention arm after they have completed three distinct surveys on separate days, demonstrating the willingness and ability to participate in a longer study.

Participants were recruited using Facebook and Instagram ads specifying that the study is about meditation, and that it includes a Headspace subscription and 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.[17] All told, 10,615 respondents completed our screening survey.[18]

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 excluded 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.[19] 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).[20]

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, 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 per-

[17]These clicks include Facebook reactions (e.g., "liking" the ads), comments, visits to the associated Facebook page, and opening the link to the screening survey.

[18]Seventy-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. 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.

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

[20]We 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).

formance 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.

2.3 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) App No Incentive, (ii) App Plus Short Incentives, intended to encourage at least some experimentation with the app among most users, (iii) App Plus Long Incentives, intended to generate lasting habits through frequent initial usage, (iv) Pure Waitlist control, or (v) Waitlist 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.

Three-fifths of participants receive a free three-month subscription to Headspace in the form of a voucher code immediately at the end of the third baseline survey. They are advised to start with a specific series called "Basics" that trains novice users in the technique of mindfulness meditation. Participants who receive the license are further randomized into one of three groups with equal probability: (i) App No Incentive, (ii) App Plus Short Incentive, or (iii) App Plus 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 Incentive) or at least 10 separate days (Long Incentive) 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 Short Incentive increases the value of trying the app for a few days and is designed to encourage experimentation among a broad set of participants, including skeptical ones who believe the app will deliver fewer benefits. The Long Incentive may increase usage among fewer participants but may create longer-lasting habits through sustained use.

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.[21]

We divide the waitlist participants evenly between a Pure Waitlist control and a Waitlist Cash Transfer. The Pure Waitlist receives no subsequent intervention. To test whether any effects from offering the app access may be generated by reciprocity or wealth effects due to the value of the app,

[21] We 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. 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.

the Waitlist Cash Transfer group receives a $15 cash transfer as a highly fungible online gift card. This amounts to a little more than the cost of a Headspace license for four weeks.[22]

2.4 Further Randomization

Two weeks after randomization, participants complete an online experimental session containing our main economic outcomes. This module includes two further randomized treatments designed to (i) encourage a subset of the App Access group to meditate immediately before engaging in the economic tasks, and (ii) experimentally increase stress in a subset of participants. We assign these treatments using the same strata as in the assignment among the app versus waitlist groups. We defer details to Section 5, where they are most relevant to interpreting results.

2.5 Sample Composition, Balance, and Attrition

Not unexpectedly given our recruitment strategy, our sample differs from the general population in a few important ways. 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 is majority female, more educated than the general population, and more likely to politically identify as a Democrat or independent. The screening criteria mean that few participants have household incomes over $150,000 or more children at home than the average. Individuals identifying as Black or Hispanic are under-represented in our sample, as are those older than 60 years. 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.[23]

Table 1 reports estimated differences between the Waitlist and App Access groups obtained by regressing each variable on a treatment dummy, pooling the two Waitlist arms and three App Access arms for brevity.[24] We observe statistically detectable imbalance in the distribution of income, with the treatment arms having slightly more participants with income in the $35,000-$74,999 bin and fewer in the adjacent bins. This is somewhat an artifact of our binned income measure: the treatment group having more participants in the $35,000-$74,999 bin trivially means it will have fewer participants in other bins. It is also unsurprising, as we assess balance on many covariates. We view this imbalance as small in absolute terms and our main regression specifications will not adjust for it. 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 A.3 we employ the debiased machine learning approach of Chernozhukov et al. (2018) to

[22]We observe near-perfect compliance with treatment assignment in the waitlist groups. Two participants out of 479 in the Pure Waitlist group and 1 in the Waitlist Cash Transfer obtained voucher codes prematurely by retaking our baseline survey. To address this, we use their initial responses to the baseline survey and report intent-to-treat analyses throughout.

[23]We describe our elicitation of these symptoms in Section 4

[24]Table B.1 separates the five treatment arms.

flexibly adjust for potential imbalance, finding that point estimates and standard errors are relatively unchanged.

Our sample size and retention rates redefine the frontier of mindfulness research in adults. Figure 2 shows these statistics for our 4-week and 3-4 month endlines, as well as for comparable studies in recent meta-analyses (Galante et al., 2018; Sommers-Spijkerman et al., 2021). We recruit 2,384 participants, retaining 97.7% of participants at two weeks and 94.8% at four weeks. Experiments with comparable retention rates have fewer than 500 respondents, while those with over 1,000 participants have retention rates below 50 percent. The size and stability of our sample enable us to measure the effects mindfulness with minimal concern about attrition polluting our estimates. We can also consider a suite of outcomes, building a comprehensive set of findings while maintaining power after a multiple testing correction.

That said, individuals in the App Access group are slightly less likely to participate in the main endline surveys (1-3 percentage points, $p < 0.05$). Table B.2 presents retention rates for each of our surveys. We view these differences as small, especially relative to response rates of 97.7% and 94.8%. Statistical significance aside, attrition would be most problematic if it created imbalance between the treatment and control arms. In Appendix A.4 we show that attrition does not create concerning imbalance, does not appear to be "worst case", and that even so, worst-case bounds on treatment effects are similar to those in the main text under a common model of selective attrition (Lee, 2009).

3 App Usage and Effects of Incentives

We receive administrative usage data from Headspace for each voucher code we distribute in the study. The data describe the title, start time, and duration of each recording users listen to during their 90-day voucher period.[25] We find that usage is high even among participants assigned to the App 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 engage with it at a high rate. We then show that the App Plus Short Incentives and App Plus Long Incentives treatments generate additional usage in the period of time when they are active, with small effects once they expire.

We estimate effects of receiving any usage incentive using the following regression:

$$Y_i = \delta_s + \beta_1 \text{AppAccess} + \beta_2 \text{ShortIncentive} + \beta_3 \text{LongIncentive} + \epsilon_i, \tag{1}$$

where Y_i is the outcome for individual i, δ_s are fixed effects for randomization strata, ϵ_i is an error term, and the remaining terms are indicators for treatment assignment. The "AppAccess" term compares the App No Incentive group to the pooled waitlist groups, and the "ShortIncentive" and "LongIncentive" terms describe the marginal effects of each incentive relative to the App No Incentive

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

group. Table 2 presents treatment effects on an indicator for completing any meditation session, on the number of days with a meditation session, and on the number of meditation sessions, at several points in the experiment. Figure 3 complements these estimates by depicting engagement in continuous time, as the proportion of participants who use the app within a sliding three-day window.

Overall, most participants use the app frequently at the beginning of the study period even without incentives. Eighty percent of participants in the App No Incentive group record at least one session in the first two weeks, using the app for an average of 5.2 days out of 16. Usage steadily declines in the following weeks, but 46.7 percent of participants still log at least one session more than four weeks after randomization.

Incentives increase initial usage of the app, especially when they require more usage to qualify payment. Both Short and Long Incentives have a similar effect on propensity to try the app in the first two weeks, raising it by 8 to 9 points relative to the App No Incentive group (panel A, column 1). Long Incentives increase the number of days with app usage by 51.7 percent during this period, compared to a 25 percent increase from Short Incentives (panel B, column 1).

Incentives have little lasting effect after they expire on both the extensive and intensive margins. Participants in the incentive groups remain 7-8 percentage points more likely to use the app in the two weeks after incentives expire (panel A, column 2). This difference diminishes in the ensuing weeks (panel A, column 3). In the long term, they may even use the app more sporadically than those in the No Incentive group, but these differences are nearly indistinguishable from zero (panels B and C, column 4).

These patterns contrast with previous findings that incentivizing health behaviors, such as going to the gym or washing hands, generates lasting habits (Charness and Gneezy, 2009; Acland and Levy, 2015; Hussam et al., 2019). One interpretation is that unlike exercise or hand washing, which require continued practice to reap rewards, 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 our final followup survey three to four months after randomization, roughly 46 percent of treatment participants and 36 percent of waitlist participants report meditating without the app.

4 Effects on Mental Health

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, in the two weeks after randomization, and three to four months after randomization.

Our anxiety measure is 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. The GAD 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 follow mental health in the short term, we administer the GAD-2 scale in brief surveys 4, 7, and 11 days after randomization. This allows us to sketch general changes in anxiety levels and keep participants engaged with short tasks early in the experiment.

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).

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.

To improve power and reduce the number of hypotheses we test, we combine these measurements into a mental health index. In each followup survery (e.g., in the two-week followup), we standardize each elicited scale by subtracting the Pure Waitlist mean and dividing by the Pure Waitlist standard deviation. We then take the simple average of the standardized scales, so that lower values of the index indicate fewer reported symptoms of mental distress.

4.1 Estimation

We now estimate the intent-to-treat effects of providing the mindfulness app and incentivizing its usage. In doing so, we take advantage of baseline measures of mental health to improve precision. Our regression specification is:

$$Y_i^{\text{post}} = \delta_s + \beta_1 \text{AppAccess} + \beta_2 \text{AnyIncentive} + \gamma Y_i^{\text{pre}} + \epsilon_i \tag{2}$$

where Y_i is the outcome for individual i, δ_s are fixed effects for randomization strata, ϵ_i is an error term, and the remaining terms are indicators for treatment assignment. The "AppAccess" term compares the App No Incentive group to the pooled waitlist groups, and the "AnyIncentive" describes the marginal effect of the pooled incentive groups relative to the App No Incentive group. Pooling the App Plus Short Incentive and App Plus Long Incentive provides a succinct summary of their effects. Table 3 presents treatment effects on an index of the mental health scales we elicit. Results are similar if we separate the treatment arms (Table B.6), and if we drop the pre-treatment mental health index or flexibly control for it and other covariates with debiased machine learning (Figure B.2).

4.2 Effect of App Access on Mental Health

Providing access to the app leads to marked improvements in mental health. Two weeks after randomization, the App Access group reports an improvement of 0.38 standard deviations (SDs) on the mental health index. These effects become larger at four weeks, at 0.46 SDs. Once the Waitlist group gains access to the app they report reduced anxiety, and the App-Waitlist gap shrinks to 0.2 SDs at the three to four month mark. The persistent effect of initial app access after the waitlist ends are consistent with higher uptake among the App No Incentive group. In sum, we find that offering app-based mindfulness meditation improves self-reported mental health for at least three months after intervention.

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. However, as Sommers-Spijkerman et al. (2021) emphasize and Figure 2 shows, past studies of digital mindfulness interventions suffer from smaller sample sizes and high attrition rates. In addition, fewer studies evaluate self-directed programs such as Headspace. Our estimates are thus a significant contribution to the evidence base for widely used mindfulness interventions.

The effects are smaller than but comparable to previously measured effects of in-person therapy. For instance, Cuijpers et al. (2013) conduct a meta-analysis of in-person cognitive behavioral therapy (CBT) and estimate an effect size of 0.53 SDs on depression. Directly comparing these to our estimated effects requires nuance. Studies of CBT and pharmacotherapy typically target participants with diagnosed mental illnesses, conditions that are challenging to treat. While we conduct our study during a time of heightened mental illness (the Covid-19 pandemic), we do not explicitly recruit participants based on their mental health.[26] We expect app-based mindfulness to have a smaller effect if tested in a clinical population.

Unsurprisingly, app access has a larger effect on mental health than economic policies that do not directly target this kind of well-being. The effects we find are about ten times larger than previously estimated effects from receiving health insurance on depression in the Oregon Medicaid experiment (0.05 SDs in Finkelstein et al. (2012)). They are also about as large as the mental health improvements that a 0.2 SD increase in household income would generate, according to structural

[26] Using responses to our baseline survey and common clinical cutoffs, 32 percent of our sample screens positive for anxiety (GAD-7 $\geq$ 10) and 26 percent for depression (PHQ-2 $\geq$ 3). These are not proper diagnoses, but higher scores on GAD-7 and PHQ-2 are highly predictive of diagnoses based on longer interviews with a medical professional (Plummer et al., 2016; Arroll et al., 2010).

estimations that use cross-sectional variations (Alloush and Wu, 2023). Of course, health insurance and increased income confer many other benefits and operate at a longer time horizon than our experiment; we provide these comparisons as another benchmark.

The effects we measure are not due to improvements among only certain subgroups, nor are they artifacts of treating mental health measurements as continuous variables. Figure 4 presents the cumulative distributions of the anxiety, depression and stress scores, and shows that improvements are distributed across the range of symptom severity. Table B.5 shows treatment effects on the proportion of participants who report at least mild, moderate or severe symptoms of anxiety and depression. Effects at all levels are highly statistically significant.

4.3 Effect of Usage Incentives on Mental Health

Incentives appear to increase mental health improvements moderately, if at all. While the incentive groups consistently report better mental health than the App No Incentive group, these effects tend to be smaller than we can resolve with our sample. The Short Incentives group does register a statistically significant improvement in mental health at the 16-day mark, where we estimate an additional 0.08 SD improvement to the index of mental health scores. This is about 19 percent of the effect of app access alone.

This small effect may appear surprising given our finding that incentives induce roughly 25 percent and 50 percent increases in days with app usage. One interpretation is that meditation is most effective when one wants to do it, and individuals are well equipped to choose how and when to use the app. Forcing more sessions past this point may result in low-effort or low-quality meditation. In other words, compliance with guided meditation may involve more than the physical act of listening to a recording on an app. This would set meditation apart from more easily measured health behaviors, such as taking steps, exercising, or washing hands (Charness and Gneezy, 2009; Acland and Levy, 2015; Hussam et al., 2019; Aggarwal et al., 2020). Developing better measures of compliance than app usage 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 on Mental Health

Participants in our experiment are free to choose when and how they use the app, creating natural opportunities for heterogeneous treatment effects. We now turn to estimating heterogeneity along baseline covariates that enrich our understanding of average effects. To do so, we fully interact treatment indicators in Equation 2 with dummies for whether each respondent is strictly below or weakly above the median value of the baseline covariate. The regression specification is

$$\begin{aligned} Y_i^{\text{post}} = &\delta_s^{\text{lo}} \times L_i + \beta_1^{\text{lo}} \text{AppAccess} \times L_i + \beta_2^{\text{lo}} \text{AnyIncentive} \times L_i + \\ &\delta_s^{\text{hi}} \times H_i + \beta_1^{\text{hi}} \text{AppAccess} \times H_i + \beta_2^{\text{hi}} \text{AnyIncentive} \times H_i + \gamma Y_i^{\text{pre}} + \epsilon_i \end{aligned} \quad (3)$$

where L_i and H_i are dummies for being strictly below or weakly above the median, respectively. Figure 5 presents estimates, and Table B.7 formally tests for differences in the below- and above-median subgroups.

We first present heterogeneity along four axes: baseline anxiety, baseline mindfulness, prior beliefs about the effect of meditation on anxiety, and a measure of social desirability. Our measure of baseline mindfulness is the Five Facet Mindfulness Questionnaire, a 15-item self-reported scale capturing various aspects of emotional awareness and mental habits.[27] We elicit prior beliefs about the effects of meditation by asking participants to forecast anxiety rates among other subjects who are randomized to treatment and control, with incentives for accurate predictions.[28] We defer discussion of social desirability to the next section, where we address possible demand effects.

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

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 more skeptical of the benefits are two such groups. Panel B of Figure 5 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 extrapolating to individuals who are uninterested in meditation.

4.5 Experimenter Demand Effects

An important concern with self-reported scales is that participants may not honestly report their mental states. Our estimates would be biased if participants respond based on their beliefs about the study's purpose (de Quidt et al., 2019), or respond strategically to manipulate study outcomes.

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

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

We wish to take these concerns seriously, and will now assess potential sources of bias. In doing so we emphasize that demand effects, like treatment effects, are a positive explanation for observed results. A compelling demand effects story ought be plausible given our study design and consistent with our empirical findings.

One possibility is that treated participants may exaggerate the effects of the app. Perhaps they say what they think we'd like to hear, or they prefer to represent themselves positively. To investigate, we can ask whether treatment effects are driven by participants who are more likely to overstate their good qualities. With this in mind, we included the 13-item Marlow-Crowne Social Desirability Scale in our baseline survey. Higher scores on this scale indicate a tendency to endorse unrealistically positive statements about oneself, which can capture desires to distort one's self-image and to ingratiate others.[29] It would be concerning if participants who exaggerate these virtues also downplay their anxieties, producing a spurious treatment effect. However, in Table B.7 Panel D, we find that this is not the case. Treated participants with above median social desirability scores do not systematically report larger improvements.

Similarly, we can examine heterogeneity on baseline beliefs about the app's effectiveness. Participants who forecast smaller effects are unlikely to be motivated by experimenter demand: they express beliefs that are unlikely to please the researchers. We find large improvements in mental health even among these skeptical respondents.

Rather than attempting to appease experimenters, these effects on mental health may reflect respondents internalizing our experimental instructions, which we wrote to mitigate such concerns. Before measuring mental health, we inform participants that our sole intention was to understand how they were feeling, and that we did not expect any pattern of responses.[30]

Another threat would be participants reporting improved mental health in reciprocity for receiving a Headspace license from us. This is a curious concern, because we expect participants to behave reciprocally if they actually benefit from the app; there is no "favor" to return otherwise. Regardless, we can test for reciprocity effects by comparing participants in the Waitlist Cash Transfer arm to those in the Pure Waitlist. The Cash Transfer arm received an unconditional \$15 at the start of the study as recompense for being placed on a waitlist. If these participants were thinking strategically and behaving reciprocally, they could report worse mental health to increase the average treatment effect in our study. Other patterns of misreporting are possible as well. However, in Table B.6, we show that the cash transfer has no appreciable effect on reported mental health. Finally, the license is a one-time transfer, which would be more consistent with a constant treatment effect rather than the growing one we observe. Empirically and conceptually, reciprocity appears unlikely to explain away the effects we measure.

A more remote possibility is that Waitlist participants believe that they can receive the app sooner by reporting poorer mental health. This mechanism cannot explain why treatment effects persist in the long-term followup, well after all participants have received a license (Table 3, column

[29] For example, "No matter who I'm talking to, I'm always a good listener."

[30] This 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."

6). It is also unlikely that participants believe they can influence how we administer the experiment. Survey instructions clearly state that "answers to the questionnaires have no impact on [their] chances to receive the license now or later, which is determined by a computerized lottery." In addition, we administer several surveys during the waitlist period. If participants were attempting to influence our actions as experimenters, they would quickly observe that their responses do not change the course of the study.

Of course, we cannot anticipate every source of demand effects, and an engaged reader may posit a mechanism we have not discussed. Such a theory would have to square with our basic empirical findings. First, treatment group participants take up the app quickly, with over 75% beginning their first session within the first week of the study. This rules out explanations that rely on a steady stream of new meditators injecting bias into later surveys. Second, we observe a treatment effect that steadily grows over the first four weeks of the study and persists at the three-four month mark. This rules out explanations in which treated participants slightly bias each report of their mental health. Such biases would produce a constant treatment effect, not an increasing one. These patterns give us some confidence that our results reflect the legitimate ability of self-directed app-based mindfulness meditation to capture some benefits of previously studied clinical mindfulness therapies.

5 Effects on Productivity and Decision-Making

We now turn to the effects on productivity, cognitive function and decision-making, that we capture in our main online experimental session that involves additional cross-randomizations. Two weeks after randomization into receiving App Access, participants complete an online experimental session that contains our main economic tasks. This module follows the two-week mental health survey, and its goal is to measure productivity and decision making in the presence of distractions and emotions. Figure 1b presents the experiment's structure.

The survey begins with tasks that have built-in distracting or emotion-inducing stimuli: an incentivized Stroop cognitive test that measures attention control; decisions to acquire or avoid distressing information; and a decision to take a risky gamble with salient low probability losses. The survey concludes by randomizing participants to think of neutral or stressful thoughts before completing an incentivized proofreading task and making a set of decisions between lotteries. We pre-registered the proofreading task, the Stroop test, and the final set of risk-taking choices as our primary outcomes.

While our primary variation is the two weeks of unstructured meditation allowed by App Access, we also incentivize half of the App Access group to meditate before engaging with the survey proper. This allows us to test both the accumulated effects of mindfulness training, as well as the short-term effects of a single meditation session. Such short-lived effects and their relationship with longer-term App Access effects may interest practitioners who view meditation as a decision aid, they may help understand how meditation sessions work, and they may have methodological implications for future research. We discuss the immediate effects of a meditation session in Section 5.3, after first describing the effects of App Access on participants who don't receive this incentive.

5.1 Productivity

We design an easily explained attention task based on proofreading paragraphs of text to measure productivity. The general nature of this task is well-suited to our participant population, which has a variety of skills and backgrounds. The task requires no complicated training, making it a natural part of a longer survey where participants undertake other tasks as well.

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 non-errors. Participants begin with an endowment of 20 cents, so that initial mistakes are costly. A practice paragraph presents these incentives clearly. The median participant spends two minutes on the task and earns 90 cents out of a possible $1.05, for an effective wage of roughly $30 per hour.

Mindfulness meditation may affect a practitioner's ability to attend to a task at hand, especially in the presence of unrelated worries. To investigate this, we experimentally assign all participants to one of two conditions before they proofread: Neutral or Stressful. In the Neutral condition, participants describe an act or routine they use to stay grounded. The Stressful condition asks participants to recall and describe an unresolved source of stress. In both cases, we incentivize engagement with the task: 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. Participants proceed with the proofreading task after the Stressful/Neutral treatment.

One salient mechanism that could be responsible for effects on 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: the Stroop test (Jensen, 1965). In each item on the test, a word appears on the screen for three seconds, and the participant must select the name of the color that the word is printed in from one of five choices. However, the word that appear is also the name of a color, complicating the task by capturing attention. For example, if the word "yellow" appears in blue font, participants must submit "blue" and not "yellow" as their answer. Our implementation contains forty iterations under time pressure. We incentivize both speed and accuracy by paying a bonus that increases in how quickly participants click the correct answer, but decreases each time they select an incorrect answer. The median participant spends 66 seconds on the task and earns $1.14 out of a possible $1.50 from it, for an effective wage of roughly $61 per hour.

5.1.1 Estimation

We pool the three App Access arms when studying effects on economic outcomes, to increase power and in line with our pre-registration. For interpretability, we present outcomes on a percentage point scale by dividing out the maximum possible score and multiplying by 100. Our first regression is of the form

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

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 performance on Stroop or proofreading task in the baseline survey. The coefficient of interest is β_1, which estimates the average effect of App Access on task performance.

Next, to investigate whether effects on task performance are robust to inducing unrelated worries, we modify the regression to

$$\begin{aligned} Y_i^{post} =& \beta_1 \text{App Access}_i \times \text{Neutral Task}_i + \beta_2 \text{Stressful Task}_i + \beta_3 \text{App}_i \times \text{Stressful}_i \\ &+ \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i, \end{aligned} \tag{5}$$

where Neutral Task$_i$ and Stressful Task$_i$ are indicators for assignment to the neutral and stressful conditions, respectively. The coefficients of interest are β_1 and β_3, which estimate the average treatment effect of App Access for participants in the Stressful and Neutral arms, respectively. Table 4 presents treatment effects from both regressions.

5.1.2 Effect of App Access on Productivity

App Access increases performance on the proofreading task, overall as well as in both the Stressful and Neutral conditions. Treated participants earn 1.8 percentage points (2.0 percent) more from proofreading (Table 4, column 2). This effect comes largely from spotting more errors in the paragraphs, not from making fewer mistakes, nor spending more time on the task; formal tests of these sub-measures apepar in Appendix Table B.9. While the Stressful task itself does not degrade proofreading performance, participants with app access outperform their waitlisted counterparts in both the Stressful and Neutral conditions (column 3) by similar amounts.

By contrast, app access has no appreciable effect on performance in the Stroop test. Participants with App Access earn 0.3 percentage points (.4 percent) more from this task, which is indistinguishable from zero. Past work on mindfulness has suggested that the practice can improve cognitive control, as measured by the Stroop test (Jha et al., 2007), although a recent meta-analysis describes the quality of existing empirical work as inconclusive (Whitfield et al., 2022). Our results suggest that mindfulness has nearly no effect on the ability to exert cognitive control over such visual distractions.

We interpret these results as indicating that, at the levels of stress we induce and focus we require, mindfulness practice can improve performance on attention-demanding tasks. We do not find evidence for three mechanisms based on the psychology literature: improved cognitive control (measured by the Stroop task), insulation from worrying thoughts (induced by the Stressful treatment), or persistence (measured by time taken). One possible mechanism is that reduced levels of depression, anxiety, or stress are driving the effect, which is consistent with findings that mental illness reduces task productivity (Ridley et al., 2020). Another is that meditation trains practitioners to disengage from distractions, making it similar to interventions that train cognitive endurance (Brown et al., 2022).

5.2 Risk-Taking Decisions

We next study whether access to the app mitigates the interference of stressful thoughts on risk-taking and information acquisition. The first set of decision tasks comes at the beginning of the economic tasks survey, before the Stroop and proofreading tasks. We elicit willingness to accept potentially distressing information, as well as the choice to take a risky gamble with salient low probability losses. These are pre-registered as secondary outcomes. After the proofreading tasks, participants encounter a second dose of the Stressful/Neutral treatment and make several choices between lotteries. A summary of their choices is pre-registered as our primary outcome. On each task, as well as pooling information across tasks, we find no effect of app access on decision making. This null result is informative because it contrasts with a meaningful variation that we measure with the same outcomes: participants are more information avoidant and less willing to risk salient losses when they are incentivized to meditate immediately prior to the decision.

5.2.1 Primary Outcome: Certainty Premium

We focus on whether mindfulness affects the interplay of emotion and decision making. Our experiment adapts a task from Callen et al. (2014), who find that study participants in Afghanistan make different lottery choices after recalling violent experiences. Our first Stressful treatment resembles this induction, asking participants to dwell on an unresolved source of stress. However, engaging in the proofreading task may diminish this stress. We therefore assign a "second dose" of the Stressful / Neutral treatment, asking participants to describe in detail how they would handle an unexpected medical bill. In the Neutral condition this is a bill for \$100, while in the Stressful condition it is for \$8,900. We designed this based on the treatment in Mani et al. (2013), where considering a large financial cost reduced cognitive function for poorer participants. As before, answers deemed thoughtful by an independent reader are eligible for a bonus.

Participants then make two choices over menus of binary lotteries. The two menus differ in whether the prospects are risky or certain, and by comparing choices between them we can determine whether participants have a preference for certain outcomes over uncertain ones. The first menu elicits the probability P_{certain} such that a participant is indifferent between receiving \$10 for sure or playing a lottery that pays \$30 with P_{certain} and \$0 otherwise. The second menu elicits $P_{\text{uncertain}}$ such that the participant is indifferent between a lottery that pays \$30 with $P_{\text{uncertain}}$ and \$0 otherwise, versus a 50-50 lottery with \$30 and \$10 payouts.[31] Assuming expected utility and that the utility values of \$30 and \$0 are constant across choices, we can derive the implied utility of \$10 in both

[31] The questionnaire prevents dominated choices and enforces monotonic preferences. For example, in the first menu we prevent participants from choosing the \$10 payout over a degenerate lottery that always pays \$30, and we require that if they select the \$30-\$0 lottery at probability P, they also do at $P' > P$. 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.

cases, $u_{\text{certain}}(10)$ and $u_{\text{uncertain}}(10)$. The certainty premium is then defined as $CP = u_{\text{certain}}(10) - u_{\text{uncertain}}(10)$, the implied difference in utility from receiving \$10 as a sure payment versus as a risky prospect.

The certainty premium is zero under expected utility, but 70 percent of respondents in our sample have positive certainty premia, implying that they prefer to get \$10 from a guarantee rather than a gamble. This is consistent with the sample in Callen et al. (2014), where the certainty premium also tends to be positive. The Stressful treatment decreases the certainty premium (Table 5 column 4), primarily by making participants more risk averse when choosing between two risky prospects.[32]

Our focus, however, is not to document how risk-taking reacts to emotions in general, but specifically whether mindfulness mitigates such interference. Our main finding is that access to the app has no large effect on the certainty premium, both overall and in either of the Neutral or Stressful conditions (Table 5, columns 3 and 4). While app access tends to undo the effect of the Stressful task, the effect is indistinguishable from zero.

5.2.2 Secondary Outcomes

Next we broaden our analysis to include information avoidance and the reluctance to risk a salient, low probability loss. Individuals have documented tendencies to avoid useful but unpleasant information (Golman et al., 2017), presumably because they prefer to avoid facing negative emotions from anticipating and receiving the information (Bénabou and Tirole, 2016). In some cases, people also behave as if they evaluate a risky prospect based on how it makes them feel rather than on the associated payoffs and probabilities (Loewenstein et al., 2001). Because mindfulness meditation is hypothesized to regulate emotions and reduce reactivity to them, we investigate its effects on both of these behaviors. Though we discuss these secondary outcomes after the certainty premium, we note that participants engaged in the corresponding tasks at the start of the economic tasks survey, before any interference with the stressful/neutral randomization (see Figure 1b).

To measure the effects on information avoidance, we offer participants optional informational links, which are ultimately delivered at the end of the survey. They can choose to receive up to four links: (i) a life expectancy calculator; (ii) risk factors for developing dementia; (iii) the risk of one's job of being replaced by automation; and (iv) a calculator of financial risk in retirement. We use the proportion of links the participant refuses as our measure of information avoidance.

We also offer participants a small-stakes bet with excellent expected value but a salient low probability loss. The bet earns \$1 with 99 percent probability but loses \$10 otherwise. This decision mirrors documented real-life choices to purchase high-premium insurance against low probability

[32] This differs from the finding in Callen et al. (2014), where recalling violent memories *increases* the certainty premium. These opposing signs may reflect differing risk preferences in our study sample and theirs, or the stakes in the lotteries. Callen et al. (2014) study a population with an average of 10 years' formal education. Most of our participants have completed at least a college degree, which may translate to different behavior on lab-style lottery tasks. In addition, while we view our lottery outcomes, ranging from \$0 to \$30, as reasonable, they are lower than the stakes in Callen et al. (2014)'s experiment where participants stand to gain between one and three days' wages. The relationship between risk and emotion may reasonably depend on stakes, which could also explain the differing signs on our effects.

moderate risks (Sydnor, 2010).

Overall, app access has nearly no effect on these choices (Table 5, columns 1 and 2). Waitlist participants decline the information we offer 45 percent of the time and decline the bet 32 percent of the time. The corresponding figures for the App Access group are 43 percent and 33 percent, respectively; neither differs meaningfully from the waitlist. While practicing mindfulness may take longer than two weeks to affect choices, or have effects on decisions we do not test, it does not fundamentally alter how our participants approach risk and information in our setting.

One explanation for this result would be that our outcomes are poor measures of attention and preferences. We can rule this out by turning to our last treatment, which does affect productivity and decision making. The null effects from app access stand in sharp contrast with the immediate effects of a meditation session, which we turn to now.

5.3 Immediate Effects of Meditation Sessions

Incentivizing pre-survey meditation in the App Access group allows us to compare the immediate effects of a mindfulness session to the longer-term effects of being a regular practitioner. Habitual actions, like exercising and studying, often have distinct short-term and long-term effects. Distinguishing these effects helps to inform practitioners about the advisability of meditating as a short-term decision aid, and researchers about whether experimentally inducing a meditation session is a valid approach to studying the effects of a sustained mindfulness practice.

We randomize half of participants in the App Access group to receive a bonus if they meditate before beginning the economic tasks module. Participants are told they will receive an additional $3 if they complete an app session of at least 10 minutes at the start of the survey. Compliance is strong: 60 percent of incentivized participants meditate on the app at this time, compared to 3 percent of unincentivized participants.

One concern is that this treatment causes other differences in the survey-taking experience, such as participants postponing the survey to a quieter time when they can meditate. We wrote our invitation to this survey module to limit such effects, instructing all participants to plan an hour of time (longer than actually necessary) to complete the module 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.

We estimate effects of immediate meditation by comparing participants in the App Access group who received the incentive to those who did not. Comparisons to the Waitlist group would conflate the effects of regular practice with the immediate effects of meditation, which is not our goal. We now repeat the analyses of the productivity and decision making sections, estimating a regression of the form:

$$Y_i^{post} = \beta_1 \text{Immediate Meditation}_i + \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i, \tag{6}$$

where Immediate Meditation$_i$ is an indicator for receiving an incentive to meditate immediately before the survey. The omitted group is the set of App Access participants who did not receive the incentive, so that β_1 can be interpreted as an intent-to-treat effect of encouraging a meditation session among participants who have practiced mindfulness.

For the proofreading and certainty premium tasks, where participants encountered Stressful/Neutral treatments, we estimate a regression of the form:

$$\begin{aligned} Y_i^{post} =& \beta_1 \text{Immediate Meditation}_i + \beta_2 \text{Stressful Task}_i \\ &+ \beta_3 \text{Stressful Task}_i \times \text{Immediate Meditation}_i + \delta_{\text{stratum}} + \gamma Y_i^{pre} + \epsilon_i. \end{aligned} \tag{7}$$

Here, β_1 and β_3 are the coefficients of interest, as estimates for the effect of encouraging a meditation session for the subsets of participants who subsequently encounter neutral and stressful memories, respectively.

We observe moderate evidence that an immediate meditation session reduces performance on productivity tasks. Table 6 presents the results. Encouraging a meditation session reduces performance on the proofreading task by 1 percent overall ($p = 0.084$). The effect is largest in the Stressful condition, where the Immediate Meditation group earns 2 percentage points less (-2.3%). It is unlikely that this reflects reduced marginal value of effort on the task due to the $3 bonus or time spent meditating. The immediate meditation treatment does not affect time taken, particularly in the Stressful condition where performance falls the most (Table B.10, column 6). Effects are smaller and indistinguishable from zero for proofreading in the Neutral condition, and on the Stroop task. While these results do not give definitive proof that meditation sessions degrade short-term productivity, they are inconsistent with the view that meditating boosts performance on such tasks in the short run, and they suggest that meditation sessions may increase immediate (but not long-run) sensitivity to stressful thoughts.

Meditation sessions also affect decision making, increasing the tendency to avoid unpleasant information and salient losses. Table 7 presents the results. Participants in the Immediate Meditation arm are 3.8 percentage points (8.8%) more likely to avoid potentially distressing information ($p = 0.042$), and 5.1 percentage points (15.5%) more likely to decline a lottery with high expected value but a salient low-probability loss ($p = 0.053$). These effects are unlikely to reflect changes in the marginal value of time, as we deliver information via links that participants can follow at their leisure. There are no clear effects on the certainty premium in either the Neutral or Stressful conditions. Overall, these results suggest that a meditation session can change behavior on some tasks that call up negative emotions.

A classical reading of these results is that meditating induces mistakes in the form of reduced productivity, disinterest in information, and aversion to take a small-stakes high-value lottery. However, another reading would be that meditation directs attention inward, and that participants are simply reacting in light of what they attend to. It is not so strange for choices to respond to mental states, such as when doctors avoid prolonging work past the end of a shift (Chan, 2018). Even information avoidance can stem from awareness of emotional costs, such as when investors postpone checking a stock portfolio or patients delay medical tests (Karlsson et al., 2009; Oster et al., 2013;

Bolte and Raymond, 2023). In short, these behaviors can still reflect preferences at the individual level. Practically, our results suggest that meditating during work or before decisions is more a matter of the mental state individuals wish to experience, than attaining some dramatic improvement in their performance.

6 Discussion

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

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

Usage of the app declines over time, and incentives have nearly no long-term effect on meditation behavior. This may reflect a combination of our participants' high baseline interest making the incentives unnecessary, long-term effects of mindfulness rendering future sessions less useful, and the difficulties of adopting a new habit. Still, incentives may prove effective for encouraging adoption of mindfulness in a less engaged sample. Future work can study methods of increasing interest in mindfulness in a general population. In addition, the growing popularity of wearable health devices, like exercise trackers and smart watches, presents an opportunity to design mindful nudges and develop a richer understanding of treatment compliance and self-directed meditation.

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

The short-term effects of meditation are less statistically precise than long-term effects on mental health and productivity. Still, they are inconsistent with the hypothesis that a meditation session improves performance in a classical economic sense. This suggests maintaining skepticism for the view that meditating is a surefire shortcut to “making better decisions”. It also serves as a caution to future research in this area, which should distinguish between whether it studies the long-run effects of practicing mindfulness, and short-run effects that might arise in one-shot experiments that randomize meditation before lab tasks.

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

References

Acland, Dan and Matthew R. Levy, "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," *IZA Discussion Paper*, 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, Mo and Stephen Wu, "Income, Psychological Well-Being, and the Dynamics of Poverty," *Economic Development and Cultural Change*, 2023, p. 76.

American Society of Clinical Psychology, "Mindfulness-Based Cognitive Therapy | Society of Clinical Psychology," https://div12.org/treatment/mindfulness-based-cognitive-therapy/ April 2019.

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.

Bénabou, Roland and Jean Tirole, "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

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 Speca, 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," *Working Paper*, October 2023.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, "Memory, attention, and choice," *The Quarterly journal of economics*, 2020, *135* (3), 1399–1442.

Braghieri, Luca, Roee Levy, and Alexey Makarin, "Social Media and Mental Health," SSRN Scholarly Paper ID 3919760, Social Science Research Network, Rochester, NY August 2021.

Brown, Christina L, Supreet Kaur, Geeta Kingdon, and Heather Schofield, "Cognitive endurance as human capital," Technical Report, National Bureau of Economic Research 2022.

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," *Working Paper*, 2023, p. 75.

Chan, David C, "The efficiency of slacking off: Evidence from the emergency department," *Econometrica*, 2018, *86* (3), 997–1030.

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.

DellaVigna, Stefano and Devin Pope, "What Motivates Effort? Evidence and Expert Forecasts," *The Review of Economic Studies*, April 2018, *85* (2), 1029–1069.

Duquennois, 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, "Behavioral Inattention," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 2, Elsevier, 2019, pp. 261–343.

Galante, Julieta, Géraldine Dufour, Maris Vainre, Adam P Wagner, Jan Stochl, Alice Benton, Neal Lathia, Emma Howarth, and Peter B Jones, "A Mindfulness-Based Intervention to Increase Resilience to Stress in University Students (the Mindful Student Study): A Pragmatic Randomised Controlled Trial," *The Lancet Public Health*, February 2018, *3* (2), e72–e81.

Ghanem, Dalia, Sarojini Hirshleifer, and Karen Ortiz-Beccera, "Testing attrition bias in field experiments," *Journal of Human Resources*, 2023.

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.

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.

Karlsson, Niklas, George Loewenstein, and Duane Seppi, "The ostrich effect: Selective attention to information," *Journal of Risk and uncertainty*, 2009, *38*, 95–115.

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.

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.

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.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao, "Poverty Impedes Cognitive Function," *Science*, August 2013, *341* (6149), 976–980.

Oster, Emily, Ira Shoulson, and E. Ray Dorsey, "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease," *American Economic Review*, April 2013, *103* (2), 804–830.

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.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel, "Poverty, Depression, and Anxiety: Causal Evidence and Mechanisms," *Science*, December 2020, *370* (6522).

Romano, Joseph P and Michael Wolf, "Efficient Computation of Adjusted P-Values for Resampling-Based Stepdown Multiple Testing," *Statistics & Probability Letters*, 2016, p. 3.

Santomauro, Damian F, Ana M Mantilla Herrera, Jamileh Shadid, Peng Zheng, Charlie Ashbaugh, David M Pigott, Cristiana Abbafati, Christopher Adolph, Joanne O Amlag, Aleksandr Y Aravkin, Bree L Bang-Jensen, Gregory J Bertolacci, Sabina S Bloom, Rachel Castellano, Emma Castro, Suman Chakrabarti, Jhilik Chattopadhyay, Rebecca M Cogen, James K Collins, Xiaochen Dai, William James Dangel, Carolyn Dapper, Amanda Deen, Megan Erickson, Samuel B Ewald, Abraham D Flaxman, Joseph Jon Frostad, Nancy Fullman, John R Giles, Ababi Zergaw Giref, Gaorui Guo, Jiawei He, Monika Helak, Erin N Hulland, Bulat Idrisov, Akiaja Lindstrom, Emily Linebarger, Paulo A Lotufo, Rafael Lozano, Beatrice Magistro, Deborah Carvalho Malta, Johan C Månsson, Fatima Marinho, Ali H Mokdad, Lorenzo Monasta, Paulami Naik, Shuhei Nomura, James Kevin O'Halloran, Samuel M Ostroff, Maja Pasovic, Louise Penberthy, Robert C Reiner Jr, Grace Reinke, Antonio Luiz P Ribeiro, Aleksei Sholokhov, Reed J D Sorensen, Elena Varavikova, Anh Truc Vo, Rebecca Walcott, Stefanie Watson, Charles Shey Wiysonge, Bethany Zigler, Simon I Hay, Theo Vos, Christopher J L Murray, Harvey A Whiteford, and Alize J Ferrari, "Global Prevalence and Burden of Depressive and Anxiety Disorders in 204 Countries and Territories in 2020 Due to the COVID-19 Pandemic," *The Lancet*, October 2021, p. S0140673621021437.

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.

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.

Vos, Theo, Stephen S Lim, Cristiana Abbafati, Kaja M Abbas, Mohammad Abbasi, Mitra Abbasifard, Mohsen Abbasi-Kangevari, Hedayat Abbastabar, Foad Abd-Allah, Ahmed Abdelalim et al., "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019," *The lancet*, 2020, *396* (10258), 1204–1222.

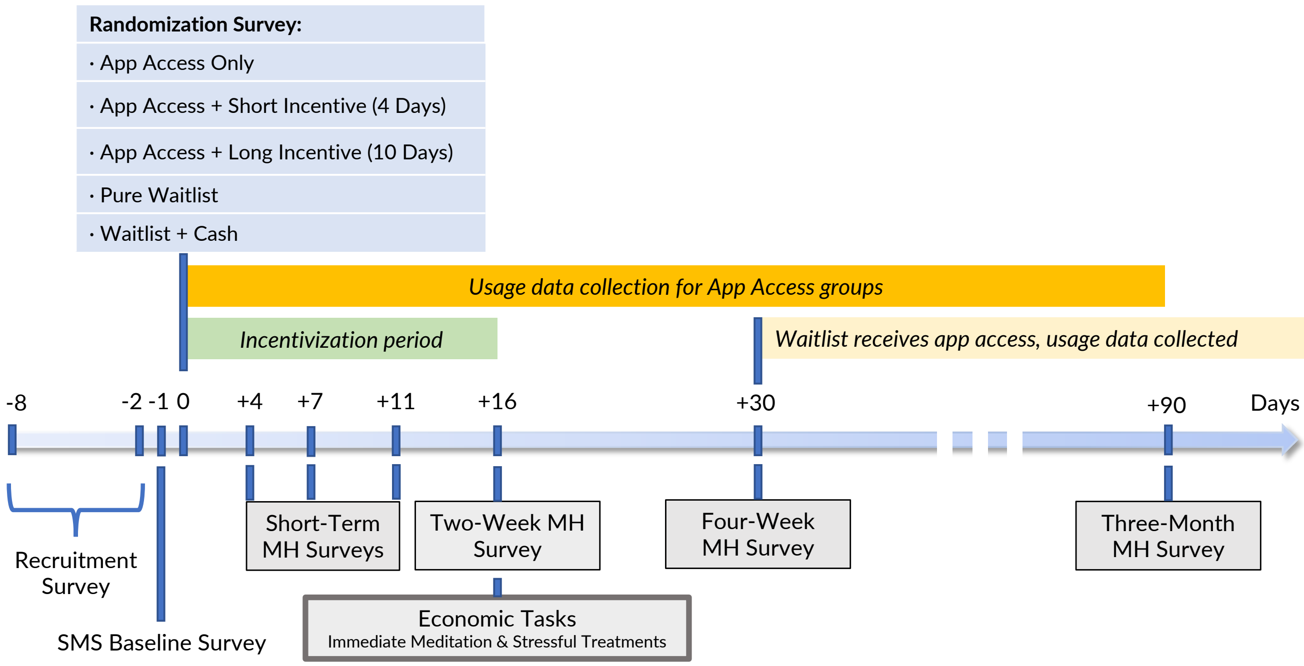
Whitfield, Tim, Thorsten Barnhofer, Rebecca Acabchuk, Avi Cohen, Michael Lee, Marco Schlosser, Eider M Arenaza-Urquijo, Adriana Böttcher, Willoughby Britton, Nina Coll-Padros et al., "The effect of mindfulness-based programs on cognitive function in adults: A systematic review and meta-analysis," *Neuropsychology Review*, 2022, pp. 1–26.

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, Mitsuhiko 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.

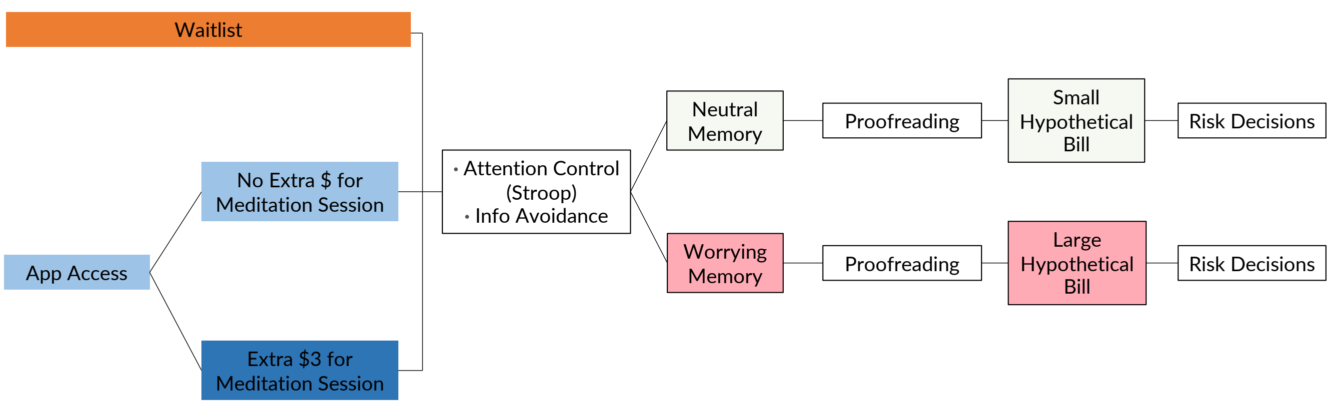
7 Figures and Tables

Figure 1: Experiment Overview

(a) Timeline



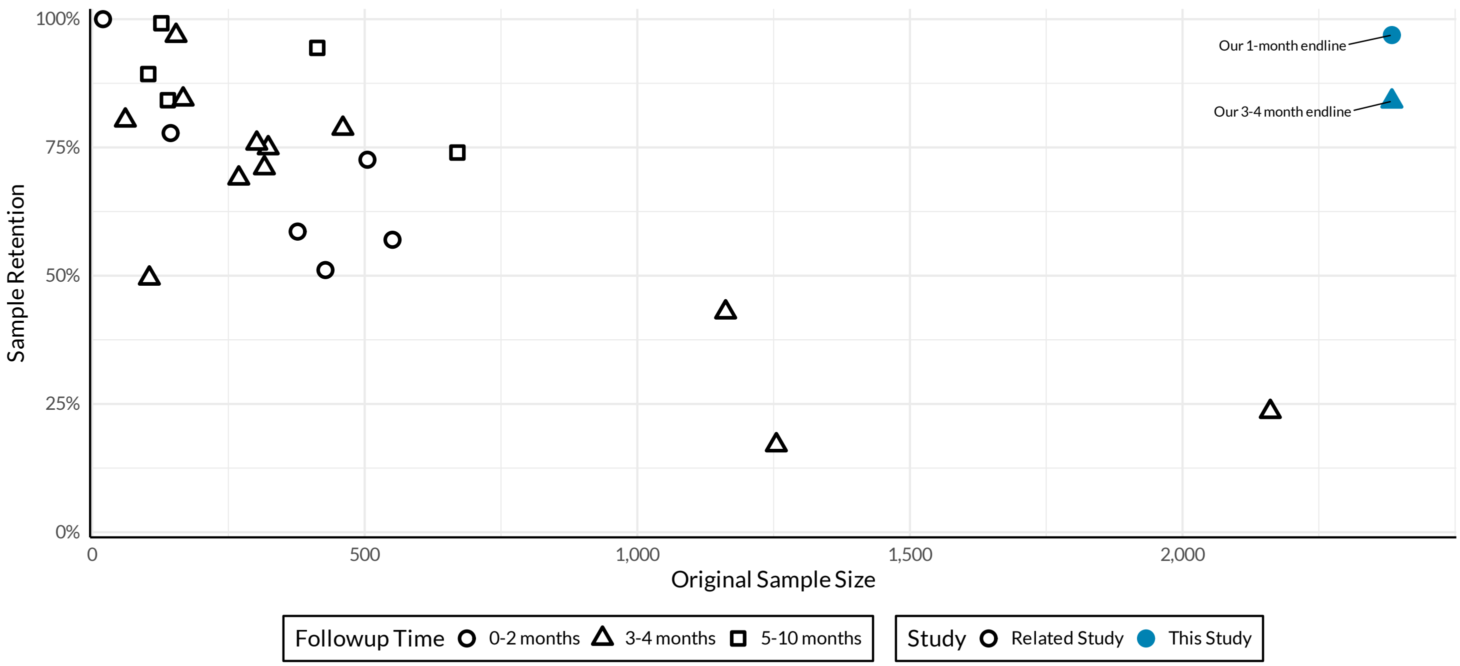
(b) Economic Tasks



Notes: This figure depicts the structure of our experiment. Panel A presents the timeline of the entire experiment. We recruited participants in July and August 2021 and conduct three baseline surveys. Randomization occurs weekly in the last baseline across one of five arms: (i) App Access Only, immediately receiving a 90-day license for the Headspace app; (ii + iii) App Access + Short (Long) Incentives, additionally being offered a $10 bonus for using the app on at least 4 (at least 10) 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.

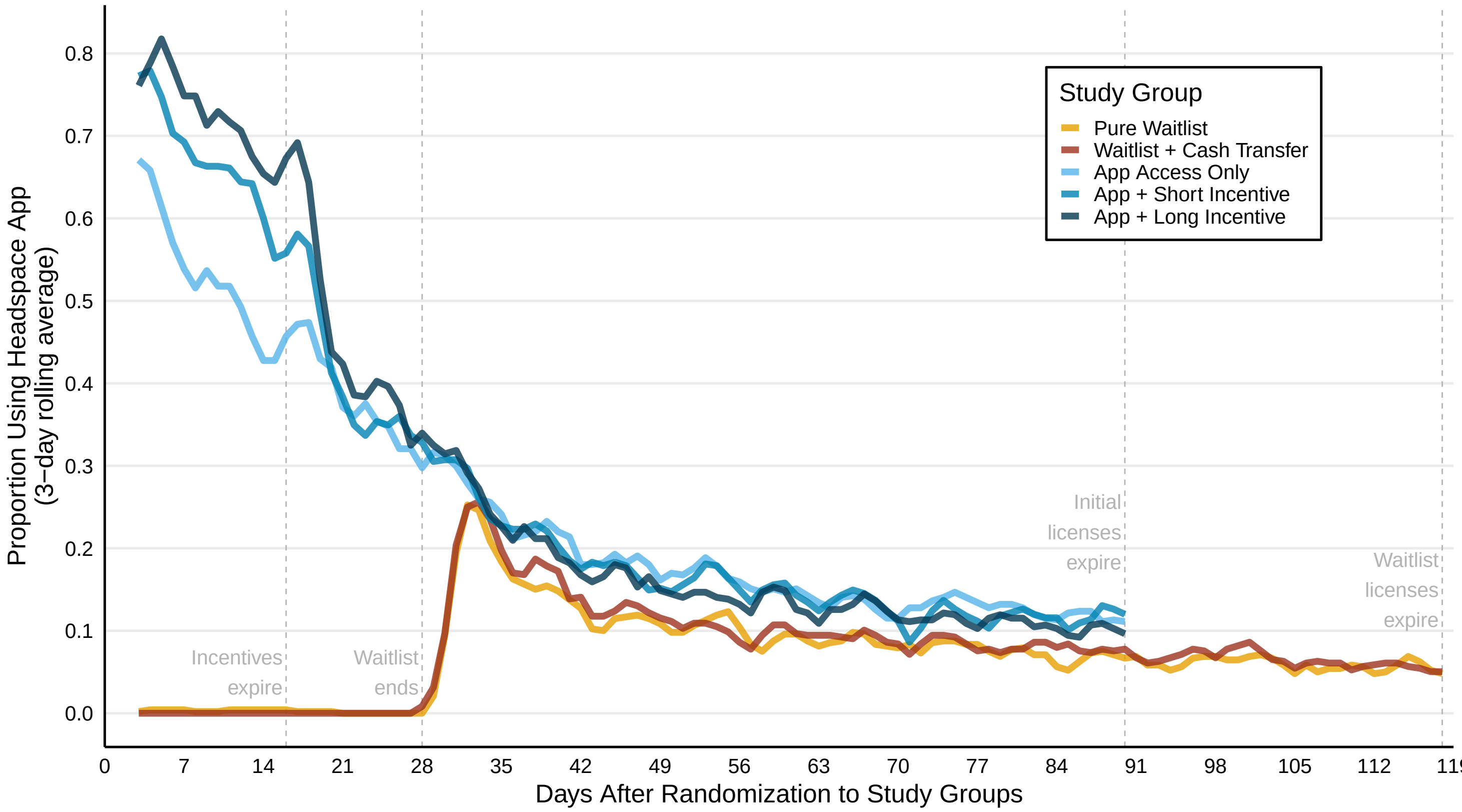
Panel B summarizes the economic tasks completed at two weeks. Those with app access are randomly assigned to receive incentives to meditate using the app right before continuing into the survey. All participants then complete tasks that have built-in distracting or emotion-inducing elements: a Stroop 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: Recent Studies of Mindfulness Interventions



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

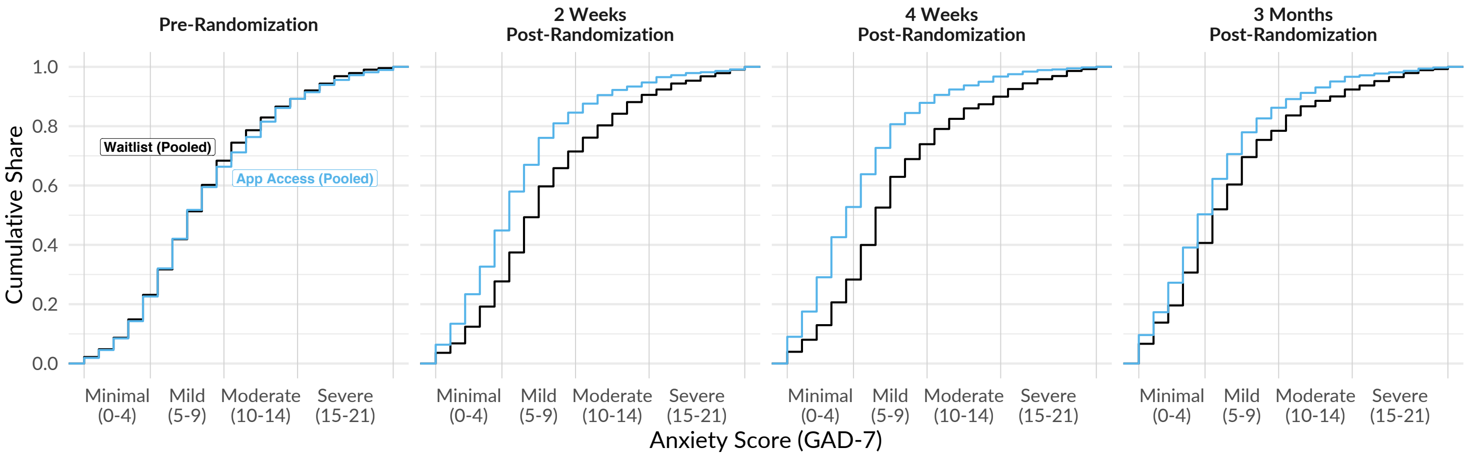
Figure 3: Proportion of Participants Using Headspace in a 3-Days Rolling Window In Each Intervention Arm



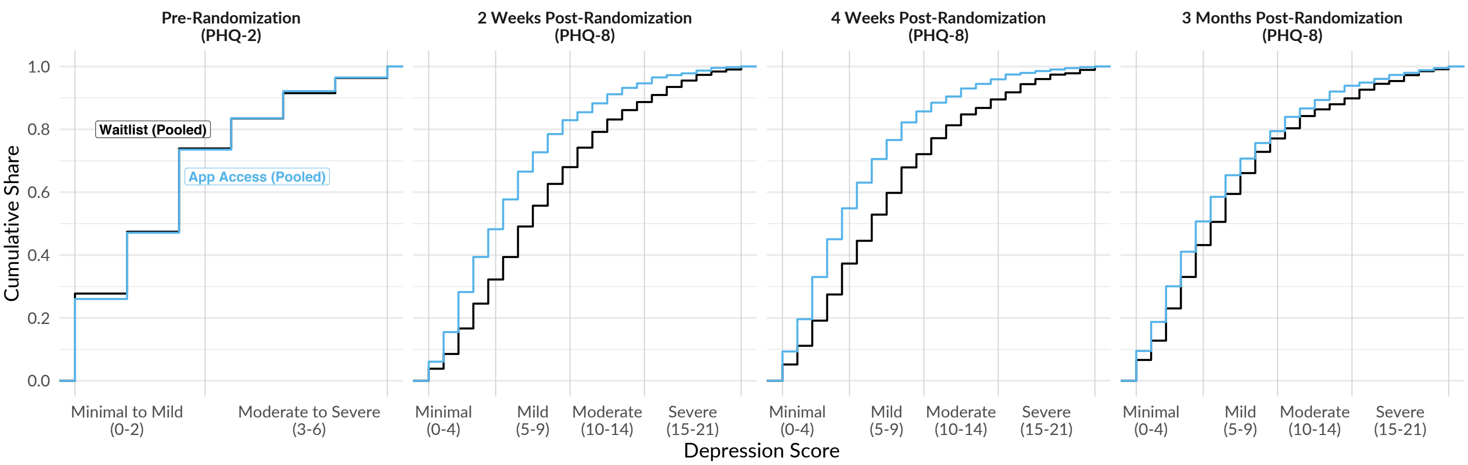
Notes: This figure shows the proportion, over time, of participants who recorded at least one session on the Headspace app within the last 3 days. Participants are randomized within one of five arms: (i) App Access Only, receiving free access to the Headspace app; (ii) App + Short Incentive, additionally 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 4: Long-Term Distribution of Mental Health Scores

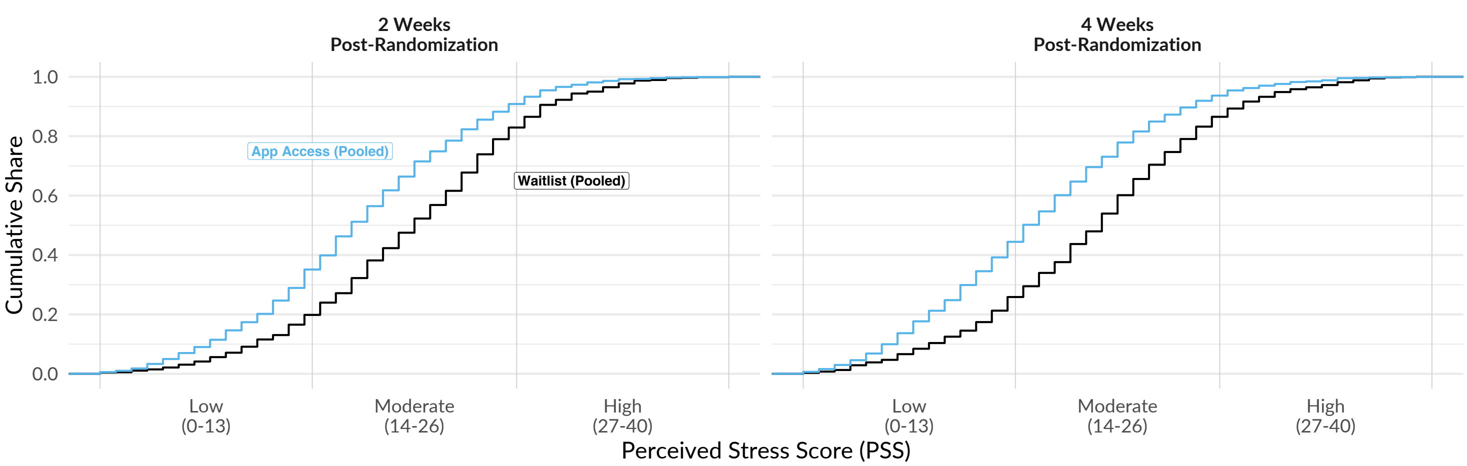
(a) Anxiety



(b) Depression



(c) Stress



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

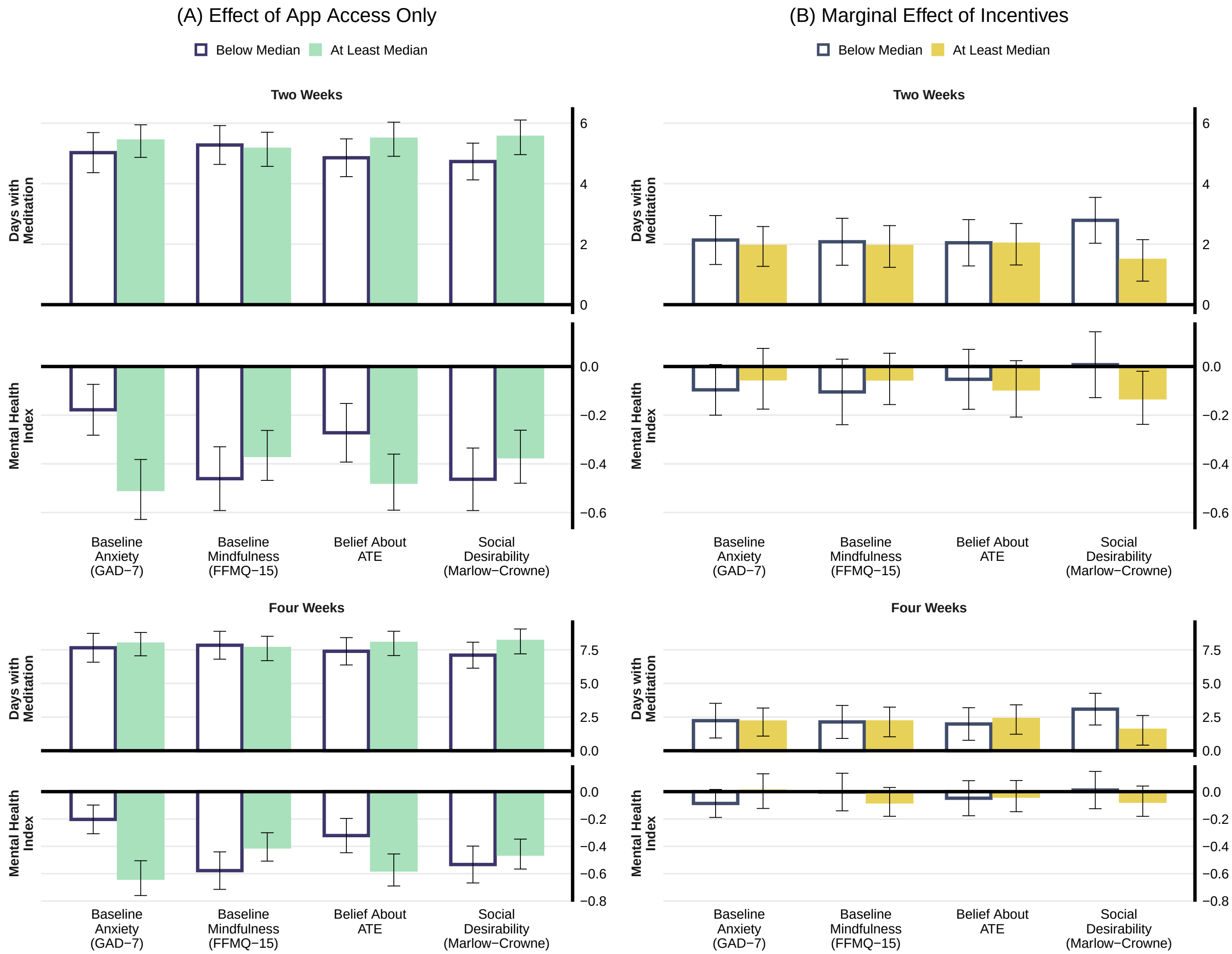


Figure 5: Heterogeneous Effects of License Provision and Incentives

Notes: This figure presents total effect of 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. The estimating equation is Equation 3. Panel A presents estimates and 95% confidence intervals for the coefficient on App Access, and Panel B presents the same for Any Incentive. Social Desirability refers to a 13-item version of the Marlow-Crowne Social Desirability Scale, which measures the tendency to portray oneself favorably in a survey. Days with Meditation counts the number of distinct days a users completed a meditation session on the Headspace app. Table B.7 presents formal hypothesis tests for equality of effects between the below and above median groups.

Table 1: Sample Characteristics and Balance

	(1)	(2)	(3)	(4)	(5)
		Study Sample			
	US Adults	Waitlist Group		App Access Group	
	Mean	Mean	(St. Dev.)	Diff.	(St. Err)
Age Group					
18-29	0.213	0.202	(0.402)	−0.010	(0.011)
30-39	0.172	0.204	(0.403)	0.014	(0.016)
40-49	0.158	0.250	(0.433)	−0.010	(0.017)
50-59	0.163	0.209	(0.407)	−0.016	(0.016)
60-69	0.150	0.109	(0.312)	0.016	(0.013)
70+	0.143	0.025	(0.157)	0.006	(0.007)
Female	0.492	0.848	(0.359)	0.019	(0.015)
Education					
No Bachelor's degree	0.640	0.166	(0.373)	−0.023	(0.015)
Bachelor's degree	0.225	0.345	(0.475)	0.020	(0.020)
Graduate or professional degree	0.135	0.441	(0.497)	0.003	(0.020)
Household Size	2.261	2.874	(1.328)	−0.087	(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	(0.008)
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)	−0.001	(0.011)
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					
Anxiety Score (GAD-7)		8.012	(4.505)	0.168	(0.123)
Depression Score (PHQ-2)		1.796	(1.645)	0.018	(0.061)
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. In Figures B.2, B.3, and B.4 we adjust for potential covariate imbalance using debiased machine learning.

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

	Time from Randomization				
	Days 1-16	Days 17-28	Days 1-28	Day 29+	Cumulative
	(1)	(2)	(3)	(4)	(5)
A. Any Meditation Session					
App Access	0.805***	0.580***	0.835***	0.031	0.408***
(S.E.)	(0.018)	(0.023)	(0.017)	(0.028)	(0.023)
Short Incentive	0.083***	0.066**	0.060***	0.059*	0.056***
(S.E.)	(0.023)	(0.031)	(0.022)	(0.032)	(0.021)
Long Incentive	0.094***	0.071**	0.067***	0.015	0.061***
(S.E.)	(0.023)	(0.031)	(0.021)	(0.032)	(0.021)
Waitlist Mean	0.002	0.004	0.005	0.437	0.439
N	2384	2384	2384	2384	2384
B. Days with Any Meditation					
App Access	5.291***	2.574***	7.821***	1.720**	9.526***
(S.E.)	(0.215)	(0.157)	(0.344)	(0.721)	(0.955)
Short Incentive	1.319***	0.023	1.333***	−0.474	0.863
(S.E.)	(0.291)	(0.214)	(0.464)	(0.843)	(1.193)
Long Incentive	2.728***	0.289	3.004***	−1.036	1.963*
(S.E.)	(0.312)	(0.218)	(0.482)	(0.825)	(1.175)
Waitlist Mean	0.015	0.006	0.021	4.497	4.518
N	2384	2384	2384	2384	2384
C. Meditation Sessions					
App Access	11.362***	5.147***	16.508***	3.859**	20.367***
(S.E.)	(0.628)	(0.419)	(0.985)	(1.765)	(2.440)
Short Incentive	2.761***	−0.081	2.680*	−1.451	1.228
(S.E.)	(0.888)	(0.598)	(1.378)	(2.078)	(3.147)
Long Incentive	5.501***	−0.225	5.276***	−3.361*	1.915
(S.E.)	(0.932)	(0.547)	(1.380)	(1.924)	(2.978)
Waitlist Mean	0.016	0.007	0.023	8.232	8.255
N	2384	2384	2384	2384	2384

Notes: This table presents the average treatment effects of app access and usage incentives on app usage during various time windows. We calculate usage based on administrative data associated with each participant's unique voucher code. Panel A presents effects on an indicator for completing any meditation session in the specified period. Panel B describes the effect on the number of days with a meditation session, and Panel C does the same for the total number of meditation sessions. We randomize some participants to receive incentives for app usage in the first 2 weeks of their voucher period (column 1), and the waitlist remains in effect for 2 more weeks after incentives expire (column 2). Column 3 describes the entire waitlist period. After the waitlist ends, all participants have full app access (column 4). Our administrative data ends 90 days after participants activate their vouchers, and we report cumulative usage in this period in column 5. The estimating equation is Equation 1, which includes stratum fixed effects. The reference group combines the Pure Waitlist and Waitlist Cash Transfer arms. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Time from Randomization					
	Four Days	Seven Days	Eleven Days	Two Weeks	Four Weeks	Three-Four Months
App Access	−0.089**	−0.157***	−0.226***	−0.381***	−0.456***	−0.198***
(S.E.)	(0.040)	(0.042)	(0.042)	(0.036)	(0.039)	(0.046)
Any Usage Incentive	−0.063	−0.037	−0.069	−0.075**	−0.046	−0.027
(S.E.)	(0.040)	(0.043)	(0.043)	(0.037)	(0.039)	(0.045)
Waitlist Mean	0.017	0.017	0.034	0.035	0.056	0.046
N	2305	2145	2191	2330	2311	2004
Index Components:						
Anxiety	GAD-2	GAD-2	GAD-2	GAD-7	GAD-7	GAD-7
Depression				PHQ-8	PHQ-8	PHQ-8
Stress				PSS-10	PSS-10	

Notes: This table presents average treatment effects of app access and usage incentives on reported symptoms of metal distress over time. We measure symptoms of anxiety using the two- and seven-item Generalized Anxiety Disorder scales (GAD-2 and GAD-7, respectively); symptoms of depression using the eight-item Patient Health Questionnaire (PHQ-8); and stress using the ten-item Perceived Stress Scale (PSS-10). The outcome at each timepoint is a standardized index that combines the mental health scales measured at that time. We first standardize each scale in each time period by subtracting the Pure Waitlist mean and dividing by the Pure Waitlist standard deviation. The index is the average of these standardized scales. Lower scores indicate lower reported levels of distress. The estimating equation is Equation 2, which includes stratum fixed effects and the baseline mental health index. The reference group combines the Pure Waitlist and Waitlist Cash Transfer arms. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. Table B.6 presents a version of this table that separates the two waitlist and incentive groups, and Figure B.2 presents a version of this analysis that flexibly adjusts for baseline covariates using debiased machine learning. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effect of App Access on Earnings in Productivity Tasks

	(1)	(2)	(3)
	Before Stressful/Neutral Task	**After Stressful/Neutral Task**	
	Stroop Earnings	**Proofreading Earnings**	**Proofreading Earnings**
App Access	0.311	1.785***	
(S.E.)	(0.512)	(0.600)	
App × Neutral			1.854**
(S.E.)			(0.863)
App × Stressful			1.718**
(S.E.)			(0.840)
Stressful Memory			−0.297
(S.E.)			(0.856)
Waitlist Mean	74.120	87.928	88.362
N	1592	1592	1592

Notes: This table presents average treatment effects of app access on earnings on two incentivized tasks: a Stroop test (column 1) and a proofreading task (columns 2 and 3). We report earnings relative to the maximum possible amount in percentage points. For example, the average waitlist participant earned 74.12% of the maximum on the Stroop task. Participants are randomized to recall either a stressful or neutral memory after the Stroop task, but before the proofreading task. Column 2 reports effects on proofreading performance combining both groups, and column 3 reports on them separately. The estimating equations are Equations 4 (columns 1 and 2) and Equation 5 (column 3). All regressions include stratum fixed effects and control for performance on these tasks in the baseline survey. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. Figure B.3 presents a version of this analysis that flexibly adjusts for baseline covariates using debiased machine learning. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Impacts of App Access and Stressful Treatment on Decision Making

	(1)	(2)	(3)	(4)
	Before Stressful/Neutral Task		**After Stressful/Neutral Task**	
	Avoid Info	**Avoid Salient Loss**	**Certainty Premium**	**Certainty Premium**
App Access	−0.021	0.010	−0.003	
(S.E.)	(0.017)	(0.024)	(0.015)	
App × Neutral				−0.019
(S.E.)				(0.021)
App × Stressful				0.014
(S.E.)				(0.022)
Stressful Task				−0.038**
(S.E.)				(0.020)
Waitlist Mean	0.454	0.319	0.143	0.163
N	1592	1592	1585	1585

Notes: This table presents average treatment effects on choices made in several decision making tasks: an index for avoiding potentially unpleasant useful information (column 1), the decision to reject a small-stakes lottery with a salient loss (column 2), and certainty premium of Callen et al. (2014) (column 3). After the first two choices, but before the certainty premium task, participants are randomized between a more and a less stressful task. The estimating equations are Equation 4 (columns 1-3) and Equation 5 (column 4). All regressions include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. Figure B.4 presents a version of this analysis that flexibly adjusts for baseline covariates using debiased machine learning. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Effects of Immediate Meditation on Attention

	(1)	(2)	(3)
	Before Stressful/Neutral Task	**After Stressful/Neutral Task**	
	Stroop Earnings	**Proofreading Earnings**	**Proofreading Earnings**
Immediate Meditation	−0.475	−1.081*	
(S.E.)	(0.553)	(0.625)	
Immediate × Neutral			−0.131
(S.E.)			(0.830)
Immediate × Stressful			−2.053**
(S.E.)			(0.933)
Stressful Memory			−0.435
(S.E.)			(0.851)
App Access Mean	74.229	89.691	89.890
N	1331	1332	1332

Notes: This table presents the intent to treat effect of encouraging a meditation session shortly before the incentivized attention tasks. Half of the App Access group received a cash incentive to meditate at the beginning of the survey. We compare their performance to App Access participants who did not receive this incentive. All of these participants are randomized to recall either a stressful or neutral memory after the Stroop task, but before the proofreading task. Column 2 reports effects on proofreading performance combining both groups, and column 3 reports on them separately. The estimating equations are Equations 6 (columns 1 and 2) and Equation 7 (column 3). All regressions include stratum fixed effects and performance on a baseline version of the given task. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Effects of Immediate Meditation on Decision Making

	(1)	(2)	(3)	(4)
	Before Stressful/Neutral Task		**After Stressful/Neutral Task**	
	Avoid Info	**Avoid Salient Loss**	**Certainty Premium**	**Certainty Premium**
Immediate Meditation	0.038**	0.051*	0.002	
(S.E.)	(0.019)	(0.026)	(0.017)	
Immediate × Neutral				−0.024
(S.E.)				(0.023)
Immediate × Stressful				0.029
(S.E.)				(0.024)
Stressful Task				−0.006
(S.E.)				(0.024)
App Access Mean	0.433	0.328	0.141	0.143
N	1332	1332	1327	1327

Notes: This table presents the intent to treat effect of encouraging a meditation session shortly before the decision making tasks. Half of the App Access group received a cash incentive to meditate at the beginning of the survey. We compare their performance to App Access participants who did not receive this incentive. The estimating equations are Equation 6 (columns 1-3) and Equation 7 (column 4). All regressions include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Supplementary Materials

Table of Contents

A Supplementary Materials

A.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.15 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.

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

In the main text, we discuss the effects of mindfulness on information avoidance. This section offers a supplementary test of mechanisms through which mindfulness may affect preferences over information.

In principle, mindfulness may both (i) change the utility of experienceing emotions by training individuals to react less to them, and (ii) reduce the utility of future payoffs by encouraging one to focus on the present. To help isolate effects from (ii), we design a short, incentivized game where information has instrumental value but is emotionally neutral. This contrasts with the information avoidance task, where information has the potential to trigger negative emotions.

In the emotionally neutral task, participants begin with a digital wallet that randomly contains either \$0 or \$1 with equal probability. They can then acquire 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 \$1 to their bonus. But, two or more tails will subtract the same amount.

Before choosing to purchase the stock, participants receive up to three pieces of information. The first is the outcome of one of the coin tosses, which is relevant to the decision and emotionally neutral. The second and third pieces of information are irrelevant: the starting value of their digital wallet or the age of the oldest tree in the world.[33]

We ask participants to select one piece of information they would especially like to know. By default we present all three, but with a small probability present only the piece they choose. This makes information acquisition costly, so that participants should select the piece of information they would most prefer to know. If participants do not prefer to learn the outcome of the first coin toss, this indicates that they derive less utility from future payoffs than from immediate payoffs (warm glow from learning the initial value of their digital wallet, intrigue from learning a new fact).

We present the effects of app access and immediate meditation on information choice in column 5 of Table B.13. Twenty-nine percent of waitlist participants choose to avoid the coin flip. App access has a negligible positive effect relative to the waitlist, and an immediate meditation session has a negligible positive effect relative to app access. These results are consistent with mindfulness and meditation sessions having little effect on the relative utiltiy of future versus present payoffs. They are more consistent with mindfulness affecting the role of emotions in utility.

This test is far from conclusive. Future work could directly and repeatedly test how practitioners of mindfulness respond to present versus future payoffs, as well as to a variety of emotion-inducing stimuli.

[33] At the time of the survey, the oldest known tree was 5,071 years old.

A.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. Randomization implies that any relationship between the remaining variation in the outcome and treatment assignment reflects a treatment effect.

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. We choose relaxed elasticnet, rather than just lasso, because we have no basis to believe that the regression coefficients in any specification are approximately sparse, as assumed by lasso.

- 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 License $\times$ AnyIncentive $\times$ $1(Anxiety > Median)$, 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:

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

   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:

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

We present DML-adjusted treatment effects in Appendix Figures B.2, B.3, and B.3. Each figure mirrors the structure of a table in the main text. 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 large randomized experiment, treatment is independent of the covariates leaving little room for omitted variables bias.

A.4 Attrition

We reduce attrition in our study through extensive pre-screening and routine reminders. Still, some participants do not respond to every followup survey. In this section, we investigate whether attrition limits our ability to infer causal effects for participants who respond to the survey. We show that there are small differences in attrition rates between the treatment and control arms, that these differences are unlikely to cause imbalance in the experiment, and present partial identification bounds on treatment effects under a widely used model for selective attrition (Lee, 2009).

Overall attrition at our main two and four week endline surveys remains below 5 percent; it rises to about 15 percent at three months. We present exact sample size and attrition rates in Table B.2. Attrition tends to be 1-3 percentage points higher in the App Access Group than in the Waitlist, perhaps because some Waitlist participants remain in the study to receive license codes in the fourth week. Although these differences are statistically distinguishable from zero — in large part due to this study's sample size — they are small in absolute terms. We expect that the typical experiment would not consider a differential attrition rate of 3 percentage points to be a major concern.

Attrition would be especially concerning if different participants drop out from treatment and control arms, creating imbalance and biasing treatment effect estimates. In such a case, we could not credibly estimate treatment effects for respondents. We assess this possibility by determining whether attrition induces imbalance on baseline outcomes. For example, if App Access participants with poor mental health dropped out at higher rates than comparable Waitlist participants, the treatment would spuriously appear to improve mental health. We adopt the simple regression test for imbalance from Ghanem et al. (2023), estimating the equation

$$Y_i^{pre} = \delta_s \times R_i + \delta_s \times (1 - R_i) + \beta_1 T_i \times R_i + \beta_2 T_i \times (1 - R_i) + \epsilon_i \tag{A.1}$$

where Y_i^{pre} is the baseline outcome, R_i is an indicator for whether a participant responded to an endline survey, T_i is an indicator for treatment, and δ_s are stratum fixed effects. The coefficient β_1 estimates the difference between treated respondents and control respondents, while β_2 measures the difference between treated attritors and control attritors. Jointly testing $\beta_1 = \beta_2 = 0$ tells us whether attrition has created significant imbalance on our baseline outcome. We present estimates in Table A.4.

We find no strong evidence that attrition contaminates our study of effects on mental health or proofreading performance. That is, treatment and control respondents (and treatment and control attritors) have statistically similar baseline mental health and proofreading scores. We expect these null findings are reasonably precise, as the sample is large enough to detect 1.6 percentage point differences in attrition between treatment and control. This suggests the treatment effects in the main text are valid estimates for participants who respond to followup surveys (roughly 95% of the sample at two and four weeks).

Finally, although attrition rates and composition are similar for the study's treatment arms, we conclude by showing that a worst-case model of selective attrition does not significantly alter our main conclusions. We take the popular approach in Lee (2009), which involves manually equalizing the attrition rates in treatment and control by eliminating the most extreme values from the more responsive group. Eliminating extreme responses from the top of the distribution provides one bound, and eliminating from the bottom of the distribution provides the other. It is possible to tighten these bounds by conditioning on discrete covariates, and aggregating the conditional bounds. Under the assumption that propensity to respond is monotonic in treatment given the conditioning variables, the resulting estimates are worst-case bounds on the treatment effect for the "always responder" population (i.e., respondents who would complete followup surveys whether they received app access or were placed on the waitlist).

Table B.4 present both the identified set for the treatment effect, and a p-value testing whether the effect is zero. The identified set contains all parameter values that fit the data under monotonic attrition. Note that the identified set is *not* a confidence interval. Rather, it's a generalization of the point estimates in the main text. The corresponding p-value describes how well we can distinguish the identified set from 0 given sampling uncertainty; this can be interpreted as one would interpret a confidence interval. To tighten the bounds for mental health, the Stroop task, and the proofreading task, we condition on baseline values of these outcomes subdivided into 12 roughly evenly-sized cells. The resulting estimates may differ in decimal places from the regression estimates in the main text, which adjust linearly for baseline outcomes and include randomization stratum fixed effects. We calculate the implied p-value by inverting the Imbens and Manski (2004) confidence interval for partially identified parameters.

The main positive findings of this study are relatively unchanged under a worst-case selective attrition model. App access improves mental health at two weeks, four weeks, and three months, and improves performance on the proofreading task. Incentivizing an immediate meditation session reduces performance on the proofreading task and increases information avoidance. Although these bounds contain values closer to zero than in the main text, they do not raise concern that attrition drives our estimated treatment effects, especially as our post-attrition balance tests indicate we are unlikely to be in a worst case scenario.

We offer two caveats in interpreting these bounds. First, they are worst-case bounds, while attrition in our study does not appear to be worst-case. If it were, we would expect attrition to remove participants with the lowest (or highest) baseline mental health and proofreading scores from the App Access group, relative to the Waitlist. Table A.4 shows that this is not the case. Second, the bounding procedure assumes that treatment has a monotonic effect on attrition conditional on covariates. We view this as a tenuous. It seems possible that some waitlisted participants will be more likely to respond to followup surveys because they wish to receive the license, while other waitlisted participants will be less likely to respond if they are discouraged by not receiving the app right away. Declines in mental health may also reduce their likelihood of responding. Each of these factors likely make these bounds conservative (too wide); we therefore view them as illustrative but not definitive.

A.5 Usage And Beliefs

A.5.1 Local Average Treatment Effects

Our focus on 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 incremental improvement does additional meditation deliver? Both questions relate to the local average treatment effect (LATE) of the app on compliers, which we now discuss.

Estimating the LATE requires us to take a stance on how to measure meditation. Economically, this means making assumptions about the production function that relates meditation to outcomes. Do we believe treatment effects are due to total time spent meditating? Does regularity of practice matter? Moreover, when investigating these questions we need to consider that participants whom we induce to meditate more may fundamentally differ from those who meditate less. They may also engage in different kinds of meditation. We offer a preliminary exploration, but a fuller resolution is beyond this paper's scope.

First, we describe app usage in more detail. We categorize Headspace sessions based on how they are presented in the app, corresponding roughly to the menus or search terms a user would use to access each session. These categories often observe a common theme. For example, the "Basics" category introduces users to the core tenets of mindfulness. Figure B.5 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 than the Waitlist groups, who tended to use sleep-focused sessions and assorted mindful activities, such as deep breathing.

Second, we estimate the LATE of the app under several definitions of app usage. Tables B.19, B.20, and B.21, present the TSLS estimates where we measure usage with a binary indicator, as the number of days with a meditation session, and the number of minutes meditated. Our second-stage regression is of the form:

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

The endogenous regressor 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:

$$\begin{aligned}\text{Meditation Amount}_i = &\beta_1^{(1)}\text{App Access Only}_i + \beta_2^{(1)}\text{Short Incentives}_i + \beta_3^{(1)}\text{Long Incentives}_i \\ &+ \delta_{\text{stratum}}^{(1)} + \gamma^{(1)} Y_i^{\text{pre}} + \epsilon_i^{(1)}.\end{aligned} \tag{A.3}$$

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.20 and B.21 as the marginal effect of another day's meditation or another 10 minutes of meditation. However, we stress that these estimates are linear approximations of the 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 discern this functional form.

A.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 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 missing values (say, if they skipped a survey with the WTP elicitation).

- 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 quantities, we adapt Equation 2, separating all treatment arms:

$$Y_{it} = \delta_s + \beta_1 \text{CashTransfer} + \beta_2 \text{AppAccess}_i + \beta_3 \text{ShortIncentive}_i + \beta_4 \text{ShortIncentive}_i + \gamma Y_i^{\text{pre}} + \epsilon_{it} \tag{A.4}$$

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

$$Y_{it} = \delta_{\text{stratum}} + \beta_1 \text{AnyIncentive}_i + \beta_2 \text{LongIncentive}_i \epsilon_{it} \tag{A.5}$$

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 separate the treatment arms in this regression to investigate whether the additional meditation induced by the short incentive (relative to no incentive) and the long incentive (relative to the short incentive) affected the difficulty of maintaining a meditation practice.

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.16, row 2). However, all participants anchor heavily to the midpoint of the scale, reporting average WTPs roughly 50 dollars, suggesting that many of them may not have understood our elicitation. We therefore do not read too deeply into these findings.

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 B.17, 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.

A.6 Multiple Hypothesis Testing

We test a number of hypotheses in this project. 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. However, the immediate effects of a meditation session are less pronounced.

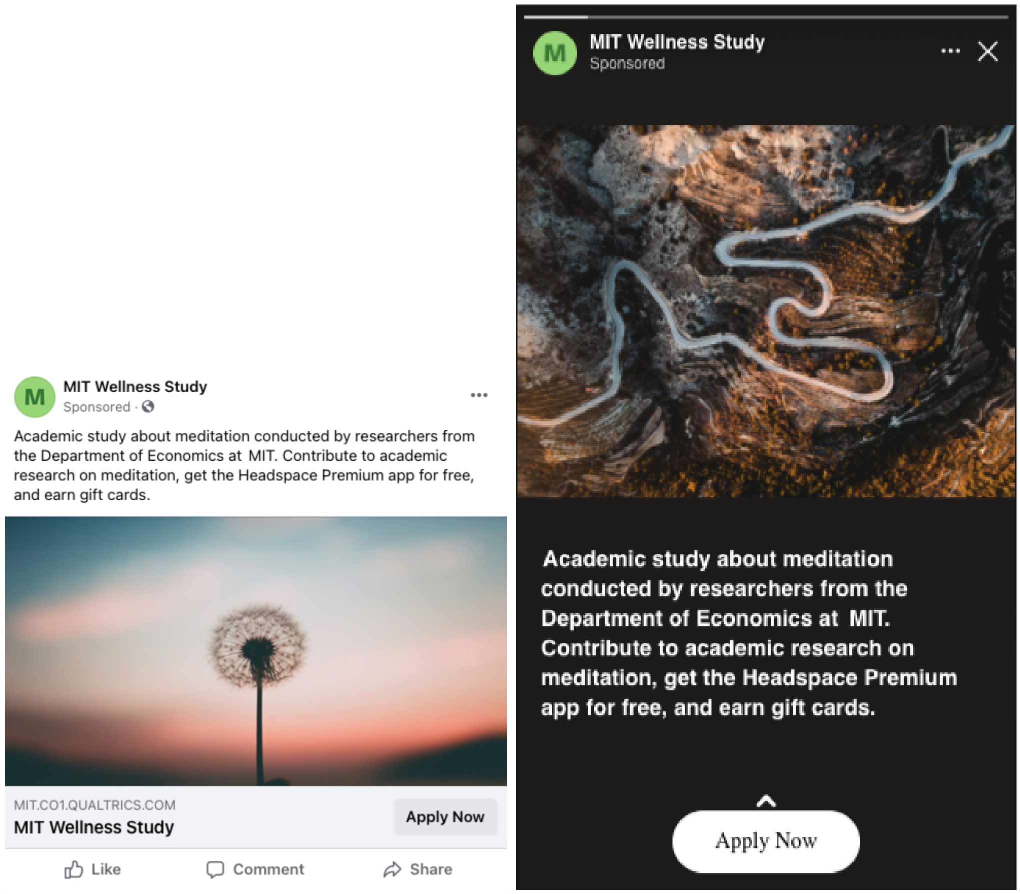
We follow the procedure from Romano and Wolf (2016), which asymptotically controls the probability of falsely rejecting one null hypothesis within a set of tested hypotheses. By using bootstrap resampling to estimate the joint sampling distribution of the test statistics in question, this procedure attains higher power than approaches like the Bonferroni or Bonferroni-Holm corrections. The corrected p-values we present are based on $9,999$ resamples, respecting stratified random assignment. We present results in Table B.22. In keeping with our pre-analysis plan, we group outcomes into families based on our pre-specified Key Contrasts and randomized treatments. The randomized treatments are app access, long-term usage incentives, and the short-term incentive to meditate before the productivity and decision-making survey. The contrasts are effects on (1) habit formation (app usage), (2) mental health, (3) attention allocation, and (4) decision-making.

Column 1 of Table B.22 contains unadjusted p-values from the main text. We then present two tiers of multiplicity correction. First, when investigating the effect of a particular treatment, we group outcomes into families (e.g., collecting all mental health outcomes into a mental health family). Column 2 contains p-values adjusted for multiplicity within treatment and within family. For example, when reporting the effects of app access on mental health at 2 weeks, it corrects for the fact that we measure 3 key mental health outcomes (2 weeks, 4 weeks, 3 months). Column 3 contains a stricter correction that adjusts for the fact that we evaluate the App Access treatment on all 10 outcomes in Panel A. 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 app access on attention could consult Column 3.

Correcting for multiplicity does not alter our main findings. App access improves mental health from 2 weeks to three months as well as proofreading performance, but not Stroop scores or decision making. Incentives have little long-term effect on meditation behavior or mental health. We find some evidence that immediate meditation affects decision making; a joint test for the effect of immediate meditation on any outcome would likely return a stronger signal than implied by the multiplicity correction here.

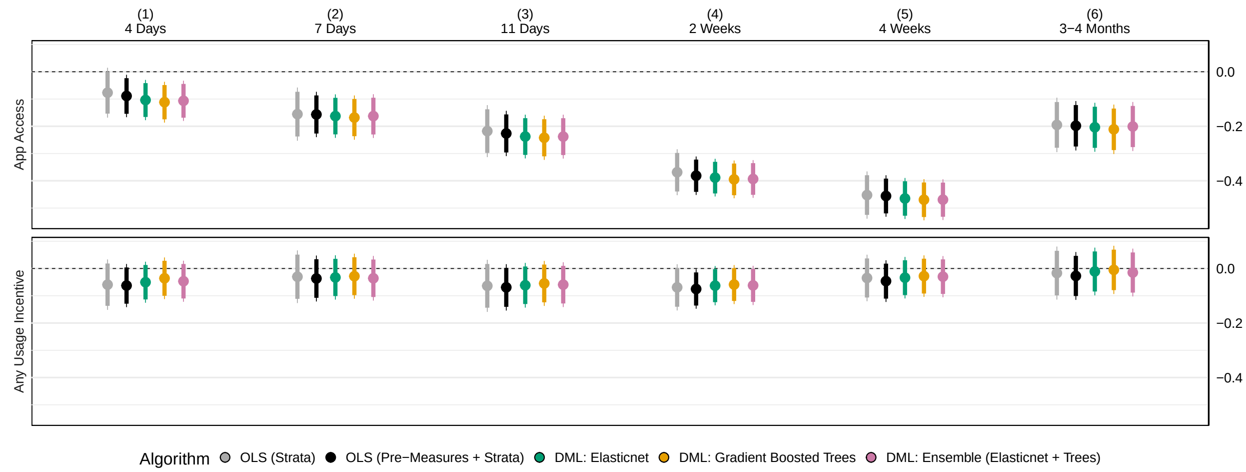
A.7 Appendix Figures

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



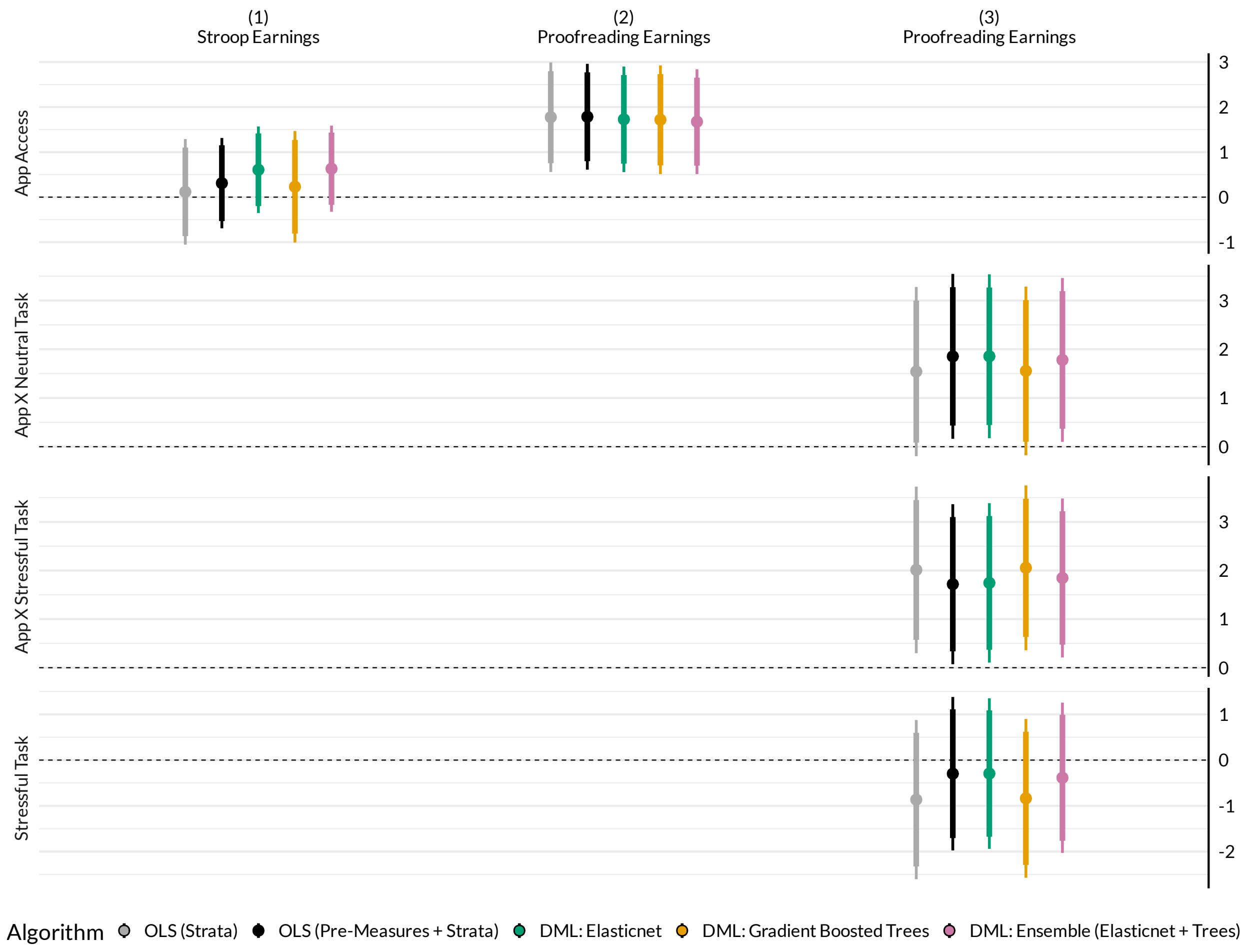
Notes: This figure shows examples of ads that were used to recruit participants for the study. A variety of images were used, with the most effective being automatically selected to be distributed more widely. The ad text was always the same.

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



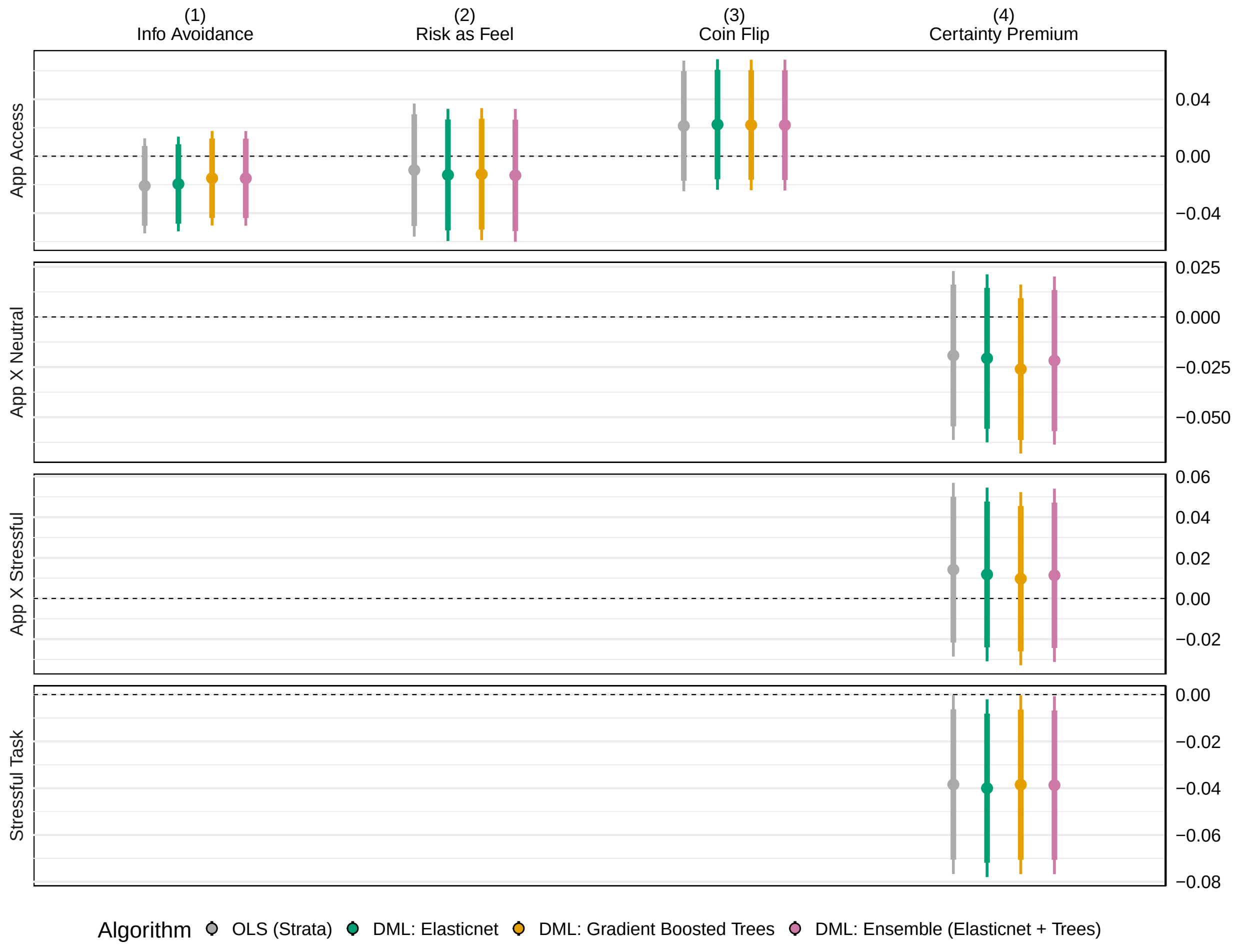
Notes: This figure presents average treatment effects of app access and usage incentives on reported symptoms of metal distress over time, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The figure's layout mirrors that of Table 3. In each column, the black point corresponds to the point estimate reported in Table 3, the thick bar corresponds to a 90% confidence interval, and the thin bar a 95% confidence interval. This point estimate comes from Equation 2, which includes stratum fixed effects and a baseline measure of the outcome. The remaining points present the debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of three algorithms: elasticnet, gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from elasticnet. Our debiased machine learning approach generalizes Equation 2, by including a broader set of baseline covariates and allowing them to enter the regression nonlinearly and interactively. Details are in Section A.3.

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



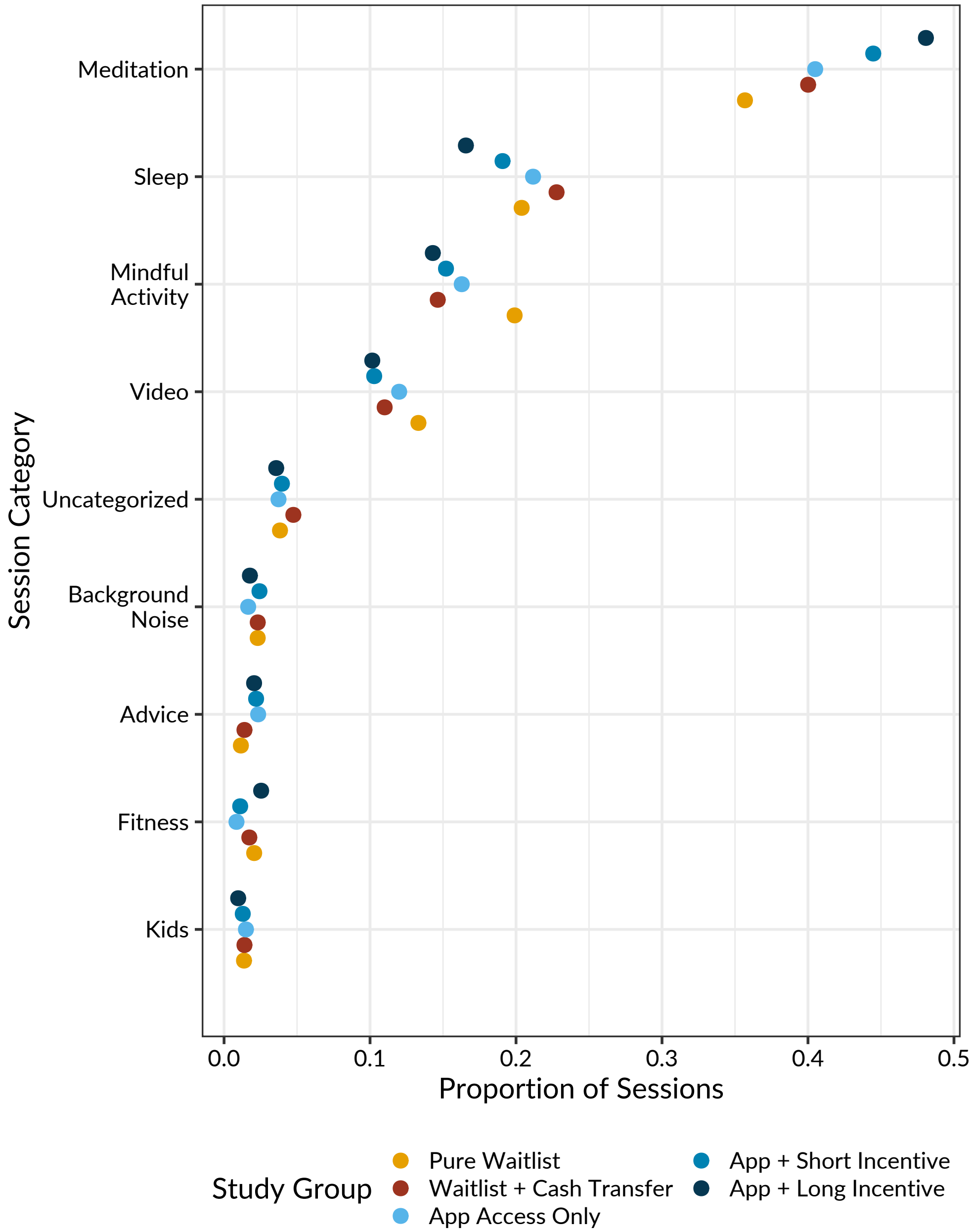
Notes: This figure presents average treatment effects of app access on on earnings in a Stroop test and a proofreading task, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The figure's layout mirrors that of Table 4. In each column, the black point corresponds to the point estimate reported in Table 4, the thick bar corresponds to a 90% confidence interval, and the thin bar a 95% confidence interval. These point estimates comes from Equations 4 and 5, which include stratum fixed effects and a baseline measure of the outcome. The remaining points present the debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of three algorithms: elasticnet, gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from elasticnet. Our debiased machine learning approach generalizes Equations 4 and 5, by including a broader set of baseline covariates and allowing them to enter the regression nonlinearly and interactively. Details are in Section A.3.

Figure B.4: Effect of App Access on Decision Making, Adjusting for Covariates



Notes: This figure presents average treatment effects of app access on choices made in several decision making tasks, adjusting for covariates using debiased machine learning (Chernozhukov et al., 2018). The figure's layout mirrors that of Table 5. In each column, the black point corresponds to the point estimate reported in Table 5, the thick bar corresponds to a 90% confidence interval, and the thin bar a 95% confidence interval. This point estimate comes from Equation 4 and 5, which includes stratum fixed effects. The remaining points present the debiased machine learning estimates of the same parameters, using 10-fold cross validation and one of three algorithms: elasticnet, gradient boosted trees, and ensemble that uses gradient-boosted trees to predict the residuals from elasticnet. Our debiased machine learning approach generalizes Equation 4 and 5, by including a broader set of baseline covariates and allowing them to enter the regression nonlinearly and interactively. Details are in Section A.3.

Figure B.5: 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.

A.8 Appendix Tables

Table B.1: Balance Table (All Arms)

	(1)	(2)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			vs. Pure Waitlist				vs. App Access Only			
	Pure Waitlist		Cash Transfer		App Access Only		Short Incentive		Long Incentive	
	Mean	(St. Dev.)	Diff.	(St. Err)	Diff.	(St. Err)	Diff.	(St. Err)	Diff.	(St. Err)
Age Group										
18-29	0.21	(0.41)	−0.01	(0.02)	−0.02	(0.02)	0.01	(0.02)	0.01	(0.02)
30-39	0.20	(0.40)	0.00	(0.03)	0.01	(0.03)	0.01	(0.03)	0.01	(0.03)
40-49	0.27	(0.44)	−0.04	(0.03)	−0.02	(0.03)	−0.02	(0.03)	0.01	(0.03)
50-59	0.19	(0.39)	0.03	(0.02)	0.01	(0.02)	−0.01	(0.02)	0.00	(0.02)
60-69	0.11	(0.31)	0.00	(0.02)	0.01	(0.02)	0.02	(0.02)	0.01	(0.02)
70+	0.02	(0.15)	0.00	(0.01)	0.02*	(0.01)	−0.01	(0.01)	−0.02**	(0.01)
Female	0.83	(0.37)	0.03	(0.02)	0.04*	(0.02)	−0.01	(0.02)	−0.01	(0.02)
Education										
No Bachelor's degree	0.18	(0.38)	−0.02	(0.02)	−0.02	(0.02)	−0.01	(0.02)	−0.01	(0.02)
Bachelor's degree	0.33	(0.47)	0.04	(0.03)	0.04	(0.03)	0.02	(0.03)	−0.01	(0.03)
Graduate or professional degree	0.44	(0.50)	−0.01	(0.03)	−0.01	(0.03)	0.00	(0.03)	0.01	(0.03)
Household Size	2.87	(1.34)	0.00	(0.09)	−0.06	(0.08)	0.00	(0.08)	−0.08	(0.08)
Household Income										
$34,999 or less	0.23	(0.42)	0.00	(0.03)	−0.01	(0.03)	−0.01	(0.03)	−0.02	(0.03)
$35,000-$74,999	0.32	(0.47)	0.00	(0.03)	0.00	(0.03)	0.02	(0.03)	0.10***	(0.03)
$75,000-$149,000	0.33	(0.47)	0.01	(0.03)	−0.02	(0.03)	0.01	(0.03)	−0.05*	(0.03)
$150,000 or more	0.04	(0.20)	0.00	(0.01)	0.02	(0.01)	−0.02	(0.01)	−0.02	(0.01)
Prefer not to answer	0.08	(0.28)	−0.02	(0.02)	0.00	(0.02)	0.00	(0.02)	0.00	(0.02)
Race & Ethnicity										
White	0.83	(0.38)	0.01	(0.02)	0.00	(0.02)	0.03	(0.02)	0.04*	(0.02)
Black	0.03	(0.17)	−0.01	(0.01)	0.00	(0.01)	0.00	(0.01)	0.00	(0.01)
Hispanic	0.08	(0.26)	−0.02	(0.02)	−0.01	(0.02)	−0.01	(0.02)	−0.02	(0.01)
Asian	0.09	(0.28)	0.00	(0.02)	−0.01	(0.02)	0.01	(0.02)	0.00	(0.02)
Other race	0.06	(0.23)	0.00	(0.01)	0.01	(0.02)	−0.03**	(0.01)	−0.04***	(0.01)
Political Party										
Democrat	0.61	(0.49)	0.04	(0.03)	0.00	(0.03)	0.01	(0.03)	0.04	(0.03)
Republican	0.03	(0.17)	0.00	(0.01)	0.02	(0.01)	−0.01	(0.01)	−0.01	(0.01)
Other	0.37	(0.48)	−0.03	(0.03)	−0.02	(0.03)	0.00	(0.03)	−0.02	(0.03)
Mental Health at Baseline										
Anxiety Score (GAD-7)	7.94	(4.39)	0.14	(0.19)	0.24	(0.19)	0.13	(0.19)	−0.16	(0.19)
Depression Score (PHQ-2)	1.67	(1.55)	0.24**	(0.09)	0.12	(0.09)	0.02	(0.09)	0.05	(0.09)
Sample Size	479		476		477		475		477	

Notes: This table is an extension of Table 1 that separates the five treatment arms in the study. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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
p-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: Post-attrition Balance on Baseline Outcomes

	(1) MH Two Weeks App Access Only vs. Waitlist	(2) MH Four Weeks App Access Only vs. Waitlist	(3) MH Three Months App Access Only vs. Waitlist	(4) Proofreading App × No Immediate vs. Waitlist	(5) Proofreading App × Immediate Meditation vs. App × No Immediate
Treated × Responded	0.021	0.006	0.007	−0.069	1.612
(S.E.)	(0.038)	(0.038)	(0.040)	(1.216)	(1.289)
Treated × Missing	0.025	0.211	0.088	−6.332	−0.088
(S.E.)	(0.430)	(0.235)	(0.100)	(6.851)	(4.018)
Stratum × Responded FE	✓	✓	✓	✓	✓
Stratum × Miss FE	✓	✓	✓	✓	✓
Test for Valid Inference (p-value): Treated × Responded = 0, Treated × Missing = 0	0.859	0.659	0.668	0.651	0.457

Notes: This table presents a post-attrition balance check, to determine whether attrition is selective on baseline outcomes. Columns 1-3 compares baseline mental health index for Waitlist and App Access participants who respond to the survey at two weeks, four weeks, and three-four months, respectively. Column 4 compares baseline proofreading scores for Waitlist and App Access (No Immediate Meditation Incentive), and column 5 does the same for Waitlist versus App Access (Immediate Meditation Incentive). The table also present a joint hypothesis test for whether attrition induces imbalance among respondents (if Treated × Responded $\neq 0$) and non-respondents (if Treated × Missing $\neq 0$). We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. The estimating equation is Equation A.1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Bounds on Population Intent to Treat Effects, Given Attrition

		ATE for Always Responders	
Treatment Group	**Control Group**	Estimate	*p*-value
A. Mental Health Index (2 weeks)			
App Access Only	Pure Waitlist	[-0.39, -0.33]	<0.001
Any Incentive	Pure Waitlist	[-0.09, -0.05]	0.094
B. Mental Health Index (4 weeks)			
App Access Only	Pure Waitlist	[-0.48, -0.42]	<0.001
Any Incentive	Pure Waitlist	[-0.07, 0.03]	>0.999
C. Mental Health Index (3-4 months)			
App Access Only	Pure Waitlist	[-0.34, -0.10]	0.022
Any Incentive	Pure Waitlist	[-0.06, 0.10]	>0.999
D. Stroop Earnings			
App × No Immediate Meditation	Waitlist	[-0.66, 0.93]	>0.999
App × Immediate Meditation	App × No Immediate	[-1.04, -0.20]	0.391
E. Proofreading Earnings			
App Access × No Immediate Meditation	Waitlist	[0.86, 2.15]	0.073
App × No Immediate × Stressful Memory	Waitlist × Stressful	[1.32, 2.49]	0.052
App × No Immediate × Neutral Memory	Waitlist × Neutral	[1.23, 2.27]	0.071
App × Immediate	App × No Immediate	[-1.41, -1.07]	0.046
App × Immediate × Stressful	App × No Immediate × Stressful	[-2.26, -2.16]	0.012
App × Immediate × Neutral	App × No Immediate × Neutral	[-0.72, -0.07]	0.650
F. Avoid Information			
App Access × No Immediate Meditation	Waitlist	[-0.02, -0.02]	0.185
App × Immediate Meditation	App × No Immediate	[0.04, 0.04]	0.026
G. Avoid Salient Loss			
App Access × No Immediate Meditation	Waitlist	[0.01, 0.01]	0.670
App × Immediate Meditation	App × No Immediate	[0.05, 0.05]	0.034

Notes: this table presents worst-case bounds on the main study's treatment effects, based on the procedure outlined in Lee (2009). This procedure assumes that treatment assignment has a monotonic effect on attrition. The estimand is the treatment effect on the "always responder" population (participants who would have responded to the survey regardless of the treatment group they were assigned to). The "Estimate" column presents the identified set: treatment effects that are consistent with the data, under the worst-case selection model. We also report a *p*-value, formed by inverting Imbens and Manski (2004)'s confidence intervals for partially identified parameters.

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Two Weeks			Four Weeks		
	Mild or Worse (Score ≥ 5)	Moderate or Worse (Score ≥ 10)	Severe (Score ≥ 15)	Mild or Worse (Score ≥ 5)	Moderate or Worse (Score ≥ 10)	Severe (Score ≥ 15)
A. Anxiety (GAD-7)						
App Access	−0.147***	−0.112***	−0.036***	−0.224***	−0.129***	−0.063***
(S.E.)	(0.024)	(0.020)	(0.012)	(0.025)	(0.019)	(0.012)
Any Usage Incentive	−0.037	−0.037*	−0.009	−0.028	−0.022	−0.005
(S.E.)	(0.025)	(0.019)	(0.011)	(0.026)	(0.018)	(0.010)
Waitlist Mean	0.723	0.285	0.094	0.717	0.261	0.100
N	2330	2330	2330	2312	2312	2312
B. Depression (PHQ-8)						
App Access	−0.133***	−0.139***	−0.054***	−0.154***	−0.136***	−0.074***
(S.E.)	(0.024)	(0.021)	(0.014)	(0.025)	(0.020)	(0.013)
Any Usage Incentive	−0.042*	−0.018	−0.011	−0.036	−0.009	0.003
(S.E.)	(0.025)	(0.020)	(0.013)	(0.025)	(0.019)	(0.011)
Waitlist Mean	0.680	0.324	0.119	0.631	0.287	0.115
N	2330	2330	2330	2311	2311	2311

Notes: This table presents average treatment effects of app access and usage incentives on reported symptoms of anxiety and depression at two and four weeks after randomization. It supplements Table 3. Anxiety is measured using the GAD-7 scale, where scores range from 0 (no symptoms) to 21 (every symptom is severe). Depression is measured using PHQ-8, where scores range from 0 (no symptoms) to 24 (every symptom is severe). This table discretizes these scales into indicators for obtaining a score above commonly used thresholds. The estimating equation is Equation 2. All regressions include stratum fixed effects and control for the similarly binarized baseline value of the outcome. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Effects of App Access and Usage Incentives on Mental Health, Separating Treatment Arms

	Time from Randomization					
	Four Days (1)	Seven Days (2)	Eleven Days (3)	Two Weeks (4)	Four Weeks (5)	Three-Four Months (6)
Cash Transfer	−0.030	−0.028	0.009	0.011	0.052	0.031
(S.E.)	(0.046)	(0.048)	(0.048)	(0.039)	(0.044)	(0.054)
App Access	−0.104**	−0.171***	−0.222***	−0.376***	−0.430***	−0.183***
(S.E.)	(0.046)	(0.049)	(0.049)	(0.040)	(0.044)	(0.054)
Short Incentive	−0.015	−0.010	−0.061	−0.081*	−0.047	−0.017
(S.E.)	(0.048)	(0.049)	(0.050)	(0.043)	(0.045)	(0.051)
Long Incentive	−0.110**	−0.064	−0.078	−0.070*	−0.046	−0.038
(S.E.)	(0.046)	(0.050)	(0.051)	(0.042)	(0.045)	(0.051)
Pure Waitlist Mean	0.000	0.000	0.000	0.000	0.000	0.000
N	2305	2145	2191	2330	2311	2004
Index Components:						
Anxiety	GAD-2	GAD-2	GAD-2	GAD-7	GAD-7	GAD-7
Depression				PHQ-8	PHQ-8	PHQ-8
Stress				PSS-10	PSS-10	

Notes: This table presents average treatment effects of app access and usage incentives on reported symptoms of mental distress over time. It supplements Table 3 by fully separating the treatment arms. We measure symptoms of anxiety using the two- and seven-item Generalized Anxiety Disorder scales (GAD-2 and GAD-7, respectively); symptoms of depression using the eight-item Patient Health Questionnaire (PHQ-8); and stress using the ten-item Perceived Stress Scale (PSS-10). The outcome at each timepoint is a standardized index that combines the mental health scales measured at that time. We first standardize each scale in each time period by subtracting the Pure Waitlist mean and dividing by the Pure Waitlist standard deviation. The index is the average of these standardized scales. Lower scores indicate lower reported levels of distress. The estimating equation adapts Equation 2 to include indicators for each treatment arm, omitting the Pure Waitlist as the reference group. It includes stratum fixed effects and the baseline mental health index. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.7: Heterogeneous Effects of App Access and Usage Incentives

	(1)	(2)	(3)	(4)
	2 Weeks		**4 Weeks**	
	Days Used	Mental Health	Days Used	Mental Health
A. Baseline Anxiety				
App Access	5.052***	−0.178***	7.670***	−0.203***
(S.E.)	(0.340)	(0.053)	(0.548)	(0.054)
App Access × At Least Median	0.411	−0.328***	0.261	−0.429***
(S.E.)	(0.439)	(0.082)	(0.704)	(0.084)
Any Incentive	2.158***	−0.096*	2.228***	−0.086*
(S.E.)	(0.415)	(0.053)	(0.659)	(0.052)
Any Incentive × At Least Median	−0.228	0.046	−0.099	0.090
(S.E.)	(0.536)	(0.083)	(0.847)	(0.083)
B. Baseline Mindfulness				
App Access	5.338***	−0.426***	7.867***	−0.547***
(S.E.)	(0.330)	(0.066)	(0.524)	(0.068)
App Access × At Least Median	−0.060	0.123	−0.043	0.190**
(S.E.)	(0.437)	(0.083)	(0.693)	(0.085)
Any Incentive	2.084***	−0.108	2.134***	−0.006
(S.E.)	(0.399)	(0.069)	(0.624)	(0.070)
Any Incentive × At Least Median	−0.138	0.062	0.021	−0.065
(S.E.)	(0.533)	(0.087)	(0.840)	(0.088)
C. Belief about ATE				
App Access	4.907***	−0.277***	7.433***	−0.333***
(S.E.)	(0.320)	(0.060)	(0.514)	(0.062)
App Access × At Least Median	0.657	−0.160*	0.670	−0.210**
(S.E.)	(0.432)	(0.083)	(0.687)	(0.085)
Any Incentive	2.073***	−0.052	1.987***	−0.047
(S.E.)	(0.394)	(0.063)	(0.618)	(0.065)
Any Incentive × At Least Median	−0.070	−0.034	0.340	0.019
(S.E.)	(0.529)	(0.086)	(0.833)	(0.088)
D. Social Desirability				
App Access	4.687***	−0.438***	6.998***	−0.496***
(S.E.)	(0.312)	(0.064)	(0.489)	(0.067)
App Access × At Least Median	0.980**	0.124	1.332**	0.082
(S.E.)	(0.427)	(0.084)	(0.676)	(0.086)
Any Incentive	2.814***	0.005	3.103***	0.008
(S.E.)	(0.389)	(0.069)	(0.603)	(0.070)
Any Incentive × At Least Median	−1.337**	−0.131	−1.582*	−0.076
(S.E.)	(0.525)	(0.089)	(0.824)	(0.090)

Notes: This table presents the heterogeneous treatment effects of app access and usage incentives on app usage and mental health, along several baseline covariates. Each panel contrasts participants at or above the median baseline value of the covariate with those who are below the baseline. Panel A separates participants by their baseline anxiety score on the GAD-7 scale, and Panel B does the same for baseline mindfulness on the FFMQ-15 scale. Panel C splits participants based on their beliefs about the average treatment effect of app access on anxiety. Above-median beliefs indicate that the participant believes app access will cause a larger decline in anxiety. Panel D presents heterogeneity by social desirability, measured on the Marlow-Crowne scale. In all cases, the estimating equation is: $Y_i^{\text{post}} = \delta_s^{\text{lo}} + \beta_1^{\text{lo}}\text{AppAccess} + \beta_2^{\text{lo}}\text{AnyIncentive} + \delta_s^{\text{hi}} \times H_i + \beta_1^{\text{hi}}\text{AppAccess} \times H_i + \beta_2^{\text{hi}}\text{AnyIncentive} \times H_i + \epsilon_i$. This modifies Equation 3 so that β_1^{hi} and β_2^{hi} are marginal effects of being weakly above versus strictly below the median value. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: Effect of Stressful Tasks on Self-Reported Mood

	(1)	(2)
	Mood After ...	
	First Dose	Second Dose
A. Long-term Practice Sample		
Stressful Task	1.047***	0.306***
(S.E.)	(0.068)	(0.065)
App Access × Stressful Task	−0.096	−0.209***
(S.E.)	(0.076)	(0.070)
App Access × Neutral Task	−0.454***	−0.355***
(S.E.)	(0.072)	(0.067)
Waitlist Mean (Neutral Task)	1.966	2.032
N	1592	1592
B. Short-term Meditation Sample		
Stressful Task	1.408***	0.453***
(S.E.)	(0.080)	(0.071)
Immediate × Stressful Task	−0.006	0.002
(S.E.)	(0.086)	(0.076)
Immediate × Neutral Task	−0.169**	−0.132*
(S.E.)	(0.073)	(0.068)
App Access Mean (Neutral Task)	1.498	1.664
N	1332	1332

Notes: This table presents average treatment effects of the Stressful tasks on self-reported mood. It supplements Tables 4, 5, 6, and 7 by measuring the subjective effect of the emotion induction treatments. The first dose involves recounting either an unresolved worry (Stressful) or a daily routine (Neutral). The second dose involves describing how one would respond to a large medical bill (Stressful) or a small one (Neutral). After each dose, we participants' mood on a 6-point scale from "very calm, very relaxed" (1) to "very upset, very stressed" (6). Panel A presents the effects on the app access group, relative to the waitlist. The estimating equation is Equation 4. Panel B presents the effects for the immediate meditation group, relative to the app access group. Both regressions includes stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.9: Effects of App Access on Detailed Measures of Performance in the Proofreading Task

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Errors Found		False Positives		Time Taken (sec)		Hourly Earnings	
App Access	0.301**		−0.050		3.323		0.563	
(S.E.)	(0.120)		(0.035)		(3.027)		(0.588)	
App × Neutral		0.365**		−0.007		5.708		0.763
(S.E.)		(0.173)		(0.051)		(4.383)		(0.907)
App × Stressful		0.238		−0.094*		0.943		0.364
(S.E.)		(0.168)		(0.048)		(4.187)		(0.756)
Stressful Memory		−0.055		0.017		0.499		−0.112
(S.E.)		(0.168)		(0.051)		(3.967)		(0.726)
Waitlist Mean	14.756	14.830	0.291	0.274	130.544	128.219	30.988	31.760
N	1592	1592	1592	1592	1592	1592	1592	1592

Notes: This table presents average treatment effects of app access on additional outcomes in the proofreading task. It supplements Table 4. 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. The estimating equations are Equation 4 (columns 1, 3, 5, 7) and 5 (columns 2, 4, 6, 8). All regressions include stratum fixed effects and control for performance on these tasks in the baseline survey. The sample excludes participants in the App Access group who received incentives to meditate immediately before the survey. The reference group is the combined Waitlist group. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.10: Effects of Immediate Meditation on Measures of Performance in the Proofreading Task

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Errors Found		False Positives		Time Taken (sec)		Hourly Earnings	
Immediate Meditation	−0.130		0.078**		−4.268		−0.478	
(S.E.)	(0.123)		(0.039)		(3.424)		(0.657)	
Immediate × Neutral		−0.026		−0.007		−7.174		−1.201
(S.E.)		(0.167)		(0.049)		(4.985)		(0.970)
Immediate × Stressful		−0.240		0.164***		−1.360		0.252
(S.E.)		(0.182)		(0.060)		(4.993)		(0.894)
Stressful Memory		−0.172		−0.070		−5.176		−0.511
(S.E.)		(0.173)		(0.048)		(4.770)		(0.924)
App Access Mean	15.079	15.153	0.244	0.276	133.702	136.924	31.205	31.407
N	1332	1332	1332	1332	1332	1332	1332	1332

Notes: This table presents average treatment effects of encouraging an immediate meditation session on additional outcomes in the proofreading task. It supplements Table 6. 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. The estimating equations are Equation 6 (columns 1, 3, 5, 7) and 7 (columns 2, 4, 6, 8). All regressions include stratum fixed effects and control for performance on these tasks in the baseline survey. The sample excludes participants in the Waitlist group. The reference group is subset of the App Access group that did not receive incentives to meditate immediately before the survey. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.11: Effects of Immediate Meditation on Time Taken in Stroop Task

	Time Taken (sec)
A. Long-term Practice Sample	
App Access	-0.214
(S.E.)	(0.301)
Waitlist Mean	67.817
N	1590
B. Short-term Meditation Sample	
Immediate Meditation	0.159
(S.E.)	(0.513)
App Access Mean	67.921
N	1329

Notes: This table presents average treatment effects on time taken in the Stroop task. It supplements Tables 4 and 6. Panel A presents the effect of app access, relative to the waitlist. The estimating equation is Equation 4. Panel B presents the effect of encouraging an immediate meditation session, among participants with app access. The estimating equation is Equation 6. Both equations include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.12: Effects of App Access on Risk Aversion Under Certainty and Uncertainty

	(1)	(2)	(3)	(4)
	$P_{\text{uncertainty}}$		$P_{\text{certainty}}$	
App Access	0.735		1.220	
(S.E.)	(0.829)		(1.034)	
App × Neutral		1.808		1.626
(S.E.)		(1.153)		(1.465)
App × Stressful		−0.345		0.812
(S.E.)		(1.192)		(1.462)
Stressful Memory		2.326**		0.658
(S.E.)		(1.047)		(1.310)
Waitlist Mean	72.047	70.887	58.409	58.089
N	1587	1587	1587	1587

Notes: This table presents average treatment effects of app access and a stress-inducing task on decisions between risky prospects. It supplements Table 5 by decomposing the certainty premium into its components. $P_{uncertain}$ is the probability (times 100) that makes a participant indifferent between a 50-50 lottery with \$30 and \$10 payouts, and a a lottery that pays \$30 with probability $P_{uncertain}$ and \$0 otherwise. A value of ≈ 67 indicates risk neutrality in this choice, and higher values indicate greater risk aversion. $P_{certain}$ is the probability (times 100) that makes a participant indifferent between receiving \$10 for sure or playing a lottery that pays \$30 with probability $P_{certain}$ but 0 otherwise. A value of ≈ 33 indicates risk neutrality, and higher values of $P_{certain}$ indicate greater risk aversion. The estimating equations are Equation 4 (columns 1 and 3) and Equation 5 (columns 2 and 4). We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. Related tables and figures. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.13: Effects of App Access and Immediate Meditation Incentives on Information Avoidance

	(1)	(2)	(3)	(4)	(5)
		Avoid Info About ...			Mechanism
	Life Expectancy	Dementia	Job Loss	Retirement Finances	Avoid Coin
A. Long-term Practice					
App Access	0.020	−0.038	−0.033	−0.033	0.021
(S.E.)	(0.024)	(0.023)	(0.025)	(0.025)	(0.023)
Waitlist Mean	0.319	0.328	0.595	0.576	0.294
N	1592	1592	1592	1592	1592
B. Short-term Meditation					
Immediate Meditation	−0.005	0.037	0.039	0.081***	0.033
(S.E.)	(0.026)	(0.025)	(0.027)	(0.027)	(0.026)
App Access Mean	0.339	0.289	0.562	0.543	0.315
N	1332	1332	1332	1332	1332

Notes: This table presents the average treatment effects on avoidance of potentially distressing information. It supplements Tables 5 and 7 by decomposing the average information avoidance outcome into its four component decisions. Panel A presents the effect of app access, relative to the waitlist. The estimating equation is Equation 4. Panel B presents the effect of encouraging an immediate meditation session, among participants with app access. The estimating equation is Equation 6. Both equations include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.14: Independent Meditation Activities

	(1)	(2)	(3)	(4)
	2 Weeks	**3-4 Months**		
	Any Non-Headspace	Any Non-Headspace	Other App	Non-App
Cash Transfer				
(S.E.)				
App Access	0.259***	0.097***	0.054**	0.071***
(S.E.)	(0.026)	(0.030)	(0.025)	(0.027)
Short Incentive	−0.003	0.039	−0.024	0.042
(S.E.)	(0.030)	(0.035)	(0.029)	(0.032)
Long Incentive	−0.013	0.022	−0.002	0.022
(S.E.)	(0.030)	(0.035)	(0.030)	(0.032)
Pure Waitlist Mean	0.484	0.341	0.171	0.218
N	2242	1994	1994	1994

Notes: This table presents average treatment effects of app access and usage incentives on self-reported meditation activities outside the mindfulness app. Two weeks after randomization, we ask participants to recall whether they practiced any form of meditation in the preceding two weeks. Three-four months after randomization, we ask them to indicate separately whether they have used Headspace, another app, or meditated without an app in the previous month. The terms Cash Transfer and App Access express effects relative to the Pure Waitlist group, while App × Short Incentive and App × Long Incentive express the marginal effects of incentives relative to the App Access Only group. All regressions include stratum fixed effects. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.15: Five Facet Mindfulness Questionnaire

	(1) Total	(2) Observing	(3) Describing	(4) Acting with Awareness	(5) Non-judgment	(6) Non-reactivity
App Access	0.194***	0.228***	0.150***	0.184***	0.240***	0.172***
(S.E.)	(0.022)	(0.036)	(0.034)	(0.033)	(0.039)	(0.036)
Any Usage Incentive	0.008	−0.020	0.016	0.022	0.015	0.012
(S.E.)	(0.023)	(0.037)	(0.034)	(0.033)	(0.039)	(0.037)
Waitlist Mean	2.232	2.164	2.484	1.961	2.472	2.080
N	2330	2330	2330	2330	2330	2330

Notes: This table presents average treatment effects of app access and usage incentives on mindfulness, as measured by the Five Facet Mindfulness Questionnaire (FFMQ-15). The FFMQ contains fifteen items, each of which is scored on an integer scale of 1 (lower mindfulness) to 5 (higher mindfulness). Column 1 presents the average score across all 15 items. Columns 2 through 6 present scores on subsets of questions that focus on specific aspects of mindfulness. All regressions include stratum fixed effects and control for a baseline value of the outcome. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.16: Willingness to Pay

	(1) 4 Days	(2) 7 Days	(3) 11 Days	(4) 4 Weeks
Cash Transfer	2.213*	2.908**	2.169*	1.122
(S.E.)	(1.184)	(1.245)	(1.268)	(1.427)
App Access	4.049***	3.224**	2.719**	0.544
(S.E.)	(1.238)	(1.341)	(1.366)	(1.570)
Short Incentive	0.579	1.835	1.166	0.326
(S.E.)	(1.306)	(1.389)	(1.481)	(1.640)
Long Incentive	0.229	2.301*	1.633	0.364
(S.E.)	(1.233)	(1.351)	(1.426)	(1.660)
Pure Waitlist Mean	49.133	48.998	49.204	49.189
N	2260	2140	2189	2310

Notes: This table presents average treatment effects of app access and usage incentives on willingness to pay for a 90-day extension of the Headspace license. The cash value of the license was approximately $39 at the time of the experiment. We elicit willingness to pay with a probabilistic Becker–DeGroot–Marschak mechanism, allowing participants to indicate a valuation for the extension of between $0 and $100 using a sliding scale. All regressions include stratum fixed effects and control for a baseline value of the outcome. We calculate standard errors that are robust to heteroskedasticity and misspecification with the HC3 estimator. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. See Section A.5.2 for more details on the elicitation.

Table B.17: Subjective ATE

	(1) Baseline	(2) 4 Days	(3) 7 Days	(4) 11 Days	(5) 4 Weeks
Cash Transfer	0.021	−0.047	0.003	0.078	0.125
(S.E.)	(0.135)	(0.079)	(0.107)	(0.100)	(0.112)
App Access	−0.088	−0.083	−0.068	0.093	0.346***
(S.E.)	(0.138)	(0.085)	(0.114)	(0.113)	(0.119)
Short Incentive	−0.091	0.058	−0.005	−0.049	−0.143
(S.E.)	(0.132)	(0.090)	(0.109)	(0.115)	(0.123)
Long Incentive	−0.070	0.071	0.194*	0.101	0.124
(S.E.)	(0.127)	(0.087)	(0.111)	(0.119)	(0.118)
Pure Waitlist Mean	−3.635	−3.576	−3.577	−3.611	−3.658
N	2384	2289	2131	2185	2309

Notes: This table presents average treatment effects of app access and usage incentives on participants' predictions about the average treatment effect in the study. See Section A.5.2 for more details on the elicitation.

Table B.18: Ease of Meditating

	(1) 4 Days	(2) 7 Days	(3) 11 Days	(4) 30 Days
A. Finding Time and Space				
Short Incentive	0.498***	0.413**	0.659***	0.438**
(S.E.)	(0.159)	(0.176)	(0.177)	(0.185)
Long Incentive	0.450***	0.268	0.106	0.068
(S.E.)	(0.159)	(0.172)	(0.173)	(0.186)
App Access Only Mean	4.696	4.919	4.905	5.012
Sample Size	1098	825	816	956
B. Focusing for 10 Minutes				
Short Incentive	0.044	0.088***	0.067**	0.053*
(S.E.)	(0.032)	(0.032)	(0.032)	(0.030)
Long Incentive	0.062**	0.071**	0.066**	0.044
(S.E.)	(0.031)	(0.033)	(0.032)	(0.030)
App Access Only Mean	0.603	0.618	0.634	0.672
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 A.5. 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 A.5.2 for more details on the elicitation.

Table B.19: LATE of Any App Usage on Key Outcomes

	(1) Two Weeks	(2) Four Weeks	(3)	(4) Three-Four Months
	Mental Health Index	Mental Health Index	Proofreading	Mental Health Index
A. App Access (Pooled)				
Any App Usage	-0.492***	-0.542***	1.366**	-0.472***
(S.E.)	(0.029)	(0.031)	(0.588)	(0.074)
[95% AR Interval]	[-0.572, -0.412]	[-0.631, -0.454]	[-0.239, 2.975]	[-0.681, -0.270]
B. App Access Only				
Any App Usage	-0.468***	-0.534***	1.545*	-0.472***
(S.E.)	(0.044)	(0.046)	(0.807)	(0.111)
[95% AR Interval]	[-0.552, -0.385]	[-0.623, -0.447]	[-0.143, 3.236]	[-0.700, -0.254]
C. Short Incentives				
Any App Usage	-0.510***	-0.548***	1.271*	-0.450***
(S.E.)	(0.039)	(0.042)	(0.759)	(0.094)
[95% AR Interval]	[-0.585, -0.435]	[-0.629, -0.467]	[-0.278, 2.821]	[-0.646, -0.261]
D. Long Incentives				
Any App Usage	-0.491***	-0.542***	1.283*	-0.496***
(S.E.)	(0.038)	(0.041)	(0.739)	(0.093)
[95% AR Interval]	[-0.565, -0.418]	[-0.621, -0.462]	[-0.239, 2.809]	[-0.691, -0.308]
E. Supplementary Information				
Sargan-Hansen Overidentification Test				
J-statistic	0.73	0.07	0.10	0.16
p value	0.694	0.967	0.951	0.925
First-Stage F				
Pooled	2379	2800	2470	228
App Access Only	4075	5208	4230	233
Short Incentives	9039	9870	9553	323
Long Incentives	10227	10921	10394	321
Sample Size				
Pooled	2330	2311	2257	2004
App Access Only	1408	1398	1373	1209
Short Incentives	1400	1389	1366	1209
Long Incentives	1408	1396	1368	1202

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 A.3 for the first stage and A.2 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.20: LATE of Days Meditated on Key Outcomes

	(1) Two Weeks	(2) Four Weeks	(3)	(4) Three-Four Months
	Mental Health Index	Mental Health Index	Proofreading	Mental Health Index
A. App Access (Pooled)				
Days Per Week	-0.426***	-0.332***	1.110**	-0.125***
(S.E.)	(0.026)	(0.020)	(0.494)	(0.020)
[95% AR Interval]	[-0.438, -0.415]	[-0.361, -0.306]	[-0.209, 2.432]	[-0.184, -0.071]
B. App Access Only				
Days Per Week	-0.493***	-0.377***	1.559*	-0.124***
(S.E.)	(0.047)	(0.034)	(0.817)	(0.030)
[95% AR Interval]	[-0.583, -0.405]	[-0.442, -0.314]	[-0.145, 3.275]	[-0.185, -0.066]
C. Short Incentives				
Days Per Week	-0.474***	-0.357***	1.136*	-0.128***
(S.E.)	(0.037)	(0.028)	(0.677)	(0.027)
[95% AR Interval]	[-0.544, -0.404]	[-0.411, -0.304]	[-0.249, 2.521]	[-0.184, -0.075]
D. Long Incentives				
Days Per Week	-0.385***	-0.303***	0.964*	-0.124***
(S.E.)	(0.030)	(0.023)	(0.554)	(0.024)
[95% AR Interval]	[-0.443, -0.328]	[-0.348, -0.258]	[-0.180, 2.111]	[-0.173, -0.077]
E. Supplementary Information				
Sargan-Hansen Overidentification Test				
J-statistic	7.66	6.01	0.51	0.03
p value	0.022	0.050	0.775	0.985
First-Stage F				
Pooled	729	542	720	108
App Access Only	1216	1003	1196	184
Short Incentives	2395	1715	2282	243
Long Incentives	2635	2064	2712	335
Sample Size				
Pooled	2330	2311	2257	2004
App Access Only	1408	1398	1373	1209
Short Incentives	1400	1389	1366	1209
Long Incentives	1408	1396	1368	1202

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 A.3 for the first stage and A.2 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.21: LATE of 10 Minutes of Meditation on Key Outcomes

	(1) Two Weeks	(2) Four Weeks	(3)	(4) Three-Four Months
	Mental Health Index	Mental Health Index	Proofreading	Mental Health Index
A. App Access (Pooled)				
10 Minutes Per Day	-0.006***	-0.008***	0.016**	-0.008***
(S.E.)	(0.000)	(0.001)	(0.007)	(0.001)
[95% AR Interval]	[-0.006, -0.006]	[-0.009, -0.007]	[-0.003, 0.035]	[-0.013, -0.005]
B. App Access Only				
10 Minutes Per Day	-0.007***	-0.010***	0.024*	-0.008***
(S.E.)	(0.001)	(0.001)	(0.013)	(0.002)
[95% AR Interval]	[-0.009, -0.006]	[-0.012, -0.008]	[-0.002, 0.050]	[-0.012, -0.004]
C. Short Incentives				
10 Minutes Per Day	-0.007***	-0.009***	0.016*	-0.008***
(S.E.)	(0.001)	(0.001)	(0.010)	(0.002)
[95% AR Interval]	[-0.008, -0.006]	[-0.010, -0.007]	[-0.004, 0.036]	[-0.012, -0.004]
D. Long Incentives				
10 Minutes Per Day	-0.005***	-0.008***	0.014*	-0.009***
(S.E.)	(0.000)	(0.001)	(0.008)	(0.002)
[95% AR Interval]	[-0.006, -0.005]	[-0.009, -0.006]	[-0.003, 0.031]	[-0.013, -0.005]
E. Supplementary Information				
Sargan-Hansen Overidentification Test				
J-statistic	7.59	4.94	0.61	0.27
p value	0.022	0.085	0.736	0.873
First-Stage F				
Pooled	214	179	210	37
App Access Only	285	323	281	75
Short Incentives	778	579	751	95
Long Incentives	860	677	868	110
Sample Size				
Pooled	2330	2311	2257	2004
App Access Only	1408	1398	1373	1209
Short Incentives	1400	1389	1366	1209
Long Incentives	1408	1396	1368	1202

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 A.3 for the first stage and A.2 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.22: Multiple Hypothesis Testing

		(1)	(2)	(3)
		Unadjusted	Adjusted within ...	
Family	Outcome	p-value	Family	Treatment
A. Effect of App Access				
Mental Health	2 Week Index	<0.001	<0.001	<0.001
	4 Week Index	<0.001	<0.001	<0.001
	3-4 Month Index	<0.001	<0.001	<0.001
Attention	Proofreading Score	0.003	0.010	0.019
	Proofreading Wage	0.339	0.684	0.860
	Stroop Score	0.544	0.690	0.916
	Stroop Time	0.478	0.690	0.916
Decisions	Certainty Premium	0.864	0.900	0.916
	Avoid Info	0.221	0.521	0.757
	Avoid Salient Loss	0.682	0.900	0.916
B. Effect of Usage Incentives				
Usage	After Short Incentive	0.066	0.117	0.226
	After Long Incentive	0.437	0.428	0.676
Mental Health	2 Week Index	0.042	0.104	0.172
	4 Week Index	0.236	0.390	0.527
	3-4 Month Index	0.542	0.542	0.676
C. Effect of Immediate Meditation				
Attention	Proofreading Score	0.084	0.290	0.351
	Proofreading Wage	0.467	0.746	0.846
	Stroop Score	0.390	0.746	0.846
	Stroop Time	0.757	0.759	0.941
Decisions	Certainty Premium	0.896	0.896	0.941
	Avoid Info	0.042	0.130	0.268
	Avoid Salient Loss	0.053	0.130	0.278

Notes: Multiple testing corrections using the Romano and Wolf (2016) and 9,999 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.