Does promoting one healthy behavior detract from others? Evidence from a field experiment*

Hannah Trachtman†

February 23, 2023

Abstract

Impact evaluations of behavioral interventions typically focus on target outcomes. Might interventions induce negative spillovers on other behaviors? I run a large field experiment in which individuals receive combinations of messages and incentives promoting two healthy behaviors, meditation and meal logging. I find that the interventions reduce completion rates of the opposite behavior by 19-29%. I find that interventions with larger target effects do not necessarily generate larger negative spillovers, and demonstrate implications for cost-effectiveness analysis. I investigate the mechanisms behind the observed spillovers.

JEL classification: D62, D91, I12

*I am so grateful to my dissertation committee for their help at every stage of this project: Jason Abaluck, Dean Karlan, Costas Meghir, and Chris Udry. For excellent research assistance I thank Osman Tuunteeyah. I thank several anonymous referees for their insightful comments. For helpful discussions, I thank Joseph Altonji, Gaurav Chiplunkar, Taha Choukhmane, Marvin Chun, Ori Heffetz, John Eric Humphries, Ro'ee Levy, Cormac McCarthy, Mushfiq Mobarak, Mark Rosenzweig, and Jaya Wen. Financial support from the Yale Economics Department and Economic Growth Center, the Sylff Foundation, the National Science Foundation, the Russell Sage Foundation, and the Institution for Social and Policy Studies is greatly appreciated. This research was approved by the Yale Institutional Review Board (Protocol ID 2000021379). The experiment is registered at the AEA RCT Registry (Trachtman, 2019). All errors are my own.

†Department of Economics, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem, 9190501, Israel. hannah.trachtman@mail.huji.ac.il

Seemingly small changes in behaviors like routine childhood immunization, healthy eating, and the use of clean cookstoves—among others—can lead to big economic benefits.[1] Given their large returns, low take-up of such behaviors is often seen as a puzzle that merits intervention. In recent years, both traditional interventions like "sin taxes" (O'Donoghue and Rabin, 2006; Allcott et al., 2019; Farhi and Gabaix, 2020), as well as non-traditional instruments like "nudges" (Thaler and Sunstein, 2009), have been widely used. Many have been shown to be effective and cost-effective with respect to their target outcomes.[2]

Most evaluations measure the behavior targeted by the intervention and little else unless there are specific reasons to expect spillovers.[3] But there may be general reasons to worry about spillovers. Of course, time-consuming behaviors must displace other activities. Additionally, limits to attention or preferences to maintain one's self-image could result in intervention spillovers (Dolan and Galizzi, 2015; Persson, 2018; Nafziger, 2020). Several recent empirical papers have discovered unintended side-effects of real interventions that appear to be consistent with these ideas (Medina, 2020; Hussam and Oh, 2021; Hall and Madsen, 2022).

How concerned should we be about spillovers when designing interventions, in general? I address this question in two ways. First, I test for the existence of spillovers in a relatively standard context: messaging and incentive interventions that encourage small daily changes in health and wellness behavior. The pervasiveness of similar interventions in the real world makes this context highly relevant, but one where traditional economic models would not predict large spillovers. I implement a large, pre-registered field experiment and collect objective, high-frequency data using smartphone apps. Because the experiment design involves the promotion and measurement of two behaviors, I can assess the costliness of spillovers from the perspective of a social planner who cares about them both. Second, I analyze the mechanisms behind spillovers. I discuss several classes of mechanisms that can potentially drive spillovers, and then leverage additional data and

[1]For example, see Afshin et al. (2019), Institute for Health Metrics and Evaluation (2018), and Zhou et al. (2014).

[2]For example, soda taxes have shown to reduce sugar consumption due to soda by 18% on average, and by 40% among young people age 13-21 (Dubois et al., 2020). Prompting people to write down a plan for getting the flu vaccine raised immunization rates by 13% (Milkman et al., 2011). Other interventions have proven less effective: encouraging the use of clean cookstoves has proved to be much more difficult than expected (Hanna et al., 2016).

[3]See for example the literature how retirement savings policies affect other financial behavior (Chetty et al., 2014; Beshears et al., 2021) and the literature on the general equilibrium effects of cash transfers and other anti-poverty programs (Muralidharan et al., 2017; Egger et al., 2019).

analysis to assess the importance of each.

I run an online experiment, recruiting 3,818 individuals via Facebook Ads that promote a study about daily meditation (behavior x) and meal logging (behavior y). Participants took a baseline survey and downloaded two smartphone applications, one for tracking meditation, and the other for logging meals. Upon verifying that participants downloaded the apps, they were enrolled and randomized to one of five interventions: control (no intervention), messages about behavior x, messages about behavior y, messages about both behaviors x and y, and incentives for behavior y. The treatment period lasted four weeks. Participants in groups with only x or y messages received prompts twice a day targeting the corresponding behavior. Participants in the group with both x and y messages received the union of both message sets, resulting in four daily messages. Lastly, participants in the y incentive group received an expected reward for every day they successfully logged their meals. Data on both behaviors was collected from the apps during the treatment period and for an additional four weeks after the end of treatment.[4]

The results show large target effects of both message and incentive interventions. Meditation messages raised the rate of meditation by 8.8 percentage points (almost double the control rate) and nutrition messages raised the rates of meal logging by 16.6 percentage points (more than double the control rate). Incentives for meal logging had an even larger effect, raising rates of meal logging by 38.1 percentage points (more than triple the control rate). But all three treatments imposed substantial spillovers on the opposite behavior, as measured by comparisons with the control group. Messages about meditation reduced meal logging by 2.4 percentage points (19%), messages about nutrition reduced meditation by 2.8 percentage points (29%), and incentives for meal logging reduced meditation by 2.5 percentage points (27%). The group with both target and non-target messages also did worse on target behaviors than the group that received just target messages, by 2.2 and 5.0 percentage points for meditation and meal logging, respectively. Taking all the results together, it becomes clear that interventions with dramatically different effects on target behaviors generated similarly sized spillover effects. With respect to dynamics, I find that treatment effects on target behaviors exhibited substantial

[4]This is not technically a natural field experiment in the framework of Harrison and List (2004), as subjects filled out a consent form and knew they were taking part in a study. They also knew that assignment to programs was random, for reasons that I explain in Section VI.C. However, it is close to a natural field experiment since the subjects undertake the tasks in their natural environments (and the commodity, task, and information set all have a field context).

decay, and yet persisted in the post-treatment period, as did spillover effects.

To assess the implications my findings, I consider a social planner who evaluates interventions on the basis of cost-effectiveness with respect to both behaviors. First, I show that due to spillovers, interventions are likely to be more costly than they appear. Second, I draw out implications of the finding that high-impact interventions do not generate comparatively larger spillovers, and demonstrate how it can produce unappreciated differences in cost-effectiveness across interventions.

Finally, I examine the evidence for different mechanisms, including time and effort constraints, limited attention, and moral licensing. First, I take advantage of the fact that under time or effort constraints, spillovers on non-target behaviors would be induced by positive effects on target behaviors. I check for this in four ways and find no supportive evidence. Second, I look at several potential indicators of limited attention. I find that participants with both message sets were less likely to read messages than those receiving just one. I find no evidence of interaction effects between message sets, differential opt-out behavior across message treatments, or heightened spillovers at the time of treatment. Finally, to look for evidence of moral licensing, I check whether participants, after being assigned to treatment, expected or hoped to do less of non-targeted behaviors. I do not find evidence of this. Overall, I find some supportive evidence of limited attention, and cannot rule out spontaneous or sub-conscious moral licensing. I discuss alternative explanations in Section VI.

This paper contributes to several strands of literature. The first is the small but growing literature on behavioral spillovers rooted in the costs of attention (Nafziger, 2020) and/or cognitive effort (Altmann et al., 2022). Motivated in part by the vast empirical research on the effectiveness of marketing-inspired tactics to capture attention and draw it to important economic choices (e.g. Taubinsky (2013); Allcott and Rogers (2014); Karlan et al. (2016); Rogers and Milkman (2016)), researchers have begun to document that such campaigns can also divert attention. Medina (2020) found that sending SMS reminders to bank clients effectively reduced late fees paid by 14%, but it also increases overdraft fees paid by 9%, resulting in a net loss for some. Hall and Madsen (2022) found that highway safety campaigns displaying roadside fatality counts was so effective at seizing attention that it actually increased the number of traffic crashes. Altmann et al. (2022) found that in the lab, choice-promoting interventions improved performance on a target task, but it came at the expense of performance on a background task. These studies fall within the broader, rich literature on the implications of limited attention and costly mental effort for economics.[5].

Second, this paper contributes to the more established literature on behavioral spillovers rooted in preferences. Dolan and Galizzi (2015) provide a useful framework, organizing spillovers by whether they are positive and "promote" other good behaviors, or whether they are negative and "permit" negligence.[6] Spillovers may be positive if good behavior provides a "positive signal" for one's future self (Bénabou and Tirole, 2004), building a sense of identity around the behavior (Galizzi and Whitmarsh, 2019). On the other hand, if people aim to simply maintain a positive self image, then engaging in something "good" can potentially license them to subsequently engage in something "bad," or vice versa (moral licensing). Empirically, some studies find interventions promoting good behaviors to have negative spillovers on other behaviors (Wertenbroch, 1998; Khan and Dhar, 2006; Tiefenbeck et al., 2013; Dolan and Galizzi, 2014) and even on the same behavior in a different context (Hussam and Oh, 2021). But other studies find positive spillovers (Jessoe et al., 2017; Ek and Miliute-Plepiene, 2018; Brandon et al., 2019).[7] A related literature looks at whether fundraising appeals for one charity have negative spillovers on donations to others (Donkers et al., 2017; Meer, 2017; Scharf et al., 2017; Deryugina and Marx, 2021).

Finally, this paper contributes to a growing literature that attempts to explore unanticipated consequences of nudges, both in general equilibrium (Spiegler, 2015) and for welfare more broadly. Several recent papers have taken more seriously the psychological costs of nudges that exploit social comparisons, shaming, or social pressure, in order to measure total welfare effects (DellaVigna et al., 2012; Jimenez-Gomez, 2018; Allcott and Kessler, 2019; Butera et al., 2019). Another recent literature has documented substantial opt-out behavior in text message programs, indicating that messages can be costly (Damgaard and Gravert, 2018; Fricke et al., 2018). Nudges have also been shown to exacerbate adverse selection in health insurance markets (Handel, 2013), impose hidden time costs (Taylor, 2020), and weaken support for costlier policies by offering a promise of a "quick fix" (Hagmann et al.,

[5] See for example Sims (2003); Chetty et al. (2009); Bordalo et al. (2012); Gabaix (2014); Taubinsky and Rees-Jones (2018); Farhi and Gabaix (2020); Bronchetti et al. (2020); Morrison and Taubinsky (2019)

[6] In the case of a bad initial behavior, spillovers can be similarly be negative and "precipitate" other bad behaviors, or positive and "purge" them.

[7] Galizzi and Whitmarsh (2019) provide a useful review of the empirical methods used to measure spillovers.

2019). Most of these studies do not dispute the benefits of nudges, but rather aim to expose their costs for the sake of a fuller picture.

The literature has found spillovers across diverse contexts, accompanied by a variety of explanations. But we have little sense of whether negative spillovers should be a general consideration in the design of behavioral interventions. The present study takes on this question in two ways. First, it assesses the costliness of spillovers in an environment that resembles behavior change interventions prevalent in the real world. The context is more natural than Altmann et al. (2022), but also more generic than existing natural experiments.[8] Second, I assess the importance of different mechanisms, exploiting experimental variation in conjunction with rich data on behavior, opt-out, the reading of messages, and expectations.

I begin in Section II by describing the experiment and presenting orthogonality tests. In Section III I report the main results on spillovers. In Section IV I discuss implications for cost-effectiveness analysis. In Section V I present evidence on mechanisms. In Section VI I discuss alternative explanations for the observed spillovers, and in Section VII I conclude.

II Experiment Design

II.A Design and Protocol

The experiment design is displayed in Table 1.[9] The control group received no intervention. Group 2 received only messages about behavior x, and Group 3 received only messages about behavior y. Group 4 received messages about behavior x as well as messages about behavior y. Group 5 received incentives for behavior y. Behaviors x and y were daily meditation and nutritional self-monitoring via meal logging, respectively. These behaviors were chosen for three reasons. First, they

[8] For example, in Medina (2020), the observed spillovers can be attributed to the budget constraint. The findings in Hall and Madsen (2022) are quite specific to billboards and traffic accidents.

[9] The study was pre-registered at the AEA RCT registry (Trachtman, 2019); see Appendix F for details.

are important health behaviors for the sample frame (young Americans).[10,11] Second, both behaviors can be measured objectively at high frequency via pre-existing smartphone applications.[12] Third, both behaviors require only small amounts of time. The average meditation session was 21 minutes, and meal logging took 11 minutes per day on average.

Participants were recruited using Facebook advertisements, which targeted adults age 18-35 living in the U.S. (see Appendix Figure A1). Upon clicking the link, participants underwent a brief screening that ensured they (1) had an iPhone or Android phone; (2) were over 18; (3) were interested in working on wellness habits like meditation and tracking nutrition; and (4) were willing to download the two free applications. They then provided informed consent and proceeded to Survey 1, which took about 15 minutes (see Appendix G.1). The first part of Survey 1 provided instructions for downloading the two apps (which were also emailed upon survey completion). Participants were instructed that in order to enroll, they would need to download both apps within 24 hours. Participants were then asked questions on demographics, electronic notifications, and preferences/experiences surrounding meditation and meal logging.

Participants who were verified to have downloaded both apps (3,884 of the 5,845 who filled out Survey 1, or 66%) were then randomized to treatments, re-randomizing on gender, age, whether or not they had a college degree, daily notifications, whether or not they meditated in the last month, and whether or not they tracked their meals in the last month. These participants then received an enrollment confirmation email with their treatment assignment, a link to Survey 2, and other details about the study (see Appendix G.2). Survey 2 required about five minutes and contained questions about participants' expectations of each behavior, conditional on their treatment assignment (see Appendix G.3).

[10] A recent meta-analysis in the Journal of the American Medical Association found that meditation programs improved anxiety by 0.38 SDs at 8 weeks (and 0.22 at 3-6 months), improved depression by 0.30 SDs at 8 weeks (and 0.23 at 3-6 months), and reduced pain by 0.33 SDs at 8 weeks (Goyal et al., 2014). The use of smartphone apps for nutritional self-monitoring and feedback have been linked to weight loss (Wharton et al., 2014), which is associated with many health benefits.

[11] In the screening process, I asked participants if they would feel comfortable using a nutrition tracking app, and asked those who have struggled with eating disorders or body image issues in the past to consider this carefully.

[12] The meditation application allowed participants to access a wide variety of guided meditations or meditate on their own, and recorded details about each meditation session including the time spent meditating. In the meal logging application, participants inputted information about the meals they ate and then tracked various measures of the nutritional quality of their diet.

Importantly, when informed of their treatment assignment, participants were told that ”this assignment was completely random, and has nothing to do with your survey responses or the relative importance of meditation, exercise, nutrition, and sleep.” The purpose of this was to rule out the possibility that participants infer relative benefits of the behaviors from their treatment assignment.[13] In order to avoid experimenter demand effects, I also tell participants that, ”Regardless of any programs you were or were not assigned above, your ultimate use of [meditation app] and [meal logging app] is entirely up to you.” Finally, and relatedly, I give the meditation and meal logging programs separate names (Remindful and eNOMerate, respectively), and send the messages from different phone numbers, in order to “unbundle” the two behaviors as much as possible. I discuss these concerns further in Section VI.

Each message program included twice-daily text messages: one simple reminder to do the behavior, and one longer message with information about some proven benefits to the behavior, as demonstrated in Appendix Table A1. As the table demonstrates, the timing of meditation vs. nutrition messages and information vs. reminder messages alternated in a balanced fashion. Messages were sent at either 7am and 7pm or at 8am and 8pm, alternating on a daily basis. They were scheduled so that the group that received both meditation and nutrition messages never received them at exactly the same time (it was always the case that one message was at 7 and the other at 8). There were 14 distinct messages, and 27 days of treatment, so each message (save one) was sent twice over the course of the program. The full set of messages is shown in Appendix Table A2. Participants were told during consent, in the enrollment email, and at the start of treatment that they could opt out any time by replying “STOP” to the relevant number.

Incentives took the form of a raffle. Participants were informed in their enrollment email that they would earn one green lottery ticket for every day they successfully do the behavior, and one red lottery ticket every day that they do not. We informed them that at the end of the four weeks, we would draw one of their tickets, and a green ticket would be worth a $10 Amazon gift certificate. Every Sunday during the program, participants received an email updating them about the tickets earned the previous week. They also received an email informing them when the program ended.

[13] This mechanism is potentially important, but cannot be well studied in this kind of experimental context, since many participants already assume assignment is random.

Four weeks after the end of treatment, participants received Survey 3 via email. Survey 3 included questions about meditation and meal logging outside of the assigned apps, the timing of behaviors, some measures of mental health and diet, and quizzes about the information content of any message program they received (see Appendix G.4). For further details about the experiment protocol see Appendix A.

Of the 3,884 who were initially enrolled in the study, 40 dropped out, and for an additional 27 data was lost due to a technical issue, resulting in a final analysis sample of 3,818. There is no evidence of differential attrition (see Appendix Table A3). In terms of compliance with messaging programs, of the 1,585 participants assigned to meditation messages, 13% opted out, and of the 1,625 assigned to meal logging messages, 18% opted out. I analyze this data in Section V.B.2.

II.B Data

I have four sources of data. The first is data on meditation and meal logging from the two apps. The meditation app records minutes meditated on the app, and the meal logging app provides data on whether at least one meal was logged as well as total calories logged.[14] For much of the analysis, I focus on binary outcomes—whether or not individuals meditated each day, and whether or not they logged at least one meal—because this is what the treatments promoted. Messages focused on daily behavior at the extensive margin, sometimes even downplaying the importance of intensity (i.e., "find a few minutes to meditate / log your meals"). Incentives were explicitly tied to the extensive margin, as described above.[15]

There are a few limitations to the data collected from the apps. First, with the meditation app, it is possible to record meditation without actually meditating (the person can start the timer and do something else). For this reason, I did not include incentives for meditation as a treatment. Still, it is possible that some part of the meditation message treatment effect is driven by participants induced to falsely report meditation as a result of messages. To the extent that this happened, I would be measuring both effort to meditate, and effort to falsify meditation, presumably motivated by experimenter demand effects. Second, neither the meditation nor meal logging measures capture activity done outside of the apps. Finally, the mea-

[14] Due to technical constraints, I do not have data on the number of meals or items logged or what foods were logged.

[15] Correspondingly, these binary indicators were the principal outcomes specified in the pre-analysis plan.

sure of calories logged is not an unbiased measure of meal logging, since it conflates meal logging and dietary changes.

The second source of data comes from the three surveys described above. The third source of data is opt-out data for those assigned to messaging treatments. The fourth source of data is participation in a surprise raffle that I conducted with those assigned to messaging treatments, described in Section V.B.2, which I use as a proxy for message reading.

Table 2 shows means and standard deviations of key variables across treatments, as well as an F-test of the joint significance all treatment variables. The re-randomization procedure ensured that the first variables were balanced across treatments, and the rest of the variables are highly balanced as well. Overall, the sample was overwhelmingly female (93%), mostly college educated (71%), with an average age of 27. Participants received on average 51 notifications daily, 36 of which were messages, 10 of which were updates, 4 of which were reminders, and 1 of which was classified as "other." (See Appendix Figure A2 for details.) Most participants had experience with both meditation and meal logging. With respect to meditation, 90% had meditated before, 57% had done so on a daily basis, and 46% had done so in the last month. With respect to meal logging, 90% had logged their meals before, 87% had done so on a daily basis, and 32% had done so in the last month. (See Appendix Figure A3 for details.) These are people who have strong prior interest and experience in both behaviors, but who, for whatever reason, have not been engaged in them recently.

The analysis sample is obviously particular and select, dominated by relatively young college educated women with prior experience and interest in these behaviors, but who have not been engaged with them recently. This was by design: I recruited people who were likely to be compliers in order to raise the likelihood of having strong target treatment effects. I discuss external validity concerns more generally in Section VI.D.

III Results on Spillovers

III.A Specification

For the main specification, I estimate linear probability models of the outcome on treatments at the individual-day level, where the outcome is 1 if the participant did the behavior on a given day. For meditation, the behavior is having meditated a positive amount. For meal logging, the behavior is having logged at least one meal and 0 otherwise. I define m_x (m_y) to be 1 if the individual received x (y) messages and 0 otherwise; $m_x m_y$ is 1 if the individual received both sets of messages. I use the below specifications (where i indexes the individual and t indexes the day):

$$x_{it} = \beta_0^x + \beta_1^x m_x + \beta_2^x m_y + \beta_3^x m_x m_y + \beta_4^x z_y + Z_i + \psi_t + \epsilon_{it} \tag{1}$$

$$y_{it} = \beta_0^y + \beta_1^y m_y + \beta_2^y m_x + \beta_3^y m_x m_y + \beta_4^y z_y + Z_i + \psi_t + \epsilon_{it} \tag{2}$$

$\hat{\beta}_2^x$ is an estimate of the spillover of m_y on x; $\hat{\beta}_2^y$ is an estimate of the spillover of m_x on y, and $\hat{\beta}_4^x$ is an estimate of the spillover of z_y on x. $\hat{\beta}_3^x$ and $\hat{\beta}_3^y$ are interaction effects that capture whether the two message sets interfered with one another. I include a vector of controls Z_i that consists of the variables used for re-randomization: whether or not the participant is female, whether or not they completed college, daily notifications, whether or not they meditated in the month prior to the study, and whether or not they logged a meal in the month prior to the study. I also include day fixed effects ψ_t, and cluster standard errors at the individual level. Coefficients on message treatments represent intent-to-treat effects, as some participants chose to stop receiving messages. In all tables, I report q-values calculated according to the Benjamini Hochberg step-down procedure, considering all tests in the table (but excluding linear combinations of coefficients).

III.B Treatment Period Effects

I begin by discussing binary outcomes during the treatment period. The raw data is shown in Figure 1 and the regression results are shown in columns 1 and 2 of Table 3. Recall that I report main effects of m_x, m_y, and z_y as well as an interaction term between m_x and m_y. In the middle panel of the table I report linear combinations that capture the difference between being assigned both messaging treatments versus just one, and at the bottom of the table I report control means.

All three treatments—m_x, m_y, and z_y—had large and significant target effects. The coefficient on m_x in column 1 shows that meditation messages raised the meditation rate from 9.4% to 18.2%, and that on m_y in column 2 shows that nutrition messages raised the meal logging rate from 11.8% to 28.4%. Incentives for meal logging had large target effects, more than quadrupling the rate of meal logging (from 11.8% to 50%) as evidenced by the coefficient on z_y in column 2.

All three treatments also produced negative spillovers on non-target behaviors. Participants getting only nutrition messages meditated 29% less than the control group (6.6% relative to 9.4%), as demonstrated by the coefficient on m_y in column 1. Participants getting only meditation messages logged meals 19% less than the control group (9.4% relative to 11.8%), as demonstrated by the coefficient on m_x in column 2. The group with both m_x and m_y did 2.2 percentage points worse on meditation than the group with just m_x, as captured by the linear combination $m_y + m_x m_y$. Similarly, the group with both m_x and m_y did 5.0 percentage points worse on meal logging than the group with just m_y, as captured by the linear combination $m_x + m_x m_y$. The incentive treatment z_y also had negative spillover effects, reducing meditation by 27% (6.9% relative to 9.4%) as evidenced by the coefficient on z_y in column 2.[16,17,18]

In Appendix Table A7, I estimate the same specification for two continuous measures of each behavior: minutes meditated, and daily calories logged.[19] The effects on continuous outcomes are generally consistent with those on binary outcomes. I find negative spillovers of m_y and z_y on meditation (though conditional on $m_x = 1$, the negative effect of m_y is not statistically significant). I also find negative spillovers of m_x on meal logging, both conditional on $m_y = 1$ and on $m_y = 0$. In Appendix Table A8 I show treatment effects on three additional health outcomes, including self-reported mental health, fraction of participants' weight goal achieved, and self-reported diet scores. I find no significant effects.[20]

[16] See Appendix Table A4 for estimates reported as mutually exclusive treatment effects.

[17] Because of the re-randomization procedure used, in Appendix Table A5 I report p-values computed using the bootstrap procedure from Bertsimas et al. (2015). I find no meaningful differences in statistical significance.

[18] In Survey 3 I asked participants how many days they did each behavior outside of the apps; in Appendix Table A6 I inflate meditation and meal logging rates accordingly and find similar results.

[19] Recall that the measure of daily calories logged confounds meal logging effort and dietary changes. However, given that I see minimal effects on diet or weight loss in Appendix Table A8, I view total calories logged as indicative mainly of meal logging effort.

[20] This is not necessarily inconsistent with prior research on meditation and meal logging, which typically evaluate much more intensive programs. For example, the meta-analysis of studies on meditation looked at TOT effects of meditation programs that lasted between eight weeks and six months; here I am estimating ITT effects of encouraging people to meditate over four weeks.

III.C Dynamics

The dynamics portrayed in Figure 1 illuminate several things. First, we see that all treatment groups had positive rates of both meditation and meal logging at the start of the study, which decayed substantially over time. The fact that in the first week of the treatment period, the control group meditated and logged their meals at rates of 14% and 24%, respectively, suggests that the initiation of the study was motivating in and of itself. But this motivation clearly wore off over time, at rates that do not seem to differ substantially by treatment. Computing treatment-level averages over participants by day, I find daily decay rates of 2.6% and 4.7% for meditation and meal logging, respectively, and no evidence that they vary across treatment groups. In other words, the initial excitement at the start of the study wore off over time, and neither messages nor incentives were strong enough to counteract it.

In columns 3 and 4 of Table 3, I show estimates of treatment effects from the post-treatment period. Since decay appears to be somewhat smooth over the full study period, without sharp drops at the end of treatment (with the exception of the z_y group), we should view the post-treatment period as capturing not only the end of treatment, but also the accumulated effects of decay. Control group means confirm the decline in engagement for both behaviors. On average, between the treatment and post-treatment periods, participation in meditation for the control group drops by 43% (from 9.4% to 5.4%), and meditation length drops by 29% (from 2.1 to 1.5 minutes). Both participation in meal logging (logging at least one meal a day) and the calories logged drop by about 72% (from 11.8% to 3.3%, and 153 to 41 calories, respectively).

Though treatment effects also fall in the post-treatment period, they do persist, for both target and non-target behaviors. The effects of messages on target behaviors are 28% and 18% the size of their treatment-period magnitude for meditation and meal logging, respectively. Target effects of incentives also persisted, at 10% the size of their treatment period effect. Interestingly, the negative spillover effects on meditation of meal logging messages and incentives persisted at almost 100% of their treatment period sizes.

We thus have three important facts to explain: the constant rate of decay across treatment groups; the persistence of both treatment effects and spillover effects after the end of treatment, and the differential persistence of spillovers. The first two facts are consistent with patterns of action, backsliding, and persistence that have been documented in a number of settings (e.g. Allcott and Rogers (2014); Agarwal et al. (2013)). A model of learning or habit formation, where attention or effort also wane over time, can explain these two facts.

In such a model, during the treatment period, two factors contribute to spillovers: (1) the contemporaneous effect of the target treatment on the non-target behavior, and (2) the cumulative effects of the differences in habit stocks that result from (1). In the post-treatment period, (2) can explain why spillovers persist. Namely, groups with non-target interventions would be expected to experience higher rates of decay of the target behavior than the control group, due to lower habit stocks. However, the magnitude of spillovers should be lower in the post-treatment period since (1) is no longer active.

Empirically, there is not sufficient power to test either prediction. I do not detect significant differences in rates of decay across treatment groups. Moreover, even though the spillovers generated by m_y and z_y appear to persist at their treatment period levels (-0.025 and -0.024, respectively), the 95% confidence interval includes spillovers as small as -0.01, so we certainly cannot rule out the possibility that spillover magnitudes fell.

What is clear empirically is that the post-treatment drop in *levels* of meditation in the m_x group was much more dramatic than that in the control or m_y groups. This is to be expected, given that the m_x group had dramatically higher treatment-period levels of meditation (slightly slower decay rates notwithstanding). This fact is sufficient to produce the differential persistence between treatment and spillover effects that we observe.

III.D Comparing Target and Spillover Effects

An immediate implication of the results in Table 3 is that treatments that generated larger target effects do not seem to have generated comparatively larger spillovers. m_x, m_y, and z_y all produced spillover effects with similar magnitudes despite having target effects with very different magnitudes. In Section IV, below, I will explore the policy implications of this. In Section V, I will explore what classes of mechanisms are consistent with this fact, and others. Finally, in Section VI, I will address the concern that this fact is a product of low rates of behavior in the control group.

IV Implications for Cost Effectiveness Analysis

The results have two implications for cost-effectiveness analysis. First, the existence of spillovers implies that interventions are costlier than they appear. Second, the fact that the magnitude of spillover effects appears to be independent of the magnitude of target effects suggests that high-impact interventions will not necessarily create larger spillovers than low-impact ones. More generally, there may be unappreciated differences in cost-effectiveness across interventions.

In this section, I will conduct a simple cost-effectiveness analysis to quantify these implications, exploiting the comparison between interventions m_y and z_y. Suppose a policymaker cares about the total cost-effectiveness of a set of interventions, taking multiple behaviors into account. Let the total effect of a set of interventions on multiple behaviors be simply the sum of the normalized effects on the set of behaviors. In our context, the total effect of a set of interventions will be equal to their net effect on meditation (in standard deviations), plus their net effect on meal logging (in standard deviations). I will consider two sets of interventions. Policy A consists of messages about both x and y, and costs $2p_m$, where p_m is the cost of one message set. Policy B consists of messages about x and incentives for y, and costs $p_m + p_z$, where p_z is the cost of incentives. I will assume that there are no interaction effects between interventions.[21]

Standardizing effects, I find that m_x, generated a target effect of 0.30 S.D.'s (0.088) and a spillover effect of -0.08 S.D.'s (-0.024); m_y generated a target effect of 0.52 S.D.'s (0.166) and a spillover effect of -0.09 S.D.'s (-0.028); and z_y generated a target effect of 1.19 S.D.'s (0.381) and a spillover effect of -0.08 S.D.'s (-0.025).[22]

In a hypothetical world without spillovers, Policy A would achieve a total effect of $0.3 + 0.52 = 0.82$ S.D.'s, and Policy B would achieve a total effect of $0.3 + 1.19 = 1.49$ S.D.'s. But with spillovers, Policy A achieves a total effect of $0.3 + 0.52 - 0.08 - 0.09 = 0.65$ S.D.'s, and Policy B achieves a total effect of $0.3 + 1.19 - 0.08 - 0.08 = 1.33$ S.D.'s. This means that under Policy A, which costs $2p_m$, the cost per standard deviation effect rises from $2p_m/0.82 = 1.22p_m$ to $2p_m/0.65 = 1.54p_m$ once spillovers are taken into account, or by 26%. In contrast, under Policy B, which costs p_z, the cost per standard deviation rises from $(p_m + p_z)/1.49 = 0.67(p_m + p_z)$

[21] This is indeed the case for m_x and m_y. I did not measure it for m_x and z_y but I believe it is a reasonable assumption, since z_y is unlikely to induce information overload.

[22] For meditation, the standard deviation in the control group was 0.29, and for meal logging, the standard deviation in the control group was 0.32.

to $(p_m + p_z)/1.33 = 0.75(p_m + p_z)$ once spillovers are taken into account, or by 12%. Taking spillovers into account thus has harsher cost-effectiveness implications for Policy A than it does for Policy B, which generates a large effect on y without a correspondingly large spillover effect on x.

In order to compare the cost-effectiveness of Policy A and Policy B, we can ask what p_m and p_z would need to be in order for the total cost-effectiveness of Policy A to equal that of Policy B. In the absence of spillovers, Policy A has cost per standard deviation of $2p_m/0.82$, and Policy B has cost per standard deviation of $(p_m + p_z)/1.49$, so they are equally cost-effective when $p_z = 3.6p_m$. This means that Policy B will be more cost-effective as long as $p_z < 2.6p_m$. In the presence of spillovers, Policy A has cost per standard deviation of $2p_m/0.65$, and Policy B has cost per standard deviation of $(p_m + p_z)/1.33$, so they are equally effective when $p_z = 3.1p_m$. This means that Policy B will be more cost-effective as long as $p_z < 3.1p_m$. In the experiment conducted here, incentives cost \$5.23 per capita, and messages cost \$1.89 per capita, so $p_z = 2.8p_m$. Failing to consider spillovers would thus have resulted in the choice of the less cost-effective policy. More generally, in the presence of spillovers, we see a shift in the relative cost-effectiveness of different interventions. Specifically, the higher-impact intervention was more cost-effective than the low-impact one, once multiple behaviors were considered. I consider the external validity of this implication in Section VI.D.

V Results on Mechanisms

In this section, I exploit additional data and analysis to examine potential mechanisms underlying spillovers. I focus on four classes of mechanisms. The first mechanism is *time and/or effort constraints*, or more generally, some kind of convexity in the costs of meditating and meal logging. This captures the possibility that individuals induced to do the target behavior as a result of the intervention had less time, or costlier effort, for doing the non-target behavior. Neither of the behaviors in this experiment required much physical effort, but may well have required cognitive effort, which has also been shown to be costly.[23]

[23] Botvinick and Braver (2015) found that performance on difficult tasks tends to increase with incentives, and Dunn et al. (2016) showed that when given a choice between tasks that require high and low cognitive effort, participants tend to prefer the latter. Typically, the more automatic the task, the less effort it requires (Shenhav et al., 2017), so engaging in new behaviors may well require cognitive effort.

The second mechanism is *limited attention*. One way limited attention might operate in this context is that non-target messages might cause individuals to pay less attention to target messages (I will call this "information overload"). The possibility of such overload is well-supported in psychology, marketing, and economics.[24,25,26] With respect to messaging, specifically, there is evidence that people become habituated or desensitized to alerts, designed to promote some behavior, over time.[27]

Another way that limited attention might operate in this context is that a stimulus promoting some target behavior might distract individuals from non-target behaviors (I will call this "diversion"). This is distinct from information overload in the sense that non-target messages do not reduce the effectiveness of target messages; rather, they reduce target behaviors directly. Studies on "attentional capture" show that irrelevant stimuli can easily draw people's attention away from a task at hand (Yantis and Jonides, 1984).[28,29] More generally, working memory is limited; there is only so much space at the "top of mind."[30]

The third and final mechanism is *moral licensing*, or more generally, some kind of substitutability in preferences. This captures the possibility that doing more of one good health behavior gave people "license" to do less of the other. Moral licensing has been well established in diverse contexts (Wertenbroch, 1998; Khan and Dhar, 2006; Tiefenbeck et al., 2013; Dolan and Galizzi, 2014).

[24] For example, people are unable to attend to two visual or two auditory streams (Neisser and Becklen, 1975), but better able to attend to one of each (Duncan et al., 1997).

[25] The information overload literature in marketing has documented a hump-shaped relationship between the quantity of information that consumers have about products, and the "decision quality" of their purchases (Hwang and Lin, 1999; Edmunds and Morris, 2000; Eppler and Mengis, 2004).

[26] In the economics arena, we have evidence that people have difficulty processing all of the information about the products they buy (Lacetera et al., 2012), all of the choices available to them (Chernev et al., 2015), and all of the dimensions of their production processes (Hanna et al., 2014).

[27] For example, in medicine, as the use of electronic medical records and attendant automatic alerts to provide "decision support" have become widespread, there has been extensive discussion of "alert fatigue," the idea that physicians become habituated to alerts over time. One SMS program designed to alert physicians to new clinical trials found that response rates declined 2.7% every two weeks (Embi and Leonard, 2012).

[28] The types of stimuli most likely to achieve attentional capture are novel stimuli (i.e. an unexpected SMS), emotionally salient stimuli (i.e. footage of a humanitarian crisis) and stimuli associated with rewards (i.e. a plate of cookies placed in front of you) (Chun et al., 2011; Fawcett et al., 2015).

[29] One relevant example is the phenomenon of "intention cost," or reduced performance (and brain activity) in a current task as a result of thinking ahead to a future task (Burgess et al., 2003; Gonen-Yaacovi and Burgess, 2012).

[30] For example, people are capable of holding only limited sets of digits or words at a time (Miller, 1956; Luck and Vogel, 1997).

V.A Time and/or Effort Constraints

If negative effects of target interventions on non-target behaviors are driven by time or energy constraints, then they must be induced by higher levels of the target behavior. I search for evidence of this in two main ways, detailed below, and conduct two additional exercises in the appendix.

V.A.1 Comparing Spillover to Target Effects

First, I compare spillover/target ratios across interventions m_y and z_y. If spillovers operate through the target action, we should expect interventions with large positive target effects to also have large negative spillovers, and interventions with small positive target effects to have small negative spillovers. We have already observed that the results in Table 3 are not consistent with this, and explored its implications for cost effectiveness analysis. Here, I run a more formal test of whether spillover/target effect ratios are constant across different interventions.

For the binary outcomes (meditating or logging at least one meal), I find that the z_y intervention has a spillover to target effect ratio of 0.07, and the m_y intervention has a spillover to target effect ratio of 0.17. I can reject that the ratios are equal (p=0.03). If I do the same exercise using continuous outcomes, I find a difference in spillover/target ratios across meal logging incentives and messages, but it is only marginally significant at the 10% level (p=0.097).[31]

V.A.1 Covariance Across Treatments

Second, I compare the covariance of meditation and meal logging across treatments. In the presence of time/effort constraints, doing x should negatively covary with doing y. With heterogeneous agents, however, the covariance between x and y also reflects any underlying covariance in preferences about meditation and meal logging. I can separate the two by comparing the covariance between x and y across treatments. Specifically, if spillovers are driven by time/effort constraints,

[31] This is because while meal logging incentives had more than double the effect of messages on the extensive margin of logging at least one meal, they were only 65% more effective than messages on the intensive margin of calories logged. This makes sense: the meal logging incentives were explicitly tied to logging at least one meal. Indeed we see in Appendix Table A7 that conditional on logging at least one meal per day, participants with messages (and even control participants) record more calories on average then participants with incentives.

then treatments that induce higher levels of the target behavior induce lower levels of the non-target behavior, pushing the covariance downward. If spillovers are not driven by time/effort constraints, then treatments induce lower levels of the non-target behavior independently of increases in the target behavior, leaving the covariance unchanged. I look at both the covariance of x and y across individuals as well as within individuals over time. (Meditating might make meditators less likely to log their meals, but it might also make someone who meditates on one day less likely to log meals on the same day.)

Table 4 shows the effects of treatments on the covariance between x and y. Specifically, column 1 shows treatment effects on the covariance over individuals (within days), and column 2 shows treatment effects on the covariance over days (within individuals). If time/effort constraints fully drive spillovers, we should expect both sets of messages, as well as incentives, to have negative effects on the covariance. The data show no evidence of this.

In the appendix, I report the results from two additional exercises. First, I use an instrumental variables approach to estimate the effect of each behavior on the other, and report the results in Table A9. I do not find evidence of negative effects, though the confidence intervals are large. Second, I run the original specification of binary target outcomes on treatments, but adding controls for non-target behaviors as well as interactions between non-target behaviors and treatments. I report the results in Appendix Table A10. I do not find that spillovers are mediated by target behaviors: all of the spillovers documented in Table 3 remain strong and significant; if anything, they are stronger.

V.B Limited Attention

Information overload could help explain negative spillovers if those who received messages about just one behavior were better able to pay more attention to them, relative to those who received messages about two behaviors.[32] Diversion could also explain negative spillovers, if those who received messages about a target behavior were distracted from the non-target behavior.

[32] In theory, overload could also drive negative effects of non-target messages conditional on *not* receiving target messages, if individuals set up their own notifications for a given target behavior. In fact, this is unlikely, since only 4% of participants reported that they set up their own notifications for meditation, and similarly for meal logging.

V.B.2 Interaction Effects

An obvious indicator of overload would be a negative interaction between m_x and m_y. This would imply that target messages reduce the effectiveness of non-target messages. Table III shows no evidence of negative interaction effects between m_x and m_y—neither for meditation nor for meal logging.

Even if information overload did not generate detectable spillovers at the level of behavior, it may have generated spillovers on intermediate outcomes such as opt-out and message reading, which I explore below.

V.B.2 Opt-Out Behavior

If participants receiving two sets of messages experienced information overload or diversion, and were conscious of it, they might have decided to opt-out of one of the message sets. To opt-out, participants simply had to reply ”STOP” to the same number from which they were receiving messages. They were informed of this in the consent form, by email upon treatment assignment, as well as in the first text message they received.

Table 5 examines whether participants receiving meditation messages ever opted out (column 1), and whether participants receiving nutrition messages ever opted out (column 2), where the omitted group has one set of messages, and the treatment group has both. Because the analysis in each column is restricted to those who received a given set of messages (m_x in column 1, m_y in column 2), I estimate only one effect: that of having the additional set of messages. Because I do not have measures of opt-out for control participants, I cannot estimate the interaction term (I can only estimate the equivalent of the linear combinations reported in Table 3, $m_x + m_x m_y$ and $m_y + m_x m_y$).

Baseline levels of opt-out were relatively high, at 14.5% for the meditation-only group and 16.8% for the nutrition-only group. There is no evidence that participants receiving m_x and m_y were more likely than those receiving m_x to opt out of the meditation messages. Participants receiving m_x and m_y were somewhat more likely to opt-out of the nutrition messages than the group receiving just m_y, but the difference is not quite significant ($p = 0.11$).

V.B.2 Reading and Recall of Messages

If participants receiving two sets of messages experienced information overload or diversion, they might have been less likely to read and/or attend to messages. To understand the extent to which participants actually read and recalled the content of messages, I use two additional measures.

The first measure is whether or not participants responded to a surprise raffle sent to participants through the messaging program(s) they were assigned.[33] The message said: "Hi from Remindful/eNOMerate! We are offering a surprise raffle for a USD 20 Amazon gift card. To enter, tap [link] and press send. Msg & dta rates may apply." Each message participant received a maximum of one raffle message over the course of the experiment. Roughly half of message participants received the raffle on day 10 or 11 (halfway through the second week) and the other half received the raffle on day 20 or 21 (at the end of the third week). Participants receiving both messages were randomly assigned to receive either the eNOMerate raffle or the Remindful raffle.

Table 6, displays the results in columns 1 and 3. Again, because the analysis in each column is restricted to those who received a given set of messages (m_x in column 1, m_y in column 3), I estimate only the effect of having the additional set of messages. I find that raffle response rates were low even in the groups with one set of messages, at 31% and 26% for meditation and meal logging raffles, respectively (but higher conditional on not opting-out, at 36% and 31% respectively). The group with both sets of messages was about 30% less likely to respond to both raffles, suggesting that they were reading messages at a lower rate.

The second indicator is knowledge about meditation and nutrition, as measured by the percentage of questions answered correctly on a quiz administered at the end of the study, one month after the end of treatment. The quiz consisted of true/false questions on information provided in the messages, with additional options to answer "I remember seeing this message but I do not remember the details" or "I do not remember seeing this message." Participants received 1 point for every correct answer, 0 point for every incorrect or "I do not remember seeing this message" answer, and 0.5 points for answering "I remember seeing this message but I do not remember the details." They were unaware of this scoring system,

[33]Due to an implementation error, we are missing this data from 592 participants with messages. These participants accidentally received messages with a broken link, and so we do not know whether they responded or not.

and had no explicit incentives to perform well. Both raffle response rates and quiz scores are restricted to participants with messages (I do not quiz groups on information they did not receive), so the omitted group has one set of messages, and the treatment group will have both.

The group with both sets of messages may have done slightly worse on the knowledge quizzes, as demonstrated in columns 2 and 4 of Table 6, but these effects are small and not significant. The coefficient of interest does not change substantially when I restrict the sample to those who did not opt out, suggesting that the bulk of the effect is driven by participants who continued to receive messages throughout the treatment period.

The fact that participants with both sets of messages are less likely to read, say, messages about x, has two potential implications. First, lower reading rates of message x may be the reason why participants engage less in behavior x, so that message spillovers are in part driven by reading spillovers. This is consistent with an overload story, in which y messages make x messages less effective. Reading spillovers, however, can only explain differences between the group that got both messages sets and the groups that got one. It cannot explain differences between the group that got one message set and the control group, or differences between the group that got y incentives and the control group. A second potential implication is that lower reading rates of message x might *indicate* that participants were already paying less attention to behavior x (and doing less x). In other words, the data on message reading are also potentially consistent with a diversion story, in which treatments targeting y divert attention away from x.

V.B.2 Treatment Timing

Finally, if participants experienced diversion, so that interventions targeting one behavior diverted limited attention away from the other, we might expect this to occur not only in the cross-section but over time, as well. To check this, I exploit two variations in the timing of treatments.

First, for all messaging groups, messages were sent in the morning, at 7am or 8am, and in the evening, at 7pm or 8pm. I can look at data on the timing of meditation to check whether the m_y group shows a dip in meditation rates at these times, relative to the control group.[34] In Appendix Figure A4 I plot the likelihood

[34]Unfortunately I do not have data on the timing of meal logging.

of meditating in each hour of the day, by treatment group. We do see that meditation rates for the m_x group jump when messages were sent, relative to the control group, indicating that target messages operate in part through attentional capture. However, meditation rates for the m_y group do not show a significant dip at the times that messages were sent, relative to the control group.

Second, for the group that received meal logging incentives, emails were sent on Sundays providing feedback on the previous week. I can look at meal logging rates and meditation rates over the week to check whether the z_y group shows a dip in meditation on Sundays or Mondays, relative to the control group. In Appendix Figure A5a I plot the rate of meal logging over calendar dates, and in Appendix Figure A5b I plot the rate of meditation over calendar dates. Again, we do see that meal logging rates for the z_y group jump on Sundays and Mondays, suggesting that even for the incentive treatment, attentional capture played a role. However, again, meditation rates for the z_y group do not dip significantly on Sundays and Mondays relative to the control group.

V.C Moral Licensing

If moral licensing drove spillovers, we might expect to see decreases in the non-target behavior being driven by increases in the target behavior, as with time and effort constraints. However, we can also imagine a kind of "pseudo" moral licensing in which people feel license to not engage in one good health behavior because they *intend* to do another good health behavior, but ultimately do not. One way to check for any type of deliberate moral licensing would be to check whether participants who know they will receive some target intervention expect, or hope, to do the non-target behavior any less.

During Survey 2, immediately after treatment assignment, but before treatment began, I collected data on the expectations of participants. Table 7 shows parameter estimates from regressions of individual-level expectations about rates of meditation (column 1) and meal logging (column 3) on treatment assignment. Columns 2 and 4 show the differences between expected and actual rates of meditation and meal logging, respectively.

Participants in the control group grossly over-predict their rates of both behaviors: they expect to meditate 38% of the time when actually they do so 9.4% of the time, and they expect to log their meals 52% of the time when they actually do so 11.8% of the time. Participants who receive only meditation messages predict their meditation rates to be even higher than the control group, but the level of over-prediction is similar, resulting in an "expected target effect" (11%) that is actually close to the true target effect (8.8%). Participants who receive only nutrition messages also predict their meal-logging rates to be higher than the control group, but here they over-predict less, resulting in an "expected target effect" (12%) that is significantly lower than the true target effect (16.6%).

"Expected spillover effects" of nutrition messages and incentives on meditation are small and positive (though neither is significant). Participants thus significantly over-estimate spillover effects by 5 percentage points in both cases (i.e. they under-estimate negative spillovers by 5 percentage points). In the case of the spillover generated by meditation messages on meal logging, we cannot conclude that participants correctly predicted the negative spillover (the expected effect is about 70% of the true spillover, at -0.017, but the standard error is high), but there is no evidence that they under-estimated spillovers either. In Panel B of Appendix Table A11, columns 1 and 2, I additionally show the effects of treatment assignment on self-reported hopes, and find similar results.

V.D Summary

What have we learned from the above evidence? First, reduced levels of the non-target behavior do not seem to be mediated by higher levels of the target behavior, suggesting that time/effort constraints are unlikely to explain the entirety of spillovers. Second, participants with both message sets were less likely to read messages than those receiving just one. These reading spillovers may help explain message spillovers, but they may also indicate that receiving a message diverts attention and message-reading-effort away from non-target behaviors. Moreover, target behaviors are responsive in time to treatments, suggesting that treatments operate in part by capturing attention. That said, there was no evidence of interaction effects, differential opt-out behavior, or responsiveness of non-target behaviors to treatment timing. Finally, there is no evidence that participants anticipating treatment for a target behavior felt moral license to expect or hope to do the non-target behavior any less. However, this does not rule out spontaneous or subconscious "pseudo" moral licensing.

VI Alternative Explanations

VI.A Is the measurement of spillovers constrained by low rates of behavior in the control group?

Naturally, the spillovers that we estimate are bounded by the levels of participation in the control group. This gives rise to two concerns. The first is that spillovers may be under-estimated (in absolute value). The second is that spillover sizes may be fixed across different intervention types because of these bounds. Then, the policy implication described in Section IV would be specific to a context with low baseline rates of behavior.

I address this by dividing the data into various sub-groups with different baseline rates of behavior. I focus on spillovers on meditation so that I can check whether they continue to be fixed across interventions generating different target effects (m_y and z_y). First, I compare spillovers in the first half of the treatment period, when control rates were higher (meditation at 11.6% and meal logging at 37.6%), to the second half of the treatment period, when control rates were low (meditation at 7.2% and meal logging at 7.0%). I show the results in Appendix Table A12. All three spillovers continue to hover around 2.5 percentage points, both early and late in the treatment period. There are no statistical differences between the two time periods for any of the spillovers. Moreover, the difference in the magnitude of the spillover between m_y and z_y is minimal in both time periods.

Second, I predict meditation rates using variables collected in the baseline survey, and then compare spillovers among high predicted meditators with high control rates (39%), to low predicted meditators with low control rates (1.7%). I show the results in Appendix Table A13. Unsurprisingly, spillovers generated by m_y and z_y do collapse when the control rate of meditation is just 1.7% among low predicted meditators. Moreover, spillovers appear to be larger in magnitude among high predicted meditators: 6.4 percentage points for m_y and 4.1 percentage points for z_y (the latter effect is not significant, but this test is under-powered). The two are not statistically different, and if anything, the z_y spillover is smaller in magnitude than the m_y spillover, despite its target effect being nearly double.

Finally, I revisit estimates of spillovers on meditation conditional on receiving m_x versus conditional on not receiving m_x. Recall that the group that received only meditation messages meditated at a rate of 18.2%, and the group that received meal logging messages in addition meditated only 2.2 percentage points less. Unfortunately we cannot do this exercise for meal logging incentives. We can, however, do the same exercise for spillovers on meal logging. The group that received only meal logging messages logged their meals at a rate of 28.4%, and the group that received meditation messages in addition logged their meals 5.0 percentage points less. The spillover indeed appears to be a bit larger than that conditional on $m_y = 0$ (2.4 percentage points), but the difference is not significant, as indicated by the interaction term.

In summary, the fact that spillover estimates are larger among high predicted meditators means it is possible that the primary spillover estimates are be bounded above by low control-group rates. However, low control-group rates are unlikely to explain why spillover magnitudes do not appear to vary with the intervention's target effect.

VI.B Can experimenter demand account for spillovers?

An important concern is that spillovers might be driven by a "negative" experimenter demand effect. If participants were not explicitly encouraged to do a behavior, they might feel as if they were "not supposed" to do it. For example, someone who was randomly assigned to receive messages about meditation might have believed that they were not supposed to log their meals. This was indeed a concern: in piloting, several participants expressed confusion about what they were "supposed" to do. For this reason, when I rolled out the experiment (and as described in Section II), I added to the enrollment confirmation email, "Regardless of your treatment assignment, your ultimate use of the apps is entirely up to you" in attempt to make it clear that participants were not "supposed" to do anything in particular for the purpose of the experiment or experimenter.

If experimenter demand effects caused people to feel they were not "supposed" to the non-target behavior, then this should be reflected in their self-reported expectations and hopes after treatment assignment. We already saw in Tables 7 and A11 that there is little evidence that participants expected or hoped to do non-targeted behaviors any less. Indeed, participants assigned to meal logging messages actually "hoped" to meditate more than the control group.

VI.C Can the bundling of behaviors account for spillovers?

A related question is whether the bundling of behaviors encouraged a trade-off mindset or mental account that led to spillovers.[35] Indeed, if people were substituting away from *every* other activity, we would be unlikely to observe such large spillovers.

There are three possibilities to consider. The first is that meditation and meal logging were bundled together, and separately from other health behaviors, because they were targeted and measured in the experiment. The second possibility is that meditation, meal logging, sleep, and exercise were all bundled together because they were all associated with the experiment. Finally, it is of course also possible that health behaviors are naturally bundled together by individuals, independently of the intervention context, and that this bundling was an essential precondition for spillovers. I will focus on the first possibility, since the others would imply that spillovers are specific either to a context where multiple behaviors are promoted by the same entity—a context that is not so uncommon—or to the health context more generally. The results would be only slightly less generalizable.

The experiment design attempts to address the first possibility in two ways. First, in the baseline survey, I present the experiment as a study on "wellness behaviors," and every time I ask about meditation or meal logging, I also ask about sleep and exercise. Moreover, at the end of the baseline survey when the assignment of treatments is explained, participants are told that, "we'll randomly assign you to one or several (or none) of our messaging or incentive programs for one or several wellness behaviors," so participants have no reason to believe that treatments are limited to meditation and meal logging (as they ultimately are). Thus, the only sense in which meditation and meal logging were "bundled" together separately from exercise and sleep was through the association with required phone apps. Second, as discussed in Section II, I give the meditation and meal logging programs separate names and send messages from separate numbers, to keep them distinct for those assigned both m_x and m_y.

The best way to check whether these measures worked would have been to collect high frequency data on other health behaviors without associating them with

[35]This could be consistent with (real or pseudo) moral licensing as a mechanism, but it could also be consistent with diversion, if when attention is drawn to a particular behavior, it is withdrawn from a behavior within the same mental account.

the experiment, which was not possible in this context.[36] Another way to check for bundling is to look at whether participants' hopes for sleep and exercise look similar to those for meditation and meal logging at the outset of the experiment. In Table A11, columns 3 and 4, I look additionally at hopes regarding exercise and sleep, which were also elicited in Survey 2. I can compare spillover effects on hope between each of the four behaviors, amounting to six pairwise comparisons. I find just one significant negative difference: those assigned to m_x are less hopeful to meal log than they are to exercise. In summary, treatments did not generate strong negative spillovers on hopes to meditate and meal log, and hopes to exercise and sleep seem to be no different.

VI.D Are the results specific to this context?

In Sections VI.B and VI.C I have argued that the results are not likely to be tied to experimenter demand effects or the artificial bundling of meal logging and meditation. More generally, are the results externally valid to other behaviors, interventions, and populations?

I did not collect data on behaviors other than meditation and meal logging, and did not test interventions other than messages and incentives. Moreover, the recruitment procedure targeted individuals interested in both behaviors, so that the sample is far from representative. However, the context is relevant: messages and incentives promoting small-scale health-related behavior change are implemented with willing individuals, at scale, by increasing numbers of corporations and governments across the globe. The context is also generic in the sense that it is not obviously conducive to spillovers. I thus interpret the main results as evidence that negative spillovers can exist in a relevant and generic context, and can have important implications for cost-effectiveness. Since I did not measure other behaviors, I cannot make claims about net spillovers or cost-effectiveness in general.

VII Conclusion

A growing literature has documented a wide variety of interventions that shift behavior and generate meaningful economic impacts. This paper has shown that in the health context, two standard interventions—messages and incentives—can

[36] Unfortunately, the endline survey also did not ask about directly about sleep and exercise.

have negative effects on non-targeted behaviors. The results suggest that in general, designers of interventions should be attentive to, and potentially measuring, non-target behaviors. An important agenda for future research is to identify what kinds of behaviors, in what kinds of contexts, are susceptible to spillovers.

References

Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., Mullany, E. C., Abate, K. H., Abbafati, C., Abebe, Z., et al. (2019). Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*.

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2013). Learning in the credit card market. Working paper, National Bureau of Economic Research.

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.

Allcott, H., Lockwood, B. B., and Taubinsky, D. (2019). Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics*, 134(3):1557–1626.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–37.

Altmann, S., Grunewald, A., and Radbruch, J. (2022). Interventions and cognitive spillovers. *Review of Economic Studies*, 89(5):2293–2328.

Bénabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy*, 112(4):848–886.

Bertsimas, D., Johnson, M., and Kallus, N. (2015). The power of optimization over randomization in designing experiments involving small samples. *Operations Research*, 63(4):868–876.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Skimmyhorn, W. L. (2021). Borrowing to save? The impact of automatic enrollment on debt. Working paper, SSRN.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, 127(3):1243–1285.

Botvinick, M. and Braver, T. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology*, 66:83–113.

Brandon, A., List, J. A., Metcalfe, R. D., Price, M. K., and Rundhammer, F. (2019). Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proceedings of the National Academy of Sciences*, 116(12):5293–5298.

Bronchetti, E. T., Kessler, J. B., Magenheim, E. B., Taubinsky, D., and Zwick, E. (2020). Is Attention Produced Rationally? Working paper, National Bureau of Economic Research.

Burgess, P. W., Scott, S. K., and Frith, C. D. (2003). The role of the rostral frontal cortex (area 10) in prospective memory: A lateral versus medial dissociation. *Neuropsychologia*, 41(8):906–918.

Butera, L., Metcalfe, R., Morrison, W., and Taubinsky, D. (2019). Measuring the welfare effects of shame and pride. *American Economic Review*, Forthcoming.

Chernev, A., Böckenholt, U., and Goodman, J. (2015). Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25(2):333–358.

Chetty, R., Friedman, J. N., Leth-Petersen, S., Nielsen, T. H., and Olsen, T. (2014). Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark. *The Quarterly Journal of Economics*, 129(3):1141–1219.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

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

Damgaard, M. T. and Gravert, C. (2018). The hidden costs of nudging: Experimental evidence from reminders in fundraising. *Journal of Public Economics*, 157:15–26.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, 127(1):1–56.

Deryugina, T. and Marx, B. M. (2021). Is the Supply of Charitable Donations Fixed? Evidence from Deadly Tornadoes. *American Economic Review: Insights*, 3(3):383–98.

Dolan, P. and Galizzi, M. M. (2014). Because I'm worth it: a lab-field experiment on the spillover effects of incentives in health.

Dolan, P. and Galizzi, M. M. (2015). Like ripples on a pond: behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47:1–16.

Donkers, B., van Diepen, M., and Franses, P. H. (2017). Do charities get more when they ask more often? Evidence from a unique field experiment. *Journal of behavioral and experimental economics*, 66:58–65.

Dubois, P., Griffith, R., and O'Connell, M. (2020). How well targeted are soda taxes? Technical Report 11.

Duncan, J., Martens, S., and Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, 387(6635):808.

Dunn, T. L., Lutes, D. J., and Risko, E. F. (2016). Metacognitive evaluation in the avoidance of demand. *Journal of Experimental Psychology: Human Perception and Performance*, 42(9):1372.

Edmunds, A. and Morris, A. (2000). The problem of information overload in business organisations: A review of the literature. *International Journal of Information Management*, 20(1):17–28.

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., and Walker, M. W. (2019). General equilibrium effects of cash transfers: Experimental evidence from Kenya. Working paper, National Bureau of Economic Research.

Ek, C. and Miliute-Plepiene, J. (2018). Behavioral spillovers from food-waste collection in Swedish municipalities. *Journal of Environmental Economics and Management*, 89:168–186.

Embi, P. J. and Leonard, A. C. (2012). Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study. *Journal of the American Medical Informatics Association*, 19(e1):e145–e148.

Eppler, M. J. and Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, 20(5):325–344.

Farhi, E. and Gabaix, X. (2020). Optimal taxation with behavioral agents. *American Economic Review*, 110(1):298–336.

Fawcett, J., Kingstone, A., and Risko, E. (2015). *The Handbook of Attention*. MIT Press.

Fricke, H., Kalogrides, D., and Loeb, S. (2018). It's too annoying: Who drops out of educational text messaging programs and why. *Economics letters*, 173:39–43.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710.

Galizzi, M. M. and Whitmarsh, L. (2019). How to measure behavioral spillovers: A methodological review and checklist. *Frontiers in Psychology*, 10:342.

Gonen-Yaacovi, G. and Burgess, P. (2012). Prospective memory: The future for future intentions. *Psychologica Belgica*, 52(2-3).

Goyal, M., Singh, S., Sibinga, E. M., Gould, N. F., Rowland-Seymour, A., Sharma, R., Berger, Z., Sleicher, D., Maron, D. D., Shihab, H. M., et al. (2014). Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. *JAMA Internal Medicine*, 174(3):357–368.

Hagmann, D., Ho, E. H., and Loewenstein, G. (2019). Nudging out support for a carbon tax. *Nature Climate Change*, 9(6):484–489.

Hall, J. D. and Madsen, J. (2022). Can behavioral interventions be too salient? Evidence from traffic safety messages. *Science*, 376(6591):eabm3427.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–82.

Hanna, R., Duflo, E., and Greenstone, M. (2016). Up in smoke: the influence of household behavior on the long-run impact of improved cooking stoves. *American Economic Journal: Economic Policy*, 8(1):80–114.

Hanna, R., Mullainathan, S., and Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics*, 129(3):1311–1353.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Hussam, R. and Oh, D. (2021). Behavioral Transmission: Evidence from a Public Health Campaign in Bangladesh. Working paper.

Hwang, M. I. and Lin, J. W. (1999). Information dimension, information overload and decision quality. *Journal of Information Science*, 25(3):213–218.

Institute for Health Metrics and Evaluation (2018). Findings from the Global Burden of Disease Study 2017. Technical report, Seattle, Washington.

Jessoe, K., Lade, G., Loge, F., and Spang, E. (2017). Spillovers from behavioral interventions: Experimental evidence from water and energy use.

Jimenez-Gomez, D. (2018). Nudging and phishing: A theory of behavioral welfare economics. *Available at SSRN 3248503*.

Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016). Getting to the top of mind: How reminders increase saving. *Management Science*, 62(12):3393–3411.

Khan, U. and Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research*, 43(2):259–266.

Lacetera, N., Pope, D. G., and Sydnor, J. R. (2012). Heuristic thinking and limited attention in the car market. *American Economic Review*, 102(5):2206–36.

Luck, S. J. and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279.

Medina, P. C. (2020). Side Effects of Nudging: Evidence from a Randomized Intervention in the Credit Card Market. *The Review of Financial Studies*, 34(5):2580–2607.

Meer, J. (2017). Does fundraising create new giving? *Journal of Public Economics*, 145:82–93.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., and Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, 108(26):10415–10420.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81.

Morrison, W. and Taubinsky, D. (2019). Rules of thumb and attention elasticities: Evidence from under- and over-reaction to taxes. *Review of Economics and Statistics*, Forthcoming.

Muralidharan, K., Niehaus, P., and Sukhtankar, S. (2017). General equilibrium effects of (improving) public employment programs: Experimental evidence from India. Working paper, National Bureau of Economic Research.

Nafziger, J. (2020). Spillover effects of nudges. *Economics Letters*, 190:109086.

Neisser, U. and Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive psychology*, 7(4):480–494.

O'Donoghue, T. and Rabin, M. (2006). Optimal sin taxes. *Journal of Public Economics*, 90(10-11):1825–1849.

Persson, P. (2018). Attention manipulation and information overload. *Behavioural Public Policy*, 2(1):78–106.

Rogers, T. and Milkman, K. L. (2016). Reminders Through Association. *Psychological Science*, 27(7):973–986.

Scharf, K. A., Smith, S., and Wilhelm, M. (2017). Lift and shift: the effect of fundraising interventions in charity space and time. Technical report.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., and Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40:99–124.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690.

Spiegler, R. (2015). On the equilibrium effects of nudging. *The Journal of Legal Studies*, 44(2):389–416.

Taubinsky, D. (2013). From intentions to actions: A model and experimental evidence of inattentive choice. Working paper.

Taubinsky, D. and Rees-Jones, A. (2018). Attention variation and welfare: theory and evidence from a tax salience experiment. *The Review of Economic Studies*, 85(4):2462–2496.

Taylor, R. L. (2020). A mixed bag: The hidden time costs of regulating consumer behavior. *Journal of the Association of Environmental and Resource Economists*, 7(2):345–378.

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Tiefenbeck, V., Staake, T., Roth, K., and Sachs, O. (2013). For better or for worse? empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57:160–171.

Trachtman, H. (2019). Nudges in Equilibrium. *AEA RCT Registry*. https://doi.org/10.1257/rct.2435.

Trachtman, H. (2022). Replication data for: Does promoting one healthy behavior detract from others? Evidence from a field experiment. *American Economic Association, Inter-University Consortium for Political and Social Research*. http://doi.org/10.3886/E181963V1.

Wertenbroch, K. (1998). Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Science*, 17(4):317–337.

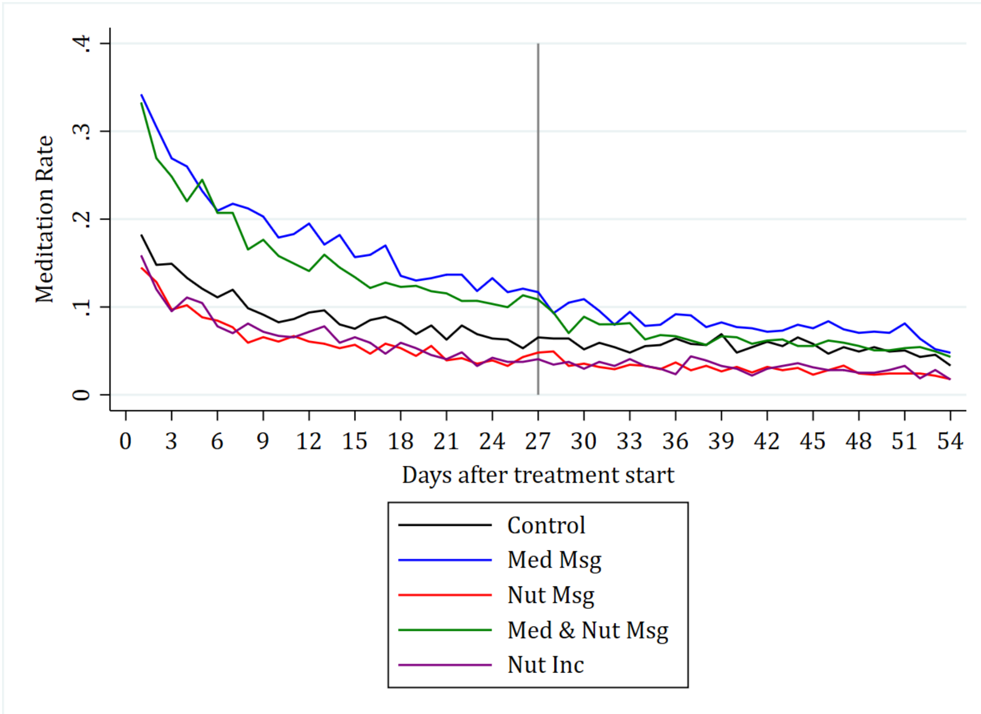
Wharton, C. M., Johnston, C. S., Cunningham, B. K., and Sterner, D. (2014). Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. *Journal of Nutrition Education and Behavior*, 46(5):440–444.

Yantis, S. and Jonides, J. (1984). Abrupt visual onsets and selective attention: evidence from visual search. *Journal of Experimental Psychology: Human perception and performance*, 10(5):601.

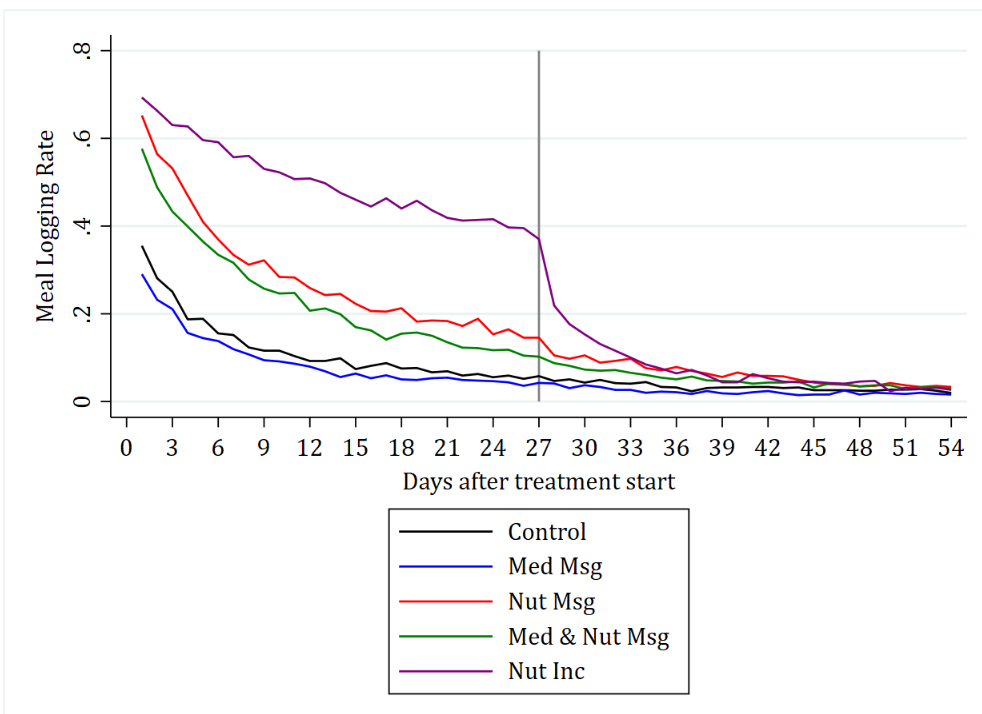
Zhou, F., Shefer, A., Wenger, J., Messonnier, M., Wang, L. Y., Lopez, A., Moore, M., Murphy, T. V., Cortese, M., and Rodewald, L. (2014). Economic evaluation of the routine childhood immunization program in the United States, 2009. *Pediatrics*, 133(4):577–585.

Figure 1: Rates of Meditation and Meal Logging by Treatment

(a) Meditation Rate by Treatment



(b) Meal Logging Rate by Treatment



Notes: Rates of meditation (a) and meal-logging (b) by treatment group, over the treatment period (days 0-27) and post-treatment period (days 28 onward). The outcome is 1 if the participant did the behavior on the given day (meditated, logged at least one meal) and 0 otherwise. This data includes the full analysis sample of 3,818 participants.

Table 1: Experiment Design

Group	Description	Messages $m_x = 1$	$m_y = 1$	Incentives $z_y = 1$	N
1	ctrl				811
2	x messages	x			754
3	y messages		x		794
4	x & y messages	x	x		817
5	y incentives			x	642
					3818

Notes: The experiment included five treatment groups: a control group, a group that received only messages about meditation (x), a group that received only messages about meal logging (y), a group that received messages about both meditation (x) and meal logging (y), and a group that received incentives to log meals (y). This table shows the sample used in all analyses. This sample excludes 67 participants who were randomized but for whom we do not have data, including 40 participants who dropped out and 27 for whom we lost data due to technical issues. See Appendix Table A3 for details.

Table 2: Orthogonality Check

	control	mx	my	mx & my	zy	F-test, joint sig
female†	0.93	0.93	0.93	0.93	0.93	0.99
	0.25	0.25	0.25	0.25	0.26	
went to college†	0.71	0.72	0.71	0.71	0.72	0.99
	0.45	0.45	0.45	0.45	0.45	
age	27.47	27.37	27.21	27.74	27.11	0.19
	5.72	5.90	5.23	5.48	4.91	
daily notifications†	52.42	53.35	52.83	54.38	53.44	0.99
	69.90	84.01	78.91	74.43	70.27	
meditated daily, ever	0.56	0.59	0.56	0.57	0.57	0.83
	0.50	0.49	0.50	0.50	0.49	
meditated daily, last month†	0.46	0.46	0.46	0.47	0.46	0.99
	0.50	0.50	0.50	0.50	0.50	
logged meals, ever	0.88	0.86	0.86	0.87	0.89	0.48
	0.33	0.34	0.35	0.34	0.31	
logged meals, last month†	0.33	0.32	0.33	0.33	0.33	0.99
	0.47	0.47	0.47	0.47	0.47	
importance, x - y	-0.45	-0.49	-0.45	-0.44	-0.48	1.00
	3.36	3.31	3.39	3.28	3.26	
difficulty - fun, x - y	-2.47	-2.80	-2.59	-2.56	-2.54	0.78
	4.93	5.06	5.34	5.16	5.10	

Notes: Means and standard deviations of ten variables measured in the baseline survey. † indicates variables used in the re-randomization procedure. p-value from F-test of the joint significance of treatments is reported in last column. *Daily notifications* includes all notifications the participant receives across all devices and all applications, where a notification is defined as anything that generates an alert (including SMS and email). *Importance*, $x - y$ is the "importance" of meditation, self-reported on a scale from 1 to 10, minus that of meal logging. *Difficulty*, $x - y$ is the "difficulty" of meditation, self-reported on a scale from 1 to 10, minus that of meal logging. *Fun*, $x - y$ is the "fun" of meditation, self-reported on a scale from 1 to 10, minus that of meal logging. I report only the difference between difficulty and fun.

Table 3: Treatment Effects on Binary Outcomes

	Treatment Period		*Post-Treatment Period*	
	Meditated (0/1) (1)	Logged Meal (0/1) (2)	Meditated (0/1) (3)	Logged Meal (0/1) (4)
mx	0.088***	-0.024**	0.024***	-0.010
	(0.011)	(0.010)	(0.009)	(0.006)
my	-0.028***	0.166***	-0.025***	0.029***
	(0.008)	(0.013)	(0.007)	(0.008)
mx X my	0.006	-0.026	0.009	-0.002
	(0.014)	(0.018)	(0.011)	(0.010)
zy	-0.025***	0.381***	-0.024***	0.037***
	(0.009)	(0.016)	(0.007)	(0.008)
mx + mxmy	0.094	-0.050	0.034	-0.013
	(0.009)	(0.014)	(0.007)	(0.008)
my + mxmy	-0.022	0.140	-0.016	0.027
	(0.011)	(0.012)	(0.009)	(0.007)
Ctrl Mean	0.094	0.118	0.054	0.033
Ctrl SD	0.291	0.323	0.227	0.178
Obs	102905	102905	102499	102499

Notes: OLS regressions of outcomes on treatments during the treatment period (columns 1 and 2) and the post-treatment period (columns 3 and 4). mx is 1 if the individual received meditation messages and 0 otherwise; my is 1 if the individual received meal logging messages and 0 otherwise; mx X my is the interaction between mx and my, zy is 1 if the individual received meal logging incentives. Specifications include controls for the five baseline variables on which re-randomization was based (female, college, daily notifications, whether individual meditated in last month, whether individual logged meal in last month) as well as day fixed effects. Standard errors clustered at individual level. One, two, and three stars indicate q-values of 1\%, 5\%, and 10\% respectively; q-values calculated according to the Benjamini Hochberg step-down procedure, considering all tests in the table (but excluding linear combinations of coefficients).

Table 4: Treatment Effects on Covariances

	Treatment Period	
	Cov(x,y) over people (within day) (1)	Cov(x,y) over time (within people) (2)
mx	0.001	0.002
	(0.004)	(0.001)
my	-0.002	0.001
	(0.004)	(0.001)
mx X my	0.021***	0.007***
	(0.005)	(0.002)
zy	-0.003	-0.001
	(0.004)	(0.001)
Ctrl Mean	0.019	0.008
Ctrl SD	0.138	0.084
Obs	102905	102905

Notes: OLS regressions at the individual-day level of the covariance between daily meditation (0/1) and daily logging of at least one meal (0/1) on treatments. Column 1 reports the covariance over people within a day (do people who meditate on a particular day also tend to log their meals?). Column 2 reports the covariance over days within a person (does someone who meditates a lot over the treatment period also log their meals a lot over the treatment period?). mx (my) is 1 if the individual received x (y) messages and 0 otherwise; mx*my is 1 if the individual received both sets of messages. The specification includes controls for the five baseline variables on which re-randomization was based (female, college, daily notifications, whether individual meditated in last month, whether individual logged meal in last month) as well as day fixed effects. Standard errors clustered at individual level. One, two, and three stars indicate q-values of 1%, 5%, and 10% respectively; q-values calculated according to the Benjamini Hochberg step-down procedure, considering all tests in the table.

Table 5: Treatment Effects on Opting Out of Message Programs

	Opted Out Ever, Meditation Msgs (X) (1)	Opted Out Ever, Nutrition Msgs (Y) (2)
mx & my	-0.026	0.034
	(0.017)	(0.019)
mx-only Mean	0.142	
mx-only SD	0.349	
my-only Mean		0.162
my-only SD		0.369
Obs	1571	1611

Notes: OLS regressions of whether individual ever opted out of each messaging program on treatments. Restricted to individuals with a message treatment. Omitted groups are mx-only (Column 1) and my-only (Column 2). mx & my is 1 if the individual received both sets of messages. The specification includes controls for the five baseline variables on which re-randomization was based (female, college, daily notifications, whether individual meditated in last month, whether individual logged meal in last month). One, two, and three stars indicate q-values of 1%, 5%, and 10% respectively; q-values calculated according to the Benjamini Hochberg step-down procedure, considering all tests in the table.

Table 6: Treatment Effects on Attention to Messages: Reading Rate and Memory of Content

	Raffle Response Meditation (X) (1)	Knowledge Score Meditation (X) (2)	Raffle Response Nutrition (Y) (3)	Knowledge Score Nutrition (Y) (4)
Panel A: Unconditional				
mx & my	-0.087**	-0.010	-0.081**	-0.008
	(0.028)	(0.006)	(0.029)	(0.007)
mx-only Mean	0.306	0.299		
mx-only SD	0.461	0.086		
my-only Mean			0.264	0.300
my-only SD			0.441	0.100
Obs	990	849	924	839
Panel B: Conditional on Not Opting-Out				
mx & my	-0.082**	-0.005	-0.084**	-0.005
	(0.034)	(0.006)	(0.034)	(0.007)
mx-only Mean	0.353	0.308		
mx-only SD	0.478	0.079		
my-only Mean			0.310	0.308
my-only SD			0.463	0.098
Obs	821	755	768	743

Notes: OLS regressions of whether individual responded to surprise raffle (Columns 1 and 3) and score on knowledge quiz (Columns 2 and 4) on treatments. *Raffle Response* is 1 if the individual responded to the raffle and 0 otherwise. *Knowledge Score* is calculated using answers to a true/false quiz administered in the endline survey on the information provided in the messaging program. Participants received 1 point for every correct answer, 0 point for every incorrect or "I do not remember seeing this message" answer, and 0.5 points for answering "I remember seeing this message but I do not remember the details." The score is the fraction of 14 possible points received. Columns 1 and 3 are restricted to individuals with a message treatment who received a functional raffle message (592 did not due to an implementation error). Columns 2 and 4 are restricted to individuals with a message treatment who took the endline survey. Omitted groups are mx-only (Column 1) and my-only (Column 2). mx & my is 1 if the individual received both sets of messages. The specification includes controls for the five baseline variables on which re-randomization was based (female, college, daily notifications, whether individual meditated in last month, whether individual logged meal in last month). One, two, and three stars indicate q-values of 1%, 5%, and 10% respectively; q-values calculated according to the Benjamini Hochberg step-down procedure, considering all tests in the table.

Table 7: Baseline Expectations of Behaviors by Treatment

	Expected Meditation (X) (1)	Expected - Actual Meditation (X) (2)	Expected Meal Logging (Y) (3)	Expected - Actual Meal Logging (Y) (4)
mx	0.113***	0.009	-0.017	0.012
	(0.012)	(0.015)	(0.017)	(0.018)
my	0.017	0.053***	0.125***	-0.041*
	(0.012)	(0.013)	(0.014)	(0.018)
mx X my	-0.027	-0.028	0.034	0.049
	(0.017)	(0.021)	(0.021)	(0.025)
zy	0.022	0.053***	0.214***	-0.223***
	(0.013)	(0.014)	(0.015)	(0.021)
mx + mxmy	0.086	-0.020	0.017	0.061
	(0.012)	(0.014)	(0.013)	(0.018)
my + mxmy	-0.009	0.024	0.159	0.008
	(0.012)	(0.016)	(0.016)	(0.018)
Ctrl Mean	0.384	0.278	0.519	0.381
Ctrl Mean S.E.	(0.009)	(0.010)	(0.012)	(0.012)
Obs	2871	2871	2871	2871

Notes: OLS regressions of expected rates of behavior over treatment period, measured at baseline (Columns 1 and 3) and difference between individual's expected and actual rate (Columns 2 and 4) on treatments. The specification includes controls for the five baseline variables on which re-randomization was based (female, college, daily notifications, whether individual meditated in last month, whether individual logged meal in last month) as well as day fixed effects. Standard errors clustered at individual level. One, two, and three stars indicate q-values of 1%, 5%, and 10% respectively; q-values calculated according to the Benjamini Hochberg step-down procedure, considering all tests in the table (but excluding linear combinations of coefficients).