Step into Action: How Reminders Shape Engagement in Fitness Apps

Braut B.* Mollier C.† Zaccagni S.‡

February 27, 2025

Abstract

Regular exercise offers well-documented health benefits, and mobile applications are increasingly used to promote physical activity. In this study, we conduct a large-scale randomized controlled trial with 20,187 users of the *WeWard* app in France to evaluate the effectiveness of different message framings and intervention durations on app engagement and walking behavior. Participants were randomly assigned to receive one of three types of messages, each leveraging distinct behavioral features such as peer comparison, self-comparison, or a generic reminder, over a period of one or three weeks. Our results show that peer-comparison messages positively influence the extensive margin of app utilization. However, despite these increases in app usage, neither type of message led to a significant increase in the number of steps walked. These findings highlight the limitations of relying on engagement metrics alone when evaluating app-based interventions aimed at behavior change.

Keywords: Reminders, framing, field-experiment, health, app
JEL classification: C930, D910, I120
Conflicting interests: None
Ethics Clearance: Ethics Committee of the University of Copenhagen
Pre-Registration: AEARCTR-0010467

*Department of Economics, University of Genoa, Italy; email: `beatrice.braut@unige.it`
†EconomiX, Paris Nanterre University, France; email : `cmollier@parisnanterre.fr`
‡Department of Economics and Business Economics, Aarhus University, Denmark; email: `sarah.zaccagni@econ.au.dk`

1 Introduction

Regular physical activity is widely recognized for its significant health benefits. According to the World Health Organization,[1] exercise helps prevent and manage cardiovascular diseases, diabetes, and certain cancers, while insufficient activity increases the risk of mortality by 20% to 30%. Physical activity also plays a crucial role in reducing stress and depression symptoms, enhancing mental well-being, and improving cognitive abilities (Sharon-David and Tenenbaum, 2017; Matzer et al., 2018). Furthermore, it positively influences other health behaviors, such as self-regulated eating, contributing both directly and indirectly to combating obesity (Baker and Brownell, 2000; Mata et al., 2009). Despite these well-documented advantages, regular exercise is often avoided. This reluctance stems from the fact that, aside from the immediate mood-enhancing effects of endorphin release, most benefits are delayed, while the associated costs—such as time, effort, and, in some cases, financial expenditures for equipment or subscriptions—are immediate. These challenges align with present bias (O'Donoghue and Rabin, 1999), where individuals prioritize immediate rewards over long-term gains. In this context, identifying strategies to encourage physical activity becomes a critical public health priority.

This paper examines the effectiveness of framed messages in promoting step tracking and walking, using a field experiment. Walking is a universally accessible form of physical activity, requiring no monetary investment, and can be easily integrated into daily routines. It is also a zero-emission mode of transportation, making it an important element in sustainable mobility transitions.[2] Although the optimal number of daily steps is debated,[3] setting step goals and tracking progress can increase awareness of the risks of sedentary behavior (Walsh et al., 2016).

[1] See WHO website: `https://www.who.int/news-room/fact-sheets/detail/physical-activity`

[2] See European Commission website: `https://transport.ec.europa.eu/news-events/news/efficient-and-green-mobility-2021-12-14_en`.

[3] See The New York Times, 2021: `https://www.nytimes.com/2021/07/06/well/move/10000-steps-health.html`

The growing use of apps and wearable technologies to monitor health metrics, especially steps, creates new opportunities for behavioral interventions aimed at fostering healthy habits. Such interventions are increasingly relevant as step counts continue to decline globally (Desine et al., 2023). In collaboration with the *WeWard* app, we designed and implemented a behavioral intervention targeting a large sample of the French population. *WeWard*, a popular step-tracking app with over 20 million users across nine countries, incentivizes walking by offering monetary rewards. Beyond financial incentives, the app employs behavioral features such as informational content, social comparisons, reminders, and gamified challenges to promote walking habits.

Digital interventions via fitness apps and wearable technologies represent a promising alternative when traditional information campaigns fail to encourage healthy behaviors (Tufano and Karras, 2005; Liu et al., 2024). Gamification, in particular, has emerged as an effective tool for influencing health behaviors (King et al., 2013). Mobile messaging has also been shown to improve a range of health behaviors, such as medication adherence, smoking cessation, and physical activity (Free et al., 2009; Lester et al., 2010). Additionally, virtual communities and peer interactions can significantly influence behaviors, including eating, smoking, and exercise (Christakis and Fowler, 2007; Centola, 2013; Franken et al., 2023). This phenomenon also helps explain why posts within the rapidly expanding sports-focused social communities tend to garner significantly more engagement and attention compared to posts on traditional social media platforms.[4] These virtual sports communities provide a space where users share detailed information about their physical activities, such as workout routines, step counts, or personal achievements. The act of sharing, combined with interactions such as likes, comments, or virtual rewards from peers, fosters a sense of accountability and motivation. This engagement reinforces users' commitment to physical activity, as they feel supported and encouraged by the community, further amplifying the app's effectiveness in promoting healthy behaviors (Russell et al., 2023).

[4]See `https://business.strava.com/blog/strava-clubs-vs-traditional-social-media`.

Despite their potential, many existing studies on app-based interventions rely on self-reported outcomes, which may introduce bias. Examples include research on diet choices (Freyne et al., 2017), smoking cessation (Rodgers et al., 2005), postnatal physical activity (Fjeldsoe et al., 2010), and compliance with testing (Nyatsanza et al., 2016), medical guidelines (Vervloet et al., 2012) and vaccination plans (Busso et al., 2015; Chen et al., 2023). In contrast, our study leverages objective, app-recorded data to evaluate behavioral interventions in a naturalistic setting.

In this experiment, we collaborated with the app *WeWard* to design and test the impact of framed messages and intervention durations on walking behavior, leveraging distinct behavioral insights. Specifically, the messages were designed to capitalize on concepts such as sunk cost fallacy, peer comparison, and self-comparison. To evaluate their effectiveness, we implemented two intervention durations: a short-term period of one week and a longer-term period of three weeks. This approach allowed us to assess whether tailored messages can serve as a scalable and low-cost tool to encourage physical activity and increase awareness of the benefits of walking. Our experimental design offers several key advantages. First, we measure actual walking behavior using app-recorded data, which reduces reliance on self-reported measures and minimizes potential biases. Second, the experiment involves a large and diverse sample of individuals who did not self-select into the intervention group, enhancing the generalizability of our findings. Third, the study design includes a two-week pre-experiment observation period, enabling us to classify participants by their baseline activity levels and analyze heterogeneous treatment effects (Charness and Gneezy, 2009). Additionally, post-intervention observations allow us to examine the persistence of effects after the intervention ends.

While there is substantial evidence that traditional economic tools, such as monetary incentives (Charness and Gneezy, 2009; Campos-Mercade et al., 2021, 2024; Arad et al., 2023; Ciccone et al., 2021), can effectively foster healthy behaviors, behavioral interventions, including tailored messages and reminders, have also been shown to influence health-related behaviors in the short term (Slaunwhite et al., 2009; Falco and Zaccagni, 2021; Patel et al., 2023; Habla and Muller, 2021). However, transforming these short-term effects into sustained behavioral changes requires habit formation, a process that may demand more sophisticated strategies. In the context of physical activity, combining various tools—such as incentives, social comparisons, and reminders—has been found to be particularly effective (Royer et al., 2015; Patel et al., 2016; Stecher et al., 2024; Adjerid et al., 2022).[5]

Our findings reveal that the peer comparison treatment increases the likelihood of app engagement. However, neither treatment results in a significant increase in the number of steps walked. Walking behavior remains largely influenced by external factors, such as the day of the week and holidays.

The remainder of this paper is structured as follows. Section 2 details the experimental design, while Section 3 outlines our hypotheses. Section 4 describes the estimation strategy, and Section 5 presents the results. We conclude with a discussion of our findings in Section 6.

2 Experimental Design

This study employs a between-subject experimental design to evaluate the effects of various message framings and treatment durations on subgroups within a representative sample of *WeWard* users in France. The experiment is structured to explore how these interventions influence walking behavior across diverse population segments. In the subsections that follow, we first provide an overview of the key features and functionality of the *WeWard* app, which serves as the platform for our intervention. Next, we outline the experimental procedure, including details of the sample composition and recruitment process. Finally, we describe the treatments in detail, highlighting the behavioral principles underlying each message and the rationale for the chosen intervention durations.

[5] Although such combinations may not be equally effective when applied to multiple health behaviors simultaneously, as highlighted by Trachtman (2024).

2.1 The *WeWard* App

Launched in 2019, *WeWard* is a mobile application designed to encourage walking by integrating financial incentives with step-tracking features.[6] Unlike conventional step-tracking apps, *WeWard* rewards users with points, termed *Wards*, for meeting daily step goals. These points can be redeemed for cash once a certain threshold is reached or used to secure discounts on online purchases. For example, users earn €0.05 for completing 10,000 steps, a commonly recommended daily goal, with a maximum monthly earning potential of €3.75. However, payouts are delayed until users accumulate at least €15 worth of *Wards*, ensuring engagement over time.[7]

The app has gained substantial attraction, with over 10 million downloads in France and Italy, and has recently expanded to other countries, including Spain, Germany, the United States, Belgium, the United Kingdom, Japan, and the Netherlands, totaling over 20 million users as of March 2024. According to the company, *WeWard* increases users' daily step counts by 24%.

Several features beyond monetary incentives contribute to the app's success. First, the gamification of walking transforms a routine physical activity into an engaging and rewarding experience. For instance, the app introduces challenges where users earn badges by completing specific tasks, such as walking a designated number of steps consecutively for several days or within a given timeframe. Second, the app leverages the concept of sunk costs. Since steps are not immediately convertible into rewards, users may perceive their initial walking efforts as investments toward future payoffs. This sunk cost mechanism has been shown to act as a self-regulation tool, encouraging continued engagement by leveraging the user's prior effort (Hong et al., 2019). Third, *WeWard* facilitates self-comparison by allowing users to track their progress over time. The ability to monitor personal performance, compare it to past achievements, and observe improvements can foster intrinsic motivation

[6] A screenshot of the app's homepage is provided in Appendix A, Figure A1.

[7] These payment details reflect the app's structure during the experimental period and remained constant throughout.

and promote the formation of walking habits (De La Torre et al., 2021). Additionally, the app fosters a sense of community by enabling users to connect with others, creating opportunities for peer comparison. Users can compare their step counts with those in their network, potentially motivating increased activity through social influence (Franken et al., 2023; Gershon et al., 2024).

The app's incentive structure is based on a tiered reward system, offering *Wards* when users meet specific thresholds of daily steps - 1,000, 3,000, 6,500, 10,000, 15,000, and 20,000. Importantly, the number of *Wards* earned increases non-linearly across these thresholds, and steps taken between two thresholds do not yield additional rewards. To validate their steps and claim rewards, users must open the app and manually confirm their step count by clicking a designated button. This feature not only ensures active engagement with the app but also rises users' awareness of their daily activity levels. Overall, *WeWard* demonstrates potential as an effective tool for habit formation by incentivizing walking and reinforcing positive behaviors through financial rewards, gamification, and social comparison. By increasing users' awareness and appreciation of walking, a low-cost and accessible form of physical activity, the app aligns well with public health goals.[8]

2.2 Procedure and Sample

To implement our field experiment, we collaborated with the start-up *WeWard*. This collaboration provided two key advantages. First, it enabled us to conduct a large-scale experiment involving over 20,000 participants. Second, since *WeWard* directly tracks users' walking data, this collaboration allowed us to measure actual behavior in participants' everyday lives, thereby mitigating self-reporting biases and experimental demand effects.

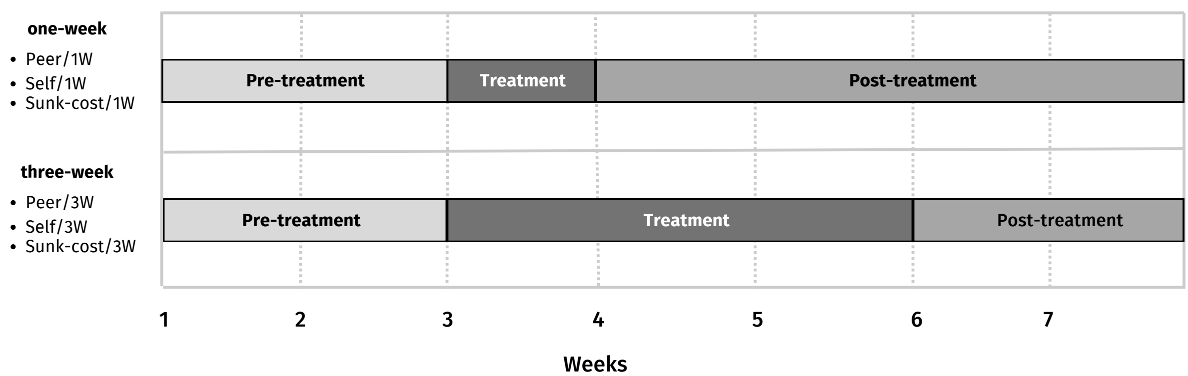
The experiment was conducted over seven weeks, spanning December 2022 to January 2023.[9] During the first two weeks, we collected pre-treatment data. In this phase,

[8]This study focuses on specific mechanisms within the app that contribute to habit formation and behavioral change.

[9]The total duration of the experiment was 50 days, comprising seven full weeks and one additional day

participants were not exposed to any new treatment; instead, they received only the app's standard reminder. Starting in the third week, we introduced our treatments. For participants in the treatment groups, one of three framed messages replaced the standard reminder. The control group continued to receive the original reminder.[10] The treatments were assigned to two durations: a short-term intervention lasting one week and a longer intervention lasting three weeks. After the treatment phase, we collected post-treatment data for four weeks for participants in the short-term group and two weeks for those in the long-term group. Thus, the total study duration was seven weeks for all participants. Figure 1 illustrates the experimental timeline. This experiment was pre-registered in the AEA RCT Registry (Registry Number: AEARCTR-0010467).[11]

Figure 1: Timeline of the project.



Our sample consisted of 20,187 individuals, totaling 422,076 observations. The sample was representative of *WeWard* users and included only participants who had consented to receive app notifications during their initial sign-up. This was a prerequisite for delivering the interventions via mobile notifications. Additionally, we restricted the sample to participants who had not customized the default time for receiving reminders (9 PM). These users neither opted out of reminders nor personalized their timing, suggesting they likely had no strong preference regarding the default reminder, minimizing selection bias

in the final week of the post-treatment observation period.

[10] A detailed description of the treatments can be found in Subsection 2.3.

[11] The pre-registration is available at `https://www.socialscienceregistry.org/trials/10467`.

based on prior engagement with the app's features.

In our sample, 52.4% of participants identified as female, 29.4% as male, and 18.2% did not disclose their gender. The sample was evenly distributed across treatment arms, with approximately 2,200 participants assigned to each message-duration combination, and twice as many participants allocated to the control group. Randomization ensured balance in pre-treatment characteristics, as shown in Table 1.

The experiment included participants from all French regions. As demonstrated in Appendix A, Figure A2, the regional distribution of users relative to the population is uniformly distributed, ensuring geographic representativeness.

Table 1: Summary statistics.

	Peer/1W	Peer/3W	Self/1W	Self/3W	Sunk-cost/1W	Sunk-cost/3W	Control	N
Female	0.65	0.64	0.65	0.62	0.65	0.64	0.64	14,215
	(0.67)	(0.82)	(0.50)	(0.22)	(0.52)	(0.87)		
% of users who do not share their location	0.42	0.45	0.46	0.43	0.44	0.44	0.45	17,377
	(0.49)	(0.50)	(0.50)	(0.49)	(0.50)	(0.50)		
Average number of steps at baseline	6,050	5,994	6,037	6,144	5,989	6,011	6,092	17,377
	(0.68)	(0.34)	(0.59)	(0.61)	(0.31)	(0.43)		
% of subjects who walk > 10,000 steps	0.12	0.11	0.12	0.13	0.12	0.13	0.12	17,377
	(0.73)	(0.32)	(1.00)	(0.75)	(0.89)	(0.71)		
% of subjects who walk < 5,000 steps	0.44	0.46	0.45	0.44	0.46	0.45	0.44	17,377
	(0.91)	(0.31)	(0.48)	(0.91)	(0.22)	(0.41)		
Average number of connections at baseline	3.35	3.30	3.28	3.34	3.26	3.46	3.36	17,377
	(0.94)	(0.60)	(0.45)	(0.88)	(0.36)	(0.37)		
Probability of opening the app at baseline	0.53	0.52	0.53	0.51	0.52	0.52	0.53	17,377
	(0.56)	(0.44)	(0.75)	(0.08)	(0.37)	(0.30)		
N	2,223	2,114	2,160	2,201	2,132	2,193	4,354	

Notes: The table reports the means of each variable across each treatment arm and p-values from t-test of the differences of means between the treatment arm and the control.

2.3 Treatments

The study leverages pre-treatment observations collected over two weeks for the entire sample to establish baseline activity levels. During this period, all users received the app's standard reminder. This same message was sent to the control group throughout the experiment and during the post-treatment period for all participants.

0. **Basic reminder:** "Don't forget to validate your steps! It's time to increase your prize

fund".

This message serves as a neutral prompt, simply reminding users of the app's functionality without emphasizing any particular behavioral motivation.

Following the pre-treatment period, participants were randomly assigned to one of six treatment arms using a 3x2 factorial design. An additional seventh group, the control group, received only the basic reminder throughout the experiment. The treatments varied along two dimensions: message content and treatment duration. The details are as follows:

1. **Sunk-Cost Message:** "You have validated your steps over the last few days, do it again to win more *Wards*".

   This message emphasizes users' prior investment of time and effort, aiming to leverage the sunk-cost fallacy to encourage continued engagement.

   1.1. **Sunk-Cost/1W:** Participants in this arm received the sunk-cost message for one week.

   1.2. **Sunk-Cost/3W:** Participants in this arm received the sunk-cost message for three weeks.

2. **Peer Comparison Message:** "Today, the WeWarders have walked XXX steps on average, what about you?"

   This message highlights the app's social features, encouraging users to compare their performance with the community. By fostering curiosity and competitive behavior, it aims to increase awareness and engagement.[12]

   2.1. **Peer/1W:** Participants in this arm received the peer comparison message for one week.

[12]In our study, participants were not provided with information about the performance of peers within their chosen group of friends. At the time of implementation, many users were not actively utilizing this feature, and those who did were likely a self-selected group with higher levels of competitiveness. Using such sub-sample could have introduced bias and reduced the generalizability of the intervention.

2.2. **Peer/3W:** Participants in this arm received the peer comparison message for three weeks.

3. **Self-Comparison Message:** “Last week you walked an average of XXX steps, how many did you walk today?”

   This message focuses on users’ prior performance, promoting self-improvement and reflection. Its effectiveness may depend on users’ potential for improvement and self-evaluation.

   3.1. **Self/1W:** Participants in this arm received the self-comparison message for one week.

   3.2. **Self/3W:** Participants in this arm received the self-comparison message for three weeks.

As outlined earlier, data collection occurred over seven weeks. Two weeks of pre-treatment data established baseline activity levels, during which all participants received the basic reminder. Treatment messages were then implemented in weeks three to five, with durations varying by treatment arm (one or three weeks). Finally, two weeks of post-treatment data were collected, during which all participants reverted to receiving the basic reminder.

3 Hypotheses

The focus of our study is the relationship between the information treatments administered to app users and their average number of steps. Given that the app tracks steps and promotes walking, we use engagement with the app as a proxy for attention to this healthy behavior. Sending reminders is a cost-effective method of encouraging physical activity (Calzolari and Nardotto, 2017; Habla and Muller, 2021). However, generic messages often yield limited impact (Costa and Kahn, 2013; Halpern, 2015; Peer et al., 2020). Tailored messages, on the other hand, have proven effective in improving outcomes in academic performance, driving behavior (O'Connell and Lang, 2018; Choudhary et al., 2022), and college program enrollment among disadvantaged populations (Castleman and Page, 2015). Such interventions can successfully induce new behaviors (e.g., attending college) and enhance existing ones. In the domain of physical activity, punctual, targeted messages have demonstrated effectiveness in promoting exercise within specific populations (Golbus et al., 2024).

Building on this evidence, our study implements specifically designed messages that emphasize particular behavioral features of the app, aiming to amplify their impact. We hypothesize that these tailored messages will outperform standard reminders in increasing the number of steps walked. Additionally, comparing the effects of different treatments allows us to identify which behavioral aspect is most effective in influencing users' behavior.

Hypothesis 1: *Walking and engagement with the app increase when users receive tailored messages leveraging behavioral features compared to standard reminders.*

The duration of the intervention is also a critical factor influencing its effectiveness. In this study, we manipulate both the content and the duration of the reminders. Evidence suggests that the timing and duration of health interventions significantly affect their impact (Head et al., 2013; Steiner et al., 2018; Hardeman et al., 2019). While previous research has examined the effect of varying the intensity and duration of physical activity requirements (Charness and Gneezy, 2009), our study focuses specifically on the duration of messaging designed to promote such behavior.

We predict that longer exposure to tailored reminders (three weeks) will have a greater impact on increasing the number of steps and app engagement compared to shorter exposure (one week). Extended exposure to behavioral messages may provide more consistent reinforcement, helping users develop the habit of consulting the app and sustaining behavioral changes over time.

Hypothesis 2: *Walking and engagement with the app increase when users receive tailored messages leveraging behavioral features for a longer period.*

Prior research often targets individuals who do not regularly exercise (Acland and Levy, 2015), or distinguishes between regular and non-regular exercisers, predicting greater treatment effects among less active individuals (Charness and Gneezy, 2009). In our study, all participants are users of the app and can be considered motivated to some extent. However, baseline data from *WeWard* allows us to differentiate between users who walk an average or above-average number of steps (active and very active) and those who walk relatively little (somewhat active and inactive).[13]

We hypothesize that the intervention will have a stronger effect among inactive users, as their potential for improvement is greater compared to regular walkers.

Hypothesis 3: *The treatment is more effective among inactive users.*

4 Empirical Strategy

To evaluate the effects of the treatments on users' attention and walking habits within the *WeWard* app, we first examine the overall impact of receiving any tailored reminder on several outcome variables. Subsequently, we compare specific outcomes within each treatment group to those in the control group, which received only the generic message throughout the intervention. Specifically, we estimate the following econometric model:

$$\begin{aligned} Y_{i,n} = \beta_0 + \beta_1 \mathit{Treated} + \beta_2 T_{i,sunk} \cdot \mathit{Treated} + \beta_3 T_{i,peer} \cdot \mathit{Treated} \\ + \beta_4 T_{i,self} \cdot \mathit{Treated} + \epsilon_i, \end{aligned} \tag{1}$$

[13]For a detailed description of the user activity categories (active, very active, somewhat active, and inactive), please refer to Subsection 5.2.2.

where $Y_{i,n}$ represents a set of n outcomes of interest, including: (1) a dummy variable indicating whether the user opened the app on a given day, (2) the number of connections to the app per day, (3) the daily number of steps (for the whole sample and for a restricted sample of individuals who opened the app every day), and (4) the weekly average number of steps. To analyze attention to steps, we examine engagement with the app by defining a dummy variable equal to 1 if the user opened the app on a given day, and 0 otherwise. Additionally, we assess the number of app connections and user-initiated actions within the app on a daily basis, using the full sample of observations for these variables. For walking habits, we first evaluate the intervention's impact on the number of steps walked across the entire sample of users. Subsequently, we calculate the weekly average number of steps (*Weekly steps*) as the mean number of steps per week per user, restricting the sample to individuals with at least one recorded observation per week. The daily number of steps (*Daily steps*) is defined as the steps walked on a given day by users who consistently opened the app each day during the intervention period.

The variable *Treated* is a dummy equal to 1 during the treatment period, specifically after week 2 and before week 4 (for one-week treatments) or week 6 (for three-week treatments). This variable identifies observations within the treatment window. Depending on the regression specification, the treatment effect is measured relative to the pre-treatment period (*Pre-Treatment*) or the post-treatment period (*Post-Treatment*). We also analyze the effects of new reminders by comparing outcomes across pre-treatment, and post-treatment periods. The variables T_{sunk}, T_{peer}, and T_{self} are dummies equal to 1 if an individual was randomly assigned to the respective control or treatment groups, where the control group received only the generic reminder. The term ϵ_i represents the error term. In this specification: β_0 captures the average baseline outcome in the control group during the pre-treatment period.

The coefficient β_1 captures the difference in the outcome for the control group between the pre-treatment period (weeks 1 and 2) and the treatment period (week 3 for the one-week treatments and weeks 3-5 for the three-week treatments). The coefficients β_2, β_3, and β_4 capture the differential impact of each treatment group (compared to the control group) on step counts or app engagement over the course of the experiment. This approach allows us to disentangle the effects of specific treatments and assess how tailored reminders influence users' walking behavior and engagement with the app.

5 Results

This section presents the results of our study, examining the effects of the treatments on app engagement and walking habits. Subsection 5.1 focuses on app-related behaviors, while Subsection 5.2 explores users' walking habits. Since opening the app is a prerequisite for *WeWard* to retrieve activity data, our analysis accounts for potential attrition by running separate analyses on walking habits using two distinct sub-samples in addition to the full sample. Subsection 5.2.1 analyzes the weekly average number of steps (*Weekly steps*), restricting the sample to participants who opened the app at least one day per week throughout the experiment. We then narrow the analysis further to users who validated their steps daily during the intervention, allowing us to study the daily number of steps (*Daily steps*), a more granular measure of walking behavior that reflects the habits of the most committed users. Finally in 5.2.2 we explore the heterogeneity of the results looking at initial differences in walking habits.

5.1 Engagement with the App

While downloading the app signals an interest in step tracking and walking, users exhibit diverse behaviors within the app. Only 3.62% of the initial sample opened the app daily throughout the study. App engagement provides valuable insights into users' attention and motivation, which are critical for understanding and influencing walking habits. Targeting app engagement may foster improvements in walking behavior, as the number of steps taken positively correlates with both the frequency of app openings and interactions (Pearson's $r = 0.27$ and $r = 0.26$, respectively; $p = 0.00$). This interdependence between physical activity and attention to its measurement may create a virtuous cycle if effectively activated.

We begin by analyzing the factors influencing whether users open the app in Subsection 5.1.1.[14] Next, we explore the number of app connections as a proxy for the intensity of user interaction with *WeWard*.

5.1.1 Opening the App

On average, users opened the app 4.37 times per day during the 50-day intervention. However, app usage varied significantly across participants. Half of the users did not open the app for 32 days, and on average, users refrained from opening the app for 3.48 consecutive days. Notably, 1,143 individuals opened *WeWard* only once during the entire study period. Figure 2 illustrates the proportion of users opening the app by day of the week. Users were significantly less likely to engage with the app during weekends (χ^2, $p = 0.00$). However, app openings were higher on Sundays compared to Saturdays (χ^2, $p = 0.00$). Gender differences in app engagement are evident in Figure A3 (Appendix), showing that women were more likely than men to open the app on any given day (χ^2, $p = 0.00$).

Figures 3a and 3b depict the evolution of app-opening behavior across different treatment groups and durations. The red lines mark the start and end of the treatment period. The data reveal temporal variations in app engagement, including weekly dips during weekends and a general downward trend influenced by the holiday season. App openings decreased significantly on key holidays, such as Christmas Eve (December 24) and New Year's Eve (December 31), as well as on Christmas Day (December 25) and New Year's Day (January 1) (χ^2, $p = 0.00$ for December 25; χ^2, $p = 0.03$ for January 1).

[14]Our sample consists of users who were active on the app at least once (N = 17,377). We excluded participants who never opened the app (N = 2,212) and those with no recorded steps during the 50-day experiment (N = 598). However, we checked the robustness of our results using the complete sample. Results are available upon request.

Figure 2: Proportion of users opening the app by day of the week.

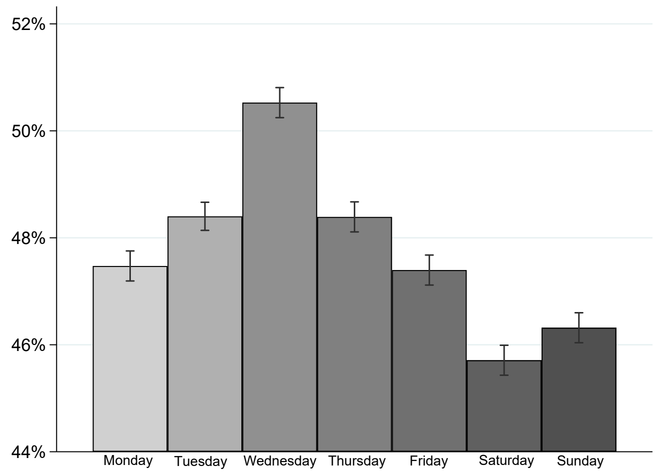
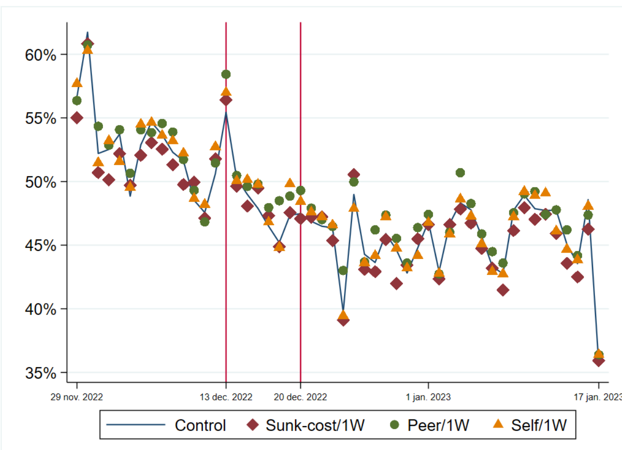
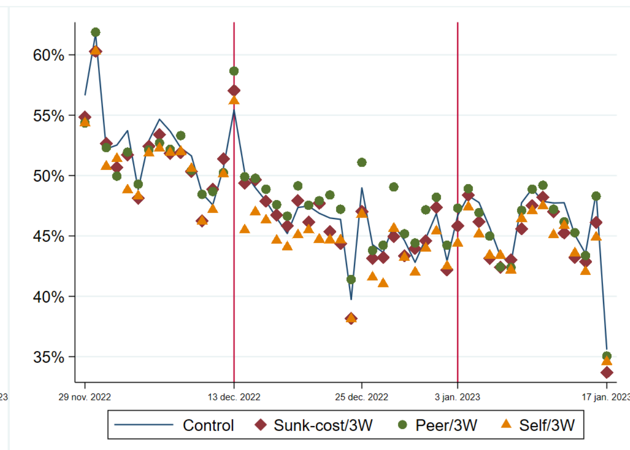


Figure 3: Time trend of the proportion of users opening the app.

(a) One-week



(b) Three-week



Tables 2 and B5 (in the Appendix) present the effects of the treatments on the probability of opening the app. Table 2 first compares the treatment period with the pre-treatment period to identify any treatment effects. Then it compares the treatment period with the post-treatment to examine potential rebound effects once the treatment is removed. Lastly, Table B5 directly compares pre- and post-treatment to assess whether the treatment produced a change that lasts.[15] Table 2 includes eight econometric specifications estimated

[15] As a robustness check, we also pooled together the treatment and post-treatment periods and tested them against the pre-treatment period. Results were consistent, but we choose to show in the main text results

using a logit model with individual fixed effects. Columns from 1 to 4 present the results of the comparison of the treatment period with the pre-treatment period, while columns from 5 to 8 present the results of the comparison of the treatment period with the post-treatment period. We present the baseline specifications (columns 1-2-5-6) without additional controls and then incorporate a time trend which allows also to account for the potential impact of the holiday period during the experiment. To incorporate the time trend we include one dummy each for the 2^{nd} through 21^{st} day of the experiment in the one-week treatments (column 3), and one dummy each for the 2^{nd} through 35^{th} day of the experiment in the three-week treatments (column 4). For the comparison of the treatment with the post-treatment, we include one dummy each for the 16^{th} through 50^{th} day of the experiment in both the one-week treatments (column 7), and the three-week treatments (column 8).

In Table 2, the treatment effect is assessed by comparing the treatment period with the pre-treatment period. The negative and significant coefficient of *Treated* in columns 1 to 4 indicates that users opened the app significantly less during the treatment period than in the pre-treatment period. The estimated coefficients range from -0.27 for the one-week treatments to -0.43 for the three-week treatments (columns 1–2) and become even more negative when accounting for the time trend, reaching -0.62 for the one-week treatment and -0.89 for the three-week treatment (columns 3–4). This evidence suggests that during the treatment period, there was a decline in the probability of opening the app among users who did not receive the intervention. When examining the marginal effect of receiving framed reminders, we find that the treatment emphasizing peer comparison significantly increases the probability of opening the app, whether administered for one or three weeks. Specifically, this treatment mitigates the downward trend by 0.08 when received for one week and by 0.14 when received for three weeks. We then compare the treatment period with the post-treatment period to assess whether the effects persist once the reminders are discontinued.

of post-treatment vs. treatment and treatment vs. pre- treatment to highlight that the effect vanishes over time. Evidence of the post-treatment vs. pre-treatment, which might be considered a more policy relevant estimate, is available in Table B5, in the Appendix.

Table 2: Effects of treatments on *Probability of opening the app.*

	Probability of opening the app							
	Treat vs. Pre				Treat vs. Post			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	-0.27***	-0.43***	-0.62***	-0.89***				
	(-14.50)	(-31.45)	(-16.00)	(-23.80)				
Peer/1W × Treated	0.08**		0.08**					
	(2.61)		(2.62)					
Self/1W × Treated	0.05		0.05					
	(1.42)		(1.43)					
Sunk-Cost/1W × Treated	0.07*		0.07*					
	(2.05)		(2.06)					
Peer/3W × Treated		0.13***		0.14***				
		(5.58)		(5.61)				
Self/3W × Treated		-0.01		-0.01				
		(-0.24)		(-0.24)				
Sunk-Cost/3W × Treated		0.04		0.04				
		(1.64)		(1.65)				
Treated					0.24***	0.04**	1.45***	1.54***
					(13.83)	(2.74)	(36.87)	(39.74)
Peer/1W × Treated					0.07*		0.07*	
					(2.35)		(2.37)	
Self/1W × Treated					0.07*		0.07*	
					(2.26)		(2.27)	
Sunk-Cost/1W × Treated					0.07*		0.07*	
					(2.22)		(2.22)	
Peer/3W × Treated						0.11***		0.11***
						(4.59)		(4.61)
Self/3W × Treated						-0.00		-0.00
						(-0.09)		(-0.09)
Sunk-Cost/3W × Treated						0.07**		0.07**
						(2.79)		(2.81)
Observations	195,468	352,240	195,468	352,240	351,504	350,964	351,504	350,964
Subjects	9,308	10,064	9,308	10,064	9,764	9,749	9,764	9,749
Day dummies			Yes	Yes			Yes	Yes

Notes: Logit model regressions with individual fixed effects. The first panel (columns 1-4) shows the comparison between the treatment and the pre-treatment periods. The second panel (columns 5-8) shows the comparison between the treatment and the post-treatment periods. Columns 1-3-5-7 correspond to the one-week treatment arms, while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively.

The positive coefficients for the *Treated* variable in this comparison suggest a reversal of the trend, potentially bringing app activity back to pre-treatment levels. Specifically, for the Peer treatment, the coefficients (0.07 and 0.11) are comparable to those observed during the treatment period, indicating that its effects do not persist after removal. This finding aligns with prior research on reminders and notifications, which suggests that while such interventions are effective during implementation, their influence often diminishes or even reverses once they are withdrawn, as individuals revert to baseline behavior or even lower engagement levels. For the other framed messages, we observe positive marginal effects, suggesting that removing these treatments and reverting to the standard message slightly reduces the probability of app usage.

To further investigate this, we compare the pre-treatment and post-treatment periods in Table B5 (Appendix). We find a decline in the probability of opening the app within the control group. Additionally, the results show no significant effects for the customized reminders, regardless of whether participants were in the one-week or three-week treatment groups. Taken together, the evidence from Tables 2 and B5 suggests that the impact of peer comparison messages does not persist once the intervention ends. These findings contribute to the literature on the short-lived nature of customized reminders, particularly those based on social or self-comparison (Slaunwhite et al., 2009).

Having examined the impact of the reminders on the probability of opening the app—an initial measure of attention to step tracking—the next section explores whether they influence the frequency of interactions with the app. We then assess their effects on walking habits. Together, these analyses provide a comprehensive understanding of how our messages shaped users' engagement with step tracking.

5.1.2 Activity on the App

In this section, we go beyond app openings to examine whether the reminders affect users' engagement with the app. To measure this, we analyze the daily number of connections to the application.

Table 3 and Table B6 (in the Appendix) present the results on the effect of the messages on the number of connections, comparing the pre-treatment, treatment, and post-treatment periods. The structure of Tables 3 and B6 mirrors that of Tables 2 and B5. We estimate an OLS model with individual fixed effects. In Table 3, which compares the treatment period to the pre-treatment period (columns 1 to 4), we observe a reduction in the number of connections for the control group. When we introduce the time trend (columns 3–4), the coefficient of *Treated* becomes more negative. These results are consistent with those observed for the probability of opening the app. However, we do not find any evidence of a marginal contribution of the customized messages to the overall effect. When comparing the post-treatment to the treatment period, we observe some marginal effects in the three-week treatments. Table B6 in the Appendix presents the comparison between the post-treatment and pre-treatment periods. As expected, none of the interventions has a statistically significant effect, while the control group experiences a decline in the number of connections.

5.2 Walking Habits

In the following sections, we examine walking behavior and the impact of our treatments in promoting healthier habits. We begin by analyzing the walking patterns of the full sample of users. To test the robustness of our findings, we then focus on the weekly average number of steps and, finally, on highly committed users who open the app daily.

Table 4 presents the effect of our reminders on the number of steps walked, first comparing the treatment period with the pre-treatment period (columns 1–4) and then with the post-treatment period (columns 5–8). We estimate an OLS model with individual fixed effects. As shown, the number of steps walked decreases during the treatment period compared to the pre-treatment period. After accounting for the time trend, the effect remains negative, ranging from -698.55 steps (for the one-week treatments) to -1403.94 steps (for the three-week treatments). None of the customized reminders stand out in influencing the number of steps walked. Similarly, we find no effect of the different messages when comparing the treatment and post-treatment periods. For a direct comparison between the post-treatment and pre-treatment periods, see Table B7 in the Appendix.

Table 3: Effects of treatments on *Number of connections*.

	Number of connections							
	Treat vs. Pre				Treat vs. Post			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	-0.32***	-0.40***	-0.49***	-0.64***				
	(-5.88)	(-6.51)	(-5.12)	(-6.47)				
Peer 1W × Treated	0.00		0.01					
	(0.04)		(0.06)					
Self 1W × Treated	0.07		0.07					
	(0.88)		(0.86)					
Sunk Cost 1W × Treated	0.06		0.06					
	(0.71)		(0.66)					
Peer 3W × Treated		0.01		0.00				
		(0.06)		(0.04)				
Self 3W × Treated		0.17		0.17				
		(1.68)		(1.68)				
Sunk Cost 3W × Treated		0.11		0.11				
		(1.16)		(1.17)				
Treated					0.05	-0.13*	0.05	0.21*
					(0.98)	(-2.44)	(0.48)	(2.01)
Peer 1W × Treated					0.06		0.06	
					(0.82)		(0.81)	
Self 1W × Treated					0.01		0.01	
					(0.10)		(0.07)	
Sunk Cost 1W × Treated					-0.03		-0.03	
					(-0.33)		(-0.35)	
Peer 3W × Treated						0.10		0.10
						(1.05)		(1.04)
Self 3W × Treated						0.26*		0.26*
						(2.42)		(2.40)
Sunk Cost 3W × Treated						0.16		0.16
						(1.65)		(1.64)
Constant	3.94***	4.14***	4.00***	4.19***	3.76***	3.92***	4.08***	3.97***
	(341.86)	(209.80)	(65.89)	(75.98)	(610.21)	(185.33)	(51.13)	(50.26)
Observations	110,034	170,101	110,034	170,101	167,191	164,857	167,191	164,857
Subjects	10,119	10,583	10,119	10,583	10,106	10,081	10,106	10,081
Day dummies			Yes	Yes			Yes	Yes

Notes: OLS model regressions with individual fixed effects. The first panel (columns 1-4) shows the comparison between the treatment and the pre-treatment periods. The second panel (columns 5-8) shows the comparison between the treatment and the post-treatment periods. Columns 1-3-5-7 correspond to the one-week treatment arms, while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

5.2.1 Weekly and Daily Steps

To leverage the temporal dimension of our data, we conduct panel analyses, restricting the sample to participants who recorded at least one observation per week throughout the

Table 4: Effects of treatments on *Steps walked*.

	Steps walked							
	Treat vs. Pre				Treat vs. Post			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	-224.44***	-648.96***	-698.55***	-1403.94***				
	(-5.16)	(-16.92)	(-8.47)	(-17.37)				
Peer 1W × Treated	-10.34		0.02					
	(-0.13)		(0.00)					
Self 1W × Treated	63.73		54.88					
	(0.90)		(0.77)					
Sunk Cost 1W × Treated	-59.87		-69.63					
	(-0.77)		(-0.90)					
Peer 3W × Treated		-35.36		-39.29				
		(-0.52)		(-0.58)				
Self 3W × Treated		60.59		55.65				
		(0.85)		(0.78)				
Sunk Cost 3W × Treated		11.98		-5.47				
		(0.17)		(-0.08)				
Treated					344.28***	-452.86***	-45.24	-52.35
					(7.60)	(-10.55)	(-0.50)	(-0.57)
Peer 1W × Treated					50.02		54.41	
					(0.66)		(0.72)	
Self 1W × Treated					63.13		63.37	
					(0.84)		(0.85)	
Sunk Cost 1W × Treated					-42.54		-43.27	
					(-0.56)		(-0.56)	
Peer 3W × Treated						39.08		37.18
						(0.52)		(0.49)
Self 3W × Treated						92.66		85.04
						(1.10)		(1.00)
Sunk Cost 3W × Treated						40.41		20.93
						(0.47)		(0.24)
Constant	6274.28***	6348.75***	6701.06***	6737.76***	5784.81***	6184.50***	6538.27***	6589.04***
	(711.24)	(445.62)	(144.64)	(140.03)	(995.43)	(357.10)	(100.09)	(91.87)
Observations	109794	169746	109794	169746	166884	164575	166884	164,575
Subjects	10,121	10,585	10,121	10,585	10,107	10,081	10,107	10,081
Day dummies			Yes	Yes			Yes	Yes

Notes: OLS model regressions with individual fixed effects. The first panel (columns 1-4) shows the comparison between the treatment and the pre-treatment periods. The second panel (columns 5-8) shows the comparison between the treatment and the post-treatment periods. Columns 1-3-5-7 correspond to the one-week treatment arms, while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

experiment. This procedure results in a final sample of 8,941 subjects, enabling a balanced panel analysis with *customerID* as the panel variable and *week* as the time variable. Our key outcome variable is the individual's average number of steps per week, which we define as *Weekly Steps*.[16]

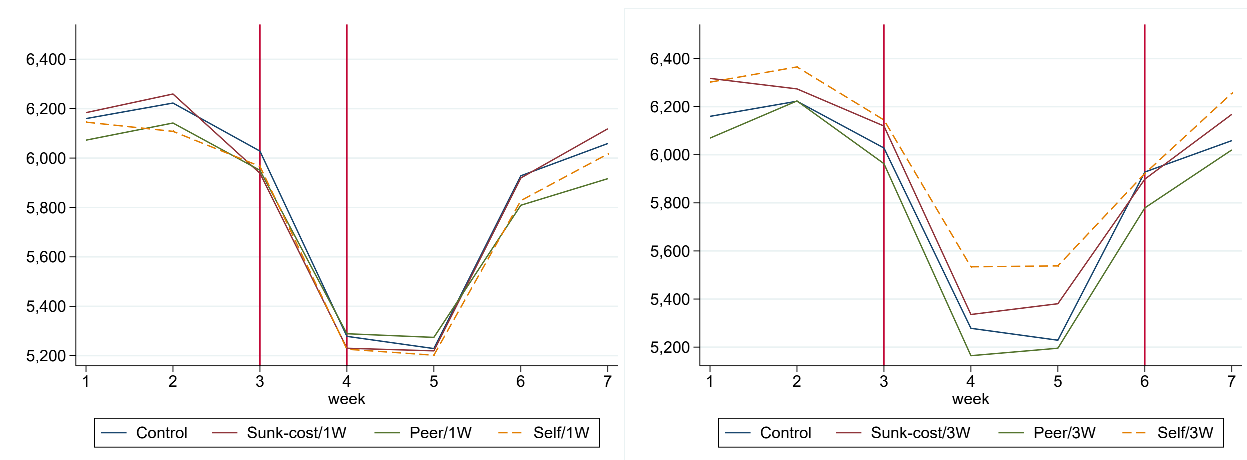
Figures 4a and 4b depict the evolution of *Weekly Steps* throughout the experiment

[16]The distribution of subjects by treatment arm in the restricted sample is reported in Table B4.

by treatment group. As shown in the figures, the average number of steps per week declines across all groups, with a pronounced drop in weeks four and five, followed by an increase. Since the fourth week begins on December 20$^{\text{th}}$ and the fifth week ends on January 2$^{\text{nd}}$, this decline is largely attributable to seasonal effects, particularly the Christmas holidays. This effect appears less pronounced for the Self/3W treatment, whereas the decline is more substantial for Peer/3W. Among participants receiving the one-week treatment, differences between groups are less evident.

Figure 4: *Weekly steps* by treatment group.

(a) One-week (b) Three-week



We further examine the effect of different message types and treatment durations on walking behavior using a panel data fixed effects model. Table 5 presents the results with *Weekly Steps* as the dependent variable. The table follows the same structure as previous tables, first comparing the treatment period with the pre-treatment period (columns 1–4) and then comparing the treatment period with the post-treatment period (columns 5–8). Odd-numbered columns report results for the one-week treatments, while even-numbered columns report results for the three-week treatments. In specifications 3–4 and 7–8, we include a time trend.

As indicated by the significant and negative coefficient of *Treated*, participants walk fewer steps per week on average following the implementation of the treatment, consistent

Table 5: Effects of treatments on *Weekly Steps*.

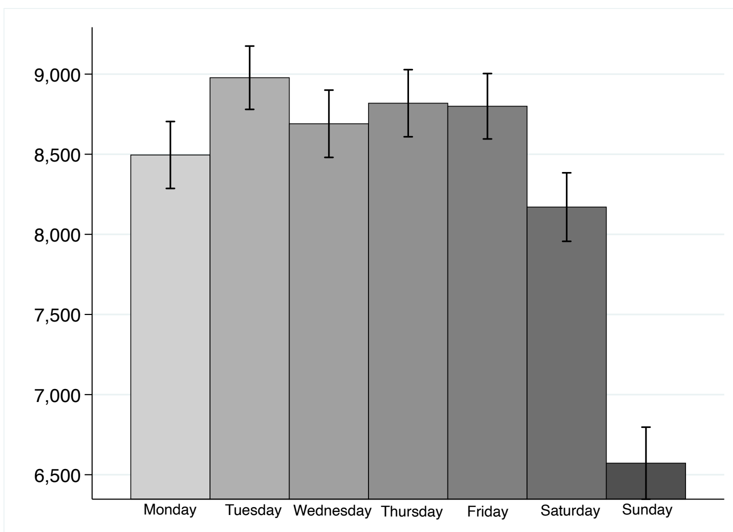
	Weekly steps							
	Treat vs. Pre				Treat vs. Post			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	-162.99***	-679.52***	-139.51**	-898.87***				
	(-3.32)	(-15.58)	(-2.65)	(-16.18)				
Peer 1W × Treated	6.33		6.33					
	(0.07)		(0.07)					
Self 1W × Treated	2.75		2.75					
	(0.03)		(0.03)					
Sunk Cost 1W × Treated	-120.09		-120.09					
	(-1.32)		(-1.32)					
Peer 3W × Treated		-26.18		-26.18				
		(-0.35)		(-0.35)				
Self 3W × Treated		84.76		84.76				
		(1.04)		(1.04)				
Sunk Cost 3W × Treated		-4.52		-4.52				
		(-0.06)		(-0.06)				
Treated					404.77***	-481.80***	-25.62	-97.30
					(8.30)	(-10.36)	(-0.47)	(-1.82)
Peer 1W × Treated					-26.34		-26.34	
					(-0.31)		(-0.31)	
Self 1W × Treated					-6.29		-6.29	
					(-0.08)		(-0.08)	
Sunk Cost 1W × Treated					-88.16		-88.16	
					(-1.02)		(-1.02)	
Peer 3W × Treated						23.07		23.07
						(0.29)		(0.29)
Self 3W × Treated						131.86		131.86
						(1.45)		(1.45)
Sunk Cost 3W × Treated						60.24		60.24
						(0.64)		(0.64)
Constant	6166.54***	6230.98***	6143.06***	6201.02***	5601.40***	6001.26***	6031.79***	6111.24***
	(578.93)	(363.63)	(302.77)	(256.43)	(912.12)	(313.28)	(248.03)	(218.37)
Observations	16,914	27,770	16,914	27,770	28,190	27,770	28,190	27,770
Subjects	5,638	5,554	5,638	5,554	5,638	5,554	5,638	5,554
Week dummies			Yes	Yes			Yes	Yes

Notes: OLS model regressions with individual fixed effects. The first panel (columns 1-4) shows the comparison between the treatment and the pre-treatment periods. The second panel (columns 5-8) shows the comparison between the treatment and the post-treatment periods. Columns 1-3-5-7 correspond to the one-week treatment arms, while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

with our findings from the full dataset. Overall, the treatments do not significantly affect the average number of weekly steps. These results align with those presented in Table 4, reinforcing that customized messages are not effective in increasing step count, even when restricting the sample to users who open the app at least once per week. When comparing the treatment and post-treatment periods (columns 5 to 8), we find no evidence of a significant rebound effect after the removal of customized messages, as expected. For a comparison between the post-treatment and pre-treatment periods, see Table B8 in the Appendix.

To further validate our findings on the weekly average number of steps and total steps walked, we refine our sample by excluding all subjects who did not open the application—and thus did not allow step data collection—every day for the entire 50-day period. This procedure substantially reduces the sample size to 553 subjects but enables a balanced panel analysis with *customerID* as the panel variable and *date* as the time variable.[17] Since this dataset contains daily observations with no missing days, we analyze the number of steps per day (*Daily Steps*) rather than per week (*Weekly Steps*). Additionally, this approach allows us to explore variations in walking behavior across different days of the week. Figure 5 illustrates the mean *Daily Steps* by day of the week, showing that participants tend to walk more on weekdays than on weekends.

Figure 5: *Daily steps* by day of the week.



Figures 6a and 6b illustrate the evolution of *Daily Steps* throughout the experiment, providing greater insight into the factors driving the decline in step count compared to Figures 4a and 4b. Specifically, we find that the drop observed in the fourth week is primarily driven by December 25$^{\text{th}}$ (Christmas Day), while a similar effect occurs in the fifth week, particularly on January 1$^{\text{st}}$ (the day after New Year's Eve).

[17]The distribution of subjects by treatment arm in the restricted sample is reported in Table B4.

The impact of the new reminders on daily step counts is not pronounced. However, participants in the Peer/1W treatment group appear to have consistently lower step counts throughout the experiment window.

Figure 6: *Daily Steps* by treatment group.

(a) One-week (b) Three-week

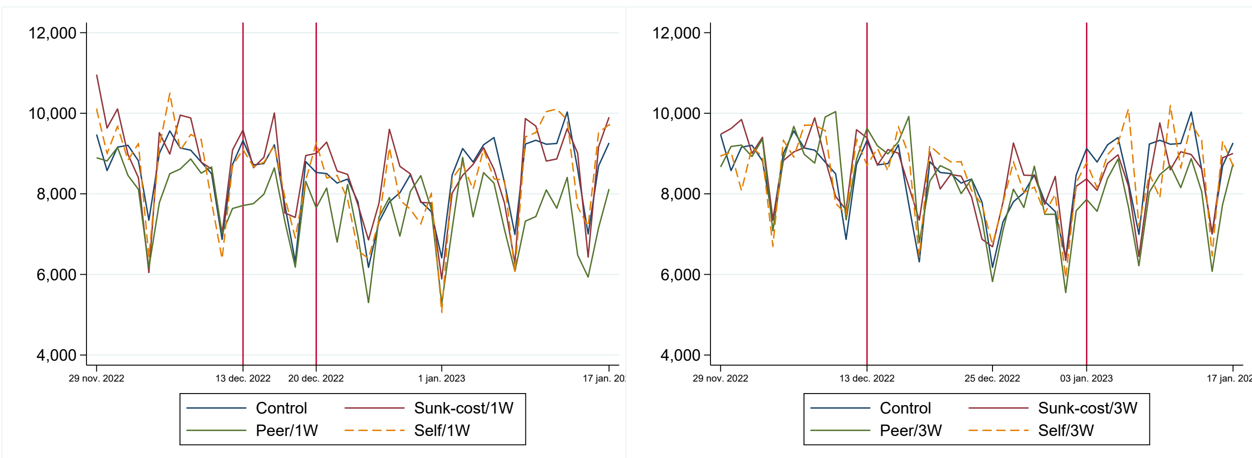


Table 6 presents the regression results for the sub-sample of users with daily step data. The first part of the table reports the comparison between the treatment and pre-treatment periods (columns 1–4), while the second part compares the treatment and post-treatment periods (columns 5–8). Odd-numbered columns (1, 3, 5, and 7) show results for the one-week treatments, whereas even-numbered columns (2, 4, 6, and 8) correspond to the three-week treatments. The effect of being *Treated* follows the same trend observed in previous tables. As with *Weekly Steps*, we find no significant impact of the customized messages on the number of steps walked per day among users who consistently used the app throughout the observation period.[18]

Given that we find a significant effect of only one treatment—the three-week peer treatment—on our first outcome, the probability of opening the app, while all subsequent analyses on walking-related outcomes across different datasets yield consistently null results,

[18] Table B9 in the Appendix reports the comparison between the post-treatment and pre-treatment periods.

Table 6: Effects of treatments on *Daily Steps*.

	Daily steps							
	Treat vs. Pre				Treat vs. Post			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	-324.49	-715.99***	-1072.31**	-1005.04**				
	(-1.86)	(-4.29)	(-3.08)	(-2.84)				
Peer 1W × Treated	-201.64		-201.64					
	(-0.66)		(-0.66)					
Self 1W × Treated	-95.30		-95.30					
	(-0.32)		(-0.32)					
Sunk Cost 1W × Treated	46.94		46.94					
	(0.16)		(0.16)					
Peer 3W × Treated		-172.53		-172.53				
		(-0.63)		(-0.63)				
Self 3W × Treated		156.21		156.21				
		(0.43)		(0.43)				
Sunk Cost 3W × Treated		-66.74		-66.74				
		(-0.20)		(-0.20)				
Treated					65.24	-813.48***	-338.73	-54.19
					(0.30)	(-3.87)	(-0.93)	(-0.15)
Peer 1W × Treated					229.29		229.29	
					(0.64)		(0.63)	
Self 1W × Treated					98.31		98.31	
					(0.28)		(0.28)	
Sunk Cost 1W × Treated					231.49		231.49	
					(0.61)		(0.61)	
Peer 3W × Treated						881.69*		881.69*
						(2.18)		(2.18)
Self 3W × Treated						267.58		267.58
						(0.76)		(0.76)
Sunk Cost 3W × Treated						530.46		530.46
						(1.47)		(1.46)
Constant	8706.62***	8826.22***	9764.76***	9230.90***	8161.64***	8576.03***	9252.88***	9008.07***
	(235.96)	(124.68)	(41.56)	(38.52)	(313.68)	(108.25)	(41.71)	(39.98)
Observations	7,224	12,425	7,224	12,425	12,384	12,780	12,384	12,780
Subjects	344	355	344	355	344	355	344	355
Day dummies			Yes	Yes			Yes	Yes

Notes: OLS model regressions with individual fixed effects. The first panel (columns 1-4) shows the comparison between the treatment and the pre-treatment periods. The second panel (columns 5-8) shows the comparison between the treatment and the post-treatment periods. Columns 1-3-5-7 correspond to the one-week treatment arms, while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

we have reasons to assert the robustness of our findings. To further strengthen this claim, we conduct several robustness checks. First, we re-run the analysis on the restricted sample of users who opened the app every day, using the lagged outcome variable (one day) to account for potential delayed effects. Second, we extend the analysis to the full sample without excluding users who never opened the app, instead assigning them a value of zero for the outcome variable. Third, we conduct the analysis excluding data from Christmas Day and

New Year's Eve to control for potential holiday-related anomalies. Across all these robustness checks, our main findings remain unchanged, reinforcing the validity of our conclusions.[19]

In the next section, we further examine potential heterogeneous effects based on users' activity levels.

5.2.2 Heterogeneity

We investigate heterogeneous effects based on subjects' pre-treatment activity levels, following the approach of Charness and Gneezy (2009). To do so, we classify users into four categories—*inactive*, *somewhat active*, *active*, and *very active*—based on the quartiles of their pre-treatment step distribution.[20] We conduct this analysis separately for two subsamples: (i) users who opened the app every day during the experiment and (ii) users with at least one recorded observation per week. For the strongly committed users, we define *inactive* users as those who walked fewer than 5,445 steps during the pre-treatment period, *somewhat active* users as those walking between 5,445 and 7,510 steps, *active* users as those walking between 7,510 and 10,765 steps, and *very active* users as those exceeding 10,765 steps. For the sub-sample with at least one weekly observation, we adjust the thresholds accordingly: *inactive* users walk fewer than 3,800 steps, *somewhat active* users walk between 3,800 and 5,570 steps, *active* users walk between 5,570 and 7,710 steps, and *very active* users walk more than 7,710 steps during the pre-treatment period.[21]

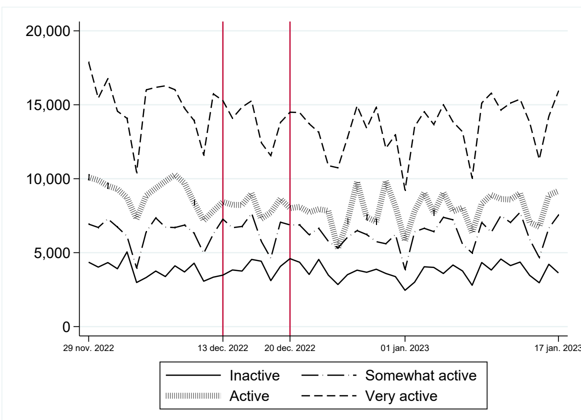
Not surprisingly, strongly committed users tend to walk more on average than those who did not open the application daily. Figures 7a and 7b illustrate the evolution of average *Daily Steps* by activity category, as defined earlier. Similarly, Figures 8a and 8b show the evolution of average *Weekly Steps* by walker category. Figure 7a depicts the evolution of *Daily Steps* for users who received the one-week treatment, while Figure 7b shows the data for those who received the three-week treatment. The trends in Figures 7a and 7b appear

[19] Results are available upon request.
[20] The distribution of the *Number of connections* in the four categories is reported in Table B3.
[21] The distribution of subjects by category and by treatment arm is reported in Table B4.

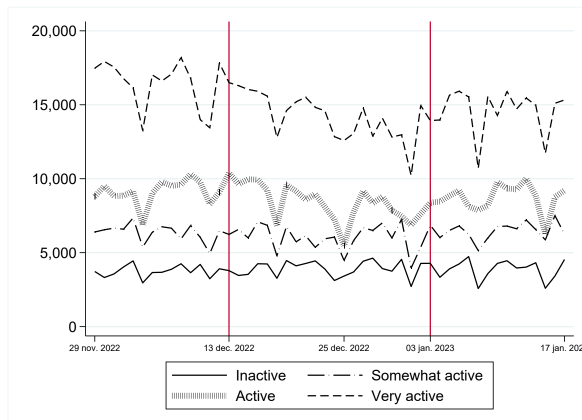
less stable over time compared to those in Figures 8a and 8b, due to the fluctuations in step counts between weekdays and weekend days in the daily data. These daily figures also provide more granular evidence of the Christmas effect, particularly driven by *very active* users. Notably, the drop in steps on December 25th is more pronounced for *very active* and *active* users than for *somewhat active* and *inactive* users. Additionally, fluctuations in daily steps appear more dependent on weekends for *very active* and *active* users.

Figure 7: *Daily Steps* by category of walkers.

(a) One-week



(b) Three-week



In the sub-sample of users who opened the app at least once per week throughout the experiment, the average number of steps per day is 5,848.[22]

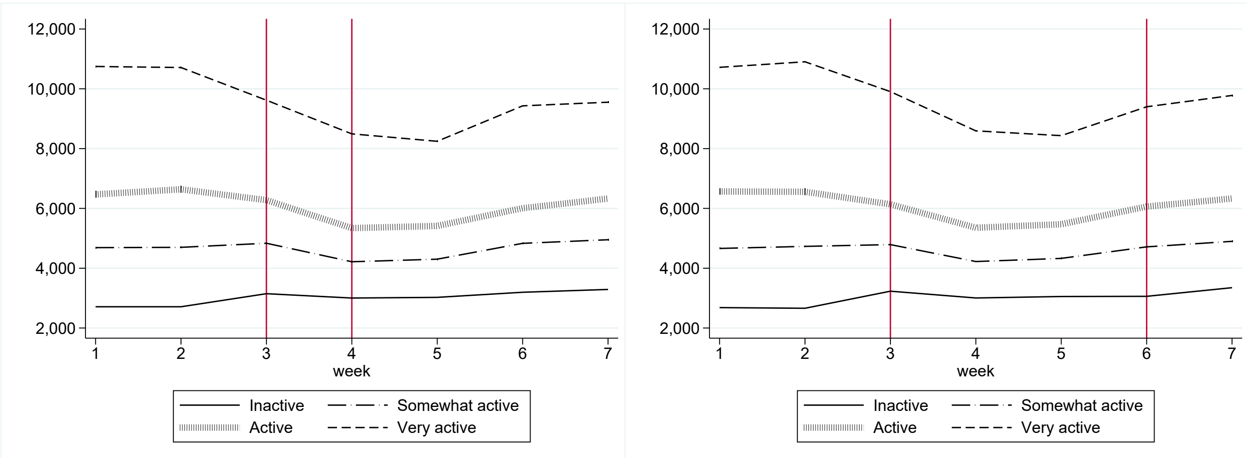
Looking at Figures 8a and 8b, the decline in the number of steps during the treatment period (as observed in Figures 4a and 4b) appears to be primarily driven by *somewhat active*, *active*, and *very active* users. While their trends show a decline during weeks four and five, the trend for *inactive* users remains relatively stable. Interestingly, users who walk the least seem to experience a slight increase in their average number of steps during week three, both in the one-week and three-week treatment groups. Figure A4 in the Appendix shows that the decrease in *Weekly Steps* for *very active* users is statistically significant, while

[22]According to the WHO, walking less than 5,000 steps per day is considered sedentary. Therefore, it is worth noting that our categorization is based on the data from our sample, where only a relatively small number of users (1,537, or 17% of the pre-treatment sample) walk in accordance with the WHO's recommended 10,000 steps per day.

inactive users experience a significant increase (Mann-Whitney, $p = 0.00$).

Figure 8: *Weekly Steps* by category of walkers.

(a) One-week (b) Three-week



To evaluate whether our treatments have a differential impact across walker categories, we split the sample and run the regression using *Weekly Steps* as the dependent variable. Table 7 presents the results of the comparison between the treatment period and the pre-treatment period. Columns 1-3-5-7 display the specifications for the one-week treatments, while columns 2-4-6-8 correspond to the three-week treatments. Specifically, columns 1-2 report results for *inactive* users, columns 3-4 for *somewhat active* users, columns 5-6 for *active* users, and columns 7-8 for *very active* users. As indicated by the significant and positive coefficient of *Treated* in specifications 1 and 2, *inactive* users exhibit a significant increase in *Weekly Steps* over time. Conversely, for more active users, the effect is reversed: the coefficient of *Treated* is negative in columns 5-8. The magnitude of the coefficients increases from columns 5-6 to 7-8, suggesting that the more active the users, the more susceptible they are to the overall negative time trend observed in previous analyses. The upward trend among *inactive* users, observed when comparing the treatment period to the pre-treatment period, persists when comparing the post-treatment period to the pre-treatment period. This pattern suggests that the observed changes may be driven by regression to the mean, where users with unusually low step counts in the pre-treatment period revert to their higher average levels, while those with unusually high step counts revert downward over time. Overall, our findings indicate that customized messages do not have a significant impact on the number of steps walked by any category of users.

Table 7: Effects on *Weekly steps* by category: **Treatment** vs. **Pre-treatment**.

	Weekly steps							
	Inactive		Somewhat active		Active		Very active	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	643.03***	308.85***	341.33***	-45.55	-263.45***	-1058.59***	-977.76***	-2439.21***
	(7.51)	(3.94)	(5.13)	(-0.68)	(-3.30)	(-12.95)	(-6.52)	(-15.75)
Peer 1W × Treated	-241.79		-24.66		73.65		42.56	
	(-1.80)		(-0.23)		(0.55)		(0.15)	
Self 1W × Treated	-205.66		-79.58		185.18		-127.18	
	(-1.60)		(-0.77)		(1.45)		(-0.60)	
Sunk Cost 1W × Treated	-246.13		-37.08		15.31		-336.71	
	(-1.85)		(-0.32)		(0.11)		(-1.30)	
Peer 3W × Treated		-6.90		-30.18		-101.64		90.68
		(-0.06)		(-0.33)		(-1.00)		(0.42)
Self 3W × Treated		65.22		127.42		286.14*		26.72
		(0.58)		(1.28)		(2.43)		(0.13)
Sunk Cost 3W × Treated		-74.43		-11.91		-3.95		146.66
		(-0.70)		(-0.13)		(-0.04)		(0.66)
Constant	2731.69***	2716.12***	3670.22***	3683.41***	6279.67***	6304.07***	11057.91***	11077.18***
	(107.65)	(87.03)	(158.80)	(130.25)	(194.53)	(172.63)	(176.04)	(155.65)
Observations	4,218	6,625	6,291	10,140	5,997	9,940	3,786	6,335
Subjects	1,406	1,325	2,097	2,028	1,999	1,988	1,262	1,267
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5-7 correspond to the one-week treatment arms, while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*. All specifications include weekly dummies.

6 Conclusion

We conduct a large-scale field experiment in collaboration with the step-tracking application *WeWard* to test the effect of tailored reminders on user engagement with the app and on actual walking behavior. Specifically, we examine the impact of messages leveraging different behavioral features—self-comparison, peer-comparison, and sunk-cost—and assess whether the duration of exposure to these interventions affects user behavior. Our results demonstrate that certain types of reminders effectively increase engagement with the app. In particular, peer-comparison messages lead to a higher likelihood of users opening *WeWard*. These findings align with recent studies on the role of peer influence in physical activity. For example, Franken et al. (2023) show that runners in Strava's virtual clubs increase their activity when they receive positive reinforcement, while Gershon et al. (2024) find that incentivizing gym visits with a friend significantly boosts attendance.

However, while our study confirms that tailored reminders successfully direct user attention to the app, they have little to no impact on the number of steps walked. Walking behavior appears more resistant to change, likely due to its strong association with individual habits and external constraints. Our data reveal that users tend to be more sedentary on weekends and holidays, suggesting that walking is often integrated into daily routines as a means of transportation rather than a discretionary health activity.

This divergence between engagement and actual behavioral change underscores a critical issue in the evaluation of digital health interventions: an over-reliance on engagement metrics can lead to an overestimation of their real-world impact. While app usage is often taken as a proxy for success, it does not necessarily translate into meaningful behavioral modifications, such as sustained increases in physical activity. The distinction between engagement and impact is particularly relevant given the prevailing mindset in digital health and behavioral technology. Many interventions celebrate high engagement as an achievement in itself, yet our findings caution against equating time spent on an app with genuine progress toward health-related goals. In some cases, the most effective interventions are those that require minimal continued engagement, helping users achieve their objectives efficiently rather than fostering perpetual interaction.

Our study contributes to this broader discussion by providing empirical evidence that challenges the assumption that engagement alone is a valid indicator of success. The effectiveness of digital health interventions should be measured by their ability to drive meaningful, long-term behavioral change rather than by the frequency of app interactions. Future research should explore alternative strategies that not only capture users' attention but also translate engagement into sustained lifestyle improvements. Additionally, studies should examine the contextual and structural barriers that make behavioral change difficult, ensuring that digital interventions address not only psychological drivers but also practical constraints. Ultimately, our findings reinforce the need for a paradigm shift in digital health research and product development: one that prioritizes measurable outcomes over mere engagement, fosters genuine behavior change rather than dependency, and supports users in achieving their health goals with minimal friction.

References

Acland, D. and M. R. Levy (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science 61*(1), 146–160.

Adjerid, I., G. Loewenstein, R. Purta, and A. Striegel (2022). Gain-loss incentives and physical activity: the role of choice and wearable health tools. *Management Science 68*(4), 2642–2667.

Arad, A., U. Gneezy, and E. Mograbi (2023). Intermittent incentives to encourage exercising in the long run. *Journal of Economic Behavior & Organization 205*, 560–573.

Baker, C. and K. Brownell (2000). Physical activity and maintenance of weight loss: Physiological and psychological mechanisms.

Busso, M., J. Cristia, and S. Humpage (2015). Did you get your shots? Experimental evidence on the role of reminders. *Journal of Health Economics 44*, 226–237.

Calzolari, G. and M. Nardotto (2017). Effective reminders. *Management Science 63*(9), 2915–2932.

Campos-Mercade, P., A. N. Meier, S. Meier, D. G. Pope, F. H. Schneider, and E. Wengström (2024). Incentives to vaccinate. Technical report, National Bureau of Economic Research.

Campos-Mercade, P., A. N. Meier, F. H. Schneider, S. Meier, D. Pope, and E. Wengström (2021). Monetary incentives increase covid-19 vaccinations. *Science 374*(6569), 879–882.

Castleman, B. L. and L. C. Page (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior & Organization 115*, 144–160.

Centola, D. (2013). Social media and the science of health behavior. *Circulation 127*(21), 2135–2144.

Charness, G. and U. Gneezy (2009). Incentives to exercise. *Econometrica 77*(3), 909–931.

Chen, N., K.-S. Trump, S. Hall, and Q. Le (2023). The effect of postcard reminders on vaccinations among the elderly: A block-randomized experiment. *Behavioural Public Policy 7*(2), 240–265.

Choudhary, V., M. Shunko, S. Netessine, and S. Koo (2022). Nudging drivers to safety: Evidence from a field experiment. *Management Science 68*(6), 4196–4214.

Christakis, N. A. and J. H. Fowler (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine 357*(4), 370–379.

Ciccone, A., A. Fyhri, and H. B. Sundfør (2021). Using behavioral insights to incentivize cycling: Results from a field experiment. *Journal of Economic Behavior & Organization 188*, 1035–1058.

Costa, D. L. and M. E. Kahn (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association 11*(3), 680–702.

De La Torre, S., D. Spruijt-Metz, A. J. Farias, et al. (2021). Associations among wearable activity tracker use, exercise motivation, and physical activity in a cohort of cancer survivors: secondary data analysis of the health information national trends survey. *JMIR cancer 7*(2), e24828.

Desine, S., H. Master, J. Annis, A. Hughes, D. M. Roden, P. A. Harris, and E. L. Brittain (2023). Daily step counts before and after the covid-19 pandemic among all of us research participants. *JAMA Network Open 6*(3), e233526–e233526.

Falco, P. and S. Zaccagni (2021). Promoting social distancing in a pandemic: Beyond good intentions. *PloS one 16*(12), e0260457.

Fjeldsoe, B. S., Y. D. Miller, and A. L. Marshall (2010). Mobilemums: A randomized controlled trial of an sms-based physical activity intervention. *Annals of behavioral medicine 39*(2), 101–111.

Franken, R., H. Bekhuis, and J. Tolsma (2023). Kudos make you run! How runners influence each other on the online social network Strava. *Social Networks 72*, 151–164.

Free, C., R. Whittaker, R. Knight, T. Abramsky, A. Rodgers, and I. Roberts (2009). Txt2stop: A pilot randomised controlled trial of mobile phone-based smoking cessation support. *Tobacco Control 18*(2), 88–91.

Freyne, J., J. Yin, E. Brindal, G. A. Hendrie, S. Berkovsky, and M. Noakes (2017). Push notifications in diet apps: influencing engagement times and tasks. *International Journal of Human–Computer Interaction 33*(10), 833–845.

Gershon, R., C. Cryder, and K. L. Milkman (2024). Friends with health benefits: A field experiment. *Management Science*.

Golbus, J. R., J. Shi, K. Gupta, R. Stevens, V. S. E. Jeganathan, E. Luff, T. Boyden, B. Mukherjee, S. Kohnstamm, V. Taralunga, et al. (2024). Text messages to promote physical activity in patients with cardiovascular disease: A micro-randomized trial of a just-in-time adaptive intervention. *Circulation: Cardiovascular Quality and Outcomes*, e010731.

Habla, W. and P. Muller (2021). Experimental evidence of limited attention at the gym. *Experimental Economics 24*(4), 1156–1184.

Halpern, D. (2015). *Inside the nudge unit: How small changes can make a big difference.* Random House.

Hardeman, W., J. Houghton, K. Lane, A. Jones, and F. Naughton (2019). A systematic review of just-in-time adaptive interventions (jitais) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity 16*, 1–21.

Head, K. J., S. M. Noar, N. T. Iannarino, and N. G. Harrington (2013). Efficacy of text messaging-based interventions for health promotion: a meta-analysis. *Social science & medicine 97*, 41–48.

Hong, F., W. Huang, and X. Zhao (2019). Sunk cost as a self-management device. *Management Science 65*(5), 2216–2230.

King, D., F. Greaves, C. Exeter, and A. Darzi (2013). 'Gamification': Influencing health behaviours with games.

Lester, R. T., P. Ritvo, E. J. Mills, A. Kariri, S. Karanja, M. H. Chung, W. Jack, J. Habyarimana, M. Sadatsafavi, M. Najafzadeh, et al. (2010). Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): A randomised trial. *The Lancet 376*(9755), 1838–1845.

Liu, C.-W., W. Wang, G. Gao, and R. Agarwal (2024). The value of virtual engagement: Evidence from a running platform. *Management Science 70*(9), 6179–6201.

Mata, J., M. N. Silva, P. N. Vieira, E. V. Carraça, A. M. Andrade, S. R. Coutinho, L. B. Sardinha, and P. J. Teixeira (2009). Motivational "spill-over" during weight control: Increased self-determination and exercise intrinsic motivation predict eating self-regulation. *Health Psychology 28*(6), 709.

Matzer, F., E. Nagele, N. Lerch, C. Vajda, and C. Fazekas (2018). Combining walking and relaxation for stress reduction—A randomized cross-over trial in healthy adults. *Stress and Health 34*(2), 266–277.

Nyatsanza, F., J. McSorley, S. Murphy, and G. Brook (2016). 'It's all in the message': the utility of personalised short message service (SMS) texts to remind patients at higher risk of STIs and HIV to reattend for testing—a repeat before and after study. *Sexually Transmitted Infections 92*(5), 393–395.

O'Connell, S. D. and G. Lang (2018). Can personalized nudges improve learning in hybrid classes? Experimental evidence from an introductory undergraduate course. *Journal of Research on Technology in Education 50*(2), 105–119.

O'Donoghue, T. and M. Rabin (1999, March). Doing it now or later. *American Economic Review 89*(1), 103–124.

Patel, M. S., K. L. Milkman, L. Gandhi, H. N. Graci, D. Gromet, H. Ho, J. S. Kay, T. W. Lee, J. Rothschild, M. Akinola, et al. (2023). A randomized trial of behavioral nudges delivered through text messages to increase influenza vaccination among patients with an upcoming primary care visit. *American Journal of Health Promotion 37*(3), 324–332.

Patel, M. S., K. G. Volpp, R. Rosin, S. L. Bellamy, D. S. Small, M. A. Fletcher, R. Osman-Koss, J. L. Brady, N. Haff, S. M. Lee, et al. (2016). A randomized trial of social comparison feedback and financial incentives to increase physical activity. *American Journal of Health Promotion 30*(6), 416–424.

Peer, E., S. Egelman, M. Harbach, N. Malkin, A. Mathur, and A. Frik (2020). Nudge me right: Personalizing online security nudges to people's decision-making styles. *Computers in Human Behavior 109*, 106347.

Rodgers, A., T. Corbett, D. Bramley, T. Riddell, M. Wills, R.-B. Lin, and M. Jones (2005). Do u smoke after txt? Results of a randomised trial of smoking cessation using mobile phone text messaging. *Tobacco Control 14*(4), 255–261.

Royer, H., M. Stehr, and J. Sydnor (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *American Economic Journal: Applied Economics 7*(3), 51–84.

Russell, H. C., C. Potts, and E. Nelson (2023). “If It’s not on Strava it Didn’t Happen”: Perceived Psychosocial Implications of Strava use in Collegiate Club Runners. *Recreational Sports Journal 47*(1), 15–25.

Sharon-David, H. and G. Tenenbaum (2017). The effectiveness of exercise interventions on coping with stress: Research synthesis. *Studies in Sport Humanities 22*, 19–29.

Slaunwhite, J. M., S. M. Smith, M. T. Fleming, and L. R. Fabrigar (2009). Using normative messages to increase healthy behaviours. *International Journal of Workplace Health Management 2*(3), 231–244.

Stecher, C., C.-H. Chen, J. Codella, S. Cloonan, and J. Hendler (2024). Combining anchoring with financial incentives to increase physical activity: a randomized controlled trial among college students. *Journal of Behavioral Medicine*, 1–19.

Steiner, J. F., M. R. Shainline, J. Z. Dahlgren, A. Kroll, and S. Xu (2018). Optimizing number and timing of appointment reminders: a randomized trial. *Am J Manag Care 24*(8), 377–384.

Trachtman, H. (2024). Does promoting one healthy behavior detract from others? evidence from a field experiment. *American Economic Journal: Applied Economics 16*(2), 249–277.

Tufano, J. T. and B. T. Karras (2005). Mobile ehealth interventions for obesity: A timely opportunity to leverage convergence trends. *Journal of Medical Internet Research 7*(5), e467.

Vervloet, M., A. J. Linn, J. C. van Weert, D. H. De Bakker, M. L. Bouvy, and L. Van Dijk (2012). The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. *Journal of the American Medical Informatics Association 19*(5), 696–704.

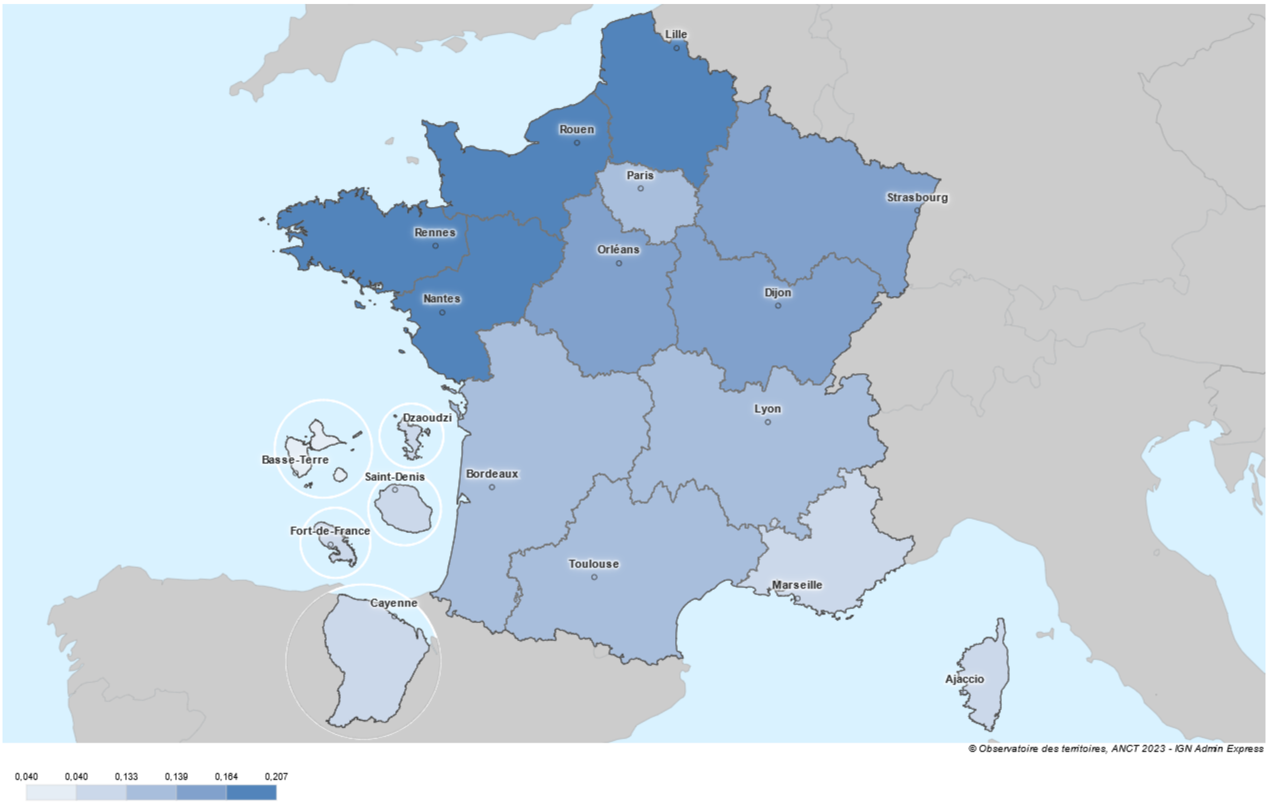
Walsh, J. C., T. Corbett, M. Hogan, J. Duggan, and A. McNamara (2016). An mhealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. *JMIR mHealth and uHealth 4*(3), e5227.

Appendix

A Additional Figures

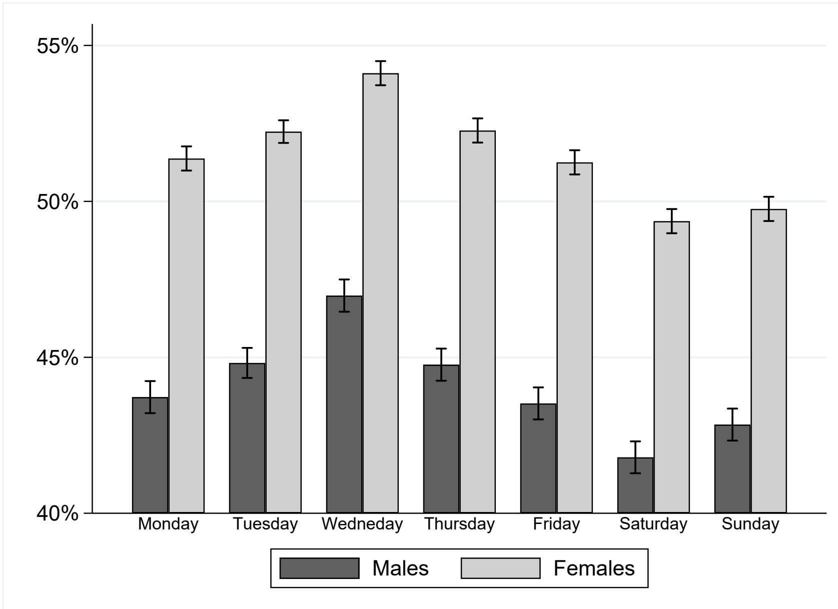
Figure A1: Screenshot of the *WeWard* app.

Figure A2: Proportion of users in the population by French region.



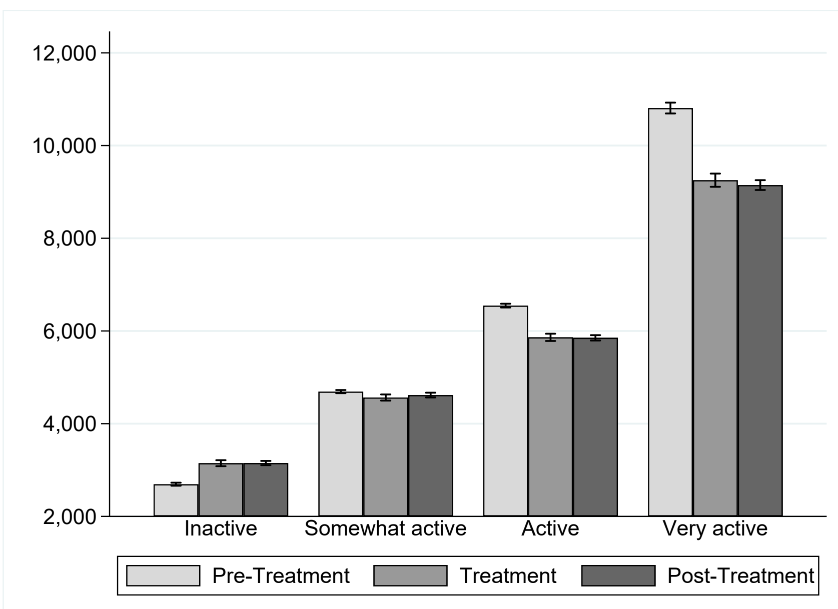
Notes: Figure A2 illustrates the fraction of users (in ‰) in the population by region. Overall, the user rate is homogeneous. The lowest user rate is in DROM-COM (i.e. 0.04‰) and the largest is in Hauts-de-France (i.e. 0.21‰).

Figure A3: Proportion of users opening the app by gender and day of the week.



Notes: Number of subjects equal to 14,215. Note that, 3,162 users of our sample do not provide their gender through the app.

Figure A4: *Weekly Steps* across the experiment window by category of walkers.



B Additional Tables

Table B1: Summary statistics.

	Peer/1W	Peer/3W	Self/1W	Self/3W	Sunk-cost/1W	Sunk-cost/3W	Control	N
Female	0.65	0.64	0.65	0.62	0.65	0.64	0.64	14,215
Average number of steps at baseline	6,050	5,994	6,037	6,144	5,989	6,011	6,092	17,377
% of subjects who walk > 10,000 steps	0.12	0.11	0.12	0.13	0.12	0.13	0.12	17,377
% of subjects who walk < 5,000 steps	0.44	0.46	0.45	0.44	0.46	0.45	0.44	17,377
Average number of connections at baseline	3.35	3.30	3.28	3.34	3.26	3.46	3.36	17,377
Probability of opening the app at baseline	0.53	0.52	0.53	0.51	0.52	0.52	0.53	17,377
N	2,223	2,114	2,160	2,201	2,132	2,193	4,354	

Table B2: Summary statistics: Strongly committed.

	Peer/1W	Peer/3W	Self/1W	Self 3W	Sunk-cost/1W	Sunk-cost/3W	Control	N
Female	0.71	0.72	0.73	0.56	0.63	0.68	0.67	495
	(0.56)	(0.52)	(0.38)	(0.17)	(0.57)	(0.87)		
Average number of steps at baseline	8,228	8,918	8,837	8,762	8,998	8,972	8,732	553
	(0.53)	(0.83)	(0.90)	(0.97)	(0.75)	(0.77)		
% of subjects who walk > 10,000 steps	0.29	0.35	0.33	0.30	0.35	0.25	0.28	553
	(0.92)	(0.29)	(0.48)	(0.81)	(0.29)	(0.62)		
% of subjects who walk < 5,000 steps	0.23	0.20	0.21	0.30	0.18	0.11	0.26	553
	(0.61)	(0.35)	(0.42)	(0.59)	(0.23)	(0.01)		
Average number of connections at baseline	12.86	14.59	12.50	14.26	11.55	15.47	14.32	553
	(0.53)	(0.93)	(0.44)	(0.98)	(0.22)	(0.61)		
N	66	65	67	64	65	80	146	

Table B3: *Number of connections* by category.

	Inactive	Somewhat active	Active	Very active
Before the treatment	2.78	3.26	3.77	5.40
In treatment				
- one-week	2.77	3.07	3.59	4.58
- three-week	2.77	3.03	3.50	5.39

Table B4: Distribution of subjects across treatment arms.

	Control	T1 peer		T2 sunk		T3 self	
		one-week	three-week	one-week	three-week	one-week	three-week
All sample	4,354	2,223	2,114	2,132	2,193	2,160	2,201
Weekly sample	2,251	1,197	1,112	1,078	1,116	1,112	1,075
- Inactive	545	308	263	269	254	284	263
- Somewhat active	587	291	288	266	295	290	265
- Active	569	304	288	267	287	266	250
- Very active	550	294	273	276	280	272	297
Daily sample	146	66	65	65	80	67	64

Table B5: Effects of treatments on *Probability of opening the app*: **Post-Treatment** vs. **Pre-Treatment**.

	Probability of opening the app			
	(1)	(2)	(3)	(4)
Post	-0.49***	-0.47***	-1.44***	-1.45***
	(-37.91)	(-31.61)	(-38.08)	(-37.65)
Peer 1W × Post	0.02		0.02	
	(0.82)		(0.82)	
Self 1W × Post	-0.02		-0.02	
	(-0.81)		(-0.81)	
Sunk-cost 1W × Post	0.00		0.00	
	(0.12)		(0.12)	
Peer 3W × Post		0.02		0.02
		(0.87)		(0.88)
Self 3W × Post		-0.01		-0.01
		(-0.22)		(-0.21)
Sunk-cost 3W × Post		-0.03		-0.03
		(-1.08)		(-1.08)
Observations	443,803	286,810	443,803	286,810
Subjects	10,321	9,890	10,321	9,890
Dummies of day			Yes	Yes

Notes: Logit model regressions with individual fixed effects. Columns 1-3 correspond to the one-week treatment arms, while columns 2-4 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively.

Table B6: Effects of treatments on *Number of connections*: **Post-treatment** vs. **Pre-treatment**.

	Number of connections			
	(1)	(2)	(3)	(4)
Post	-0.35***	-0.26**	-0.10	-0.32*
	(-4.94)	(-3.09)	(-0.79)	(-2.40)
Peer 1W × Post	-0.04		-0.04	
	(-0.36)		(-0.34)	
Self 1W × Post	0.06		0.06	
	(0.58)		(0.59)	
Sunk Cost 1W × Post	0.09		0.09	
	(0.96)		(0.94)	
Peer 3W × Post		-0.08		-0.08
		(-0.60)		(-0.62)
Self 3W × Post		-0.09		-0.09
		(-0.77)		(-0.77)
Sunk Cost 3W × Post		-0.04		-0.04
		(-0.34)		(-0.32)
Constant	4.04***	4.14***	4.08***	4.20***
	(159.31)	(183.73)	(60.58)	(73.94)
Observations	207,195	142,656	207,195	142,656
Subjects	10,736	10,398	10,736	10,398
Dummies of day			Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3 correspond to the one-week treatment arms, while columns 2-4 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

Table B7: Effects of treatments on *Steps walked*: **Post-treatment** vs. **Pre-treatment**.

	Steps walked			
	(1)	(2)	(3)	(4)
Post	-517.93***	-181.71***	-195.89*	-206.23*
	(-13.61)	(-4.16)	(-2.31)	(-2.24)
Peer 1W × Post	-41.25		-37.86	
	(-0.54)		(-0.49)	
Self 1W × Post	-11.60		-15.12	
	(-0.18)		(-0.24)	
Sunk Cost 1W × Post	-31.85		-40.49	
	(-0.47)		(-0.60)	
Peer 3W × Post		-61.69		-63.02
		(-0.81)		(-0.83)
Self 3W × Post		11.93		13.14
		(0.14)		(0.15)
Sunk Cost 3W × Post		-10.83		-2.08
		(-0.12)		(-0.02)
Constant	6311.31***	6355.87***	6724.69***	6735.17***
	(391.09)	(436.29)	(140.21)	(141.38)
Observations	206,760	142,353	206,760	142,353
Subjects	10,737	10,399	10,737	10,399
Dummies of day			Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3 correspond to the one-week treatment arms, while columns 2-4 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

Table B8: Effects on *Weekly Steps*: **Post-treatment** vs. **Pre-treatment**.

	Weekly steps			
	(1)	(2)	(3)	(4)
Treated	-567.76***	-197.73***	-113.89*	-57.79
	(-13.50)	(-4.18)	(-2.28)	(-1.09)
Peer 1W × Treated	32.67		32.67	
	(0.40)		(0.40)	
Self 1W × Treated	9.04		9.04	
	(0.13)		(0.13)	
Sunk Cost 1W × Treated	-31.93		-31.93	
	(-0.41)		(-0.41)	
Peer 3W × Treated		-49.25		-49.25
		(-0.61)		(-0.61)
Self 3W × Treated		-47.10		-47.10
		(-0.51)		(-0.51)
Sunk Cost 3W × Treated		-64.76		-64.76
		(-0.68)		(-0.68)
Constant	6166.54***	6230.98***	6143.06***	6201.02***
	(332.11)	(384.81)	(245.57)	(267.92)
Observations	33,828	22,216	33,828	22,216
Subjects	10,321	9,890	10,321	9,890
Week dummies			Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3 correspond to the one-week treatment arms, while columns 2-4 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

Table B9: Effects on *Daily Steps*: **Post-Treatment** vs. **Pre-Treatment**.

	Weekly steps			
	(1)	(2)	(3)	(4)
Post	-389.72*	97.49	-356.62	124.86
	(-2.10)	(0.43)	(-0.99)	(0.32)
Peer 1W × Post	-430.92		-430.92	
	(-1.44)		(-1.43)	
Self 1W × Post	-193.60		-193.60	
	(-0.65)		(-0.65)	
Sunk Cost 1W × Post	-184.56		-184.56	
	(-0.56)		(-0.56)	
Peer 3W × Post		-1054.22**		-1054.22**
		(-3.14)		(-3.14)
Self 3W × Post		-111.37		-111.37
		(-0.32)		(-0.32)
Sunk Cost 3W × Post		-597.20		-597.20
		(-1.30)		(-1.30)
Constant	8706.62***	8826.22***	9764.76***	9230.90***
	(113.45)	(117.82)	(38.17)	(39.37)
Observations	14,792	10,295	14,792	10,295
Subjects	344	355	344	355
Dummies of day			Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3 correspond to the one-week treatment arms, while columns 2-4 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, and 0.1% levels, respectively. Standard errors are robust and clustered at *customerID*.

Table B10: Effects on *Weekly Steps* by category: **Post-treatment** vs. **Treatment**.

	Weekly steps							
	Inactive		Somewhat active		Active		Very active	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	46.93	-84.62	-109.08	-180.71*	-116.23	-154.57	255.33	118.05
	(0.54)	(-1.01)	(-1.49)	(-2.50)	(-1.38)	(-1.90)	(1.61)	(0.75)
Peer 1W × Treated	-142.73		-19.06		-83.12		39.08	
	(-1.08)		(-0.17)		(-0.61)		(0.16)	
Self 1W × Treated	-242.34		-38.87		209.32		-245.78	
	(-1.89)		(-0.35)		(1.63)		(-1.15)	
Sunk Cost 1W × Treated	-226.50		-20.19		-6.04		-229.09	
	(-1.69)		(-0.17)		(-0.04)		(-0.96)	
Peer 3W × Treated		65.05		89.06		-189.09		305.06
		(0.60)		(0.92)		(-1.54)		(1.27)
Self 3W × Treated		58.64		99.76		283.62*		33.54
		(0.46)		(0.90)		(2.04)		(0.13)
Sunk Cost 3W × Treated		11.40		130.47		63.20		-4.28
		(0.09)		(1.24)		(0.53)		(-0.01)
Constant	3309.75***	3345.70***	4107.99***	4138.74***	6165.49***	6153.73***	9828.06***	9990.47***
	(90.95)	(82.82)	(125.73)	(122.78)	(158.30)	(160.88)	(142.42)	(110.09)
Observations	7,030	6,625	10,485	10,140	9,995	9,940	6,310	6,335
Subjects	1,406	1,325	2,097	2,028	1,999	1,988	1,262	1,267
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5-7 correspond to the one-week treatment arms while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*. All specifications include weekly dummies.

Table B11: Effects on *Weekly Steps* by category: **Post-treatment** vs. **Pre-treatment**.

	Weekly steps							
	Inactive		Somewhat active		Active		Very active	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	596.11***	659.00***	450.41***	502.31***	-147.22*	-154.21*	-1233.10***	-1074.39***
	(8.86)	(8.99)	(7.36)	(7.55)	(-2.09)	(-2.13)	(-8.48)	(-6.70)
Peer 1W × Post	-99.05		-5.60		156.77		3.48	
	(-1.04)		(-0.06)		(1.46)		(0.01)	
Self 1W × Post	36.68		-40.71		-24.13		118.60	
	(0.38)		(-0.50)		(-0.24)		(0.65)	
Sunk Cost 1W × Post	-19.63		-16.89		21.34		-107.61	
	(-0.20)		(-0.19)		(0.19)		(-0.51)	
Peer 3W × Post		-71.95		-119.24		87.45		-214.38
		(-0.62)		(-1.24)		(0.76)		(-0.93)
Self 3W × Post		6.58		27.66		2.53		-6.82
		(0.06)		(0.27)		(0.02)		(-0.02)
Sunk Cost 3W × Post		-85.82		-142.38		-67.15		150.93
		(-0.71)		(-1.43)		(-0.61)		(0.46)
Constant	2731.69***	2716.12***	3670.22***	3683.41***	6279.67***	6304.07***	11057.91***	11077.18***
	(89.57)	(92.60)	(128.88)	(138.07)	(166.37)	(186.69)	(147.91)	(152.96)
Observations	8,436	5,300	12,582	8,112	11,994	7,952	7,572	5,068
Subjects	1,406	1,325	2,097	2,028	1,999	1,988	1,262	1,267
Week dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS model regressions with individual fixed effects. Columns 1-3-5-7 correspond to the one-week treatment arms while columns 2-4-6-8 correspond to the three-week treatment arms. *, **, *** denote significance at the 5%, 1%, 0.1% level, respectively. Standard errors are robust and clustered at *customerID*. All specifications include weekly dummies.