

Discussion Paper No. 18-041

**Experimental and
Non-Experimental Evidence
on Limited Attention
and Present Bias at the Gym**

Paul Muller and Wolfgang Habla

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 18-041

**Experimental and
Non-Experimental Evidence
on Limited Attention
and Present Bias at the Gym**

Paul Muller and Wolfgang Habla

Download this ZEW Discussion Paper from our ftp server:

<http://ftp.zew.de/pub/zew-docs/dp/dp18041.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

Experimental and non-experimental evidence on limited attention and present bias at the gym

Paul Muller*

Wolfgang Habla**

September 23, 2018

Abstract

We show that limited attention and present bias contribute to low levels of exercise. First, in a large randomized experiment, we find that email reminders increase gym visits by 13 % and that they benefit nearly all types of individuals. Limited attention can explain these effects. Second, using a novel dataset, we find that many bookings for gym classes are canceled, and that bookings are made even for classes that never have a waiting list. Comparing these findings to the predictions of a dynamic discrete choice model, we conclude that many gym members use bookings to commit themselves to future attendance.

Keywords: health behavior, randomized experiment, reminders, nudging, habit formation, limited attention, time inconsistency

JEL-code: C93, D91, I12

*Department of Economics, Gothenburg University. Vasagatan 1, 405 30 Gothenburg. Paul.muller@gu.se

**Centre for European Economic Research (ZEW) and University of Mannheim. L7 1, 68161 Mannheim. habla@zew.de

We are grateful to Fysiken for their cooperation. In particular, we thank Maria Ljungqvist for agreeing to work together and granting us access to Fysiken's data; Hanna Mattson for all her help in accessing the data; and Sara Hagge for her assistance with designing and translating the email reminders. We would also like to thank Marius Alt, Anna Bindler, Fredrik Carlsson, Andreas Dzemski, Martin Dufwenberg, Christina Gravert, Mitesh Kataria, Raphael Krause and Joseph Vecchi for their many useful comments and suggestions. The paper also greatly benefited from discussions with participants at the internal ZEW Econometrics Seminar, the Labour/Applied Micro Economics Seminar at the University of Gothenburg and the Nordic Conference on Behavioral and Experimental Economics in Gothenburg. This study has been registered with the AEA RCT registry (AEARCTR-0001847).

1 Introduction

Recent research shows that the prevalence of insufficient physical exercise is over 40 % in high-income Western countries and has been increasing over the past two decades (Guthold et al., 2018). Given the overwhelming evidence on the long-term health benefits of regular exercise and the related improvement in well-being, such low levels of physical activity are both surprising and worrisome. Since exercising is costly in terms of time and effort, low levels of activity can be interpreted to be a result of individual optimization, where immediate costs are weighted against delayed benefits. There are, however, a number of reasons from both a societal and an individual perspective why the prevalence of exercise might be below its optimal level.

First, positive externalities of regular exercise have been documented, including reduced sick leave and higher work productivity (Kool et al., 2004; Van Amelsvoort et al., 2006). Moreover, most people are (to some degree) insured against medical expenses such that their behavior is prone to moral hazard. Both of these phenomena may lead to underinvestment in preventive activities such as regular exercise. Second, behavioral anomalies and biases may play an important role in investment decisions that involve immediate costs and long-term (delayed) benefits. One such investment is healthy behavior, and in this context, present bias and limited attention have been identified as the most likely causes of underinvestment (DellaVigna and Malmendier, 2006; Calzolari and Nardotto, 2017). As a result, individuals may not be able to achieve levels of exercise that they find individually optimal in the long run.

In this paper, we investigate the importance of these two popular behavioral explanations in the repeated decision to exercise, using very detailed data from a gym. First, we evaluate a randomized experiment among almost 2,500 gym members in which we sent weekly email reminders to half of the sample. We find a robust positive effect of these reminders on the number of gym visits, which we argue is most naturally explained by the prevalence of limited attention among gym members. Our contribution lies in our use of a large sample size, which allows for precise estimates, a sample of gym members who are arguably more representative of the general population than those used in previous studies, and a gym that offers a wide range of activities and sports.¹ All of these factors speak towards the strong external validity of our analysis. Second, by exploiting novel data on booking, canceling and attending gym classes, we investigate the importance of present bias. The possibility to book a spot in gym classes provides us with information on the making and changing of plans regarding future visits to the gym. We document a range of empirical observations and compare them to predictions from a simple dynamic discrete choice model. We conclude that many gym members book

¹We use data from a gym that offers activities ranging from regular weight-lifting to ball sports, yoga, cycling, swimming and dancing classes. As a result, we are likely to observe a larger share of members' exercise activities than we would find at an average gym.

classes in advance simply to commit themselves to exercising in the future. Since commitment problems stem from present bias, we infer that time-inconsistent preferences are an important factor influencing the decision to exercise. Below, we describe the two parts of the paper in more detail and briefly discuss the related literature.

Limited attention is defined slightly differently across studies (sometimes, it is called ‘limited memory’). For instance, Taubinsky (2014) defines limited attention as the probability to forget to take an action, while Ericson (2017) defines limited memory as the probability to forget about the completion of *planned* tasks. Dean et al. (2017) model inattentive consumers who tend to focus their attention on a small number of alternatives (the ‘attention set’ or ‘consideration set’) when faced with a large choice set. What these interpretations have in common is that a decision maker, due to cognitive limitations, does not take the full set of choices or actions into consideration. In the context of exercising, gym members may forget to visit the gym on a particular day, forget to book a gym class in advance or forget to attend a previously booked class. Reminders are a potential remedy for limited attention because they can direct people’s attention to a particular choice or action and thus make the execution of this choice or action more likely. To test the importance of limited attention, we performed a large-scale randomized experiment in which we sent weekly email reminders to gym members and observe these members’ subsequent gym attendance. The setup is related to the reminder study by Calzolari and Nardotto (2017) but is applied to a much larger and more representative sample and a much more diverse set of activities and individuals.² The reminders were sent during a three-month period to gym members who had recently purchased an annual membership. While the content varied across weeks, the emails were always short, containing only text encouraging regular visits to the gym.³

We find that email reminders lead to a substantial and statistically significant increase in gym visits of 13 %. The effect appears to be somewhat larger for class training (17 %) than for free training (12 %), although the difference is not significant. These estimates follow from a difference-in-differences (DiD) model in which we control for a slight pre-experiment difference in attendance between the control and treatment groups, but we find similar effects when using simple differences. These findings are robust to a range of different specifications.

²The gym members we study exercise roughly once per week, which is arguably closer to typical attendance rates than those documented in previous studies. For example, Calzolari and Nardotto (2017) consider only students that volunteered to participate in the study, and they report an average attendance of almost two visits per week, acknowledging that this figure is high compared to typical gym attendance.

³Reminders have been shown to be effective in a variety of domains. For example, they increase the amount of charitable giving (Sonntag and Zizzo, 2015; Damgaard and Gravert, 2018), monthly savings (Karlan et al., 2016a) and loan repayments (Karlan et al., 2016b); they improve rule compliance at the library (Apesteguia et al., 2013) and in traffic (Chen et al., 2017); they also increase the payment of TV license fees (Fellner et al., 2013), the claiming of tax credits (Bhargava and Manoli, 2015) and college enrollment (Castleman and Page, 2015); they decrease misdemeanor defendants’ failure-to-appear rates in court (Bornstein et al., 2012) and residential energy consumption (Allcott and Rogers, 2014). In contrast to these findings, Himmler et al. (forthcoming) show that reminders are not effective when students study for, sign up for, or take exams.

Furthermore, the positive impact of email reminders applies to nearly all types of individuals (students, non-students, men, women, etc.). Only low attenders (based on the pre-experiment period) seem to benefit substantially more than high attenders, which is not surprising, given that high attenders typically have a lower scope to increase their exercising frequency. When extending the observation period to three months after the last reminder was sent, we find evidence that the positive effects of reminders persist (visits remain 13 % higher in the treated group). We interpret this finding as a sign of habit formation, although it should be noted that these results are less precise and less robust.

The positive impact of regular reminders on gym visits clearly suggests that limited attention does indeed play a major role in the context of gym attendance and that email reminders are effective at overcoming such cognitive limitations.⁴

In the second part of the paper, we use a novel and detailed dataset on the booking, canceling and attendance of gym classes to investigate whether intertemporal choices are affected by time inconsistency and whether people use class bookings as a commitment device. Time-inconsistent preferences imply that a decision maker's preferences over current and future choices may change over time, such that a preference for a certain choice or action at one point in time can become undesirable at another point in time. More specifically, individuals have been found to assign more weight to payoffs that are closer to the present time when considering trade-offs between two future moments. This behavior is usually referred to as present bias (O'Donoghue and Rabin, 1999a,b, 2001).⁵ The decision to visit the gym is particularly vulnerable to present bias, as it entails immediate (time and effort) costs and delayed (health) benefits.

Each class offered at the gym has a limited capacity, and gym members can enroll in a class up to one week before the class. If a member changes his or her mind, he or she must cancel his or her booking (or pay a fine). Using detailed data on the timing of each booking and cancellation, we provide insights on how individuals make intertemporal choices, i.e., how and when they make plans for future gym visits and how likely they are to follow through on their plans. We first document a number of behavioral patterns in the data and then establish a simple two-period discrete choice model to demonstrate how these decisions

⁴An alternative explanation for why reminders are effective could be guilt aversion (see, e.g., Battigalli and Dufwenberg, 2009), which suggests that guilt-averse players experience a utility loss if they believe that they let someone down. In our case, gym members could feel obliged to visit the gym when they receive a reminder email because they believe that the gym – or, rather, the gym's employees – expect them to come to the gym. This explanation seems more plausible for less anonymous settings in which the gym facility is relatively small and there is a small number of employees. The health club with which we cooperated is large, and therefore, members are unlikely to see the same staff members at the gym every time they attend.

⁵Present bias is frequently modeled using (quasi-) hyperbolic (beta-delta) discounting. This type of temporal discounting assumes that discount rates decline over time (Phelps and Pollak, 1968; Laibson, 1997). Hyperbolic discounting induces an individual to consume more today and invest less in the future than she would like from an ex ante perspective (see Frederick et al., 2002 for a thorough discussion of hyperbolic discounting and time discounting, as well as time preferences more generally).

are affected by uncertainty about the future (in terms of the effort cost of attending the class), time-inconsistent preferences (present bias) and limited attention. Comparing the empirical findings with the predictions from the model, we conclude that many bookings are made only to commit oneself to attending the class.⁶ Commitment devices are useful only for individuals who have commitment problems, that is, individuals who have time-inconsistent preferences. Thus, we show that gym members suffer from time inconsistency when deciding (repeatedly) whether to visit the gym. These findings complement recent studies that demonstrate the demand for and effectiveness of commitment contracts at the stage of buying a gym membership (e.g., Goldhaber-Fiebert et al., 2010; Bhattacharya et al., 2015; Royer et al., 2015). Note that our results suggest that not only time inconsistency but also uncertainty about the future costs of attending a class play a large role in decisions with regard to booking, canceling and attending classes. Finally, the fact that a substantial share of bookings results in ‘no-shows’ (members who do not attend and do not cancel), further supports the hypothesis that limited attention is an important factor.

The rest of this paper is structured as follows. Section 2 describes the setup for the reminder experiment. In Section 3, we analyze the effects of reminders on the number of gym visits and consider robustness and heterogeneous effects. Section 4 presents the booking and canceling data, establishes a simple dynamic discrete choice model, and compares empirical findings and model predictions. Section 5 concludes.

2 Experimental setup

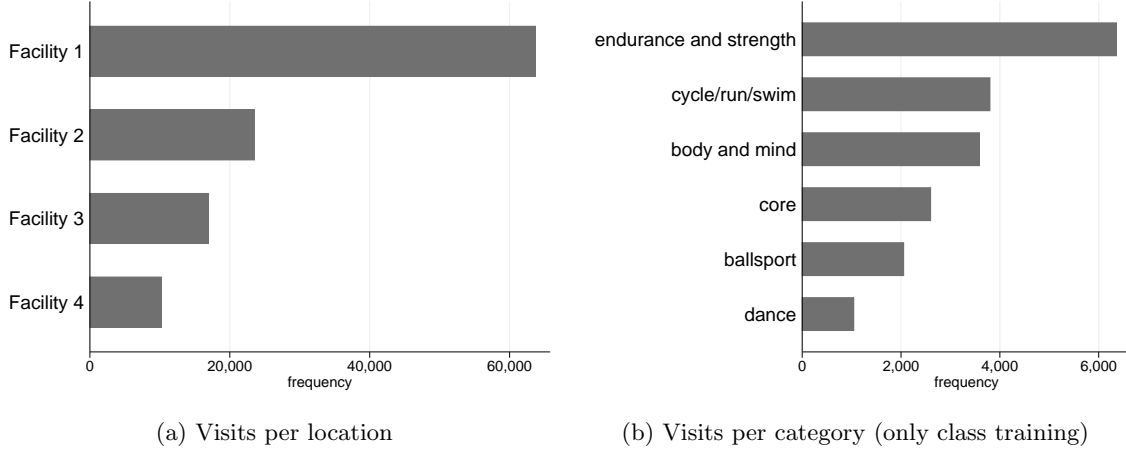
Below, we briefly describe the health club at which our study was performed, the population of members who were part of it, and the setup and design of the reminder experiment.

2.1 The health club

The study was carried out in collaboration with one of the largest health clubs in Gothenburg (Sweden). The club is a non-profit organization, has four facilities and records over 800,000 visits per year. Three facilities offer free workout areas, as well as a large range of group exercise classes led by fitness instructors (more than 200 per week). The fourth facility is a large climbing hall. Contracts can be bought for access to a single facility, or to all facilities (‘Multi card’). Access to each of the four facilities requires the scanning of an electronic membership card. Participation in a group exercise class requires an additional scan. Gym

⁶The intuition behind this finding is the following. We find that even for class types that are unpopular and never fully booked, most attendees make a booking prior to the class. This finding holds even when selecting only experienced members who have attended the same class several times and are therefore certainly aware of the unpopularity of the class. Since such bookings are clearly unnecessary in order to secure a spot, this behavior can be explained only by gym members using bookings as a commitment device. See our model in Section 4.2 for further discussion.

Figure 1: Gym attendance (July 1, 2016 - June 30, 2017)



attendance is thus almost perfectly observed. Because many classes are at risk of being crowded, each class has a participant limit, and members can sign up for classes in advance. We discuss the club’s booking and cancellation policy in Section 3.4.

From the health club, we received all data on class attendance, class bookings, class booking cancellations and class no-shows for the time period from July 1, 2016, to June 30, 2017, for all individuals who purchased a 12-month contract between July 1, 2016, and December 1, 2016 (this is the experimental sample, see Section 2.2). In Figure 1, we show descriptive statistics. Altogether, we observe 115,232 gym visits. Panel (a) shows the distribution of free training sessions (95,791) across the four facilities. Facility 1 is the most popular facility, with 50 % of all visits, while the other 50 % of the visits are spread across the other three facilities. We observe 19,441 class attendances (approximately 17 % of all visits). More than 37 different types of classes were offered, and we split them into six main categories (based on the classification used by the health club): endurance and strength, body and mind, cycle/run/swim, core, ball sports, and dance.⁷ The distribution of participation across these six categories is presented in panel (b) of Figure 1, while the distribution of visits across the 37 specific class types is shown in Figure 6 in the appendix. The most popular classes are indoor cycling (spinning), ‘krafttag’ (muscle strength training using free weights) and yoga.

⁷The actual number of different classes is much larger if we account for, e.g., differences in length, focus, instruction language etc. Classes included in the 6 main categories are: 1 (endurance and strength): Bodycombat, Bodystep, Cirkelgym, CrossFit, Fight, Fys, Gymclass, Gympa, Krafttag, Powercage, Powercircuit, Tabata; 2 (body and mind): Bodybalance, Mindfulness, Pilates, Triggerstretch, Yoga; 3 (cycle/run/swim): Spinning, Indoor walking, Swim-training; 4 (core): Cxwork, Functional training, Q-performance, TRX; 5 (ball sports): Badminton, Basketball, Football, Indoor hockey, ‘Spökboll’, Volleyball; 6 (dance): Dancing classes.

2.2 Sample selection, randomization and balance

As mentioned above, we focus on all individuals who purchased a 12-month contract (rather than a one- or three-month contract), between July 1, 2016, and December 1, 2016. This approach provides us with 2,881 individuals. We select individuals who remained active members on December 1, 2016, which was true for almost all members, as an annual membership cannot be canceled during its duration.⁸ Furthermore, we exclude members who did not provide a valid email address when registering with the gym, leaving us with a sample of 2,463 gym members.

We randomly assigned these individuals into treatment and control groups in the first week of December 2016. The experiment, which we describe in detail below, began on January 9, 2017 (see Figure 2 for a timeline of our setup). We present descriptive statistics in Table 1, which contain the mean, standard deviation, minimum and maximum for the control group (columns 1-4) and for the treatment group (columns 5-8), and the p-value of a t-test for equal means (column 9). The top panel presents individual characteristics, the second panel membership and contract characteristics. In our sample, the average age is 30, and 58 % of the gym members are males. In total, 52 % registered as students and 46 % had been members of this gym previously. A total of 40 % bought a Multi card and 5 % bought a membership that is valid only during daytime hours (a ‘daytime card’).⁹ The mean date of purchasing the contract is September 23 for both groups (Figure 7 in the Appendix shows that the full distributions of contract buying dates in the control and treatment groups are very similar). Most variables do not differ significantly between control and treatment groups, with the exception of a slightly lower share of recurring members and a slightly higher share of members who bought a daytime card in the treatment group.

The five-week period of pre-experiment gym visit observations (December 2, 2016 - January 8, 2017) allows us to compare the control and treatment groups in terms of attendance. In the lower panel of Table 1, we find that the treatment group has a significantly lower number of weekly visits in the pre-experiment period, and this difference is caused by a lower number of free training visits. In Section 3.1, we discuss how we control for these differences by including individual fixed effects when estimating the impact of the reminders. Note that the number of weekly visits is lower in this pre-experiment period than is typically the case because this time period includes the Christmas holidays, during which gym attendance is low. In Figure 8 in Appendix A, we compare the distributions of gym visits of the control and treatment groups across facilities, type of class and number of visits, and we find that the control and treatment groups are similar along these dimensions.

⁸A few exceptions occurred when newly enrolled members canceled their membership within a few days, or due to, for example, serious injury or moving.

⁹With a daytime card, members can exercise before 3:30 pm on weekdays but not on weekends.

Table 1: Characteristics of the treatment and control groups

	Control group				Treatment group				p-val. t-test
	mean	sd	min	max	mean	sd	min	max	
Age	30	10	14	79	30	11	16	80	0.77
Male	0.59	0.49			0.58	0.49			0.50
Student	0.51	0.50			0.54	0.50			0.10
Recurring member ^a	0.49	0.50			0.44	0.50			0.01
Membership type:									
Multi card (and CrossFit) ^b	0.41	0.49			0.40	0.49			0.66
Facility 1	0.32	0.47			0.34	0.47			0.43
Facility 2	0.10	0.30			0.10	0.31			0.73
Facility 3	0.12	0.32			0.10	0.30			0.34
Facility 4	0.05	0.23			0.06	0.23			0.86
Daytime card	0.04	0.19			0.06	0.23			0.05
Date of purchasing contract ^c	23sep2016	38	01jul2016	01dec2016	23sep2016	39	01jul2016	01dec2016	0.66
Weekly (pre-experiment) attendance: ^d									
Gym visits	0.97	1	0	7	0.88	0.96	0	5.8	0.02
Free training	0.83	0.98	0	6.8	0.74	0.90	0	5.8	0.03
Class training	0.15	0.39	0	3.4	0.14	0.38	0	3.2	0.51
Observations	1232				1231				

^a A 'Recurring member' is someone who had an active contract at some point during the six months prior to our sample inflow period (January 1, 2016 - June 30, 2016).

^b A CrossFit membership is similar to a 'Multi card' membership but also allows entry to CrossFit.

^c For the distribution of the dates of contract purchases in the control and treatment groups, see Figure 7 in the Appendix.

^d Attendance statistics are computed over the period from December 2, 2016, to January 8, 2017 (this is the pre-experiment period for which we have attendance data for all individuals).

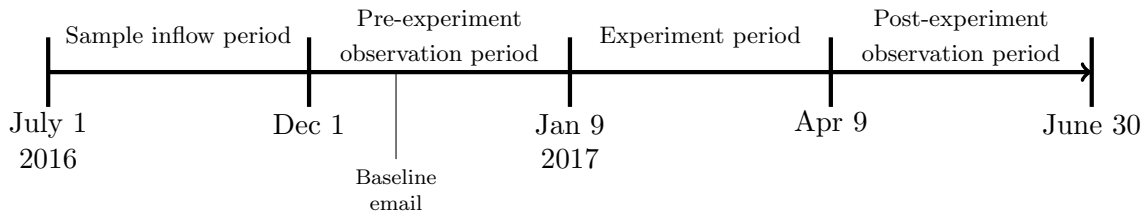


Figure 2: Timeline of the experiment

2.3 Experiment details

The intervention consisted of a series of email reminders that were dispatched to the treatment group between January 9, 2017, and April 9, 2017 (see Table 10 in Appendix B for the dispatch dates). The reminders were sent from the gym’s email account once per week, and they contained a short message (typically no longer than three sentences) encouraging the recipient to attend the gym. In addition, they contained a picture related to the gym and a link to the club’s website (for an example, see Figure 9 in Appendix B). The text and the picture differed every week in order to attract the recipients’ interest and make them more likely to open and read the reminders. Most emails referred to the various exercise options available at the club or reminded the reader about the potential benefits of attending the gym.¹⁰ An example of a reminder is the following: *Did you know that we offer more than 200 classes per week? Drop in any time to work out, or sign up for a class on www.fysiken.nu/en/book/. We hope to see you soon!* The complete list of reminders can be found in Appendix B.

Each reminder email contained a line at the bottom explaining that the readers had received this email because of their gym membership and that they could unsubscribe by clicking a link. When members unsubscribed, they were removed from our email list and did not receive any further email communication related to the study.

Receiving emails from the gym is not unusual for members. The gym has a policy of sending out an automated welcome email when a new member signs up. Furthermore, automated emails are sent when a new member has not attended the gym for 90 days (though these emails were suspended during our experiment). A few times per year, all members receive a newsletter (unless they have unsubscribed from this service). Both the control and the treatment groups received such a newsletter once during the experiment. The starting date and the end date of the email reminders were unknown to participants. Finally, the days on which the reminders were dispatched were randomized between Sundays, Mondays and Tuesdays in order to make sure that members in the treatment group did not expect to receive a reminder

¹⁰We opted for generic reminders, as several papers have shown that adding other contents to general reminders or applying gain versus loss frames does not change their effectiveness (Apesteguia et al., 2013; Altmann and Traxler, 2014; Karlan et al., 2016a). Only in the domain of electricity consumption does the inclusion of social comparisons or social norms seem to significantly alter the effects of messages (Schultz et al., 2007; Nolan et al., 2008).

on a particular weekday.

The effectiveness of email reminders obviously depends on whether the recipient receives (and notices) the reminders. An email might not be received because, for example, members have misspelled their email address or registered an address that they fail to use, or the emails have ended up in their spam folders. Even if the email appears in the user’s inbox, it may be deleted without being opened (although in that case, its impact could already have been achieved if it brought the gym to the top of the receiver’s mind). Since email reading behavior is likely to depend on individual characteristics, selecting a sample of recipients who have responded to an email, as in Calzolari and Nardotto (2017), may imply that the sample is also more responsive to the following reminders than the average gym population.

Analyzing the extent to which such selection issues drive the results is difficult because whether someone opens and reads a particular email is not typically observed. We address this issue in two ways. First, we observe a measure of whether each email has been opened using so-called ‘web beacons’, which record whether the picture included in the email has been downloaded. This measure of email opening is not perfect because not every email client downloads the picture by default.¹¹ An email can thus be read without downloading the picture, implying that our obtained measure provides a lower bound for the opening rate of emails.

Second, we sent out a ‘baseline email’ to our entire sample about five weeks prior to the start of the study. In this email, we briefly described the widely known challenge of achieving one’s exercising goals and asked the recipients whether they believed that regular email reminders could help them improve in this regard. If they thought that regular reminders could indeed be helpful, they were asked to click a button in the email.¹² The baseline email provides a measure of which members – in both the control and treatment groups – open emails. In addition, when someone clicks on the button, we know with certainty that the email was read. We will make use of the web bugs and the baseline email in the empirical analysis to identify a subgroup of individuals who are likely to open and read emails so that we can estimate heterogeneous effects for this group (see Section 3.3).

2.4 Opening of emails and unsubscribing

Before we analyze the impact of our reminders, we show some statistics on the email opening rate for our reminders and on members’ unsubscription from our email list.

Weekly unsubscribing is measured after each reminder email. It is important to note that

¹¹Gmail, for example, automatically downloads the picture once the email is opened, and downloading statistics are thus a perfect proxy for opening statistics. For clients such as Outlook, however, the user can choose whether to download the picture or not.

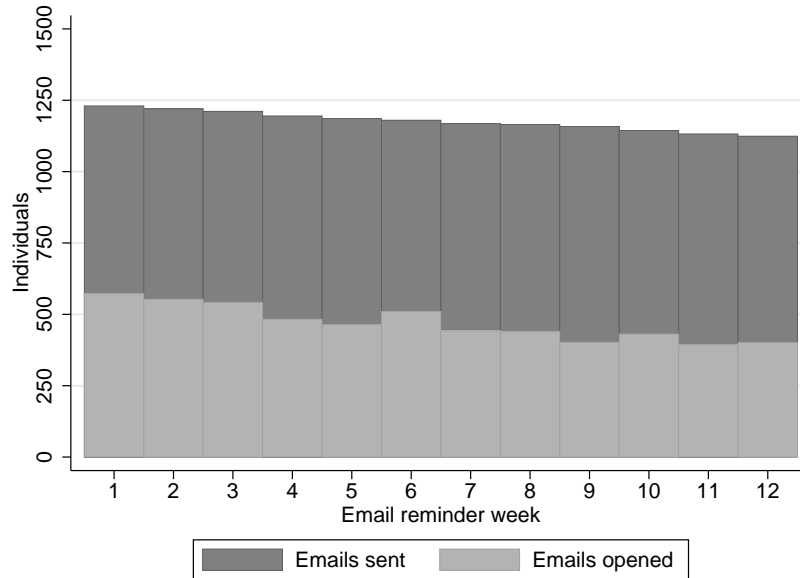
¹²The email explained that the gym was considering offering such a service in the future and would like to collect its members’ general opinion on whether they believed this practice would be useful. We made it very clear in the email that indicating interest did not imply signing up to any service or future reminders.

Table 2: Regression of the email opening rate on individual characteristics

	(1)
	Share of reminders that is opened
Male	-0.066*** (0.020)
Age <20	0.17*** (0.048)
Age 21-25	0.29*** (0.037)
Age 26-30	0.29*** (0.038)
Age 31-40	0.25*** (0.037)
Age >40	0.36*** (0.038)
Student	0.028 (0.022)
Recurring member	0.024 (0.020)
Opened the initial email	0.11*** (0.018)
Clicked link initial email	0.100** (0.046)
Observations	1111
R^2	0.166

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. In addition to the variables above, six email client dummies are also included in the regression (Gmail, Hotmail, Live, Outlook, University, Other). The sample contains only the treatment group and excludes 120 individuals for which the gender variable is missing.

Figure 3: Number of emails sent and opened



unsubscription is not equivalent to attrition in our study. We observe gym attendance (or non-attendance) for all individuals in the study. Unsubscription implies only that fewer emails are received. In Figure 3, we present the weekly number of emails that were dispatched. While the number of email addresses on our list declined steadily, 91 % of all individuals in the treatment group remained on the mailing list at the end of the intervention. Thus, weekly unsubscription is approximately 1 %. In the same figure, we show the number of delivered emails that we know with certainty were opened, i.e., those for which the picture was downloaded. Considering that this percentage is a conservative estimate, the opening rate is quite substantial, beginning at 46 % in the first week and decreasing to 36 % in the final week.

To assess whether certain types of individuals are more likely to open the email, we regress the share of reminders that are opened on several individual characteristics. The results are shown in Table 2 (obviously only the treatment group is included). We find that men are significantly less likely to open emails (-7 %-points). The same holds true for gym members who are less than 20 years old (with the difference between this group and most other age groups being more than 10 %-points). We find no differences for students or recurring members. Furthermore, we find that responding to the baseline email is, unsurprisingly, a good predictor of opening the reminder emails. Opening the baseline email is associated with an 11 %-points increase in the share of reminders that are opened during the study. Clicking the link in the baseline email adds another 10 %-points increase. Even though much unexplained variation remains in the opening rates, the included set of variables explains a

non-trivial share of the variation ($R^2 = 0.16$).¹³

3 Experimental results

In this section, we discuss the effect of our intervention on gym attendance for both the short term (during the intervention) and the long term (after the intervention). We also consider the effect on class bookings and the cancellation of bookings. Furthermore, we estimate heterogeneous effects and consider individuals' contract renewal decisions.

3.1 Impact on gym attendance

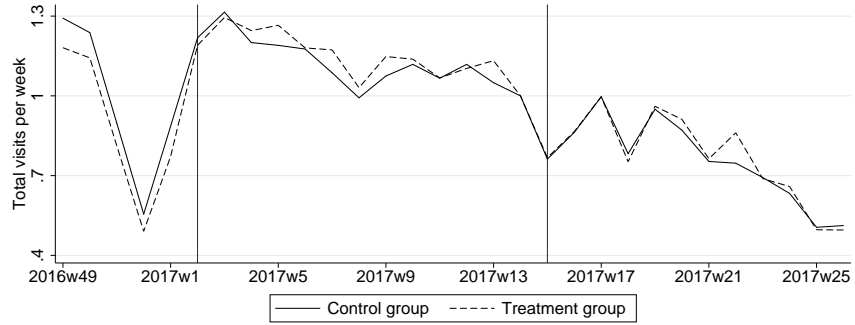
We show the average weekly number of gym visits in the control and treatment groups in panel (a) of Figure 4. On average, individuals visit the gym slightly more than once per week during the whole study period (with the exception of a large dip around the Christmas break). The time trend is negative and can have three possible explanations. First, because our experiment begins shortly after the new year, it is in line with a 'fresh start effect' at the beginning of the year and a subsequent decline in attendance.¹⁴ Second, our sample contains members with new contracts, and attendance typically declines over the course of the membership (e.g., DellaVigna and Malmendier, 2006). Finally, higher temperatures and longer hours of sunshine might drive gym members from indoor activities toward outdoor activities in spring and summer.

The first vertical line in Figure 4 indicates the start of the experiment period in which the email reminders were sent. As mentioned earlier, the control group has a slightly higher gym attendance than the treatment group in the weeks before the experiment begins. Almost directly after the reminders begin, this pattern reverses and the treatment group has a slightly higher attendance than the control group (panel a). We can split gym visits into free training and class training visits. In panel (b), we show the average number of free training sessions per week, which is just below one visit. In panel (c), we find that the impact of the reminders on class training is particularly pronounced. Class attendance is higher for the treatment group in almost all experiment period weeks, while the two groups have almost identical pre-experiment class attendance. Finally, in panel (d), the difference between the treatment and control groups for each of the three variables is depicted. The diagram clearly shows that

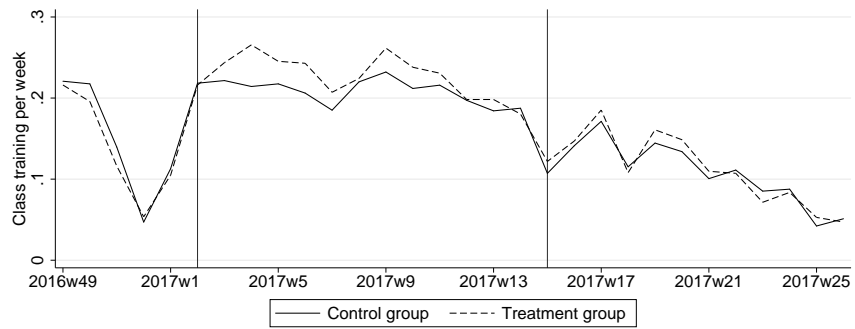
¹³We will use the estimated coefficients in a robustness analysis to obtain predicted email opening rates not only for the treatment group but also for the control group. Since these predicted values are based only on baseline (pre-experiment) variables, we can use them to classify individuals as 'likely to open emails' and 'unlikely to open emails' and subsequently estimate treatment effects separately for these two groups. This process allows us to assess whether opening the email is essential for the reminder to affect behavior. We return to this issue in Section 3.3.

¹⁴This phenomenon is found by Dai et al., 2014, who describe how salient temporal landmarks can increase aspirational behaviors such as gym visits.

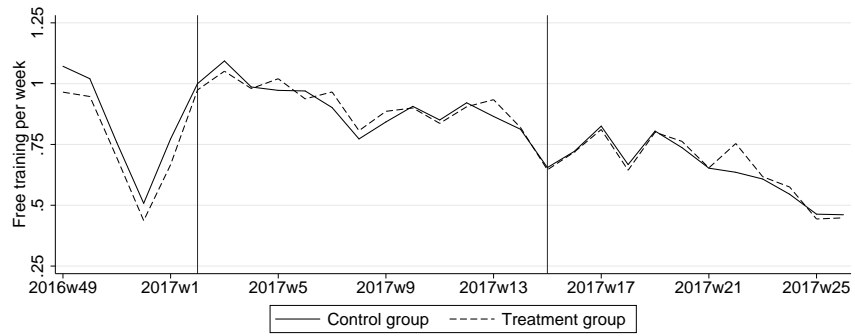
Figure 4: Weekly gym attendance



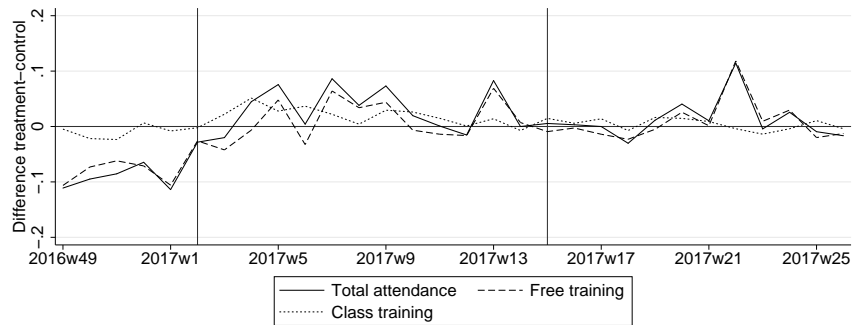
(a) Total visits



(b) Class training



(c) Free training



(d) Treatment-control difference (weekly)

the lower attendance of the treatment group in the pre-experiment period reverses to higher attendance after the email reminders begin.

The second vertical line indicates the end of the intervention, after which no further reminders are sent. We observe attendance until June 30. After that, the vacation period begins, during which gym attendance is traditionally very low and the number of offered classes is greatly reduced. Furthermore, extending the post-experiment beyond this day would imply that we would lose participants in our sample since some of them will not prolong their contracts. The negative trend in weekly attendance clearly continues after the experiment ends. While average weekly attendance is 1.25 in January, it is only 0.5 at the end of June. The difference between the control and treatment groups is smaller than it is during the experiment but still suggests a slightly higher average attendance for the treatment group (especially when taking the pre-experiment difference into account). As shown in panel (d), for each of the three lines, the difference is either (slightly) positive or close to zero during the post-experiment period.¹⁵

To obtain estimates of the impact of the reminders, we compute the average weekly attendance in the pre-experiment period and experiment period for the control and treatment groups.¹⁶ The simple difference-in-differences (DiD) estimate for the total number of gym visits is 0.12 additional visits per week as a result of the email reminders, which is equivalent to an 11 % increase. The additional visits consist of 0.03 additional classes (15 % increase) and 0.09 additional free training sessions (11 % increase). Over the 12 weeks of the experiment, this effect aggregates into 1.5 extra visits on average for each individual who received the reminders. Given that marginal costs of the intervention are close to zero, this shift can be considered a substantial effect.

We now turn to a regression framework and estimate a DiD model. The impact of the reminders is identified from the assumption that in the absence of the intervention, the control and treatment groups would have had the same time trend in weekly gym attendance. Weekly gym attendance (A_{it}) of individual i in week t depends on week fixed effects α_t , individual fixed effects γ_i and treatment assignment T_{it} , which is a dummy equal to 1 if individual i is in the treatment group and week t is in the experiment period. Furthermore, we control for membership duration in a flexible way by including a set of dummies for different membership duration intervals, which is indicated by π_d . We include membership duration because of the abovementioned negative time trend in attendance and because individuals in our sample

¹⁵There is a particularly huge spike in attendance of the treatment group in week 22, mostly as a result of an increase in free training. After checking with the health club, we found that several emails were sent by the gym on May 29, May 30 and May 31, with the latter being sent to all individuals who had a valid contract and, of course, registered a valid email address. We later perform several robustness checks with respect to this particular week.

¹⁶To take into account the correlation between an individual's multiple observations, we first compute individual averages before estimating the group mean and standard error of the mean. The resulting numbers, including standard errors, are shown in Figure 10 in the Appendix.

entered their current membership at different dates.

$$A_{it} = \alpha_t + \beta_1 T_{it}^{during} + \beta_2 T_{it}^{post} + \pi_d + \gamma_i + \epsilon_{it} \quad (1)$$

The parameters of interest are β_1 and β_2 , which measure the immediate effect of the email reminders and the long-term effect of the reminders, respectively. In panels (c) and (d) of Figure 8 in Appendix A, it is clear that weekly gym attendance is a count variable with many zeros. Thus, a Poisson regression is more suitable than a linear regression. We report the results for a linear model as a robustness check (see Section 3.2) and find that the results are very similar. We estimate standard errors clustered by individual.

The regression results are presented in Table 3, where we report $\exp(\beta) - 1$, which is the percentage effect. We first discuss the immediate impact of the reminders (β_1), which is shown in the first row. We find that total visits increase by 13 %, which is significant at the 1 % level (column 1). We split these visits into class training and free training in columns (2) and (3). In column (2), we find that attendance in class exercise training increases by 17 %, while free training visits increase by 12 % (column 3). These percentage effects suggest that the reminders led predominantly to more class training visits. However, the much lower base level for class training visits implies that the impact in terms of additional visits is greater for free training (and the standard errors imply that the difference between the two estimates is not statistically significant).¹⁷

As a robustness check, we also present regressions in which we exclude the pre-experiment period and thus do not include individual fixed effects (columns 4-6 in Table 3). Instead, we include a number of time-invariant individual covariates (a gender dummy, a student dummy, seven age category dummies and a dummy for being a recurring member of the gym).¹⁸ The results are similarly positive, although somewhat smaller in magnitude. Combined with somewhat larger standard errors, only the effect on the total number of visits is statistically significant. This finding was expected, as the slightly lower attendance level of the treatment group in the pre-experiment period is ignored in these regressions. Note that while the control and treatment groups are balanced with respect to most observed characteristics (see Table 1), the share of recurring members is 5 %-points higher in the control group. Recurring members typically exercise more frequently, which may explain the slightly higher pre-experiment number of visits within the control group. Thus, we prefer the DiD model that corrects for these (and potentially other) differences.

So far, we have looked at the immediate effect of email reminders on gym attendance.

¹⁷There could be positive spillovers from the treatment group to the control group via peer effects. Treated individuals could inspire untreated ones to meet up and exercise with them more frequently. If that is true, our study underestimates the positive effects of reminders, although we expect that this bias is rather small.

¹⁸For 230 individuals (9 % of the sample), some of these covariates are missing, such that these individuals are excluded from the regressions with covariates.

Table 3: Regressions on weekly gym attendance (Poisson models)

	DiD model			Simple difference		
	(1) Visits	(2) Class training	(3) Free training	(4) Visits	(5) Class training	(6) Free training
Experiment (during)	0.13*** (0.035)	0.17** (0.090)	0.12*** (0.038)	0.072* (0.040)	0.12 (0.097)	0.059 (0.044)
Experiment (post)	0.13*** (0.046)	0.10 (0.11)	0.13*** (0.050)	0.059 (0.053)	0.032 (0.12)	0.064 (0.058)
Recurring member				0.21*** (0.050)	0.15 (0.11)	0.22*** (0.057)
Male				0.21*** (0.048)	-0.66*** (0.034)	0.61*** (0.075)
Student				0.0020 (0.048)	0.060 (0.13)	-0.0040 (0.052)
Daytime card				-0.31*** (0.074)	-0.69*** (0.13)	-0.21** (0.087)
Week & duration FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	no	no	no
Include pre-period	yes	yes	yes	no	no	no
N	66900	28320	65250	55825	55825	55825

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Seven age dummies and six gym location dummies are included but not reported in columns (4)-(6). Note that the number of observations in columns (1)-(3) varies because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects.

One might expect that once the reminders cease, treated individuals will stop attending the gym at an increased frequency. Habit formation might, however, explain persistent effects of a short-term intervention. Habits are acquired when ‘repetition of a behavior in a consistent context progressively increases the automaticity with which the behavior is performed when the situation is encountered’ (Lally et al., 2010, p. 998).¹⁹ In our context, this phenomenon means that the email reminders would provide a stimulus upon which individuals decide to go to the gym, and the increased frequency of going makes it easier for individuals to attend. Only weak evidence of habituation has been found by studies in the field of health behavior (see, e.g., Charness and Gneezy, 2009; Acland and Levy, 2015; Calzolari and Nardotto, 2017), although this might be due partly to a lack of statistical power.

We present estimates of the impact of the reminders during the post-experiment period (β_2) in the second row of Table 3. We find the surprising result that the post-experiment impact on total visits is of a similar magnitude as the immediate impact (a 13 % increase). The same result holds for the impact on free training visits (column 3). The coefficient for

¹⁹For early work on the relationship between repetition and habit strength, see Hull (1943, 1951); more recent contributions on habits and habit formation include Verplanken (2006) and Wood and Neal (2007).

class training in column (2) is slightly smaller, which, combined with less precision due to a lower base rate, leads to the impact not being statistically significant. However, the magnitude of the class training coefficient is very similar to that of the free training coefficient. A simple difference model (columns 4-6) again leads to slightly smaller coefficients that are positive but not statistically significant.

3.2 Robustness checks

We perform a series of robustness checks to investigate whether these results are sensitive to certain choices in the empirical model. First, we estimate the baseline models with simple OLS to assess the sensitivity of the results to the Poisson model. The results are presented in Table 11 in Appendix C. The results are very robust. The statistically significant increase of 0.095 weekly visits translates into a 10 % increase, which is very similar to the Poisson regression estimate (13 %). Additionally, the effects on class training (16 %) and on free training (8 %) are very similar to the Poisson estimates. When using a simple difference model (columns 4-6), the coefficients are slightly smaller, and only the effect on total visits is statistically significant (at the 10 % level).

Second, we show that removing the Christmas holidays from the pre-experiment period hardly affects the estimates. During the last two weeks of 2016 (weeks 51 and 52) and the first week of 2017 (week 1), gym attendance was substantially below its long-run level (see Figure 4). In particular, in week 52, attendance was extremely low. Since these weeks are part of the pre-experiment period, they may affect the DiD estimates. In Table 12 in Appendix C, we exclude either week 52 or weeks 51, 52 and 1 from the analysis. Note that the total pre-experiment period consists of weeks 49, 50, 51, 52 and 1, and thus, we can estimate the DiD model even when excluding the Christmas holidays. Excluding either one or three weeks hardly affects our results. In all columns, we find statistically significant positive effects on gym visits, which are very close to our baseline estimates.

Third, we consider the length of the pre-experiment period. In all DiD estimations, the pre-experiment period consists of five weeks. This period is restricted by the fact that our sample consists of gym members who purchased their contracts between July 1, 2016, and December 1, 2016. As a robustness check, we select a subsample of individuals who purchased their contracts earlier (July 1 - November 1) and extend the pre-experiment period by four weeks. The results are presented in Appendix C in Table 13 and are similar, although the magnitude and significance of the estimates are smaller in the DiD model (but not in the simple difference model).

Fourth, we consider using daily data on gym visits (instead of aggregated weekly data). We use a specification that is similar to the one used for weekly visits, but we include additional weekday fixed effects. The results are presented in Table 16 and Figure 11 in Appendix C.

We find that the impact of reminders is very similar to our baseline model in terms of both magnitude and significance. Note that in the simple difference models, we can also include two weather dummies (‘Sunny day’ and ‘Rainy day’) and find that while sunny days lead to lower gym attendance, rain has no impact.

Finally, we perform two robustness checks with respect to the spike in attendance in the treatment group in week 22. We show that the positive post-experiment effects are robust to (i) excluding week 22 from the post-treatment period (see Table 14 in Appendix C) and (ii) restricting the post-treatment period to all weeks up to (but not including) week 22 (Table 15 in the Appendix). Thus, the positive long-term effects are not (solely) driven by this spike. In week 22, the gym sent an email to all its members (including both the treatment and control groups). We hypothesize that the spike suggests that treated gym members were more responsive to the email than control group members. This phenomenon could mean that the reminder emails sensitized individuals to respond to a similar stimulus in the same way.

To conclude, we find clear evidence of a persistent positive effect of reminders on attendance. Interestingly, the percentage effects of the reminders are quite similar for the experiment period and the post-experiment period. We interpret this finding as a sign of habit formation. Note that the average number of weekly visits is lower in the post-experiment period, and thus, the same percentage effects imply a lower increase in terms of the number of visits in the post-experiment period.

3.3 Heterogeneous effects

The impact of email reminders may vary, depending on an individual’s characteristics. Given our large sample size, we can investigate different sources of heterogeneous effects, and we focus on gender, baseline gym attendance, previous gym membership (at the same gym), student status and a pre-experiment measure of the likelihood of opening emails. We simply estimate the same model for each particular subsample. Identification follows from a similar standard DiD assumption. For example, in the case of heterogeneous effects by gender, we assume that in the absence of the intervention, men in the treatment group would have had the same trend in attendance as men in the control group (the same holds for women). We report results only for the total number of weekly gym visits (rather than distinguishing between classes and free training), and for simplicity, we consider only the short-run effect of the reminders (and thus include only observations until the end of the intervention period).

We first consider heterogeneous effects by gender. While men and women have similar average numbers of weekly visits (1.16 for men and 1.00 for women), they exercise in different ways. One-third of gym visits by women are class training, while the corresponding number is less than 10 % for men. In Table 4, we present estimation results for men in column (1) and

Table 4: DiD regressions on weekly gym attendance (Poisson models): heterogeneous effects

	(1)	(2)	(3)	(4)	(5)
	Men	Women	High attenders	Low attenders	Recurring members
Experiment (during)	0.13*** (0.046)	0.11* (0.059)	0.084*** (0.034)	0.17** (0.078)	0.13*** (0.047)
N	23104	16416	22211	19665	19418

	(6)	(7)	(8)	(9)	(10)
	New members	Student	No student	Opened baseline	Not opened baseline
Experiment (during)	0.13*** (0.053)	0.13*** (0.050)	0.13*** (0.049)	0.12*** (0.045)	0.15*** (0.057)
N	22458	22135	19741	24244	17632

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. All regressions include individual fixed effects and membership duration dummies.

those for women in column (2). We find that the estimates for the impact of the reminders are very similar, and we conclude that reminders are equally effective for both genders.

Second, we consider heterogeneous effects by baseline gym attendance. We split the sample at the median weekly attendance in the pre-experiment period, which is 0.6 visits per week. The estimation results are presented in column (3) (high attenders) and column (4) (low attenders) and suggest a greater impact of email reminders on members who attended less than the median. This finding is in line with the results from Calzolari and Nardotto (2017) and are not particularly surprising since for high attenders, there is a lower scope to increase the amount of exercising beyond the current level (e.g., due to time constraints). Note, however, that a formal test again cannot reject that the coefficients for high and low attenders are equal.

Third, we test whether recurring members (those who had a contract at the same gym prior to the current contract) are more or less affected by the reminders. Since recurring members extended their previous contract, they are likely to be more committed to exercising, and thus, we expect them to benefit less from email reminders. However, we do not find evidence supporting this hypothesis, with the coefficients for the two groups being almost identical (columns 5 and 6). This finding suggests that the group of recurring members is perhaps less selective than one might expect.²⁰

Fourth, we consider student status. As shown in Table 1, approximately half of our

²⁰A potential explanation is that many new members of this particular health club may have previously attended a different gym and are thus equally experienced and committed.

sample consists of students.²¹ While we have no strong hypothesis for differing impacts of the reminders for students and non-students (except for maybe differently binding time constraints), we still believe it is important with respect to external validity. For example, previous findings on the effectiveness of reminders by Calzolari and Nardotto (2017) were based only on a student population, and thus, we are interested in whether the impact extends to a more general population. The results are presented in columns (7) and (8), and we find that the estimates are basically identical for students and non-students. Thus, the positive effect is present even for a population that is more likely to be (full-time) employed and that has (perhaps) more stringent time restrictions for exercising.

Finally, we consider heterogeneous effects according to whether one opens the baseline email (1,405 out of the 2,463 individuals in the sample opened this email; see Section 2.3). Ideally, one would measure whether those who actually opened the reminders were more likely to visit the gym. However, since those members are a selective subset of the treatment group and because we do not observe opening rates for the control group, we have no valid control for those who opened the reminders. Thus, we use opening the baseline email as a proxy for opening gym emails in general. The results for those who opened the email are presented in column (9) and for those who did not open the email in column (10). We find the rather surprising result that there is no significant difference between the two groups. The coefficient for those who did not open the baseline email is even slightly larger than the coefficient for those who opened the email.²² Thus, even for those who are (relatively) less likely to open and/or read emails, we find a significantly positive effect. While this finding is somewhat puzzling, one explanation could be that opening the email is not a necessary condition for the message to be received. Seeing the email appear in one's inbox could be sufficient to bring the gym to one's attention and thus overcome limited attention. However, an alternative explanation is that the statistics on downloading web bugs severely underestimate the actual opening rate, meaning that our results are (to some extent) diluted by measurement error.

In summary, we do not find strong evidence for heterogeneous effects of email reminders. The impact does not appear to vary by gender, recurring membership, student status or the likelihood of opening emails. We find only some suggestive evidence that low attenders are more affected than high attenders, but the difference is not statistically significant. This finding suggests that nearly all groups benefit from reminders in terms of overcoming their limited attention, which strengthens the potential for reminder policies.

²¹Student status is recorded in the data only if one (voluntarily) reports being a student at the time he or she purchases the contract. Thus, this measure might not be a perfect measure of student status. However, given the sizable discount that students receive (around 30 %), most students will likely disclose this information.

²²We can extend this approach by predicting who opens emails, based on a range of individual characteristics (including whether the baseline email was opened). To do so, we regress (for the treatment group) the share of the 12 email reminders that were opened on individual (pre-experiment) characteristics. Splitting the sample at the median predicted opening rate, we again find that there is no significant difference between the effectiveness of the reminders for those who are likely to open the emails and those who are less likely to open the emails.

3.4 Booking and canceling

In the previous sections, we established that gym members who received email reminders attended gym classes more frequently both during the treatment period and in the three months after the reminders had ended (although the long-run effect is not significant). As we have detailed data on class bookings and booking cancellations, we now consider whether reminders lead to more bookings, fewer cancellations or both. The following analysis sheds light on whether reminders lead only to an (almost) immediate response of visiting the gym or whether they also stimulate the recipient to plan a future visit by booking a class. Before we proceed with the empirical analysis, we briefly describe the features of the booking and cancellation system in place at the health club. Note that we postpone a more thorough discussion of booking and canceling behavior to Section 4.

The classes offered at the gym have a maximum capacity. Members can book a spot in a class online or through a mobile app up to seven days prior to the class.²³ Once all spots are booked, new bookings are placed on a waiting list and result in a spot when earlier bookings are canceled. A booking can be canceled online or through the mobile app up to one hour prior to the class. In the last 60 minutes before the class, members without a booking can claim a spot with the gym’s reception staff if a spot is still available (‘drop-in attendance’). Finally, those with a prior booking must show up at least 10 minutes prior to the class; otherwise, the booking is canceled and the spot becomes available for drop-in attendance. If a booking is not canceled in time and the individual does not show up, his or her card is blocked from further bookings until a fine of SEK25 (approximately \$3) is paid. We refer to these cases as ‘no-shows’. In the analysis in this section, we label no-shows simply as cancellations for simplicity, but we discuss the difference in detail in the next section.

It is not uncommon for all spots in a class to be taken, often several days in advance. For example, of all 9,513 classes for which we have data, 30.7 % had a waiting list (at some point). This figure does not mean that all spots were eventually taken for one-third of the classes because canceling is also very common (for more details, see the next section). However, it does illustrate that booking classes beforehand is often necessary and very common. To keep the analysis simple, we use only the first and the last observable actions, i.e., the very first booking and the observation of whether the person attended or canceled, and we ignore all actions in between.²⁴

We estimate a DiD model with individual fixed effects, similar to equation (1), but with different outcome variables. First, we estimate the impact of the reminders on the number of weekly bookings using a linear model (column 1 in Table 5). We find a positive impact of

²³A member is limited to five active bookings at a time.

²⁴For the majority of observations, this makes no difference, but a non-trivial share of observations are somewhat more complicated. Many alternative scenarios are possible, for example, a member may book a class, then cancel it, and then attend through a last-minute drop-in.

Table 5: Regressions on weekly class bookings and cancellations of bookings

	(1)	(2)	(3)	(4)	(5)
	Number of bookings (linear)	Number of bookings (Poisson)	Positive # of bookings (LPM)	Number of cancellations (linear)	Share of bookings canceled
Experiment (during)	0.056* (0.031)	0.11 (0.076)	0.014* (0.0080)	0.026 (0.018)	-0.023 (0.026)
Experiment (post)	0.016 (0.031)	0.049 (0.092)	0.0013 (0.0083)	0.0028 (0.018)	-0.038 (0.028)
Mean value dep. var.	0.42	0.42	0.16	0.18	0.37
Percent effect (during)	13		9	14	-6
N	73890	33270	73890	73890	11653

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. In column (2), we report $\exp(\beta) - 1$, which is the percentage effect. Note that the number of observations is lower in columns (3)-(4) because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects. Columns (5)-(6) are linear probability models in which the outcome has been set to 1 in the case of a positive number of bookings/cancellations.

0.056 bookings, which is significant at the 10 % level and corresponds to a 13 % increase in the number of bookings per week. There is no impact after the experiment ends. Since the number of bookings is a count variable, we also estimate a Poisson regression in column (2). The result is again positive and very similar in magnitude (11 % increase), although it is not statistically significant. Given the relatively small share of gym members who attend classes, we also consider the extensive margin of making *any* positive number of bookings in column (3). We find that reminders lead to significantly more members making at least one booking (a 9 % increase).

We now consider cancellations. In column (4), we show that the number of weekly cancellations increases during the experiment; however, the coefficient is small and not statistically significant. Cancellations are obviously possible only if a booking was made earlier. Thus, we consider in column (5) the share of bookings that are canceled, and as a result, this regression includes only observations for weeks in which the individual made at least one booking. We find a negative coefficient, although it is again small in magnitude and not statistically significant. We conclude that the increase in class visits as a result of the reminders is due to (1) an increase in the number of bookings and (2) a lack of an increase (and perhaps even a decrease) in the *share* of bookings that are canceled.

The effects of the reminders for the post-experiment period all share the sign of the effect during the experiment. However, none of the effects is statistically significant. This finding is in line with the estimated impact on class attendance for the post-experiment period, which is smaller in magnitude and not statistically significant (see Table 3).

Table 6: Contract renewal after initial 12-month period

	Control group	Treatment group	p-value test equal proportions
Full sample:			
Active after 13 months	0.57	0.59	0.55
Active after 14 months	0.56	0.55	0.92
Observations	1232	1231	
Upfront (12-month) paid contract:			
Active after 13 months	0.49	0.48	0.77
Active after 14 months	0.51	0.48	0.39
Observations	579	554	
Monthly direct debit payments:			
Active after 13 months	0.65	0.67	0.38
Active after 14 months	0.60	0.61	0.59
Observations	653	677	

3.5 Impact on contract duration and renewal

We have seen that there are persistent effects of the intervention in terms of increased attendance in the post-experiment period. We now investigate whether these increases carry over to the decision on contract renewal once the 12-month contract period ends. One would expect that individuals who were induced by the intervention to exercise more regularly are more likely to remain members of the gym. Rational decision makers base this decision on projections of future gym attendance (though in practice, many may underestimate their future self-control, suffer from overconfidence, etc.; see, e.g., DellaVigna and Malmendier, 2006). The reminders might increase projections of future attendance if past experiences are used as a basis for predictions about the future. It is doubtful though whether the magnitude of the increase in attendance suffices to affect the (larger) decision of buying another contract.

We obtained data on whether individuals buy a new contract (or extend the first contract) and define as the outcome measure an indicator for being a member 13 or 14 months after purchasing the initial contract.²⁵ Table 6 shows the share of individuals in the control and treatment groups that were active members 13 and 14 months after the initial contract began. In the upper panel, we observe that slightly more than half of our sample had active contracts both 13 and 14 months after the initial contract began, while the share decreased slightly from month 13 to 14. The share of active members is very similar in the treatment and control groups, and the difference is not significant (with a p-value of 0.55 after 13 months and 0.92 after 14 months).

The type of payment of the initial contract (upfront or monthly) is likely to affect the

²⁵We include both measures (13 and 14 months) to take potential (short) gaps between the initial contract and the new contract into account.

likelihood of contract renewal, and thus, we report the same statistics, split by the type of payment, in the lower panels of Table 6. Again, we find no significant differences between the control and treatment groups for both payment methods after 13 months or after 14 months. Interestingly, the number of active contracts is much higher among those who had initially chosen direct debit payments compared to those who had paid upfront. This finding suggests that there is some kind of cancellation delay for contracts with direct debit payment, as also reported by DellaVigna and Malmendier (2006).²⁶ This delay can also be explained by limited attention, because for these contracts, an active choice (the choice to terminate the contract) must be made, and members might frequently forget to make this decision. Overall, we do not find any impact of our reminders on contract duration and renewal, which suggests that persistent effects of an intervention fade once a decision at a margin different from that targeted in the intervention is made.

4 Booking and canceling behavior

In this section, we go beyond the analysis of the experimental data pertaining to the reminders. We first document a number of observations related to behavioral patterns in the data on gym class bookings, booking cancellations and class attendance. A standard classical choice model assuming a rational individual with perfect foresight and the mental capacity to keep track of all decisions would never predict that the individual will book a class and subsequently cancel the booking. Therefore, we develop a theoretical model that deviates from this simple assumption and incorporates several factors and mechanisms that influence an individual’s intertemporal choices (limited attention, time-inconsistent preferences, i.e., present bias, and uncertainty about future attendance costs). Finally, we discuss which observations can be explained by which factors.

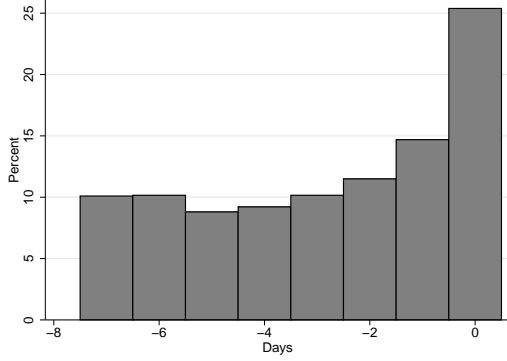
4.1 Descriptive statistics and behavioral patterns in the data

For the following analysis, we use an observational dataset that is much larger than the one for the reminder experiment. This dataset contains observations for class bookings, cancellations and no-shows among *all* gym members for the period from July 1, 2016, to June 30, 2017 (for more details on the booking system, see Section 3.4). We observe approximately 190,000 bookings and 28,000 drop-in attendances by over 7,000 gym members. Our first two observations are related to the booking of classes and cancellation of bookings.

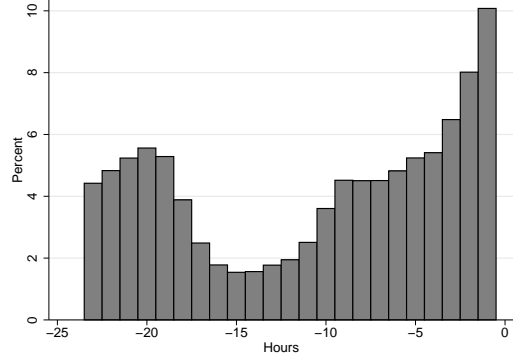
Observation 1 *Most gym class visits are booked prior to the class (78 %) and are not made through drop-in.*

²⁶Note that it cannot be excluded that (part of) the difference is due to self-selection into the two payment types.

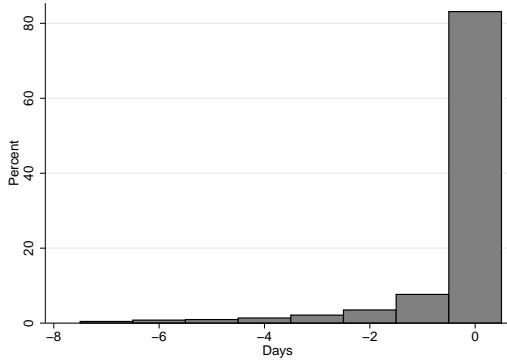
Figure 5: Distribution of bookings and cancellations



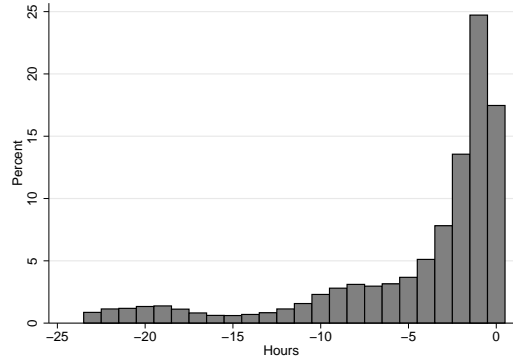
(a) Bookings in the week prior to class



(b) Bookings in the 24 hours prior to class (25.4 % of all bookings)



(c) Cancellations in the week prior to class



(d) Cancellations in the 24 hours prior to class (83.1 % of all cancellations)

In the upper panel of Figure 5, we show the distribution of class bookings during the week before a class (note that bookings can be made beginning seven days prior to a class). In panel (a), we find that the number of daily bookings is approximately uniform across the first six days prior to a class (roughly 10 % per day). Around 15 % of all bookings are made on the day before a class and another 25 % on the day of the class ('day zero'). Most of the bookings on day zero are made in the very last hours before the class (panel b).²⁷ This finding shows that there is considerable interest in booking classes quite early and that an increasing number of people book a class in the final days and even in the last hours before the class.

Observation 2 *A significant share of bookings are canceled (41 %), and most of these cancellations occur on the final day before the class (83%).*

Thus, canceling is very common, as can be seen from the first column of Table 7. In Figure

²⁷The low level of bookings between the final 10 and 18 hours before the class are likely due to this time often being nighttime, as a majority of classes are held in the late afternoon or evening hours.

Table 7: Bookings, cancellations and drop-in attendance by class popularity

	All classes	Unpopular classes (< 0.5% waiting list)	Fairly popular classes (< 5% waiting list)	Popular classes (> 5% waiting list)
Number of bookings	190454	24912	40543	124999
Of those (in %):				
Attendance (%)	0.53	0.57	0.53	0.52
Cancellation (%)	0.41	0.38	0.41	0.41
No-shows (%)	0.057	0.057	0.057	0.058
On final waiting list (%)	0.008	0.000	0.001	0.012
Number of drop-in attendance	27828	10024	7281	10523
As % of all attendance	0.22	0.42	0.25	0.14
Number of classes	238	98	44	96

The categorization of classes by popularity is made based on the share of all bookings that goes to the waiting list (which is, due to cancellations, much larger than the share that eventually remains on the list when the class occurs). The ‘Cancellation’ share includes cancellations by people on the waiting list. Being ‘on the final waiting list’ means that a person is still on the list when the class occurs (and thus is not allowed to participate).

5 (c), we show how cancellations are distributed by booking day (no-shows are included in the figure as canceling zero minutes prior to the class). We see that the vast majority (83.1 %) of all cancellations occur on the day of the class, and 62.8 % of all cancellations occur within the final six hours (panel d). Thus, we find that most people who change their minds and cancel a booking do so at the very last moment.

Our dataset contains large variation across classes in terms of their popularity. While some classes are never fully booked and thus never have a waiting list, others are (almost) always fully booked. Arguably, the popularity of a class, which is exogenous from an individual’s point of view, must have an impact on booking and cancellation decisions. For example, if a particular class is always fully booked, there is a need to book in advance because the class cannot be attended otherwise. If, by contrast, a class is never fully booked, there is no obvious need to book this class (unless the class’s unpopularity is unknown to the individual).

To define the popularity of a class, we take the following approach. First, we define a class relatively broadly by the type of activity (yoga, cycling, etc.) and its scheduled time (day of the week plus ‘daytime’ or ‘evening’). This approach leads to 238 different types of classes (see Appendix C.3 for more details). In the next step, we categorize these classes into ‘unpopular’, ‘fairly popular’ and ‘popular’ based on the share of bookings that end up on the waiting list. For a popular class, more than 5 % of all bookings appear on the waiting list. Given a typical class size of 20, there will thus always be a waiting list for this type of class on average (with at least one person on this list). For a fairly popular class, between 0.5 %

and 5 % of the bookings end up on the waiting list; for an unpopular class, less than 0.5 % of bookings end up on the waiting list. Applying this procedure implies that the likelihood of a waiting list appearing for an unpopular class is close to zero. Overall, 40 % of all classes are categorized as popular, and half of these even have a waiting list share of more than 20 %. The next two observations are related to bookings and booking cancellations for unpopular classes.

Observation 3 *Even unpopular classes, for which waiting lists are extremely rare, are booked well in advance (58 % of all visits are booked in advance).*

In columns (2)-(4) of Table 7, we present statistics on the number of bookings, cancellations and drop-in attendances for the three popularity categories. We find that, as expected, attendance through drop-in occurs more often in unpopular classes than in the other two categories. However, even in unpopular classes, where waiting lists virtually never appear, more than half of all class participants make a booking. Not only do many attendees of unpopular classes make a booking, but many do so very early. For example, approximately 30 % of the bookings are made at least three days in advance (see Figure 12 in the Appendix).

Observation 4 *For unpopular classes, a significant share of bookings are canceled (38 %).*

In Table 7, we show that of all bookings, only between 52 and 57 % are actually followed by class attendance (with unpopular classes having a slightly larger share than popular classes). The remaining share of all bookings is not followed up on (thus resulting in a fine), as the next observation states.²⁸

Observation 5 *There is a substantial share of no-shows (6 % of all bookings) across all popularity types.*

Again, this can be seen from Table 7. As a final observation, we consider the evidence from the experiment that class participation increases after receiving reminders:

Observation 6 *Gym members who receive regular reminders from the gym (i) make more bookings, (ii) are less likely to cancel their bookings and (iii) attend more classes.*

The first finding is based on the result shown in Table 5 (columns 1 and 2) that the weekly number of bookings increases during the experiment. The second finding is based on the same table (column 5) and suggests that the share of bookings that are canceled is lower when reminders are received. We acknowledge, however, that this result is far from statistically significant and suggests only weakly that canceling becomes less likely as a result of the reminder emails. The third finding, that class attendance increases when reminders are received, is most robust (Table 3, column 2).

²⁸A small share of all bookings for popular classes remain on the waiting list (1 %). Note that this small share does not necessarily mean that excess demand is rare. Members might cancel their bookings while on the waiting list or decide not to book in the first place if a waiting list exists.

4.2 A model of booking, canceling and attendance

We now proceed by laying out a simple two-period discrete choice model that describes an individual's decisions on booking, canceling and attending a class. The goal is to investigate the relevance of several behavioral biases, and the model provides a framework to think about them. Although the model is highly stylized, it contains the key elements that play a role in the decision-making process. In particular, we assume that (i) there is uncertainty about the future attractiveness of attending a class (and corresponding beliefs might be overoptimistic), (ii) there is uncertainty about whether the class will be fully booked, (iii) there is a risk of forgetting about a class that one would actually like to attend and for which one might have even made a booking, and (iv) preferences can be time-inconsistent (and individuals are either naïve or sophisticated regarding their time inconsistency).

Consider an individual who can attend a class that takes place in period $t = 2$. In period $t = 1$, she decides whether or not to book a spot in the class, and in period two, she decides whether or not to attend the class.²⁹ Her second-period utility, U_2 , depends on two (binary) state variables: whether she made a booking in the first period (B , equal to one in case of a booking and zero otherwise) and whether the class is fully booked (F , equal to one if the class is fully booked and zero otherwise). As a result, there are four possible states of the world in the second period, where the individual decides to attend ($A = 1$) or not to attend the class ($A = 0$). Only if she made no booking ($B = 0$) and the class is fully booked ($F = 1$) is there no possibility to attend and thus no decision to be made.

Attending the class is associated with time and effort cost c_A . In period two, these costs are assumed to be known (they are not known in period one; see below). The returns to attending the gym are a future stream of health benefits u_3, \dots, u_T , which are assumed to materialize in period three (one period after attending the class) and the subsequent periods. The period-two utility from attending the class is given by $-c_A + \beta \sum_{t=3}^T \delta^{t-2} u_t$. For simplicity, we denote the present value of these future (accumulated) health benefits by $\bar{U} \equiv \sum_{t=3}^T \delta^{t-2} u_t$. We apply the $\beta - \delta$ (quasi-) hyperbolic discounting framework (Phelps and Pollak, 1968; Laibson, 1997) to allow for time-inconsistent preferences. The per-period discount factor is $\delta \leq 1$, while payoffs between the present and all future periods are additionally discounted with factor β . For $\beta < 1$, discounting is quasi-hyperbolic and preferences are time-inconsistent, while for $\beta = 1$, discounting is exponential and preferences do not change over time. We will discuss the individual's knowledge about β later. Depending on the two state variables, optimal

²⁹Note that period $t = 1$ represents all days prior to the class date. The model can be extended to seven periods, with each representing one day and the last representing the day of the class or, more specifically, the time the class takes place. In each period, a booking can be made or canceled. While complicating the model significantly, this extension will not enrich the key mechanisms in any way, and thus, we opt for the simplest model that still contains all relevant mechanisms.

period-two utility is given by

$$U_2(B, F) = \begin{cases} \max_{A=0,1} [A \cdot [\beta\bar{U} - c_A]] & \text{if } B = 0, F = 0 \\ 0 & \text{if } B = 0, F = 1 \\ \max_{A=0,1} [A \cdot [\beta\bar{U} - c_A] + (1 - A) \cdot [-c_T]] & \text{if } B = 1 \end{cases} \quad (2)$$

If the individual did not book in period one and the class is not full, she attends if the discounted benefits ($\beta\bar{U}$) outweigh the costs of attending (c_A). If she did not book in period one and the class is full, she cannot attend, there is no choice to make and she obtains her outside option, which is normalized to zero. If she booked in period one, it is irrelevant whether the class is full. Again, she attends if the benefits outweigh the costs; however, not attending requires canceling the booking through the website or app, which is assumed to be associated with a transaction cost of c_T (this cost also applies to making a booking in period one, see below). If the individual does not show up to a booked class, she has to pay a fine c_F . We assume that $c_F > c_T$, so it is always optimal to cancel the booking when one does not want to attend the class.³⁰

Now, we move on to the period-one decision on whether or not to book a spot in the class. The process by which the class becomes fully booked (or not) is exogenous to the individual, and thus, we assume that there is simply a probability $(1 - \gamma)$ that between periods one and two, all available spots are booked. Furthermore, we need to specify what information the individual has in period one. We assume that the costs of attending the class in period two are unknown to the individual in period one. This assumption is reasonable, as uncertainty exists about many factors that influence the cost of attending (e.g., the weather, one's plans, one's energy and motivation level), and it is common in the literature (e.g., DellaVigna and Malmendier, 2006). Rather than knowing the precise cost, the individual obtains a signal c_A^* , which is correlated with the true value c_A .³¹ Given her prior belief about the distribution of the cost and the observed signal, the individual uses Bayesian updating to the distribution of the cost of attending the class.³²

Finally, we include limited attention in the model. In particular, we assume that in period two, there is a probability $(1 - \pi)$ that the individual forgets about the class, regardless of whether a booking has been made. If a booking was made, this individual will be a no-show; if no booking was made, she will simply not attend. This interpretation of limited attention

³⁰Given that it takes only a minute to cancel the booking and the fine for not attending a booked class is \$3, individuals would need to have a very high value of time in order for canceling to be more costly than paying the fine. This situation seems rather implausible, especially for the student population.

³¹Depending on the joint distribution of c_A^* and c_A , the signal might be unbiased. Alternatively, it might reflect systematic over- or underestimation of future attendance costs (and thereby the likelihood of attendance). We discuss overconfidence with respect to future attendance costs in more detail below.

³²Individuals who have been gym members for a while are likely to have a good sense of their true distribution of the attendance cost and use this distribution as the prior.

is similar to that of, e.g., Taubinsky (2014). Given her expectations about the future cost of attending, the likelihood of the class being fully booked and the risk of forgetting, she decides whether to book a spot during period one. Her optimal utility in period one is given by

$$U_1 = \max_{B=0,1} \left[\begin{aligned} & B \cdot [\pi\beta\delta\mathbb{E}_1[U_2(B=1, \cdot)] - (1-\pi)\beta\delta \cdot c_F] - c_T + \\ & (1-B) \cdot \pi\gamma\beta\delta\mathbb{E}_1[U_2(B=0, F=0)] \end{aligned} \right] \quad (3)$$

If she decides to book ($B = 1$) and does not forget (with probability π), she obtains the expectation of the value $U_2(B = 1, \cdot)$, discounted with factor $\beta\delta$. If she *does* forget (with probability $1 - \pi$), she will pay c_F in period two for not showing up to the booked class, discounted with $\beta\delta$. In either case, she immediately incurs the transaction cost c_T (for making the booking). The expectation \mathbb{E}_1 is taken conditional on period-one information. If she does not book, she obtains with probability $\pi\gamma$ the period-one expectation of $U_2(B = 0, F = 0)$, again discounted with $\beta\delta$. If she does not book and the class is fully booked (with probability $1 - \gamma$) or if she forgets to attend in period two without a prior booking (with probability $1 - \pi$), she obtains a payoff of zero in the second period, and thus, these terms are omitted from the above equation. The decision to book in period one simply depends on whether the additional option value that results from booking outweighs the costs in terms of booking costs, potential canceling costs and the no-show fine.

The model allows for three distinct types of time preferences. First, for $\beta = 1$, discounting is exponential and choices are consistent across periods. Second, if $\beta < 1$ and the agent is aware of the value of β , the model describes a sophisticated time-inconsistent discounter. Such an agent is present-biased and knows this. As a result, she may have a demand for commitment devices that restrict future choices or that make some choices more unlikely. Third, if $\beta < 1$ but the agent erroneously assumes that $\beta = 1$ when making decisions, the model describes a naïve time-inconsistent discounter. Such a naïve individual falsely believes that her future self will maximize her current preferences, so in period one, she would compute the period-two terms as if β were equal to one (see, e.g., Frederick et al., 2002, for an in-depth discussion of these concepts).³³

³³The model describes the single (two-period) decision regarding attendance in one particular class. Our setting is one in which this decision is repeatedly made, for example, one could imagine this situation occurring once every week. While we do not discuss what may change when modeling the repeated decision, there are some obvious possibilities. First, the prior distribution should become more accurate each time the individual observes the attendance cost in period two. As a result, overconfidence will decrease. Second, the belief about the value of γ would become more accurate. Third, one might learn about commitment problems, thereby becoming less naïve and more sophisticated about a potential present bias.

4.3 Comparing the model to observed behavior

We now elaborate on each of the mechanisms included in the model and discuss which model features can explain the six observations that we made in Section 4.1. In Table 8, the observations from the data are listed across columns, from Observations (1) to (6).

Baseline. As a baseline, we consider a model where the individual has time-consistent preferences and does not suffer from limited attention and where there is no uncertainty about future attendance costs (see the first row of Table 8). This model predicts that bookings may be made for popular classes, which might be fully booked in period two, that is, for which $\gamma < 1$, in order to secure a spot (column 1). However, such a booking would never be canceled (column 2) because the optimality of the booking does not change over time. Furthermore, making a booking for an *unpopular* class would never be optimal (columns 3 and 4), as unnecessary booking costs would be incurred. Finally, we would not expect this individual to not show up to a booked class (column 5) or behave differently after receiving reminders (column 6).

Uncertainty about attendance cost. As a first step, we enrich the baseline model with period-one uncertainty about the attendance cost in period two (see the second row in Table 8). An individual will now make a booking in period one if the option value of having a guaranteed spot in the class outweighs the cost of making the booking. Note that we still assume that preferences are time-consistent and that there is no limited attention, i.e., $\beta = 1$ and $\pi = 1$.

Booking in period one is optimal if it yields a higher expected period-two payoff than not booking, i.e., if the following condition holds:

$$\delta \mathbb{E} \max(\bar{U} - c_A, -c_T) - c_T > \gamma \delta \mathbb{E} \max(\bar{U} - c_A, 0) \quad (4)$$

The value of booking (the left-hand side of the inequality) is the discounted expected value of the maximum of either attending ($\bar{U} - c_A$) or canceling ($-c_T$), while an immediate booking cost c_T is also incurred in either case. The expectation is taken over the distribution of c_A (conditional on the observed signal in period one). The value of not booking is again the discounted expected value of the maximum of attending or not attending; however, this value is obtained only if the class is not fully booked (which happens with probability γ). From condition (4), it is clear that for sufficiently low values of γ , booking is an optimal decision. Furthermore, given a booking, a sufficiently large realization of c_A will make canceling optimal in period two. Thus, this model can explain both booking and canceling classes that are sufficiently popular. However, for classes that are unpopular and never fully booked ($\gamma = 1$), condition (4) shows that booking is never optimal. Thus, this model cannot explain booking

Table 8: How can the observed behavior of gym members be explained?

Observations:	Popular classes ($\gamma < 1$)		Unpopular classes ($\gamma = 1$)		No-shows (5)	Impact of reminders (6)
	Book	Cancel	Book	Cancel		
	(1)	(2)	(3)	(4)		
Mechanisms:						
(1) No uncertainty, TI or LA	yes	no	no	no	no	no
(2) Uncertainty	yes	yes	no	no	no	no
(3) Naïve TI	yes	yes	no	no	no	no
(4) Sophisticated TI	yes	no	yes	no	no	no
(5) Uncertainty + Naïve TI	yes	yes	no	no	no	no
(6) Uncertainty + Soph. TI	yes	yes	yes	yes	no	no
(7) Limited attention	no	no	no	no	yes	yes

Uncertainty refers to period-one uncertainty about period-two attendance cost (c_A). TI is short for time inconsistency. LA is short for limited attention.

(or canceling) unpopular classes.³⁴

As a sidestep, uncertainty also predicts that as γ decreases (as a class becomes more popular), cancellations become more likely because if a class is sufficiently popular, booking is optimal even for unfavorable signals about attendance costs. Since those who receive unfavorable signals are more likely to cancel their bookings, we should observe a larger share of cancellations for popular classes. This is indeed the case, as can be seen in Table 7.

Naïve time inconsistency. We now consider how time-inconsistent preferences affect the model predictions. To this end, we eliminate uncertainty about attendance cost and assume that the agent knows the exact value of c_A in period one. Additionally, we begin by assuming that time inconsistency is naïve (Laibson, 1997). Thus, while $\beta < 1$, the individual erroneously assumes $\beta = 1$ for future periods. In this case, the booking trade-off is between the value of having a guaranteed spot and the immediate booking cost. She will book if the attendance cost c_A is such that

$$\beta\delta(\bar{U} - c_A) - c_T > \gamma\beta\delta(\bar{U} - c_A) \iff (1 - \gamma)\beta\delta(\bar{U} - c_A) > c_T \quad (5)$$

³⁴It should be noted that the booking and canceling of popular classes can also be explained by uncertainty in combination with overconfidence. Individuals might systematically underestimate attendance costs and thus be systematically *overoptimistic* when they are confronted with the decision to make a booking. In the model, overconfidence results if the belief about the attendance cost distribution does not resemble the true distribution. Such biased beliefs can give rise to overestimation of the probability of attending in period two and thus increase the likelihood of booking (and canceling). Such overconfidence has, e.g., been identified by DellaVigna and Malmendier (2006) as a likely explanation for the observation that many people sign up at gyms but do not attend as often as they intend to. The overconfidence channel would reinforce the uncertainty channel but would not change the main conclusions.

In period two, she will change her mind and cancel her booking if the discounted returns minus the attendance costs are small enough that she prefers to incur the cancellation costs:

$$-c_T > \beta\bar{U} - c_A \quad (6)$$

If we assume that δ is close to unity for such a short period as a week, we can ignore it and combine conditions (5) and (6) into

$$\frac{c_T}{(1-\gamma)(\bar{U}-c_A)} < \beta < \frac{c_A - c_T}{\bar{U}} \quad (7)$$

If this condition holds, the individual will book in period one and cancel in period two, even in the absence of any uncertainty. The interpretation of this condition is that a low β makes booking unattractive because it puts relatively more weight on the directly incurred transaction cost, while a large β (close to unity) implies less discounting of future health benefits in period two, making canceling in period two less attractive. There exists an interval for β , dependent on c_T, c_A, γ and \bar{U} , for which time-inconsistent preferences lead to booking in period one and subsequent canceling in period two, and thus, naïve time inconsistency can explain Observations 1 and 2.³⁵ However, it cannot explain such behavior when classes are unpopular ($\gamma = 1$) since booking such classes would again never be optimal (straightforward to see from condition 5). This conclusion is summarized in row 3 of Table 8.

Sophisticated time inconsistency. If individuals are sophisticated and thus are aware of their inconsistent behavior, they understand that if they book in period one, they might cancel that booking in period two because they will discount the future benefits \bar{U} relatively more once period two arrives. Thus, they wish to commit themselves to future attendance. The simple act of booking a class may function as a commitment device and increase the likelihood of attending the class. While a booking ultimately does not commit one to attend the class, it does place a cost on not attending (in terms of the canceling cost or in terms of the fine that has to be paid in case of failing to attend a booked class). Thus, it enlarges the set of values of c_A for which attending is the optimal period-two choice.³⁶

In period one, a sophisticated individual would like to attend the gym in period two if $c_A < \bar{U}$, i.e., if attendance costs are sufficiently small. However, she knows that she will attend in period two only if $c_A < \beta\bar{U}$. By booking the class in period one, she changes the period-two condition for attending to $c_A < \beta\bar{U} + c_T$, which implies that it becomes relatively

³⁵Note that the bounds of the interval become wider if c_T is smaller. For small transaction costs, time-inconsistent preferences are thus more likely to explain booking and subsequent canceling.

³⁶Furthermore, it has been found that the simple act of planning to take an action may make its implementation more likely (Beshears et al., 2016), as one feels more guilty if an explicitly set plan is not realized. While this mechanism is not present in the model, it is a reason why the cost of canceling might be larger than what would be implied by the (small amount of) effort and the time spent canceling a booking.

Table 9: Booking and canceling by class popularity: experienced attenders that have attended at least 5 times before

	Unpopular classes ($< 0.5\%$ waiting list)	Fairly popular classes ($< 5\%$ waiting list)	Popular classes ($> 5\%$ waiting list)
Number of bookings	6251	10877	47442
Of those (in %):			
Attendance (%)	0.66	0.61	0.59
Cancellation (%)	0.29	0.35	0.36
No-shows (%)	0.04	0.04	0.04
On final waiting list (%)	0.00	0.00	0.01
Number of drop-in attendance	3864	1869	2798
As % of all attendance	0.48	0.22	0.09
Number of classes	51	40	83

more attractive to attend in period two. Of course, the commitment device comes at the cost of incurring the booking cost in period one and provides the additional benefit of having the option to attend in period two even if the class is fully booked. The commitment device is valuable for both popular and unpopular classes. Thus, as opposed to naïve time inconsistency, sophisticated time inconsistency predicts the booking of unpopular classes that will never be fully booked and can explain Observation 3. Cancellations, on the other hand, cannot be explained by sophisticated time inconsistency because a sophisticated person perfectly anticipates her choices in the second period and will never change her mind in the absence of uncertainty (see row 4 of Table 8).

One might worry that the above conclusions hinge on the agent being aware of the risk of a particular class becoming fully booked.³⁷ In particular, she would need to know that γ is (very close to) one for some classes. If she is uncertain about this risk, she might still book to be sure that she will be able to attend rather than to commit herself. We argue that this is unlikely to drive our results for two reasons. First, we have categorized a class rather broadly, which makes it much more likely that a gym member is aware of the popularity of the class. In our classification, a 45-minute spin class on Thursday at 7 pm is classified as the same type of class as a 60-minute spin class on Thursday at 8 pm. Because of this broad classification, gym members are likely to quickly learn which classes are potentially popular. Additionally, a class type is called ‘unpopular’ only if *all* versions of this class are unpopular. Second, we can focus on a subset of gym members who attend a particular type of class regularly. Those

³⁷Another concern might be that those who participate in unpopular classes systematically differ from those who attend popular classes. We show in Table 17 in the Appendix that there are some differences between the groups, but those are small compared to the variation in the sample.

individuals certainly know the popularity of the class. We restrict the sample to observations for gym members who book a certain type of class that they have attended at least five times before. In Table 9, we reproduce Table 7 for this specific sample of experienced attenders. As expected, among more experienced gym members, the share of drop-in attendance is larger for unpopular classes. However, even for very experienced attenders, we still find that a large majority book the class in advance rather than drop in. Thus, bookings indeed seem to be used to commit to attend the class. Note that the statistics in Table 9 reveal another interesting finding, which is that as the members become more experienced, they are less likely to cancel their booking (this finding holds for all popularity categories). This finding can be explained by sequential learning about the attendance cost distribution, which enables members to more precisely predict whether they will actually attend the class. Thus, potential overconfidence will decrease over time.

Since only sophisticated time inconsistency can explain Observation 3 (that unpopular classes are commonly booked), we conclude that bookings are used as commitment devices to cope with time-inconsistent preferences. This finding is complementary to results from recent field studies (e.g., Goldhaber-Fiebert et al., 2010; Bhattacharya et al., 2015; Royer et al., 2015), which show that a significant share of people are willing to sign a contract that puts a penalty on failing to achieve a specified number of workouts or gym visits over a particular time period. Significant demand for commitment devices has also been shown to exist for savings products (Ashraf et al., 2006), restricted lunch menu options (Toussaert, 2016) or self-imposed deadlines for tasks (Bisin and Hyndman, 2014). Our setting differs from these, as we consider a repeated decision on whether or not to obtain a commitment device (rather than a one-off decision). It is not obvious that the value of commitment devices in single decisions, as shown in the literature, carries over to situations where a decision is repeatedly made. Thus, we contribute by showing that even for the gym attendance decision, which is made repeatedly on a weekly basis, commitment devices play a relevant role.

Uncertainty plus time inconsistency. We have shown that booking and canceling popular classes can be explained by either uncertainty about period-two attendance cost or by naïve time inconsistency. Explaining that an unpopular class is booked requires *sophisticated* time inconsistency. However, such preferences cannot explain by themselves why a booking would be canceled. Thus, jointly explaining Observations (1)-(4) requires a combination of factors. Combining uncertainty with naïve time inconsistency (see row 5 in Table 8) still cannot explain the booking and canceling of unpopular classes. Thus, only a combination of uncertainty and sophisticated time inconsistency predicts each of these four observations in our data. In such a model, popular classes are booked to secure a spot (Observation 1), unpopular classes are booked as a commitment device (Observation 3), and each booking is canceled if the attendance cost becomes sufficiently unfavorable (Observations 2 and 4).

Of course, not all gym members need to have sophisticated time-inconsistent preferences, although the large number of unpopular class bookings suggests that at least a sizable share do. As is clear from Table 8, we cannot conclude that some share of gym members are also naïve about their time inconsistency. Nevertheless, it seems likely that this is the case, given the large share of cancellations (41 %), which otherwise would have to be fully explained by uncertainty about the future.

Limited attention. Observations 5 and 6 point more directly towards limited attention. First, we observe that 13 % of all cancellations (or 6 % of all bookings) in our data are no-shows. In the model, there is a risk $1 - \pi < 1$ of forgetting about the class in period two, which seems to be a natural explanation for no-shows. In the absence of limited attention, one would never book a class and subsequently not attend (without canceling) because, as argued earlier, the cancellation costs, c_T , are low while the costs of not showing up to a booked class, c_F , are non-trivial (around \$3). Thus, we can safely assume that $c_T < c_F$, and no-shows therefore must be the result of limited attention.³⁸ Interestingly, the share of no-shows is the same across all types of classes. This finding suggests that limited attention is a common phenomenon and does not depend on the type of class.

Furthermore, we find in the experiment that regular reminders affect decisions regarding class attendance (Observation 6). In particular, we observe that reminders increase bookings, decrease the share of cancellations and increase class attendance (where the impact on cancellations is weak and conclusive). A natural approach to incorporating reminders in the model is to assume that an individual who receives them will no longer suffer from limited attention, that is, $\pi = 1$. Compared to the situation where $\pi < 1$, this situation leads to predictions in line with Observations 6. First, if individuals anticipate the reminders (they might do so after receiving repeated emails), they realize that they will not forget in period two, which makes booking more attractive and, thus, depending on parameter values, could increase the number of bookings (as observed). Second, reminders may eliminate no-shows, decreasing the share of bookings that are not followed by class attendance (as observed). Third, reminders would increase the total number of class attendances due to both an elimination of no-shows and to a higher number of drop-in attendances. Thus, each element of Observation 6 can be explained by limited attention and its elimination through reminders.

³⁸Note that in reality (though not in our model), no-shows can also be caused by extreme present bias. Imagine that someone books a class and, one hour prior to the class, believes that she will attend. However, half an hour later, she weighs the direct effort and time costs much more highly relative to the returns and decides not to attend. Since she can no longer cancel at this point, she will not show up. It is, however, difficult to imagine that the change in discounting during the final hour is so large that it can explain the 13 % of no-shows out of all cancellations.

5 Conclusion

It is often suggested that one of the main reasons for inefficiently low frequencies of exercise are behavioral explanations, such as limited attention and time-inconsistent preferences. Our analysis of exercising behavior and planning by a large sample of gym members shows that both are indeed relevant.

Using a large-scale randomized experiment with email reminders, we find that short and simple weekly reminders lead to a significant 13 % increase in the number of weekly gym visits. The fact that reminders are so effective at increasing exercising frequencies suggests that limited attention is an important hurdle for reaching individual exercise goals. In addition, we find that gym members suffer from not only limited attention but also present bias. Analyzing a large dataset on the booking, cancellation and attendance of gym classes, we show that many members book classes simply to commit themselves to attending the class in the future. This analysis allows us to conclude that commitment problems, which stem from (sophisticated) time-inconsistent preferences, are a common phenomenon among gym members.

While our results are based on a sample of gym members (thus, individuals who decided to purchase a gym membership), we believe that the external validity of our results is relatively strong. Our large sample contains a diverse set of individuals with both students and employed members and a wide range of age groups. Since the gym offers many different sports activities, our results are not based solely on individuals interested in weightlifting but apply to all types of people interested in regular exercise and sports. Finally, average attendance in our sample (approximately once per week) is similar to the commonly found attendance rates but well below the general advice from a health perspective.

Our findings point toward the importance of policies that guide those who are interested in attaining a higher level of regular exercise. Simple reminders, which essentially have zero marginal cost, have the potential to stimulate individuals to exercise more by ensuring that the option to exercise is not forgotten during one's busy day-to-day life. Our experimental results show that the impact might even persist after reminders end, potentially due to habit formation. Additionally, offering help to commit oneself to one's exercising goals can be very effective and, as our results imply, would be welcomed by many. For example, one could consider disallowing cancellation of an exercise class booking in the 24 hours before the class in order to strengthen the commitment value of such a booking. Since many seem to be aware of their commitment problems, one option would be to offer such policies on a voluntary basis, ensuring that they are taken up specifically by those individuals who benefit most from them.

References

- Acland, D. and Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science*, 61(1):146–160.
- 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):3003–3037.
- Altmann, S. and Traxler, C. (2014). Nudges at the dentist. *European Economic Review*, 72:19–38.
- Apestequia, J., Funk, P., and Iriberry, N. (2013). Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of Barcelona. *European Economic Review*, 64:266–284.
- Ashraf, N., Karlan, D., and Yin, W. (2006). Tying odysseus to the mast: Evidence from a commitment savings product in the philippines. *The Quarterly Journal of Economics*, 121(2):635–672.
- Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35.
- Beshears, J., Milkman, K. L., and Schwartzstein, J. (2016). Beyond beta-delta: The emerging economics of personal plans. *American Economic Review*, 106(5):430–434.
- Bhargava, S. and Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review*, 105(11):3489–3529.
- Bhattacharya, J., Garber, A. M., and Goldhaber-Fiebert, J. D. (2015). Nudges in exercise commitment contracts: A randomized trial. *NBER Working Paper*, No. 21406.
- Bisin, A. and Hyndman, K. (2014). Present-bias, procrastination and deadlines in a field experiment. *NBER Working Paper*, No. 19874.
- Bornstein, B., Tomkins, A., Neeley, E., Herian, M., and Hamm, J. (2012). Reducing courts’ failure-to-appear rate by written reminders. *Psychology, Public Policy, and Law*, 19(1):70–80.
- Calzolari, G. and Nardotto, M. (2017). Effective reminders. *Management Science*, 63(9):2915–2932.

- Castleman, B. L. and Page, L. C. (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.
- Charness, G. and Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77(3):909–931.
- Chen, Y., Lu, F., and Zhang, J. (2017). Social comparisons, status and driving behavior. *Journal of Public Economics*, 155:11 – 20.
- Dai, H., Milkman, K. L., and Riis, J. (2014). The fresh start effect: Temporal landmarks motivate aspirational behavior. *Management Science*, 60(10):2563–2582.
- 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.
- Dean, M., Kıbrıs, Ö., and Masatlıoğlu, Y. (2017). Limited attention and status quo bias. *Journal of Economic Theory*, 169:93–127.
- DellaVigna, S. and Malmendier, U. (2006). Paying not to go to the gym. *The American Economic Review*, 96(3):694–719.
- Ericson, K. M. (2017). On the interaction of memory and procrastination: Implications for reminders, deadlines, and empirical estimation. *Journal of the European Economic Association*, 15.
- Fellner, G., Sausgruber, R., and Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11(3):634–660.
- Frederick, S., Loewenstein, G., and O’Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.
- Goldhaber-Fiebert, J. D., Blumenkranz, E., and Garber, A. M. (2010). Committing to exercise: Contract design for virtuous habit formation. *NBER Working Paper*, No. 16624.
- Guthold, R., Stevens, G. A., Riley, L. M., and Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 19 million participants. *The Lancet Global Health*.
- Himmler, O., Jäckle, R., and Weinschenk, P. (forthcoming). Soft commitments, reminders and academic performance. *American Economic Journal: Applied Economics*.
- Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. D. Appleton-Century Company, Incorporated, New York, London.

- Hull, C. L. (1951). *Essentials of Behavior*. Published for the Institute of Human Relations by Yale University Press, New Haven.
- Karlan, D., McConnell, M., Mullainathan, S., and Zinman, J. (2016a). Getting to the top of mind: How reminders increase saving. *Management Science*, 62(12):3393–3411.
- Karlan, D., Morten, M., and Zinman, J. (2016b). A personal touch in text messaging can improve microloan repayment. *Behavioral Science & Policy*, 1(2):25–31.
- Kool, J., de Bie, R., Oesch, P., Knusel, O., van den Brandt, P., and Bachman, S. (2004). Exercise reduces sick leave in patients with non-acute non-specific low back pain: A meta-analysis. *Journal of Rehabilitation Medicine*, 36(2):49–62.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.
- Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., and Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, 40(6):998–1009.
- Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, 34(7):913–923.
- O’Donoghue, T. and Rabin, M. (1999a). Doing it now or later. *American Economic Review*, 89(1):103–124.
- O’Donoghue, T. and Rabin, M. (1999b). Incentives for procrastinators. *The Quarterly Journal of Economics*, 114(3):769–816.
- O’Donoghue, T. and Rabin, M. (2001). Choice and procrastination. *The Quarterly Journal of Economics*, 116(1):121–160.
- Phelps, E. and Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies*, 35(2):185–199.
- Royer, H., Stehr, M., and Sydnor, J. (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.
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

- Sonntag, A. and Zizzo, D. J. (2015). On reminder effects, drop-outs and dominance: Evidence from an online experiment on charitable giving. *PLoS ONE*, 10(8):1–17.
- Taubinsky, D. (2014). From intentions to actions: A model and experimental evidence of inattentive choice. Technical report, Department of Economics, Harvard University.
- Toussaert, S. (2016). Connecting commitment to self-control: a field experiment with participants in a weight loss challenge. Mimeo.
- Van Amelsvoort, L. G., Spigt, M. G., Swaen, G. M., and Kant, I. (2006). Leisure time physical activity and sickness absenteeism; a prospective study. *Occupational Medicine*, 56(3):210–212.
- Verplanken, B. (2006). Beyond frequency: Habit as mental construct. *British Journal of Social Psychology*, 45(3):639–656.
- Wood, W. and Neal, D. T. (2007). A new look at habits and the habit-goal interface. *Psychological Review*, 114(4):843–863.

A Appendix: Additional descriptive statistics

Figure 6: Visits per class

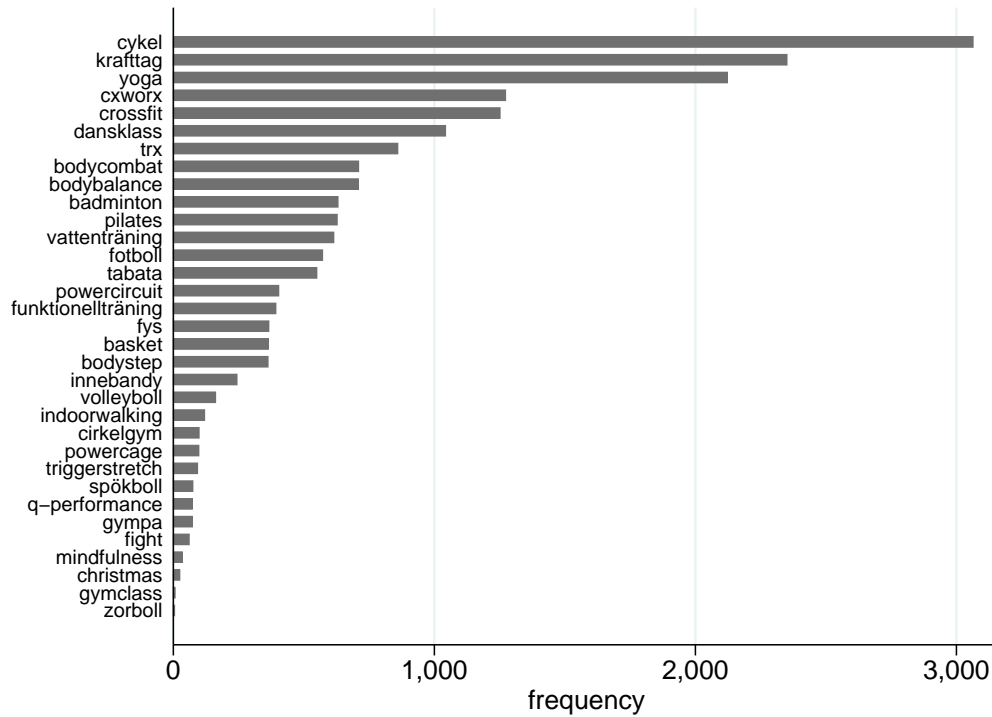


Figure 7: Distribution of dates of buying contract for control and treatment group

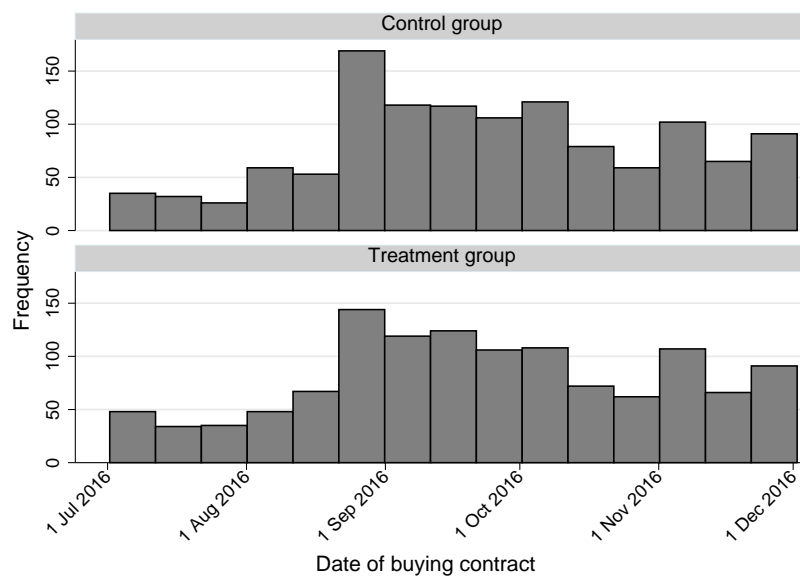
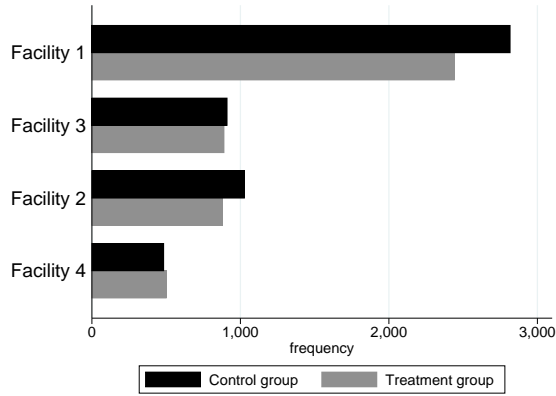
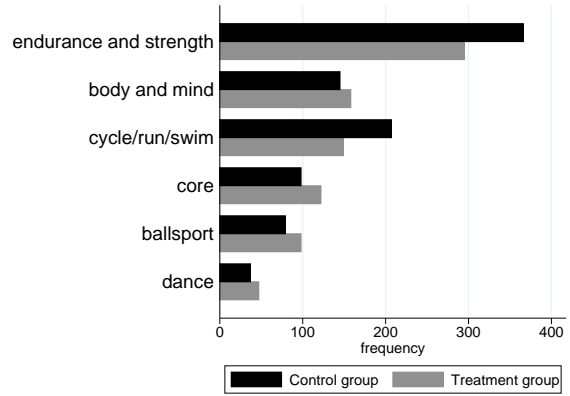


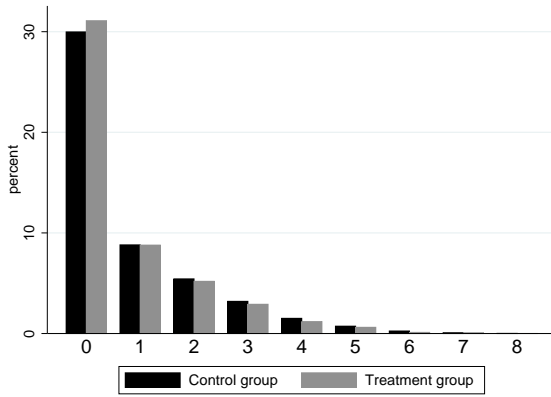
Figure 8: Comparing control and treatment: pre-experiment gym attendance



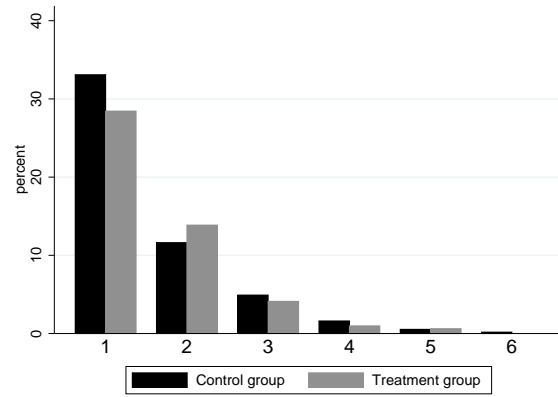
(a) Free training per facility



(b) Group classes



(c) Distribution of weekly free training visits



(d) Distribution of weekly class training visits, excluding zeros (90.9 % of all observations)

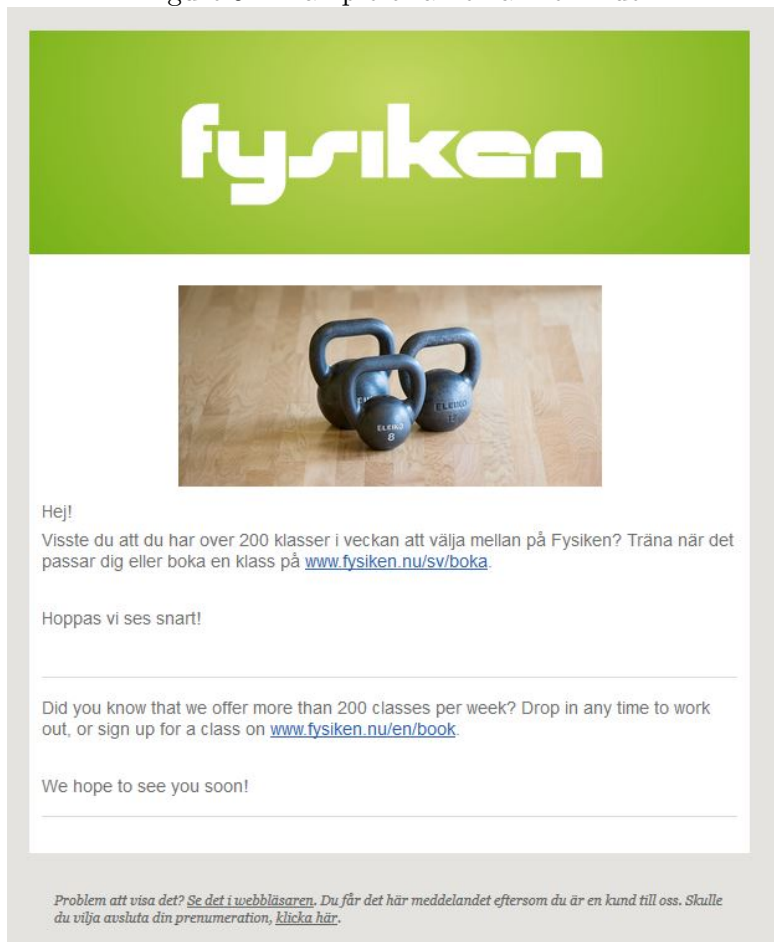
B Appendix: Email reminders

Table 10: Calendar schedule email reminders

December 2016	January 2017	February 2017	March 2017	April 2017
15 (general email)	9 (reminder)	6 (reminder)	7 (reminder)	3 (reminder)
	15 (reminder)	13 (reminder)	13 (reminder)	9 (reminder)
	24 (reminder)	19 (reminder)	21 (reminder)	
	[26 (newsletter)]		26 (reminder)	

Note: Reminders were always sent at the beginning of a week. To vary the day of dispatch, we randomized between Sundays, Mondays and Tuesdays. The newsletter on January 26 was sent to all gym members. The lack of a reminder in the fourth week of February was due to an expected newsletter, which was, eventually, not sent out by the gym.

Figure 9: Example of an email reminder



List of email reminders (in chronological order):

1. *Did you know that we offer more than 200 classes per week? Drop in any time to work out, or sign up for a class on www.fysiken.nu/en/book/. We hope to see you soon!*
2. *At Fysiken, we offer a huge range of classes and training facilities. You can find the schedule and sign up for classes at www.fysiken.nu/en/book/. Get inspired! We hope to see you soon!*
3. *Did you know that you can train at Fysiken 24/7? Our facility at Gibraltargatan is now open 24 hours a day. Also our facilities Lindholmen and Kaserntorget have long opening hours. See www.fysiken.nu/en/ for more information. See you soon!*
4. *Did you know that research has shown that working out makes you feel more energetic? Go to our website and get inspired for your next energizing workout: www.fysiken.nu/en/. We hope to see you soon.*
5. *Did you know that regular exercise has huge health benefits? Reap the benefits of regular exercise and sign up for one of Fysiken's classes at www.fysiken.nu/en/book/. We hope to see you soon!*
6. *Research has shown that regular workout reduces your stress levels significantly. Try out one of the many classes we offer. Look at www.fysiken.nu/en/ for more information. See you soon!*
7. *Fysiken offers a huge variety of training activities. For example, you can try out various ball sports, climbing activities and water training. Find more information on www.fysiken.nu/en/ and get inspired! We hope to see you soon!*
8. *Did you know that exercising helps reduce stress and improve concentration? Body and mind are connected - book your yoga, Pilates or Mindfulness class at www.fysiken.nu/en/book/. Welcome!*
9. *We all have problems with committing ourselves to a strict workout schedule. Did you know that simply putting reminders in your calendar might help you to better stick to your exercise goals? Inspired? Take a look at www.fysiken.nu/en/. We hope to see you soon.*
10. *Research has shown that regular exercise improves the quality of your sleep. Improve your life now and take a look at our wide range of gym classes at www.fysiken.nu/en/. We hope to see you soon!*

11. *Working out together is more fun - and it helps you commit to going to the gym. Why don't you contact a friend and schedule a training session right away? Go to www.fysiken.nu/en. We hope to see you soon!*
12. *At Fysiken, we offer a huge range of classes and training facilities. You can find the schedule and sign up for classes at www.fysiken.nu/en/book/. Get inspired!*

C Appendix: Extended results

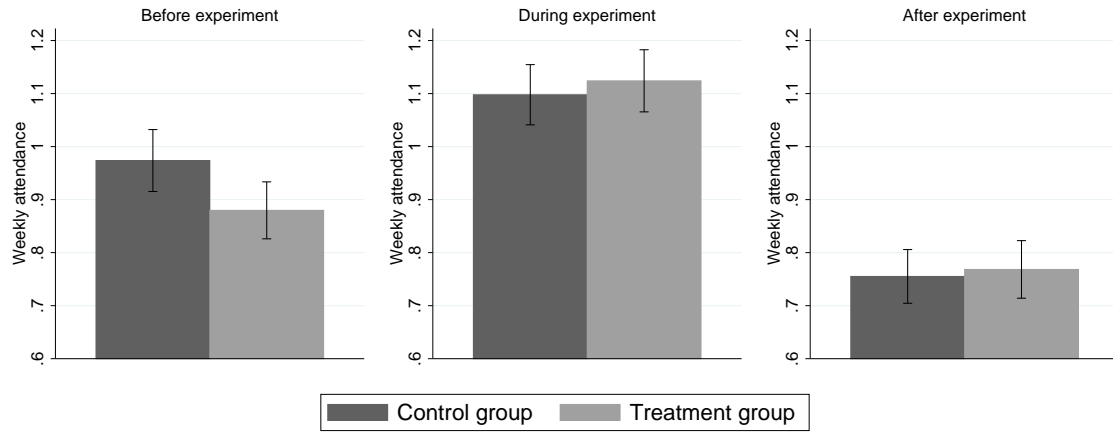
C.1 Regressions on weekly gym attendance

Table 11: Regressions on weekly gym attendance (linear models)

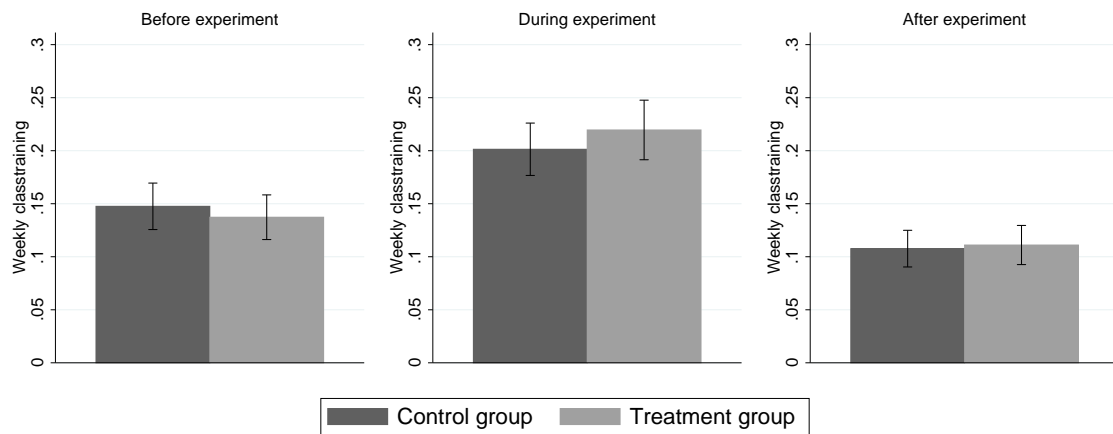
	DiD model			Simple difference		
	(1) Visits	(2) Class training	(3) Free training	(4) Visits	(5) Class training	(6) Free training
Experiment (during)	0.12*** (0.031)	0.029** (0.014)	0.091*** (0.028)	0.075* (0.043)	0.028 (0.019)	0.047 (0.039)
Experiment (post)	0.11*** (0.034)	0.014 (0.013)	0.093*** (0.032)	0.049 (0.040)	0.0053 (0.013)	0.044 (0.037)
Recurring member				0.19*** (0.042)	0.022 (0.016)	0.16*** (0.039)
Male				0.19*** (0.039)	-0.18*** (0.017)	0.37*** (0.035)
Student				0.0030 (0.047)	0.0041 (0.019)	-0.0011 (0.043)
Daytime card				-0.30*** (0.077)	-0.14*** (0.030)	-0.16** (0.071)
Week & duration FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	no	no	no
Include pre-period	yes	yes	yes	no	no	no
Mean value dep. var.	0.95	0.16	0.79	0.96	0.17	0.79
Percent effect (during)	13	18	12	8	17	6
N	73890	73890	73890	55825	55825	55825

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. Columns (4)-(6) also include seven age dummies which are not reported.

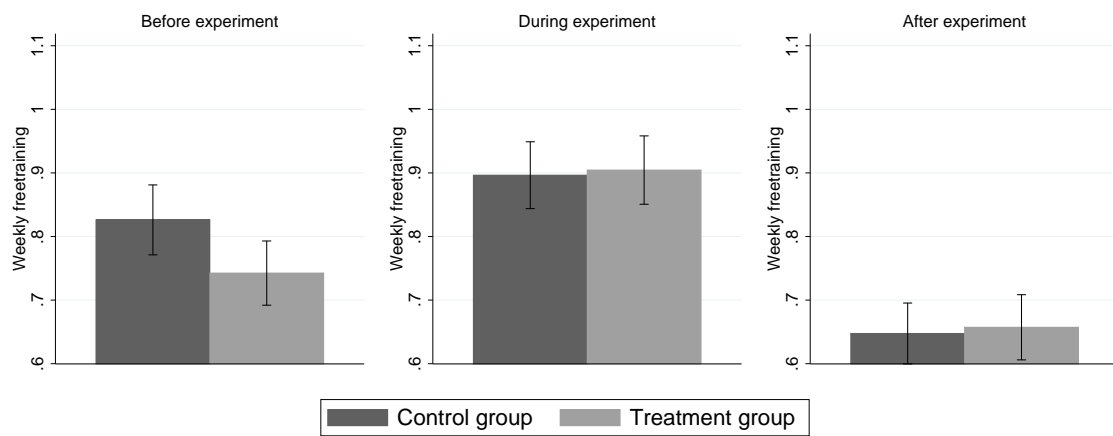
Figure 10: Weekly gym attendance: group means (with 95 % confidence interval)



(a) Total visits (DiD estimate during: 0.12, DiD estimate post: 0.11)



(b) Class training (DiD estimate during: 0.03 : DiD estimate post: 0.01)



(c) Free training (DiD estimate during: 0.09, DiD estimate post: 0.09)

Table 12: Regressions on weekly gym attendance: excluding Christmas holidays (Poisson models)

	Total visits		Class training		Free training	
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment (during)	0.13*** (0.034)	0.11*** (0.037)	0.19** (0.090)	0.15* (0.096)	0.11*** (0.036)	0.10*** (0.040)
Experiment (post)	0.12*** (0.046)	0.11** (0.049)	0.12 (0.11)	0.090 (0.12)	0.12*** (0.050)	0.11** (0.054)
Week & duration FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	yes	yes	yes
Include pre-period	yes	yes	yes	yes	yes	yes
Exclude weeks:	52	51,52,01	52	51,52,01	52	51,52,01
N	64641	60102	27289	25218	62988	58509

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Note that the number of observations varies, because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects. Columns (1), (3) and (5) exclude the last week of 2016 (week 52), in which attendance is very low. Columns (2), (4) and (6) in addition exclude week 51 (2016) and week 1 (2017) in which attendance is also below the trend. See Figure 4 for clarification.

Table 13: Regressions on weekly gym attendance: subsample with longer pre-experiment period (Poisson models)

	DiD model			Simple difference		
	(1)	(2)	(3)	(4)	(5)	(6)
	Visits	Class training	Free training	Visits	Class training	Free training
Experiment (during)	0.034 (0.032)	0.14* (0.080)	0.0084 (0.034)	0.085*** (0.017)	0.16*** (0.042)	0.066*** (0.019)
Experiment (post)	0.064 (0.048)	0.061 (0.11)	0.064 (0.052)	0.11*** (0.025)	0.044 (0.058)	0.12*** (0.028)
Recurring member				0.19*** (0.016)	0.092*** (0.035)	0.22*** (0.019)
Male				0.19*** (0.016)	-0.71*** (0.010)	0.62*** (0.025)
Student				-0.030* (0.015)	0.071* (0.040)	-0.045*** (0.016)
Daytime card				-0.31*** (0.025)	-0.69*** (0.039)	-0.21*** (0.030)
Week & duration FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	no	no	no
Include pre-period	yes	yes	yes	no	no	no
N	63070	29190	61635	42550	42550	42550

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Note that the number of observations varies, because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects.

Table 14: Regressions on weekly gym attendance (Poisson models): post-experiment effect w/o week 22 2017

	DiD model			Simple difference		
	(1) Visits	(2) Class training	(3) Free training	(4) Visits	(5) Class training	(6) Free training
Experiment (during)	0.13*** (0.035)	0.17** (0.090)	0.12*** (0.038)	0.072* (0.040)	0.12 (0.097)	0.059 (0.044)
Experiment (post)	0.11** (0.046)	0.11 (0.11)	0.11** (0.050)	0.047 (0.052)	0.040 (0.12)	0.048 (0.058)
Recurring member				0.20*** (0.049)	0.14 (0.11)	0.22*** (0.056)
Male				0.21*** (0.047)	-0.66*** (0.033)	0.61*** (0.075)
Student				0.00055 (0.048)	0.058 (0.13)	-0.0053 (0.052)
Daytime card				-0.31*** (0.074)	-0.70*** (0.12)	-0.20** (0.087)
Week & duration FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	no	no	no
Include pre-period	yes	yes	yes	no	no	no
N	64670	27347	63017	53592	53592	53592

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Seven age dummies and six gym location dummies are included but not reported in columns (4)-(6). Note that the number of observations in columns (1)-(3) vary because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects.

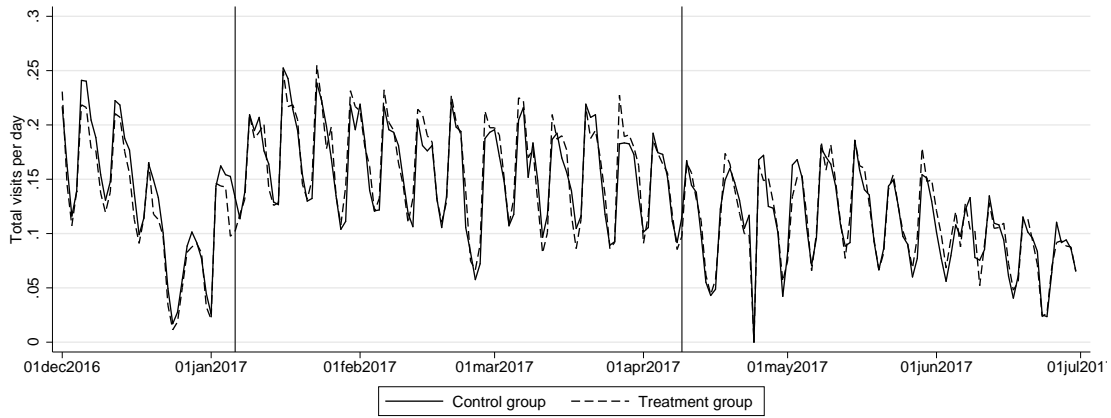
Table 15: Regressions on weekly gym attendance (Poisson models): post-experiment effect until week 21 2017

	DiD model			Simple difference		
	(1)	(2)	(3)	(4)	(5)	(6)
	Visits	Class training	Free training	Visits	Class training	Free training
Experiment (during)	0.13*** (0.035)	0.17** (0.090)	0.12*** (0.038)	0.069* (0.040)	0.12 (0.097)	0.056 (0.044)
Experiment (post)	0.11** (0.047)	0.14 (0.12)	0.11** (0.051)	0.052 (0.052)	0.068 (0.12)	0.049 (0.058)
Recurring member				0.19*** (0.048)	0.13 (0.10)	0.20*** (0.055)
Male				0.18*** (0.046)	-0.67*** (0.032)	0.60*** (0.073)
Student				0.016 (0.048)	0.072 (0.13)	0.010 (0.052)
Daytime card				-0.31*** (0.074)	-0.71*** (0.12)	-0.19** (0.086)
Week & duration FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	no	no	no
Include pre-period	yes	yes	yes	no	no	no
N	55550	23300	54000	44660	44660	44660

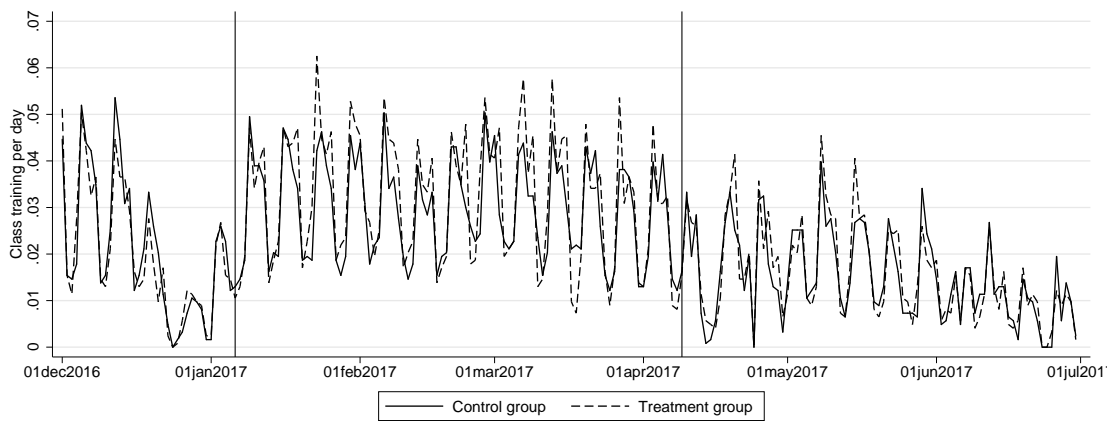
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. We report $\exp(\beta) - 1$, which is the percentage effect. Seven age dummies and six gym location dummies are included but not reported in columns (4)-(6). Note that the number of observations in columns (1)-(3) vary because individuals with zero attendance in all time periods drop out in Poisson regressions with fixed effects.

C.2 Regressions on daily gym attendance

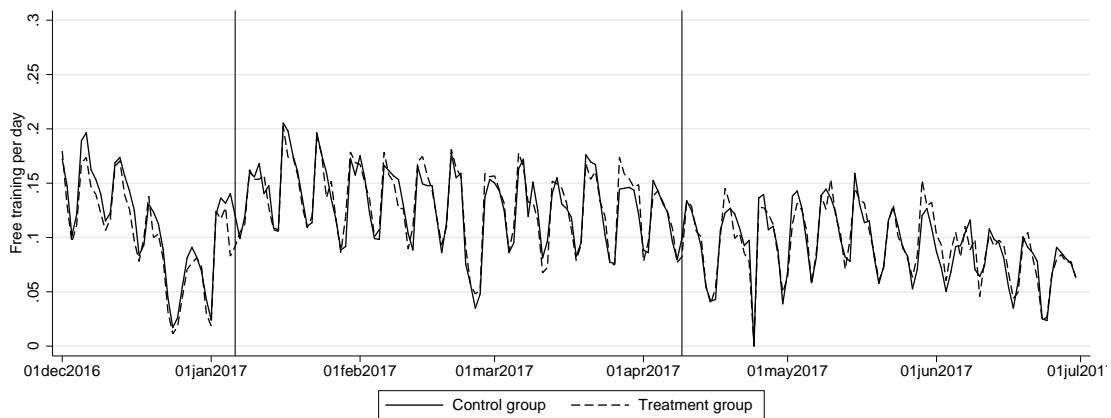
Figure 11: Gym attendance: daily data



(a) Total visits (daily)



(b) Class training (daily)



(c) Free training (daily)

Table 16: Regressions on daily gym attendance

	DiD model			Simple difference		
	(1) Visits	(2) Class training	(3) Free training	(4) Visits	(5) Class training	(6) Free training
Experiment (during)	0.014*** (0.0042)	0.0037** (0.0019)	0.011*** (0.0037)	0.010* (0.0061)	0.0038 (0.0027)	0.0063 (0.0055)
Experiment (post)	0.012** (0.0054)	0.0030 (0.0022)	0.0090* (0.0049)	0.0070 (0.0065)	0.0024 (0.0026)	0.0046 (0.0061)
Recurring member				0.025*** (0.0063)	0.0024 (0.0027)	0.023*** (0.0058)
Male				0.024*** (0.0059)	-0.032*** (0.0028)	0.056*** (0.0054)
Student				0.0017 (0.0072)	0.00075 (0.0032)	0.00091 (0.0066)
Daytime card				-0.045*** (0.012)	-0.025*** (0.0046)	-0.020* (0.011)
Sunny day				-0.0030* (0.0016)	-0.00083 (0.00069)	-0.0022 (0.0015)
Rainy day				-0.0081*** (0.0020)	-0.0026*** (0.00090)	-0.0054*** (0.0018)
Week & duration FE	yes	yes	yes	yes	yes	yes
Day of week FE	yes	yes	yes	yes	yes	yes
Individual FE	yes	yes	yes	no	no	no
Include pre-period	yes	yes	yes	no	no	no
Mean value dep. var.	0.15	0.026	0.12	0.15	0.028	0.12
Percent effect	10	14	9	7	14	5
N	384228	384228	384228	259028	259028	259028

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered by individual in parentheses. Seven age dummies and six gym location dummies are included but not reported in columns (4)-(6). A sunny day is a day with at least 5 hours of sunshine, and a rainy day is a day with at least 3 mm of rain.

C.3 Class popularity: Details of categorization and some more results

As we discuss in Section 2.1, over 200 classes are scheduled each week, with 37 main class types (and there is further variation with respect to class length, focus, instructor, instruction language, etc.). While some classes are very popular, with all places being regularly taken, other classes are less popular (or are simply offered at higher capacity) and are rarely fully booked. Below we describe how we split classes by popularity into three categories (unpopular, fairly popular and popular).

The first step to define a *class*. Simply defining classes by their activity (e.g., Yoga, Pilates, etc.) and their scheduled time (e.g. Wednesday at 8 pm) will be too narrow. For example, if someone knows that ‘30 minutes Yoga’ on Wednesday at 7 pm is very popular, she will most likely assume that ‘45 minutes Yoga’ on Wednesday at 8 pm is also very popular (even if she

never attended this particular class before). Thus, we need to define a class in a rather broad sense, such that similar classes at similar times fall into the same category. For this purpose, we first categorize each class into 37 main activities (these are the activity types, see footnote 7). Second, we assign each class to one of 14 time slots, defined by the day of the week interacted with ‘daytime’ or ‘evening’. A particular class is thus defined as the activity type and the time slot. Examples of a class would be ‘Yoga on Wednesday evening’ or ‘Football on Saturday daytime’. This approach also solves the problem that the exact scheduled time or name of a class might change during the course of one year. As a result, we observe 238 distinct classes, with between 5 and 5,617 bookings for each class. For 73 % of these classes, we observe at least 100 bookings during the 12 months.

Next, we categorize the 238 classes into three categories, based on how popular they are. We measure popularity of a class by computing the share of bookings that result in a spot on the waiting list. A share of zero means that the class is never fully booked, which is the case for 30 % of all classes. A share of 0.05 means that one in 20 bookings ends up on the waiting list. If a class has 20 available spots, the class is typically fully booked. Of course, due to cancellations there is a reasonable chance of still being admitted to the class. However, we argue that an individual’s perception of class popularity will be affected by (initially) ending up on the waiting list, and thus we opted for this particular measure. For 40 % of the classes the share is at least 0.05, while 22 % even have a share of more than 0.2 (indicating long waiting lists). We categorize the classes into unpopular (share less than 0.005), fairly popular (share between 0.005 and 0.05) and popular (share more than 0.05).

Figure 12: Cumulative distribution of the timing of class bookings by popularity of the class

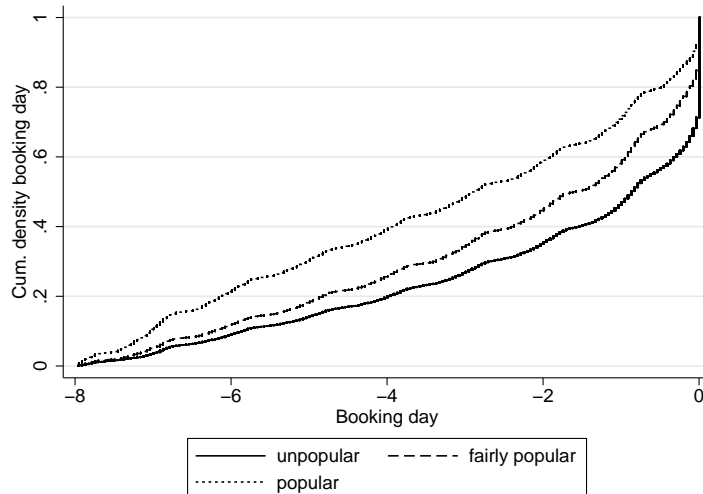


Figure 13: Cumulative distribution of the timing of canceling by popularity of the class

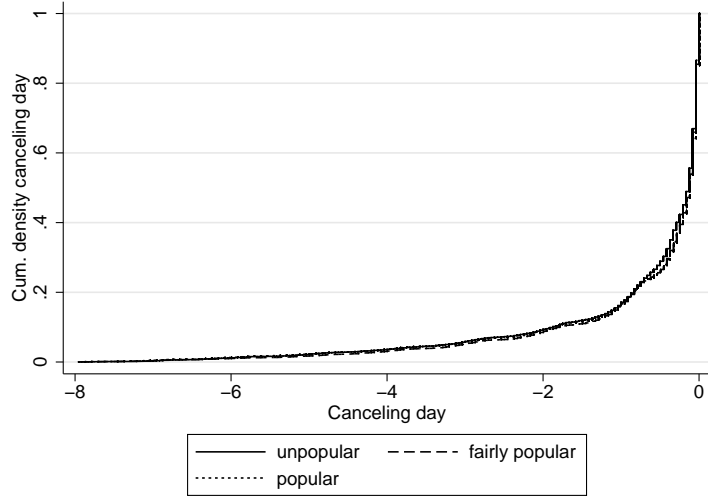


Table 17: Characteristics of attenders of classes by popularity

	Mean value			P-values		
	(1) Unpopular classes	(2) Fairly popular classes	(3) Popular classes	(1) vs (2)	(1) vs (3)	(2) vs (3)
Age	32.76	32.03	31.79	0.23	0.15	0.61
Male	0.30	0.17	0.31	0.00	0.66	0.00
Student	0.52	0.58	0.53	0.01	0.61	0.02
Recurring member	0.54	0.51	0.51	0.22	0.27	0.93
Membership type:						
Multi Card	0.50	0.47	0.53	0.29	0.30	0.01
Facility 1	0.41	0.42	0.39	0.66	0.46	0.11
Facility 2	0.07	0.09	0.07	0.10	0.96	0.04
Facility 3	0.01	0.01	0.00	0.61	0.05	0.04
Facility 4	0.01	0.01	0.01	0.16	0.93	0.17
Daytime card	0.02	0.02	0.01	0.75	0.05	0.02
Observations	7713	11004	29493			

We regress the individual characteristic on dummies for the popularity of the class and include all participants from only two categories of classes (for example all unpopular classes and all fairly popular classes). By clustering the standard error on the individual level, we take into account that each individual might appear in the sample many times as a participant.

^aThe reported p-values follow from a standard t-test for statistical significance of one of the popularity dummies relative to the other one.