Does Loss Aversion Beat Procrastination?  
A Behavioral Health Intervention at the Gym

Oliver März  
SBS-EM, ECARES, Université libre de Bruxelles

September 2017

ECARES working paper 2017-33
Does Loss Aversion Beat Procrastination?
A Behavioral Health Intervention at the Gym

By Oliver März

Financial incentives are a common tool to encourage overcoming self-control problems and developing beneficial habits. There are different means by which such incentives can be provided, yet, up to date there is little empirical evidence on the relative effectiveness of different incentive designs. In this paper, we conduct a field experiment to explore whether and how incentives that are economically equivalent but framed differently affect the likelihood of exercising at a gym. We find that framing incentives in terms of losses, meaning individuals lose cash incentives by not exercising, encourages more frequent visits to the gym than framing incentives in terms of financial gains. After removing these incentives, we observe habit formation in gym exercise only if incentives were framed as losses rather than gains. The findings are consistent with the concept of loss aversion and suggest that cost reductions and performance improvements can be achieved if opting to frame incentives in terms of losses.

JEL: C93,D03,I10
Keywords: Framing, self-control, financial incentives, habit formation, loss aversion

I. Introduction

When it comes to promoting health and preventing major diseases, many of us face self-control challenges. Such challenges generally occur either in the form of a temptation to excessively consume harmful products (e.g. eating sugary foods, smoking, drinking, etc.) or a sparse performance of activities promoting sound health (e.g. exercising, precautionary doctor visits, etc.). A common explanation for these challenges are time-inconsistent, present-biased preferences. Present-biased preferences imply that an individual uses different discount factors in the short and the long-run, such that immediate actions can be suboptimal given long-run preferences (Laibson 1997). The economic consequences of health-related self-control problems can be substantial, ranging from physical costs of illness and lower life expectancy, to reduced productivity at work and escalating healthcare costs. Because markets often offer insufficient commitment devices to help individuals to stick to their long-term goals, or worse, exploit consumers’ time-inconsistencies, it is a relevant problem for public policy to understand how to promote more preventive health-oriented behavior.

One approach to address and battle self-control problems is through financial
incentives. For example, many large companies subsidize their employees' health club membership fees, insurers offer cash paybacks upon signing up at gyms, and local authorities provide free physical activity courses for individuals at risk of physical wellness. The central target of such financial incentive programs is to encourage habit formation. This implies that individuals overcome their self-control challenges while incentives are offered, and continue to act in line with their long-run preferences after financial incentives are removed (Becker & Murphy 1988). However, a potential harmful effect of financial incentives is that self-control problems could exacerbate once incentives are removed by "crowding-out" intrinsic motivation (Deci 1971). Up to date, it is still an open question whether short-term incentives can be designed to encourage persistent long-run effects, and what the optimal incentive design should look like.

In this paper, we present the results of a randomized field experiment which sheds light on the importance of "framing" financial incentives. According to Tversky & Kahneman (1986), framing refers to describing economically equivalent outcomes in different terms, with one potential application being the presentation of outcomes in terms of gains versus losses. By implicitly shifting the reference point, relative to which incentives are perceived, framing can affect the efficacy of incentives by provoking sensations of loss aversion. The main research goal of this paper is to explore whether individuals react stronger to financial incentives that evoke sensations of losses, as compared to economically equivalent incentives that do not trigger loss aversion. If individuals indeed react stronger to incentives framed in terms of losses, the effectiveness of financial incentives could be enhanced virtually free of costs through a simple shift in presentation, thereby creating great potentials for institutions to reduce costs and/or improve the performance of incentives. Furthermore, a corresponding question is whether framing financial incentives in terms of losses also affects the likelihood of forming beneficial habits in the long run. The second central aim of this paper is thus to study whether framing in terms of losses can be helpful to encourage habit formation, or in contrast, results in a "crowding-out" effect of intrinsic motivation after financial incentives are removed.

To address these questions, we conducted a randomized field experiment in cooperation with a regional fitness chain company in Germany. Similar to previous studies, treated subjects were given financial incentives for regular visits to the gym (Charness & Gneezy 2009, Aceland & Levy 2015, Royer et al. 2015). In the recruitment process, we screened for members with potential self-control challenges that indicated that they would like to exercise more often than they currently do. 94 of these subjects were randomly assigned to three groups: one control group, one gain-treatment group and one loss-treatment group. The two treatment groups received differently framed incentives on a per-visit basis over a period of four weeks. In particular, subjects in the gain-treatment received 2.50€ for each visit to the gym (up to one visit per day), with the accumulated amount
of incentives to be paid every two weeks. Subjects in the loss-treatment group received a personal credit of 35€ at the beginning of a two-week period which was described as an investment to support a healthy lifestyle. This investment was kept at the experimenter’s bank account and was temporarily inaccessible. For each day that a participant visited the gym, no amount was deducted from the bank account. For each day that the participant did not visit the gym, she lost 2.50€ of the total entitlement. The personal credit minus the reductions for non-attendance would eventually be paid at the end of the two-week period.

These negatively framed incentives were designed to shift the reference point of participants in the loss-treatment group towards the monetary amount of 35€, so that loss aversion was triggered at each day the subject does not exercise at the gym. Due to the additional psychological influence factor, we expected exercising frequencies to be highest among subjects in the loss-treatment group, as long as incentives were offered. However, after incentives were removed it was unclear how the prolonged exposure to negative incentives affects long-run exercise patterns.

The experiment yields two main results. First, we find a strong effect of "framing" in terms of losses while incentives are offered. During the intervention period, participants who were assigned to the loss group exercise more often. We observe an increase in average weekly gym attendance by 56% and a difference in weekly attendance between subjects in the loss treatment group and subjects in the control group of 0.69 visits. Conversely, for participants assigned to the gain treatment that received economically equivalent financial incentives, we do not find a significant increase in gym attendance. Statistically, these subjects behaved no different than individuals in the control group that received no financial incentives at all. Second, framing in terms of losses results in habit formation after incentives are removed. In the post-intervention period of six weeks after the removal of incentives, participants in the loss treatment group continued to visit the gym more often than participants in the gain group, with a difference in average weekly attendance of 0.4 visits. These extended effects of negative incentives are found to slowly decay over time. Over a late post-intervention period of twelve weeks after the removal of incentives, participants in the loss treatment group continued to visit the gym more often than participants in the gain group, but the difference in average weekly attendance is no longer statistically significant.

The results are consistent with the theory of loss aversion and point to a strong importance of framing when designing financial incentives. In the current setup, even among registered fitness club members who pay a bi-weekly fee in order to exercise at a gym, small incentives can encourage more frequent exercise. This is only true, however, if they are framed in terms of losses. Since the participants

---

1These bi-weekly payoffs are chosen as members pay their membership fees every two weeks. Further information about the fitness chain and the different contract types is provided in Section II.

2Since the experiment ran over four weeks, the 35€ payment was announced twice.
in our experiment were screened by their likelihood to exhibit self-control problems and to procrastinate attendance at the gym, the results suggest that loss aversion can be an effective tool to counteract procrastination. Furthermore, the application of loss framing also seems to positively affect long-term behavior as we find an indication of habit formation with no evidence of crowding out of intrinsic motivation after incentives are removed. Loss framing thus seems to be superior to gain framing, as it triggers stronger in-treatment effects which in turn increase the likelihood to cause persistent and beneficial changes in behavior.

This work contributes to the literature that investigates the effects of financial incentives on health-related behavior. Financial incentives can be designed in a variety of ways, such as large one-time incentives that are paid after a predetermined threshold is attained (Charness & Gneezy 2009, Volpp et al. 2009), small immediate incentives that reward each instance of a behavior (Aceland & Levy 2015), or large negative incentives in the form of pre-commitment contracts that threat losses if a predetermined threshold is not met (Giné et al. 2010, Schwartz et al. 2014, Royer et al. 2015). Our study introduces small and immediate negative incentives that penalize each instance of a behavior. While the above mentioned studies indicate that health-related performance can be improved through the use of financial incentives, the relative effectiveness of various schemes has not been assessed so far. For example, even if we observe a strong response to pre-commitment contracts, we do not know whether this response was driven by the incentive design or because the environment was particularly suited for incentives in general. This paper takes a first step to close this gap by directly comparing two logically identical incentive schemes with different frames, and by measuring which scheme provokes stronger responses in health-related behavior.

The work also contributes more broadly to the literature on framing effects. Framing has recently received much attention in research on principal-agent settings. In a series of field experiments, Hossain & List (2012) and Hong et al. (2015) have shown that Chinese workers exert more effort when working under a penalty contract than under an economically equivalent bonus contract. In an education context, Fryer Jr et al. (2012) and Levitt et al. (2012) demonstrate that the performance of pupils being taught by a teacher working under a contract that penalizes low student performance is higher than the performance of pupils with teachers working under logically equivalent contracts that reward high student performance. In these environments, however, framing of financial incentives is used to correct for a mismatch between the preferences of the agent and the principal, rather than correcting for a mismatch between the agents’ short-run and long-run preferences. To the best of our knowledge, the paper by Kaur et al. (2015) is the only research that analyses the relationship between framing of financial incentives and self-control. The authors investigate the behavior of workers with self-control problems who end up slacking in

\[^{3}\text{See Gneezy et al. (2011) for a recent survey.}\]
the short-term even though working harder would be beneficial under long-run preferences. They show that, in some cases, workers actively choose penalty-contracts as a personal commitment to behave in accordance with their long-run preferences. While these findings indicate that negative incentives are effective in targeting self-control problems, they are not directly applicable to other environments. In a non-workplace environment, such as in a gym, financial incentives can generally not be part of a contractual agreement. Thus, when incentives can only be provided temporarily, it is essential to understand how they continue to influence behavior after they are removed. Furthermore, the extent of self-control problems is likely to be different in a workplace environment than in a non-workplace environment. While some workers might be willing to self-select punishment contracts to prevent time-inconsistent behavior, such behavior is not always applicable to health-related self-control problems, as the underlying benefits are more obscure and further in the future. This work attempts to shed light on the relationship between framing incentives and self-control problems in a non-workplace environment, and can therefore be seen as complementary to the work by Kaur et al. (2015).

The following chapters of this paper are structured as follows: In Section II we describe the experimental design. Next, Section III presents the empirical results. Section IV is devoted to robustness checks and Section V provides a discussion and conclusion.

II. Experimental design

A. Subject recruitment

The experiment was conducted at a regional fitness chain in Germany starting in March and ending in April 2017. At this time, the fitness chain owned twelve fitness clubs which operated in one of three categories: basic, basic-plus or women-only. Basic-plus was the most common category comprising 9 out of the 12 clubs. All three categories were similar in terms of equipment and facilities and all offered free showers, lockers and water service, as well as separate rooms for classes such as aerobics and yoga. The basic-plus and women-only gyms additionally offered small complementary services such as free coffee or sport drinks. Each member pays a bi-weekly membership fee which is automatically deducted from their bank account via direct debit authorization. This bi-weekly fee amounts to 11.90€, 16.90€ or 18.90€ for the basic, basic-plus and women-only gyms, respectively. To become a member, a contract has to be signed at one of the three gym categories at which members are then allowed to exercise at. Each contract’s duration is either set at an annual or monthly membership, and both memberships entail automatic renewal. The only difference between

---

4 Members are also permitted to exercise at clubs in lower gym categories, however, not at clubs in higher categories.
these two membership-types is that for the monthly membership members have to pay 40€ more in contract initiation fees. Each gym is open seven days a week, with opening hours from 06:00 a.m. to 11:00 p.m. during weekdays and from 08:00 a.m. to 08:00 p.m. on the weekend. Finally, in order to access a gym, members have to log-in via a magnetic swipe card. We use this log-in data as our primary information on club members’ gym attendance.

We recruited 94 registered members of the fitness chain to participate in our experiment. Our population of interest are members with self-control problems, i.e. gym members who exercise at a level less than their self-reported personal target frequency. Only for these members could financial incentives potentially facilitate a habit formation effect. In order to recruit a representative sample from this population of interest, we randomly selected a sample of members from the full list of members, excluding newly registered members who held a contract for a period of less than three months. The selected members then received an email invitation to participate in an experiment which was described as a “university project conducted in cooperation with the fitness chain”. The email invitation explained that the study aims at investigating strategies to encourage regular exercise behavior, but it did not mention that financial incentives are involved. It also outlined the intervention period of four weeks over which the experiment would take place and highlighted that each participant would receive a compensation of 20€ upon registration. To register for the experiment, members then had to complete a short online survey and confirm that they want to take part in the experiment. In the survey, we included a screening question which asked for the subjects’ current and desired exercising frequency. All subjects who stated that they are currently exercising at or above their desired exercising frequency were informed that they were not eligible to participate in the study. This recruitment procedure lasted until the prespecified number of experimental participants was successfully attained. Thereafter, the participants were randomly assigned to three groups: one control group, one gain-treatment and one loss-treatment. Subjects assigned to the treatment groups received small incentives per-visit to the gym. Participants assigned to the control group received no financial incentives. Detailed information about these interventions is provided below.

We are aware of the fact that recruitment could be performed in a variety of ways. In our case the sample consists of members with potential self-control problems that responded to the email invitation, and is thus not perfectly randomized. However, as the experiment took place in the field we had to comply with some exogenous conditions that were imposed by the fitness chain. In particular, we were required to get the explicit consent of all subjects stating that they are voluntarily participating in the experiment and that we may use their

5We excluded newly signed-up members due to two reasons. First, to obtain a balanced panel data set, as we also collect data about exercising frequencies three months prior to the start of interventions. Second, to avoid biasing our results, as newly signed-up members are still in the process of developing habits, independent of financial incentives.
exercise data for scientific research. Unfortunately, we have no information about members who did not respond to the email invitation, so we have no possibility to control for potential selection problems. Although unlikely, it might be possible that some members with infrequent exercise behavior did not want to sign up because they did not want their exercise patterns to be revealed. To minimize this form of selection bias and to generally encourage participation, we chose to reveal some broad information about the purpose of the experiment. As such, our email invitation was designed to be targeted towards members with infrequent exercise behavior and thereby was not as attractive to members with regular exercise patterns. As the population of interest lies in members with self-control challenges, the underlying recruitment process should generally be pertinent to encourage responses by these members. For the members that did not respond to the invitation we have to make the conservative assumption that they would not react to financial incentives. Finally, we decided to mention the time period of the intervention in the email invitation in order to encourage members only to sign up if they are available to attend the gym during this period. Thereby, we are able to exert increased control on compliance with our treatment and minimize attrition during the intervention period.

B. Interventions

The intervention period lasted for four weeks and began in March 2017. One week before the start, information about our treatment offers was communicated by sending emails to the participants assigned to the treatment groups. These participants received financial incentives for exercising at the gym and financial incentives were economically equivalent for both treatment groups. The only difference between treatments was how incentives were framed. In the gain-treatment, we applied a framing that rewarded frequent attendance at the gym. In particular, the email for this group stated that each participant would receive a compensation of 2.50€ for each visit at the gym (up to one visit per day). Attendance rates would be checked twice, after the first two weeks and after the second two weeks of the intervention period, and accumulated payments would automatically be transferred to the participant’s bank account after each check point. In the loss-treatment incentives were framed in such a way that infrequent attendance at the gym was penalized. The email for this group described that each subject receives a one-time personal credit of 70€. This credit is split in half and 35€ would be transferred to the subject’s bank account at the end of the first two weeks and another 35€ would be transferred at the end of the second two weeks of the intervention period. For each day that the participant does not exercise at the gym, 2.50€ would be subtracted from the credit amount for the respective two-week period. Likewise, attendance rates would be checked twice and at the end of a two-week period each participant receives his personal credit minus accumulated deductions for non-attendance.

The goal of framing incentives in terms of penalizing non-attendance was to in-
duce loss aversion among participants in the loss-treatment group. The literature in behavioral economics predicts that in this setup loss aversion would generate additional psychological incentives and promote higher attendance rates as participants want to avoid losses arising from the non-attendance penalty. Note, however, that the implemented loss-treatment is likely to provoke only an attenuated form of loss aversion. Payments are made ex-post at the end of each two-week period, and in the worst case participants would receive a transfer of 0€ if they would not exercise at all. It is likely that loss aversion would be intensified if participants would receive money ex-ante and after each day of non-attendance a penalty fee would be deducted from the participants’ bank accounts. Unfortunately, due to administrative restrictions, it was not feasible to implement such a treatment design in the present study. Our estimated treatment effect for the loss-treatment group could therefore be regarded as a lower bound of the effect of a more stringent loss framing in which financial incentives are provided upfront and non-compliant behavior is penalized immediately.

Figure 1.: Timing of the experiment

The timing of the experiment is illustrated in Figure 1. We collected gym attendance data over a period of 28 weeks, which included the four-week intervention period as well as twelve weeks before and twelve weeks after this period. Gym attendance was monitored by observing magnetic card swipe log-ins which were necessary to enter gym facilities. As common at most gyms (and similar to previous research) members only have to log-in via a magnetic card but not to log-out. We could thus not observe the length of a visit, so generally speaking participants could cheat by just logging in and not exercising to collect the rewards. To prevent such behavior, we imposed that the length of each visit had to be at least 30 minutes. While we had no possibility to directly control compliance, we told participants that the gym staff was instructed about the experiment and would check for exercise duration and suspicious log-in behavior. These precautions should mitigate cheating but ultimately we are not able to completely rule out that some participants might have visited the gym without exercising. It should be noted, however, that since financial incentives are eco-

---

6We realize that such an alternative loss-treatment would no longer be economically identical to our gain-treatment because payments would be made at different points in time. The benefit of our implemented loss-treatment is therefore that we can isolate the effect of loss aversion relative to the alternative treatment of framing outcomes in terms of gains.
nomically identical across treatments, incentives to cheat should also be identical for both groups. Hence, even in the unlikely case of dishonest conduct we would not expect cheating to explain any observed differences in behavior across treatment groups. For the estimation of post-intervention treatment effects cheating should play no role at all. After incentives are removed there is no reason for participants to log in at the gym without the intention to exercise.

Our final sample consisted of 94 participants, divided in 34 control subjects, 30 gain-treatment subjects and 30 loss-treatment subjects. 100 participants were recruited initially. After registration was closed, six treated subjects reported that they were unable to attend the gym during the intervention period due to injuries, holidays or job travel. These subjects were excluded from the sample.

### III. Results

#### A. Descriptive results

We start with a descriptive analysis of the observed exercise behavior. Figure 2 depicts the average weekly attendance rates in each group before, during and after the experiment. Note that before and after the experiment average weekly attendance rates are derived over a period of twelve weeks, while during the intervention period weekly attendance is averaged over a four week period.

![Figure 2. : Average weekly attendance per group](image)

Experimental randomization seemed to be successful. We observe similar attendance rates across groups before the intervention started. In the twelve weeks

\[As a robustness check, we included our six dropped treatment subjects back into the sample. This specification does not yield qualitatively different results.\]
prior to the experiment, subjects in the control and gain group visited the gym 1.5 times per week on average. Subjects assigned to the loss group showed a slightly lower attendance rate with an average of 1.3 visits per week. As expected, due to randomization, attendance rates across groups are not statistically different. Wilcoxon-Mann-Whitney tests do not reject the equality of attendance rates before the experiment between the control and the loss group \((p = 0.30)\), the control and the gain group \((p = 0.96)\) as well as between the gain and the loss group \((p = 0.35)\). Additionally, we can confirm that attendance in the control group was constant over the whole experimental time frame. Wilcoxon-Mann-Whitney tests do not reject the equality of attendance rates in the control group before and during the intervention \((p = 0.65)\), before and after the intervention \((p = 0.11)\) as well as during and after the intervention \((p = 0.11)\).

While incentives are offered, we observe higher attendance rates in the treatment groups. Subjects in the gain group increased weekly attendance by \(27\%\) to 1.9 visits per week on average. Subjects in the loss group increased attendance by \(56\%\) to 2.1 visits per week on average. This indicates that our treatment was successful and that participants responded to the small per-visit incentives. It also highlights that the way incentives were framed seems to be important. Economic theory predicts identical treatment effects since both treatment groups received identical incentives. What we actually observe is that the difference in average weekly attendance before and during the intervention is two times higher if incentives are framed in terms of losses.

Finally, after incentives were removed, participants in both treatment groups rapidly reduced their exercise frequencies. Relative to the intervention period, attendance in the gain group dropped by \(37\%\) to 1.2 visits per week on average. This drop is larger than the increase in attendance caused by the monetary incentives and indicates that there might be crowding-out of intrinsic motivation as a result of providing financial incentives. Likewise, attendance rates in the loss group dropped by \(36\%\) to 1.3 visits per week on average, which is exactly the average attendance frequency as before the intervention. Therefore, based on these purely descriptive results, there seems to be no indication of habit formation for participants in the loss group.

Furthermore, in Table 1 we present information about three key demographic variables: (1) age, (2) gender and (3) ex-ante exercise behavior. Previous research has shown that heterogeneity in these demographic variables could have an important impact on the effectiveness of our employed incentives. The mean age in our sample is 29 years. Across the three groups age happens to be somewhat unevenly distributed with a higher average age in the control group than in the two treatment groups. In particular, a standard t-test indicates that the mean age in the control group is statistically different from the mean age in the gain group. This can be observed in Table 1, column \((5)\) that shows a \(p\)-value of 0.01 for a test of equal means. We attribute this difference to the fact that our sample is not restricted to a specific subgroup of the population, such as a
Table 1—: Means of demographic characteristics across groups

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample group</td>
<td>Full control group</td>
<td>Gain group</td>
<td>Loss group</td>
<td>p-value</td>
<td>p-value</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>29.05</td>
<td>32.26</td>
<td>26.20</td>
<td>28.27</td>
<td>0.01</td>
<td>0.11</td>
<td>0.24</td>
</tr>
<tr>
<td>Male</td>
<td>0.56</td>
<td>0.53</td>
<td>0.57</td>
<td>0.60</td>
<td>0.77</td>
<td>0.58</td>
<td>0.80</td>
</tr>
<tr>
<td>Regular</td>
<td>0.74</td>
<td>0.74</td>
<td>0.73</td>
<td>0.77</td>
<td>0.99</td>
<td>0.78</td>
<td>0.77</td>
</tr>
<tr>
<td>Observations</td>
<td>94</td>
<td>34</td>
<td>30</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The table shows the sample averages of key demographic characteristics. Columns (5), (6) and (7) show p-values of standard t-tests for equal means assuming equal population variances.

...student sample, but rather represents an accurate approximation of the true underlying population distribution. In our sample, age ranges from 18 to 63 years and some older subjects happen to be randomly assigned to the control group. This could be problematic if our treatment is especially effective with younger (older) people as we will tend to overestimate (underestimate) the true average treatment effect. In the robustness checks in Section IV, we will therefore take explicit consideration whether treatment effects are affected by differences in age.

The other key demographic variables are very similarly distributed. Overall, 56% of our sample is male, with similar proportions in all three groups. Standard t-tests reported in Table 1 strongly support the hypothesis of equal gender proportions across all groups. Additionally, we consider whether a participant was an ex-ante regular gym-attendee. A regular attendee is defined as an individual that visits the gym at least once every two weeks prior to the start of the interventions.\(^8\) Around 74% of the participants in our sample meet this requirement. Again, this proportion is very similar in all three groups and t-tests strongly support the hypothesis of equal proportions of regular attendees across all groups.

B. Regressions

We estimate a linear difference-in-difference panel regression model to fully capture the variation in attendance rates across time and groups. The model takes the form:

\(^8\)This definition was chosen, as not to misclassify regular attendees, who might miss a week of exercise due to illness or travel. In Section IV we also considered alternative definitions of regular exercise as a robustness check.
Attendance_{it} = \beta_0 + \beta_1 During_{it} + \beta_2 Early_{it} + \beta_3 Late_{it} + \beta_4 Gain_{it} \cdot During_{it} + \beta_5 Loss_{it} \cdot During_{it} + \beta_6 Gain_{it} \cdot Early_{it} + \beta_7 Loss_{it} \cdot Early_{it} + \beta_8 Gain_{it} \cdot Late_{it} + \beta_9 Loss_{it} \cdot Late_{it} + \gamma_i + \epsilon_{it}, \tag{1}

where Attendance_{it} measures the number of visits for subject i in week t and During_{it} is a dummy variable taking the value one during the intervention period. The post-intervention period is split in half into the early post-intervention period and the late post-intervention period to allow a more accurate representation of the exercise behavior after incentives are removed. The dummy variable Early_{it} captures the post-intervention period of up to six weeks after the removal of incentives. Analogously, Late_{it} captures the late post-intention period of six until twelve weeks after the removal of incentives. The twelve week period prior to the start of the intervention serves as the omitted category. Dummies Gain_{it} and Loss_{it} indicate whether a subject is assigned to the gain or the loss group, respectively, with the control group as the omitted category. In order to increase the precision of our analysis, we further include individual fixed effects denoted by \gamma_i. Coefficients \beta_4 to \beta_9 capture the treatment effects and describe how differences in attendance across time in the treatment groups, differ from analogous differences in attendance across time in the control group. As we have repeated observations for each subject i over time, standard errors are clustered at the individual level.

The result of this exercise is presented in Table 2. The dependent variable in this regression, the number of visits per week, ranges from 0 to 7.\footnote{Despite the discrete and bounded nature of the dependent variable we present OLS estimates to facilitate the interpretation of results. A Poisson regression model would be the appropriate choice in this context to correctly account for the count data structure of the dependent variable. As a robustness check, we estimated a Poisson model and it produced very similar, qualitative equivalent results.} Column (1) shows the estimated in-treatment effects. As expected, and consistent with the descriptive analysis above, we estimate a positive in-treatment effect for both treatment groups during the intervention period. In particular, for participants assigned to the loss treatment we estimate an average of 0.686 additional visits per week in the intervention period, compared to the control group. For participants assigned to the gain group, we estimate an average of 0.344 additional visits per week compared to the control group. Only the effect for the loss treatment group is estimated to be statistically significant at a 5% level. For participants assigned to the gain treatment, we cannot reject the hypothesis that average attendance rates are equivalent to those of the control group. As a consequence, this finding confirms that financial incentives were only successful in encouraging more frequent gym attendance when they were framed in terms of losses, as opposed to when they were framed in terms of gains. Furthermore, at
Table 2—: Fixed Effects Difference-in-Difference OLS regression

**Dependent variable: Average weekly visits**

<table>
<thead>
<tr>
<th></th>
<th>In-Treatment (1)</th>
<th>Early Post-Treatment (2)</th>
<th>Late Post-Treatment (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>During</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.16)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>GainDuring</td>
<td>0.344</td>
<td>-0.236</td>
<td>-0.236</td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.21)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>LossDuring</td>
<td>0.686**</td>
<td>0.164</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.23)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Early</td>
<td>-0.059</td>
<td>-0.079</td>
<td>-0.221</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.19)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>GainEarly</td>
<td>-0.236</td>
<td>-0.079</td>
<td>-0.079</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.19)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>LossEarly</td>
<td>0.164</td>
<td>0.109</td>
<td>0.109</td>
</tr>
<tr>
<td></td>
<td>(0.23)</td>
<td>(0.24)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Subject-week observations</td>
<td>2632</td>
<td>2632</td>
<td>2632</td>
</tr>
<tr>
<td>Nr. of subjects</td>
<td>94</td>
<td>94</td>
<td>94</td>
</tr>
</tbody>
</table>

p-values of tests of equal treatment effects: Gain vs. Loss

|                  | 0.257            | 0.082                     | 0.462                   |

*Note:* Robust standard errors clustered at the individual level in parentheses. All regressions include individual fixed effects. In-Treatment corresponds to the estimated treatment effect while incentives were active. Early post-treatment refers to the the period of up to 6 weeks after incentives were removed. Late post-treatment is the period of 6-12 weeks after the removal of incentives.

*** p < 0.01, ** p < 0.05, * p < 0.1

In the bottom of the table, we additionally present p-values from tests of the equivalence of treatment effects in the gain and the loss group. These tests check whether we still observe a significant treatment effect for the loss group if we compare attendance directly with the gain group, rather than with the control group. Given this specification we find that the difference in attendance between the loss and the gain group is not statistically significant.

In column (2) we present the estimated early post-treatment effects. We find a negative effect of -0.236 visits per week on average for subjects in the gain treatment group. For participants in the loss treatment group, we estimate a
positive early post-treatment effect of 0.164 additional visits per week on average. These estimates refer to the treatment effects measuring a change in gym attendance relative to the control group and are estimated to be not statistically different from zero. Yet, the signs of these two treatment effects indicate that there are diverging post-intervention trends in exercise behavior. In the gain-treatment group, the intervention seemed to cause a "crowding-out" of intrinsic motivation, while in the loss-treatment group it seemed to have generated habit formation. This observation can be confirmed when we compare the estimated treatment effects directly with each other. The reported p-value of 0.082 at the bottom of Table 2 shows that the estimated treatment effects between the loss and the gain group are statistically different at a 10% level. This implies that when compared to the gain group, the early post-treatment effect for the loss group is estimated to be 0.4 additional visits per week on average and this effect is weakly significant. We interpret this finding as evidence of habit formation induced by the negatively framed incentives.

Finally, in column (3) the late post-treatment effects are presented. These estimates show that twelve weeks after the removal of incentives the effects of the treatments have largely disappeared. The estimated treatment effect for the loss group is still larger than that for the gain group, but the difference in treatment effects is no longer statistically significant, as indicated by the p-value of 0.462 at the bottom of the table. In the late post-treatment period we can therefore no longer assert that participants in the loss group continue to attend the gym more often than participants in the gain group, nor more often than participants in the control group. This finding highlights that the effects of loss framing are gradually decaying over time and that, in this setup, habit formation can be confirmed, however only at a 10% significance level, for an extended period of six weeks after the removal of incentives. On the other hand, this extended period of habit formation was achieved virtually free of charge through a simple change in the framing of the employed incentives.

Figure 3 offers a more thorough picture of the reported long-run differences in treatment effects and the decaying effect of loss framing. Each subfigure presents the differences in the estimated post-treatment effects between the indicated groups for different lengths of the post-intervention period. Figure 3(a) shows the differences in treatment effects between the gain and the control group over time. Since there is no treatment effect for the control group, this difference corresponds to the estimated post-treatment effect for the gain group. We can see that the post-treatment effects are estimated to be negative or very close above zero for any assumed length of the post-intervention period and at no point statistically significant at a 10% level. In Figure 3(b) we plot the post-treatment effects for the loss group. Here we could observe a decaying effect of the treatment. Within the first two weeks after incentives were removed, treatment effects were above 0.5 and statistically significant at the 10% level. This effect, however, quickly faded away and was no longer statistically significant assuming a
Figure 3. Difference in Post-treatment effects

(a) Gain vs. Control group

(b) Loss vs. Control group

(c) Loss vs. Gain group
post-intervention period of three weeks or more, and almost no longer observable after twelve weeks. Figure 3(c) plots the differences in post-treatment effects between the gain and the loss group. The difference is most distinct directly after the intervention period. In the first week after incentives were removed, we estimate one additional visit per week if participants were assigned to the loss treatment group rather than the gain treatment group. The difference quickly declines after an additional week but stays constantly elevated above zero for any length of the post-intervention period. As described above, for a length of six weeks after the removal of the incentives, we estimate a difference in treatment effects of 0.4 visits per week. Considering any post-intervention period of more than six weeks yields a difference of around 0.3 additional visits per week, but this difference is no longer significant at a 10% level.

The results of this experiment allow us to make an important distinction between different counterfactual scenarios when providing financial incentives. In particular, we find habit formation only when comparing the two treatment groups directly with each other. This is due to diverging post-intervention effects of gain framing versus loss framing, resulting in significant differences in attendance between the two groups. In general, whenever external incentives are provided, different counterfactual scenarios are perceivable. One counterfactual scenario is to provide no incentives at all. Another one, however, is to provide the same incentives but framed differently. In reality, we observe individual decisions always in one particularly framed environment and we do not know what would have happened if the decision would have been made in a differently framed environment. Our results show that if we decide to frame incentives in terms of gains, we are foregoing long-run benefits that could be obtained if the incentives would have been framed in terms of losses.

Finally, we can compare our estimated in-treatment and early post-treatment effects with a very similar recent study that investigates the effects of financial incentives on gym attendance. Royer et al. (2015) conduct an experiment in an American Fortune 500 company with a sample of company employees. Over a four week period, treated employees received incentives of $10 for each visit in the company’s on-site gym, with up to $30 per week. The authors estimate an in-treatment effect of 0.56 additional visits per week. These positive incentives can be compared to our estimated in-treatment effect for the gain-treatment group of 0.34 additional visits per week. Since our study comprises smaller per-visit incentives, it is not surprising that we estimate a smaller in-treatment effect. However, if our incentives are framed in terms of losses, our estimated in-treatment effect is slightly higher, with 0.69 additional visits per week. A comparison of the early post-treatment effects reveals a similar picture. Royer et al. (2015) estimate a positive early post-treatment effect of 0.12 additional visits per week, which is indicative of habit formation. In our study, if incentives are framed in terms of gains, we estimate a negative early post-treatment effect of -0.24 visits per week. If incentives are framed in terms of losses, however,
we observe a very similar effect of 0.16 additional visits per week. Of course, the two studies are not perfectly comparable. Yet, this juxtaposition of results offers indicative evidence that per-visit incentives framed in terms of gains must be multiple times higher than incentives framed in terms of losses in order to generate similar in-treatment and post-treatment effects.

To summarize, given the usage of small and short-termed incentives, we interpret our findings as evidence that framing in terms of losses is more effective than framing in terms of gains. By framing in terms of gains we cannot confirm that providing small incentives has a beneficial effect. When we frame in terms of losses, we unambiguously observe positive effects. These differences in responses to economically equivalent incentives are likely to be driven by psychological factors. In particular, through loss framing we can amplify the external stimuli induced by the financial incentives with additional psychological stimuli, such as loss aversion. These additional stimuli may shift incentives from being non-effective to becoming effective and create scope to generate persistent changes in habits.

IV. Robustness checks

A. Heterogeneity of Treatment Effects

In this section, we investigate the possibility of heterogeneous treatment effects with respect to three exogenous characteristics: age, gender and ex-ante exercise behavior. Charness & Gneezy (2009), Aceland & Levy (2015) and Royer et al. (2015) identify a relationship between the effectiveness of incentives and ex-ante frequency of exercise. In particular, participants who do not exercise regularly are found to be significantly more responsive to incentives. Furthermore, Royer et al. (2015) attribute their findings of a relatively low in-treatment effect to the fact that students, or in general younger people, are more responsive to financial incentives than employees, or older people. There also seems to be a gender difference in the responsiveness to incentives, as Royer et al. (2015) estimate significantly stronger post-treatment effects for males than for females.

Table 3 shows the results of the heterogeneity analysis. In each panel, the sample is cut based on one exogenous dummy variable and the estimated treatment effects, relative to the control group, are given for the gain and the loss group. Additionally, in columns (5) and (8) the differences in treatment effects between the two treatment groups are reported for the respective time period under consideration. Panel A cuts the sample based on an age dummy variable. We define a participant to be "Young" ("Old") if she has an age below (above) the median age of 28 years.\(^{10}\) The split in Panel B is based on gender. In Panel C the split is performed based on ex-ante exercise behavior, whereby we define a

\(^{10}\)This split was primarily chosen to obtain comparable sample sizes. We tried sample cuts at different values of age and the results are robust to these changes.
### Table 3—: Treatment effects based on heterogeneity cuts

**Dependent variable**: Average weekly visits

<table>
<thead>
<tr>
<th></th>
<th>In-Treatment</th>
<th>Early Post-Treatment</th>
<th>Late Post-Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gain group</td>
<td>Loss group</td>
<td>Gain group</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td><strong>Panel A: Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>0.236</td>
<td>0.642</td>
<td>-0.278</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.45)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Old</td>
<td>0.485</td>
<td>0.712*</td>
<td>-0.167</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.39)</td>
<td>(0.21)</td>
</tr>
<tr>
<td><strong>p-value: Young=Old</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.61</td>
<td>0.91</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel B: Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.192</td>
<td>0.625</td>
<td>-0.568*</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td>(0.40)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>Female</td>
<td>0.536*</td>
<td>0.762**</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.40)</td>
<td>(0.24)</td>
</tr>
<tr>
<td><strong>p-value: Male=Female</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.43</td>
<td>0.81</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Panel C: Ex-Ante Exercise</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular</td>
<td>0.417</td>
<td>0.717**</td>
<td>-0.239</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.36)</td>
<td>(0.26)</td>
</tr>
<tr>
<td>Non-Regular</td>
<td>0.141</td>
<td>0.597</td>
<td>-0.228</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.39)</td>
<td>(0.33)</td>
</tr>
<tr>
<td><strong>p-value: Regular=Non-Regular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.52</td>
<td>0.82</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note**: The table shows estimated treatment effects relative to the control group, with an exception in columns (5) and (8) which report estimated treatment effects in the loss group relative to the gain group. Robust standard error clustered at the individual level in parentheses. All regressions include individual fixed effects. Panel A presents the results of a split based on age. "Young" represents subjects with an age below the median age of 28 years. "Old" represents subjects with an age above or equal the median age. Panel C presents the results of a split based on ex-ante exercise behavior. "Regular" represents individuals who visited the gym at least once in a two-week period prior to the intervention. "Non-Regular" represents individuals who visited the gym at least once in a two-week period prior to the intervention.

**Note**: The results for the in-treatment and early post-treatment period based on these heterogeneity cuts are generally consistent with the baseline results of Section III. We estimate positive in-treatment effects for both treatment groups across

---

member to be a "Regular" attendee if she attended the gym at least once every two weeks in the pre-treatment period.\(^{11}\) Note that the heterogeneity cuts were not prespecified. The resulting sample sizes are thus generally smaller, standard errors are larger and estimates need to be more distinctly different from zero to be significant.

The results for the in-treatment and early post-treatment period based on these heterogeneity cuts are generally consistent with the baseline results of Section III. We estimate positive in-treatment effects for both treatment groups across

---

\(^{11}\)As explained above this split was performed as not to misclassify members as "Non-Regular" if they miss a week of exercise due to illness or travel. We also did an alternative split in which a subject is defined as a regular member if she attends the gym at least once every week in the pre-treatment period. The results are robust to this variation.
all sample cuts. Additionally, the in-treatment effects for the loss group are always estimated to be larger than the in-treatment effects for the gain group. Similarly, for the early post-treatment period we estimate the treatment effect for the loss group always to be positive and to be larger than the treatment effect for the gain group. The early post-treatment effects for the gain group are always estimated to be negative with the only exception being a positive estimate given a sample of only women.

Below each panel we report the p-values of Wald tests, testing for equal treatment effects across sub-populations within a treatment group. These tests check whether the reported baseline treatment effects in Section III are driven by differences in exogenous characteristics of the participants. With a single exception we find that our estimated in-treatment and early post-treatment effects are very robust and do not depend on the participants’ age or ex-ante exercise behavior. Only the difference in the estimated early post-treatment effects between males and females that received a gain-treatment is found to be significant, with a p-value of 0.07, as reported in column (3). This indicates that the overall finding of a negative early post-treatment effect for the gain group could be driven by a different response to incentives between sexes. A comparable result is obtained in Royer et al. (2015) who report a stronger early post-treatment effect for male participants. Our results suggest the opposite as we find that men show a strong and negative reaction to incentives in the first six weeks after incentives were removed, by reducing exercise frequency relative to pre-treatment levels. Women, on the other hand, do not seem to be negatively affected if incentives were framed in terms of gains. Recall, however, that these differences between genders are based on the estimated early post-intervention treatment effects for the gain group with a counterfactual scenario of providing no incentives at all. If the counterfactual scenario assumes differently framed incentives, the early post-treatment effects are not dependent on the gender of the participant. This can be observed in column (5) of Table 3 which shows a p-value of 0.2 for a Wald test of equal early post-treatment effects when the two treatment groups are compared directly with each other. Using this specification, we find no significant differences in estimated early post-treatment effects for sample cuts based on age, gender or ex-ante exercise behavior.

Finally, the last three columns of Table 3 report tests for heterogeneity in the estimated late post-treatment effects. The results generally confirm our baseline finding that the effects of the two treatments are no longer observable after twelve weeks. P-values above 0.1 at the bottom of each panel in columns (6) and (7) confirm that there are no significant differences in the late post-treatment behavior between young and old or between male and female participants. With respect to ex-ante exercise patterns, however, there seem to be important differences. Column (8) in Panel C shows that we find a strong late-post treatment effect for ex-ante non-regular attendees. In the extended period of six to twelve weeks after the removal of incentives, participants who received negatively framed incentives
have on average 0.96 additional visits per week compared to participants who received positively framed incentives. In contrast, for ex-ante regular participants we observe no differences in late post-treatment effects between treatments. This heterogeneous response to incentives, depending on whether a participant was an ex-ante regular attendee could also be found in Charness & Gneezy (2009) and Royer et al. (2015). In their studies, only ex-ante non-regular attendees displayed habit formation and increased their exercise frequency in the long run. In our study, we report habit formation formation in the first six weeks after the removal of incentives, regardless of the ex-ante exercise patterns of the participants. In the late post-treatment period, the heterogeneity analysis highlights that we continue to observe habit formation for subjects who received a loss framing but only if this subject was a ex-ante non-regular attendee. This finding further reinforces the increased effectiveness of loss framing over gain framing as it seems to induce habit formation for ex-ante non-regular attendees over an even longer time period.

B. Spillover effects

In field experimental research, a critical source of potentially mismeasured treatment effects are spillover effects. The closed environment of a health club in which our randomized controlled experiment was conducted might cause problems as participants can interact with each other. In particular, estimated treatment effects could be biased if there are interactions between participants within the same treatment group. For example, consider two treated participants that frequently exercise together. One individual is responsive to the received incentives and increases attendance at the gym. As a result, the second individual might similarly increase attendance but not because she responds to the received incentives, but rather because of the social spillover effect. Similarly, a related problem of social interaction arises if subjects assigned to the control group learn about treatments and subsequently reduce their exercise levels, potentially because they are disappointed to not have received the treatment. In both cases we would overestimate the true treatment effect.

Unfortunately, we cannot completely rule out the existence of such spillover effects, but they should generally be largely prevented in the current setup. Participation in our study is not restricted to a specific gym, but recruitment is performed using the full list of members that exercise at different clubs of the same chain. Thus, randomization should ensure that any interdependence of exercise behavior among recruited participants is minimized. As for the participants’ behavior within the control group, we could directly rule out that exercise patterns are changed as a response to potential knowledge about incentives. As presented in the descriptive results, we observe almost identical average weekly attendance rates for subjects in the control group over the entire 28 week time period.
V. Conclusion

Overcoming self-control problems and developing good habits can be a challenging undertaking. When it is socially desirable that individuals achieve their self-proclaimed long-term goals, providing financial incentives can be a useful tool to encourage the formation of beneficial habits. In this study, that is based on field-experimental research in a gym, we present evidence that the decision on how to frame financial incentives is crucial in order to maximize the underlying effectiveness. By using small incentives on a per-visit basis and framing them in terms of losses, we conclude that incentives increase attendance at the gym and result in habit formation. Providing economically equivalent incentives framed in terms of gains does not result in a significant changes in exercise behavior. We interpret this result as evidence for loss aversion. It is perceived as more painful by participants to give money back if they do not exercise than it is to receive money if they do exercise, even though the final result is identical. Consequently, exercising frequencies under loss framing are higher as the psychological incentives to avoid losses are stronger than those to acquire equal-sized gains. Ultimately, the higher level of physical activity translates into a higher probability to induce the formation of habits.

The presence of loss aversion has been confirmed in several lab experiments. This study shows that loss aversion affects intertemporal decision-making and demonstrates that framing incentives in terms of losses is superior to framing in terms of gains, when trying to target self-control problems. The results of the experiment have direct implications for practitioners contemplating to employ external incentives. Framing in terms of losses can shift incentives from being ineffective to becoming effective. This translates into cost savings, as fewer financial sources are required to achieve intended results. Put differently, in order to generate similar results it is likely that more financial resources are required when incentives are framed in terms of gains rather than framed in terms of losses.

This paper contributes to the literature about habit formation and self-control problems by quantifying the significance of framing of monetary incentives. However, it does clearly not address the list of open questions exhaustively. The focus of the present analysis is on the ability of financial incentives to facilitate the formation of "good" habits. It does not necessarily follow that loss framing would work equally well to encourage the cessation of "bad" habits. Especially in the context of addiction, the neurological underpinnings of self-control problems might be considerably different and other framing methods might be more effective to help people to abstain from harmful behavior. As Charness & Gneezy (2009) point out, there is mixed evidence about the effects of incentives on habit cessation using either losses or gains. However, to the best of our knowledge, there is no evidence about the relative effectiveness of different framing methods on encouraging the cessation of bad habits. This could be an interesting avenue for future research. Furthermore, we compare loss framing to gain framing using
relatively small incentives. It is unclear how behavior is affected if the size of the incentives is altered. Intuitively, if external incentives are strong enough, additional psychological influence factors should become less relevant. Under very high external incentives, recipients should simply put maximal effort in their attempts to obtain the incentives, regardless of the form of framing. On the other hand, if external incentives are very weak, psychological incentives should also play a minor role as punishments are too small to evoke true perceptions of losses. The relationship between the importance of loss aversion and incentive size is therefore likely to be non-linear and potentially follows an inverse U-shaped pattern. Clearly, more research in this area is required to make definite statements. However, whenever cost considerations are important, the message for practitioner is straightforward. Choosing the optimal framing method is as important as choosing the optimal incentive size.

REFERENCES


