# Using Dynamic Financial Incentives to Foster Walking Habit

Lawrence Jin
Ashish Sachdeva
Jeeva Somasundaram

**Abstract**

One major challenge of using financial incentives to induce behavioral change is that people often revert when the incentives are removed. This paper investigates if robust habits can be formed by offering personalized goals, coupled with daily incentives sustained over a long period of time. Overall, 523 adults were given wearable step trackers and were incentivized to walk 2,500 more steps per day than they normally would. Participants were randomly assigned to a no-incentive control condition, or one of four treatment conditions that offered up to SG$14 (US$10) per week if step goals were met every day. Two conditions offered a fixed SG$2 incentive per day to achieve the step goals, in either gain or loss framing. The two other conditions offered incentives that dynamically increased for each consecutive day the goal was met and for each consecutive day the goal was not met, respectively.

The incentives significantly increased the frequency of goal attainment during the 36-week intervention period. The effect sizes were similar across the four treatment conditions. Most importantly, we found almost perfect persistence of goal attainment rates during the 12-week post-intervention period. There is compelling evidence of habit formation: the dynamic incentive conditions incentivized participants to exercise in unique path-dependent patterns during the intervention period, and these participants continued to exercise in the respective patterns after the incentives were removed.

## Introduction

Habits are activities that become automatic and effortless after repeated practice over time (Pollak 1970; Becker & Murphy 1988; Wood & Runger 2016). A key aspect of habit formation is *adjacent complementarity*—the more often you engage in a particular activity, the more likely you are to continue doing it. Many behavioral change interventions rely on monetary incentives to promote the initial repetitions critical for habit formation (Charness & Gneezy 2009; Acland & Levy 2015; Royer et al. 2015, Finkelstein et al., 2016; Patel et al., 2016; Carrera et al., 2018; Carrera et al., 2019). However, while these interventions may lead to behavior change during the treatment period, participants often revert to their baseline once the intervention ends (Wood & Runger 2016).

The challenge of weak habit formation is a significant concern for researchers and policymakers, as short-term lifestyle improvements rarely lead to lasting health benefits. Moreover, such incentive-based interventions are not economically sustainable given their transient effects. In this paper, we propose financial interventions grounded in habit formation theories (Pollak 1970; Becker & Murphy 1988; Wood & Neal 2016) to encourage lasting behavioral change. Specifically, we evaluate our interventions in the domain of physical activity, aiming to foster the habit of walking more.

A seven decades of research has shown that physical activity has wide-ranging benefits for physical and mental health function. Physically active individuals live longer, due to lower rates of heart disease, stroke, cancer, and type 2 diabetes. In fact, they also have better cognitive functions and lower risk of dementia. Therefore, policy makers have generally recommended at least 150 minutes of moderate-intense physical activity -- such as brisk walking -- for better health (US Department of Health and Human Services 2018, WHO 2020). However, still many do not do the required amount of physical activity.

Previous studies have primarily used financial incentives to promote physical activity by encouraging gym attendance (Charness & Gneezy, 2009; Acland & Levy, 2015; Royer et al., 2015; Finkelstein et al., 2007, 2016; Carrera et al., 2018, 2019; Volpp et al., 2008; John et al., 2011; Cawley & Price, 2013) or weight loss (Volpp et al., 2006; Finkelstein et al., 2008; Cawley & Price, 2013). In contrast, relatively few studies have focused on increasing daily step counts (Patel et al., 2016; Omran et al., 2018; Kramer et al., 2024). Although these

studies successfully increased step counts during the intervention period, the effects were not sustained afterward—either disappearing entirely post-treatment (Patel et al., 2016) or lasting only a few weeks (Omran et al., 2018). Our study, in contrast, aims to foster a lasting habit of increased physical activity by incentivizing participants to meet daily goals, measured by step counts tracked through a wearable device.

There are several features of our approach that are designed to maximize the chances of habit formation. First, we provide daily goals (and incentives that could be achieved daily) helps reduce the ambitious lifestyle change into incremental milestones that are more achievable (Finkelstein et al., 2016; Patel et al., 2016; Chokshi et al., 2018). Second, coupling daily goals with a wearable step-tracking device provides salient real-time monitoring of each participant's progress towards the goal, which keeps participants engaged and reduces the chances of procrastination. Third, the daily goals were personalized based on each participant's baseline physical activity levels to ensure that the goals are relevant and achievable for everyone.[1] Fourth, the intervention took place for an extended period of 36 weeks – the longest field intervention conducted on physical activity.[2] This allowed for 252 instances of feedback and rewards (one each day), which are key to habit formation (Miller et al., 2019). Finally, we introduced two novel *dynamic* monetary incentives (described in detail below) to encourage people to exercise in momentum and discourage people from being inactive for an extended period of time. Habit formation theories (Becker and Murphy 1988, Pollak 1970) suggest that interventions promoting streaks are particularly effective in driving sustained behavioral change.

Five-hundred and twenty-three adults participated in a 48-week field experiment. All participants were given a wearable step-tracking device to use daily. We observed their physical activity levels for the first two weeks of the experiment to establish each participant's baseline step count. After that, we gave each participant a personalized daily step goal that was 2,500 steps more than each participant's baseline average. Participants were then randomly

[1] For example, Charness and Gneezy (2009) gave the same gym-attendance goals for each participant, and found that incentives had insignificant effect in increasing gym attendance for participants who were already regular gym attendees.

[2] Most economics studies have provided incentives for 4 to 12 weeks. A few recent medical studies have offered incentives for longer periods of time with mixed success. Patel et al. (2016) offered incentives for 13 weeks but did not find significant long-run effects. Chokshi et al. (2018) offered incentives to heart disease patients for 16 weeks. They found that, despite some level of relapse after the incentives were removed, the incentivized patients continued to exercise more than non-incentivized patients during the 8-week follow-up period.

assigned to either a non-incentivized control condition, or one of four treatment conditions. Each treatment condition offered a different incentive structure through which participants could meet the daily step goal. The intervention period was followed by a 12-week post-intervention period in which they were not offered any incentives.

We compared the control condition who received personalized daily step goals with four treatment conditions including two fixed incentives ("simple gain" and "simple loss") and two dynamic incentives ("gain streak" and "loss streak"). All conditions, including the control, significantly increased goal achievement rates and step counts during both the intervention and post-intervention periods compared to baseline. Treatment conditions also maintained a higher goal achievement rate (and step count) than control condition with minimal decline post-intervention. Notably, participants in dynamic incentive conditions (gain and loss streaks) continued exercising in path-dependent patterns, suggesting strong persistence and habit formation beyond the incentive period. The path-dependent pattern of gain and loss streak also led to varying frequencies of goal achievement days during both the treatment and post-treatment periods

.

**Experiment Design**

participants were randomly assigned to one of five conditions. The control participants did not receive any financial incentive during the experiment. They did, however, receive a personalized step goal (2,500 more steps than their baseline average) that was entered into the wearable device at the start of week 3. The other four treatment conditions provided monetary incentive for participants to meet their daily step goals.

All four treatment conditions offered subjects SG$14 (US$10) per week if they met their exercise goal every day in the week. However, the conditions differed in the following ways: (1) Simple gain condition, in which SG$2 was given each day the goal was met; (2) Simple loss condition, in which SG$14 was endowed at the beginning of the week and SG$2 was deducted each day the goal was not met; (3) Gain streak condition, in which participants earned SG$0.50 more each consecutive day the goal was met, starting with SG$0.50;[3] and (4) Loss

[3] For example, someone who exercised Monday, Tuesday, Wednesday, Thursday and Saturday would receive $5.50 ($0.50 on Monday, $1 on Tuesday, $1.50 on Wednesday, $2 on Thursday, and $0.50 on Saturday).

streak condition, in which SG$14 was endowed at the beginning of the week and participants lost SG$0.50 more each consecutive day the goal was not met, starting with -SG$0.50. The two dynamic incentive conditions—gain streak and loss streak conditions—incentivize participants to exercise in unique path-dependent patterns. The gain streak condition offered incrementally higher incentives for each consecutive day the goal was met. By contrast, the loss streak condition offered incrementally higher penalty for each consecutive day the goal was not met. Both interventions, grounded in habit formation theories (Becker & Murphy, 1988; Pollak, 1970), aim to either promote consistent streaks of goal achievement or minimize periods of non-achievement. The maximum possible earnings in each treatment is shown in Appendix: Figure S1.

736 participants completed the first visit. Of these participants, 615 completed the 2nd visit. Randomization was done during the 2nd visit. Out of 615 who completed the 2nd visit, 523 participants completed the entire study. The dropout rates for the control, gain, loss, gain streak and loss streak conditions were 20.0%, 10.2%, 12.5%, 16.1% and 15.6%, respectively. The higher dropout rate for the control condition could potentially lead to attrition bias. If less motivated participants tend to drop out, then the control participants who have completed the experiment will tend to be more motivated, which would downward-bias our estimates for the treatment effects

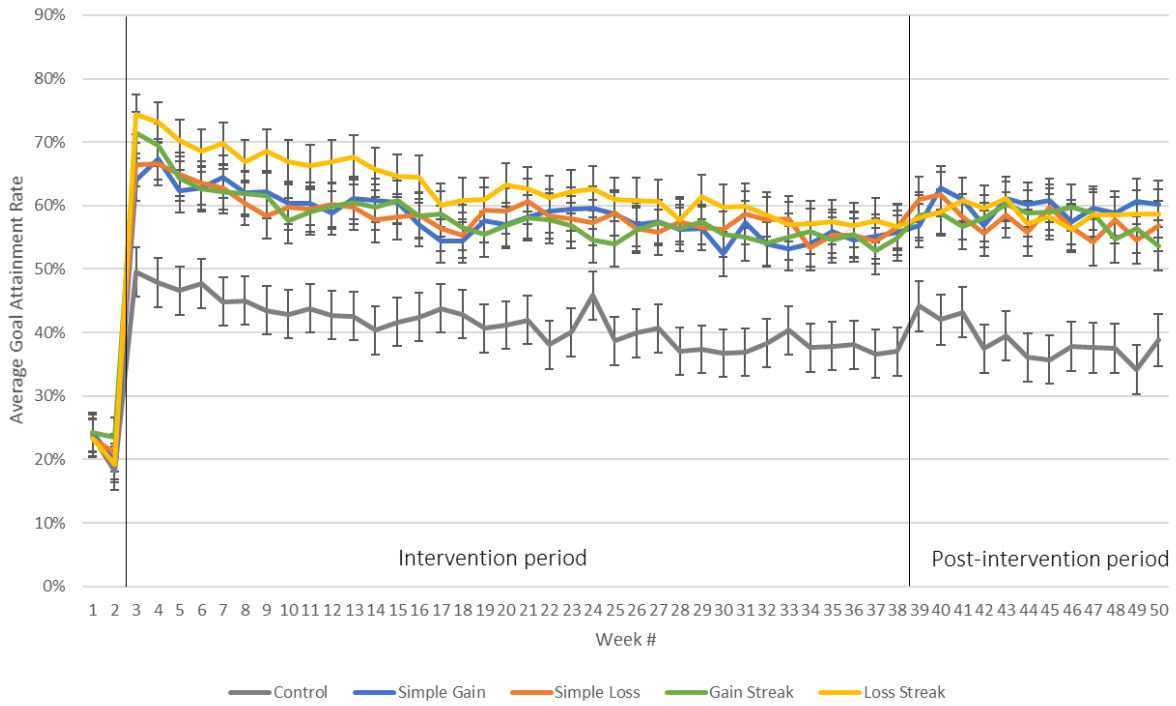
Table S1 provides summary statistics of the 523 participants who completed our study. We find no major differences in the participants' baseline characteristics across the five different conditions. The participants' average number of steps taken per day were about 10,000, which is equivalent to walking five miles. 60.4% of the participants were male, and the average age was 41.2 years old. Two-thirds of our participants received university or higher education. 62.5% were married. The average baseline BMI was 27.0, which falls under "overweight" category.

## Results

### *A. Model-Free Evidence*

Figure 2 plots the average goal attainment rates over time for each treatment condition, with 95% confidence intervals. The two vertical lines on week 3 and week 39 show the start and end of the intervention period, respectively.

**Figure 2. Average goal attainment rates over time, by treatment condition**




We begin by examining the unincentivized control condition.. Their goal attainment rate increased sharply from 18.2% in week 2 to 49.5% in week 3 ($p < 0.0001$). Over the course of 36 weeks, the goal attainment rate declined steadily at an average rate of 0.3 percentage point drop per week.[4] The average goal attainment rate during the 36-week intervention period was 41.4%, roughly double the baseline average of 21.0% ($p < 0.0001$). The observed increase in goal attainment rate can be attributed to having a personalized daily goal and being monitored by a wearable device. The average goal attainment rate in the 12-week post-intervention period was 38.7%, which remained significantly higher than the baseline average ($p < 0.0001$).

Next, we turn to the four treatment conditions. In addition to receiving personalized daily step goals, participants in a treatment condition received monetary incentives of up to SG$14 (US$10) per week to meet their daily goals. The monetary incentives were highly effective in increasing the goal attainment rate. During the 36-week intervention period, the average goal attainment rates of simple gain, simple loss, gain streak and loss streak conditions were 58.4%, 58.7%, 58.3% and 63.1%, respectively, all significantly higher than the control ($p < 0.0001$).

[4] Linear slope = -0.003 ($p < 0.0001$).

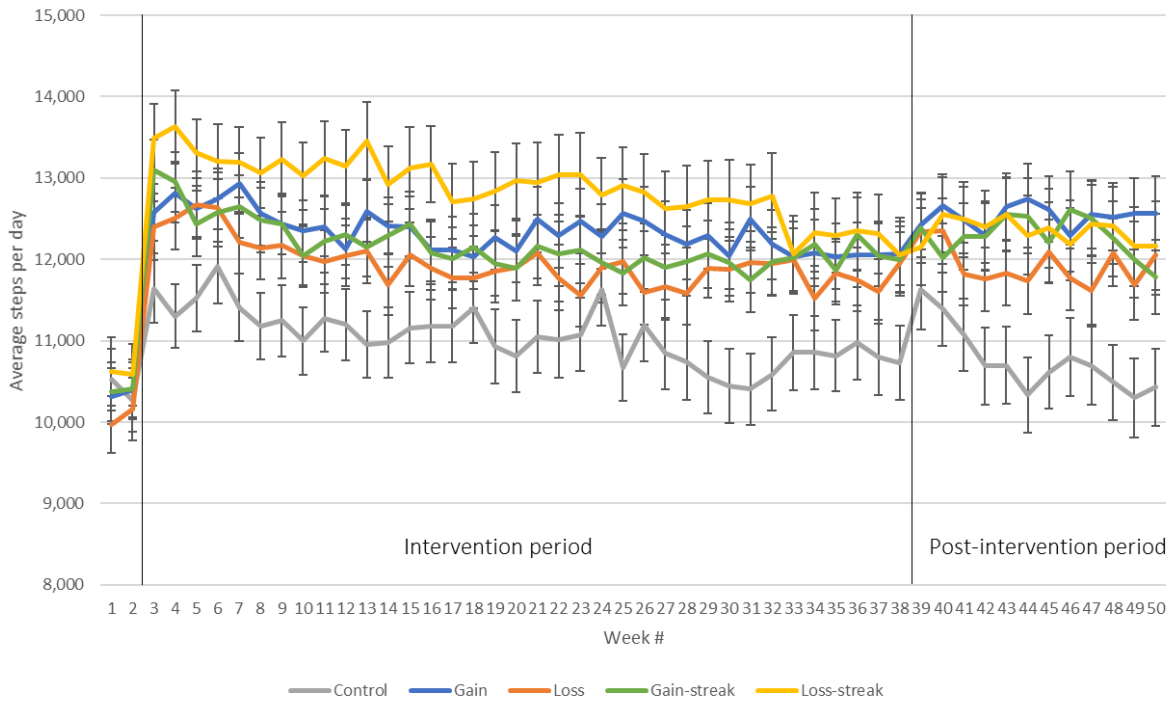
The loss streak was higher than the other three incentive conditions ($p < 0.0001$). None of the differences among the simple gain, simple loss, and gain streak conditions were statistically significant. All four treatment conditions showed steady decline in goal attainment rates during the 36-week intervention period.[5]

More importantly, we observe perfect stickiness of goal attainment rates in the post-intervention period. The average post-intervention goal attainment rates of the simple gain, simple loss, gain streak and loss streak conditions were 59.7%, 57.6%, 57.8% and 58.8%, respectively. The only significant decline was observed in the loss streak condition, from 63.1% in the intervention period to 58.8% in the post-intervention period ($p < 0.0001$). None of the other three incentive conditions had a statistically significant decline in the post-intervention period.

We find qualitatively similar results when we look at the average number of steps taken per day. Figure 4 plots the daily step count of participants in each treatment condition, with 95% confidence intervals. Although noticeably noisier, we continue to observe strong incentive effects during the intervention period, and no visible decline after the incentives are removed.

**Figure 4. Average daily steps over time, by treatment condition**

[5] Linear slopes for the simple gain, simple loss, gain streak, and loss streak conditions were -0.003 ($p < 0.0001$), -0.0002 ($p < 0.0001$), -0.0003 ($p < 0.0001$), and -0.0004 ($p < 0.0001$), respectively. These estimates imply that, on average, the goal attainment rate declined by 0.2-0.4 percentage points per week.




*B. Regression Model and Results*

We formally test our findings in a regression framework, as follows:

$$\begin{aligned} Goal_{it} = {} & (\alpha_1 + \alpha_2\, Gain_i + \alpha_3\, Loss_i + \alpha_4\, GS_i + \alpha_5\, LS_i)*INT_t \\ & + (\beta_1 + \beta_2\, Gain_i + \beta_3\, Loss_i + \beta_4\, GS_i + \beta_5\, LS_i)*POST_t \\ & + DOW_t + \sigma_i + \varepsilon_{it} \end{aligned}$$

*i* indexes the participant and *t* indexes time, measured in days. $Goal_{it}$ is an indicator that equals 1 if participant *i* has met her daily step goal on day *t*, and zero otherwise. $INT_t$ is an indicator for the intervention period, and $POST_t$ is an indicator for the post-intervention period. $Gain_i$ is an indicator that equals 1 if participant *i* is in the simple gain treatment condition; $Loss_i$ is an indicator for the simple loss condition; $GS_i$ is an indicator for the gain streak condition; and $LS_i$ is an indicator for the loss streak condition. We also include day-of-week fixed effects ($DOW_t$) and participant fixed effects ($\sigma_i$). In all regressions, the standard errors are clustered at the participant level. The sample size is 180,433.[6]

[6] There are supposed to be 523 participants x 350 days = 183,050 participant-day observations. 2,617 observations dropped out due to missing step-count data, most likely because of technical errors in the wearable device. As a robustness check, we have converted the 2,617 missing observations as zero steps and re-estimated the regressions with 183,050 observations. The results remain robust.

The coefficient $\alpha_1$ ($\beta_1$) can be interpreted as the percentage-point increase in control participants' goal attainment rate from the baseline period to the intervention (post-intervention) period. $\alpha_1$ is interpreted as the short-term effects of goal setting and monitoring, and $\beta_1$ is interpreted as long-term effects.

The coefficient $\alpha_2$ ($\alpha_3$, $\alpha_4$, and $\alpha_5$) is the percentage-point difference in goal attainment rates between the simple gain (simple loss, gain streak, and loss streak) and the control conditions during the intervention period. Coefficients $\alpha_2$, $\alpha_3$, $\alpha_4$, and $\alpha_5$ estimate the effects of giving monetary incentives on the goal attainment rate. The coefficient $\beta_2$ ($\beta_3$, $\beta_4$, and $\beta_5$) is the percentage-point difference in goal attainment rates between the simple gain (simple loss, gain streak, and loss streak) and control conditions during the post-intervention period. Coefficients $\beta_2$, $\beta_3$, $\beta_4$, and $\beta_5$ estimate the long-run effects of the intervention after the incentives are removed.

The regression results are shown in column 1 of Table 2. The estimated coefficient of 0.203 in the first row imply that, in the absence of any financial incentives, the goal attainment rates increased by 20.3 percentage points during the intervention period compared to the baseline period ($p < 0.0001$). This increase can be attributed to receiving a personalized daily step goal monitored through a wearable device. This effect remained significant in the post-intervention period. During the post-intervention period, the average goal attainment rate of the control participants remained 16.6 percentage points higher than their baseline ($p < 0.0001$). There was a 3.7 percentage point decline from the intervention to the post-intervention period (i.e., the difference between the two estimated coefficients, $p = 0.0211$).

**Table 2. Main results**

| | (1) | (2) |
|---|---|---|
| | Goal attainment | # Steps |
| Intervention | 0.2028*** | 634.99* |
| | (0.0313) | (303.88) |
| Intervention * Simple gain | 0.1616*** | 1,342.21*** |
| | (0.0439) | (399.24) |

| | | |
|---|---|---|
| Intervention * Simple loss | 0.1613*** | 1,251.16*** |
| | (0.0433) | (378.50) |
| Intervention * Gain streak | 0.1416** | 1,149.93** |
| | (0.0473) | (427.54) |
| Intervention * Loss streak | 0.2159*** | 1,645.40*** |
| | (0.0441) | (446.71) |
| Post-intervention | 0.1658*** | 294.29 |
| | (0.0327) | (359.36) |
| Post-intervention * Simple gain | 0.2013*** | 1,825.92*** |
| | (0.0470) | (475.66) |
| Post-intervention * Simple loss | 0.1834*** | 1,513.22*** |
| | (0.0469) | (460.70) |
| Post-intervention * Gain streak | 0.1641** | 1,446.12** |
| | (0.0504) | (519.57) |
| Post-intervention * Loss streak | 0.1969*** | 1,396.22** |
| | (0.0465) | (509.40) |
| Participant fixed effects | Y | Y |
| Day-of-week fixed effects | Y | Y |
| Dependent variable mean | 0.5471 | 12,020.11 |
| # Participants | 523 | 523 |
| # Observations | 180,433 | 180,433 |

Notes: OLS estimates are reported. Standard errors are clustered at the participant-level and reported in parentheses. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Next, we examine the estimated effects of financial incentives on the goal attainment rate. All four treatment conditions significantly increased the goal attainment rate during the intervention period compared to the control group. Although, the increase in goal achievement rate was not different between the conditions, the loss streak condition had directionally higher increase in goal achievement rate (21.6 percentage points higher likelihood of goal attainment than control) during the intervention period ($p < 0.0001$). While the gain streak condition had directionally the lowest increase in goal achievement rate (with a 14.2 percentage point increase in goal attainment rate compared to control).

We found that the incentivized participants continued to meet their step goals more frequently than the control participants in the post-intervention period. Participants in the simple gain condition had 20.1 percentage point higher goal attainment than unincentivized control participants in the post-intervention period ($p < 0.0001$); the simple loss treatment had an 18.3 percentage point effect ($p = 0.0001$); the gain streak treatment had a 16.4 percentage point effect ($p = 0.0012$); and the loss streak treatment had a 19.7 percentage point effect ($p < 0.0001$). Consistent with the model-free evidence, we find strong persistence of increased physical activity in the post-intervention period.

Column 2 estimates the same regression but with the number of steps taken per day as the dependent variable. During the intervention period, the incentivized participants took 1,149.9-1,645.4 more steps per day compared to unincentivized control participants. To put the magnitude into perspective, 1,000 steps is equivalent to walking about 0.5 mile.[7] During the post-intervention period, the incentivized participants took 1,396.2-1,825.9 more steps per day than the control participants. All estimates are statistically significant at the 5% level. The results confirm that participants are not simply reallocating steps across days to meet the goals, but are indeed taking more steps per day as a result of the intervention.

*C. Path dependence in goal achievement*

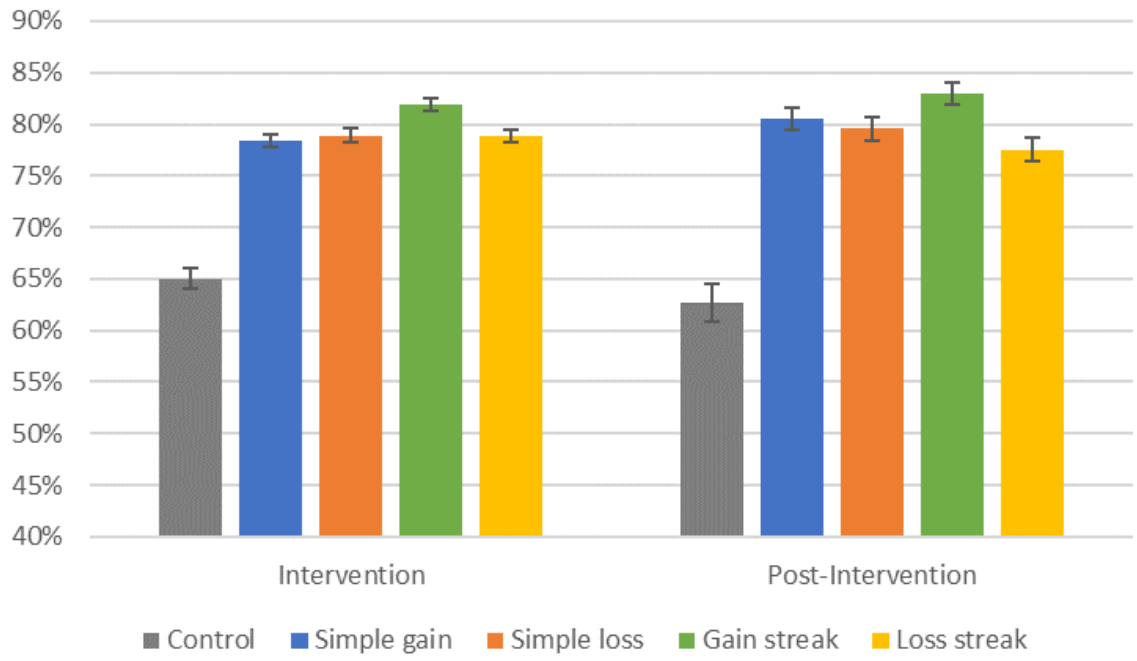
The overall strong persistence of exercising behavior in the absence of incentives suggests that the intervention may have successfully changed people's exercising habits. In this section, we conduct a novel test of habit formation. The hypothesis is that if habits were formed, the participants would not only exercise as frequently as they did before, but also continue to exercise in the same patterns that they have been exercising throughout the 36-week intervention period.

In our study, the two dynamic incentives—the gain streak and loss streak conditions—have incentivized participants to exercise in streaky and un-streaky patterns, respectively. Figure 8A shows, for each treatment condition and between the two experiment periods, the probability of goal attainment conditional on the goal being met on the previous day; 95% confidence

[7] Source: http://www.kylesconverter.com/length/steps-to-miles

intervals are also shown. Consistent with efforts to maximize earnings, the gain streak participants exhibited the highest 82.0% likelihood of goal attainment if the goal was met on the previous day. This was significantly higher than in simple gain (78.4%, $p < 0.0001$), simple loss (78.9%, $p < 0.0001$), and loss streak (78.9%, $p < 0.0001$) conditions. This streaky exercising tendency was maintained in the post-intervention period when the incentives were removed. In the post-intervention period, gain streak participants were 83.0% likely to meet the goal if the goal was met on the previous day, again significantly higher than the other three treatment conditions (vs. simple gain $p = 0.0019$; vs. simple loss $p < 0.0001$; vs. loss streak $p < 0.0001$).

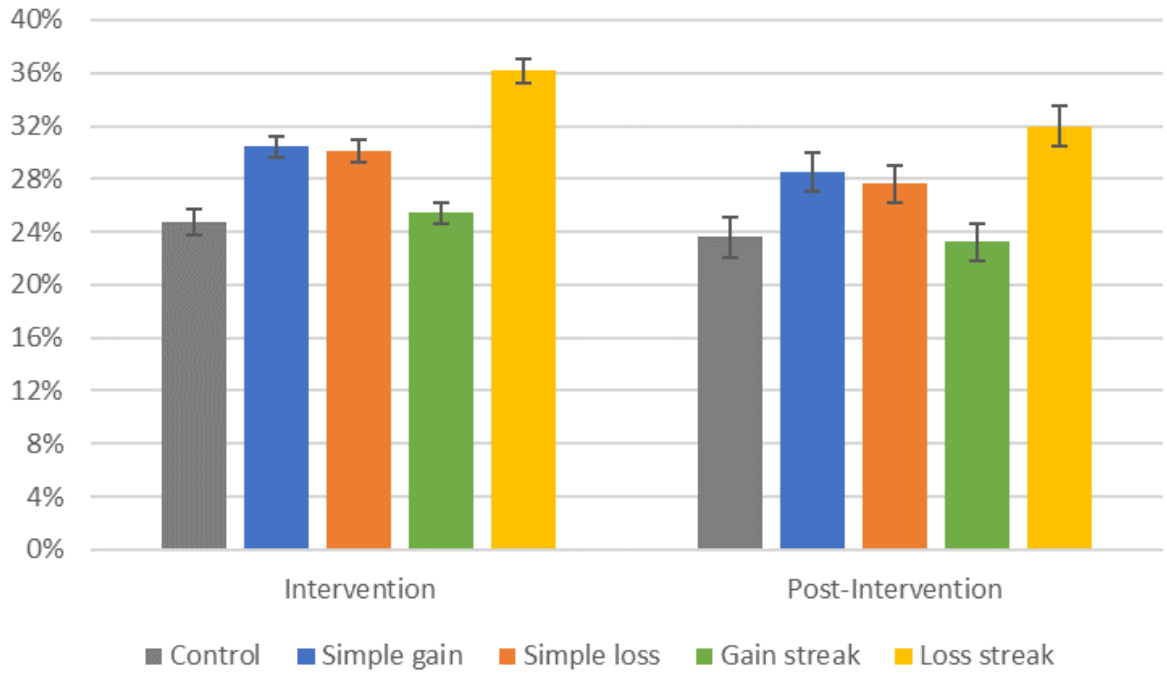
**Figure 5A. Probability of goal attainment if goal was met the day before, by treatment condition**




Similarly, Figure 8B shows the conditional probability of goal attainment if the participant has not achieved the goal on the previous day (with 95% confidence intervals). Consistent with efforts to maximize earnings, the loss streak participants exhibited the strongest tendency to meet the exercise goal after failing to meet the goal the day before. The conditional probability during the intervention period is 36.2%, significantly higher than the other three treatment conditions ($p < 0.0001$ for all pairwise comparisons). We find that this tendency persisted in the post-intervention period. The post-intervention conditional probability of goal attainment

for the loss streak condition is 32.0%, which remains highest among all treatment conditions (vs. simple gain $p = 0.0014$; vs. simple loss $p < 0.0001$; vs. gain streak $p < 0.0001$).

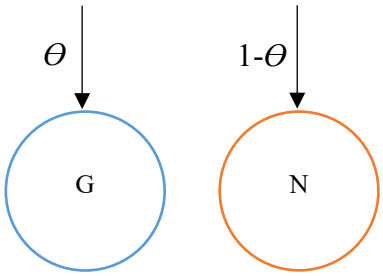
**Figure 5B. Probability of goal attainment if goal was not met the day before, by treatment condition**




We formally model the path-dependency of participants' exercising behavior with a Markov chain. Figure 9 illustrates the model. There are two states: G, a state in which the daily step goal is met; and N, a state in which the goal is not met. At the start of the week (when the incentive scheme resets), the participant has probability $\Theta$ of being in state G, and probability 1-$\Theta$ of being in state N. For the rest of the week, if the participant was in state G on the previous day, she has probability $p$ of remaining in G and probability 1-$p$ of switching to N. Similarly, if she was in state N on the previous day, she has probability 1-$q$ of remaining in N and probability $q$ of switching to G.

**Figure 6. Markov chain**

First day of the week:




Rest of the week:

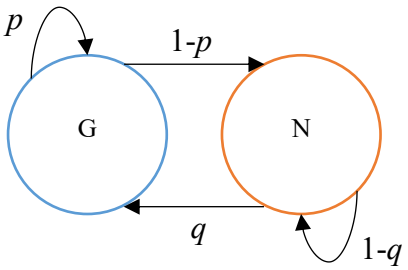



Table 4 shows the calibrated parameters for each treatment condition, and for each of the two experiment periods. During the intervention period, gain streak participants had the highest $p = 0.824$ ($p < 0.0001$), i.e., they showed the highest tendency to exercise in momentum. This is consistent with the incentive design, in which we incrementally increased the reward for exercising on consecutive days. Interestingly, the gain streak participants also had the lowest $\Theta = 0.574$ ($p = 0.0160$) and $q = 0.248$ ($p < 0.0001$) among the four incentive conditions. The lowest $\Theta$ implies that gain streak participants were least likely to meet the goal on the first day of the week.[8] The lowest $q$ implies that gain streak participants, once their exercising momentum has been broken, are least likely to start exercising again.

Consistent with exercising habits being formed, the gain streak participants retained much of these unique exercising characteristics after the monetary incentives were removed. In the post-intervention period, the gain streak participants continued to exhibit the highest $p = 0.833$ ($p = 0.0013$), i.e., the tendency to exercise in momentum. They also still had the lowest $q = 0.233$ ($p = 0.0001$)—the tendency to start exercising again after failing to meet the goal before—among the four incentive conditions. $\Theta = 0.566$ was also lowest among the incentive conditions, but the difference with the second-lowest, simple loss condition was not statistically significant ($p = 0.1349$)

**Table 4. Markov chain**

| | Intervention period | Post-intervention period |
|---|---|---|

[8] This can in part be explained by the small SG$0.50 reward (compared to e.g. SG$2 reward for simple gain participants).

| | $\Theta$ | $p$ | $q$ | $\Theta$ | $p$ | $q$ |
|---|---|---|---|---|---|---|
| Control | 0.409 | 0.658 | 0.239 | 0.398 | 0.628 | 0.228 |
| Simple gain | 0.601 | 0.781 | 0.293 | 0.595 | 0.806 | 0.286 |
| Simple loss | 0.604 | 0.787 | 0.286 | 0.579 | 0.794 | 0.276 |
| Gain streak | 0.574 | 0.824 | 0.248 | 0.566 | 0.833 | 0.233 |
| Loss streak | 0.636 | 0.788 | 0.350 | 0.581 | 0.770 | 0.319 |

Notes: The table reports calibrated Markov chain parameters.

We found similar retention of exercising patterns among loss streak participants. During the intervention period, the loss streak participants exhibited the highest $q = 0.350$ ($p < 0.0001$ for all pairwise comparisons). That is, they were most likely to start exercising after failing to meet the goal on the previous day. In the post-intervention period, the loss streak participants continued to exhibit the highest $q = 0.319$ ($p = 0.0044$), consistent with the formation of exercising habits.[9]

We conduct likelihood-ratio tests to compare the overall exercising patterns across the four incentive conditions, both during and after the intervention period. Table 5A reports the log likelihood values of fitting different parameters (shown in rows) on the different data (shown in columns) for the intervention period.[10] Table 5B shows analogous results for the post-intervention period. In each column, the bolded cells correspond to the best fit model for the data. The final row shows likelihood-ratio statistics comparing the best fit model against the second best fit model.[11] A large likelihood-ratio statistic implies that the exercising patterns of the treatment condition is unique, i.e., significantly different from the exercising patterns observed in other treatment conditions.

[9] During the intervention period, the loss streak participants' $p = 0.788$ was not lowest among the incentivized conditions. This helps explain why loss streak was most effective overall during the intervention period, as shown in column (1) of Table 2. In the post-intervention period, however, the loss streak participants' $p = 0.770$ was lowest among the incentivized groups ($p = 0.0077$).

[10] For instance, in the first row of column (1), "-16,498.64" is the log likelihood of fitting the parameters ($\Theta, p, q$) = (0.601, 0.781, 0.293), which are the parameters for the simple gain condition in the intervention period (see Table 4), onto the gain streak's intervention-period data. In the second row of column (1), "-16,501.46" is the log likelihood of fitting the simple loss parameters ($\Theta, p, q$) = (0.604, 0.787, 0.286) onto the gain streak's data.

[11] The likelihood-ratio statistic is calculated by 2*|L1-L2|, where L1 and L2 are the log likelihood values for the best-fit model and the comparison model, respectively. The statistic is $\chi^2$ distributed with 3 degrees of freedom. The critical values are 7.82 for $p = 0.05$, 11.35 for $p = 0.01$, and 16.27 for $p = 0.001$.

In Table 5A, we find that the overall exercising patterns of simple gain and simple loss participants during the intervention period are statistically indistinguishable from each other. By contrast, the overall exercising patterns of gain streak (column 3) and loss streak (column 4) are unique.

These characteristics were largely preserved in the post-intervention period (Table 5B), suggesting that exercising habits were formed.

**Table 5A. Likelihood-ratio tests (intervention period)**

| | (1) Simple gain data | (2) Simple loss data | (3) Gain streak data | (4) Loss streak data |
|---|---|---|---|---|
| Simple gain parameters | **-16,498.64** | -15,033.40 | -14,018.33 | -15,769.72 |
| Simple loss parameters | -16,501.46 | **-15,030.87** | -13,988.17 | -15,784.31 |
| Gain streak parameters | -16,645.43 | -15,133.29 | **-13,891.09** | -16,005.24 |
| Loss streak parameters | -16,583.07 | -15,123.79 | -14,197.68 | **-15,695.63** |
| Likelihood-ratio statistic (best vs. 2nd best fit) | 5.64 | 5.06 | 194.16*** | 148.18*** |

Notes: The table reports log likelihood values of fitting various parameters (shown in rows) on the different data (shown in columns). The best-fit models are bolded. The likelihood-ratio statistic is calculated by 2*|L1-L2|, where L1 and L2 are the log likelihood values for the best-fit model and the comparison model, respectively. The statistic is $\chi^2$ distributed with 3 degrees of freedom. The critical values are 7.82 for $p = 0.05$, 11.35 for $p = 0.01$, and 16.27 for $p = 0.001$. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

**Table 5B. Likelihood-ratio tests (post-intervention period)**

| | (1) Simple gain data | (2) Simple loss data | (3) Gain streak data | (4) Loss streak data |
|---|---|---|---|---|
| Simple gain parameters | **-5,092.18** | -4,837.19 | -4,287.67 | -5,157.90 |
| Simple loss parameters | -5,098.30 | **-4,835.31** | -4,292.28 | -5,154.68 |
| Gain streak parameters | -5,135.55 | -4,876.30 | **-4,261.91** | -5,251.95 |
| Loss streak parameters | -5,118.66 | -4,853.26 | -4,357.70 | **-5,132.21** |

| | | | | |
|---|---|---|---|---|
| Likelihood-ratio statistic (best vs. 2nd best fit) | 12.24** | 3.76 | 51.52*** | 44.94*** |

Notes: The table reports log likelihood values of fitting various parameters (shown in rows) on the different data (shown in columns). The best-fit models are bolded. The likelihood-ratio statistic is calculated by 2*|L1-L2|, where L1 and L2 are the log likelihood values for the best-fit model and the comparison model, respectively. The statistic is $\chi^2$ distributed with 3 degrees of freedom. The critical values are 7.82 for $p = 0.05$, 11.35 for $p = 0.01$, and 16.27 for $p = 0.001$. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

In sum, we find compelling evidence that the intervention has changed people's exercising habits. We offered dynamic monetary incentives (gain streak and loss streak conditions) that induced subjects to exercise in different path-dependent patterns. When the intervention ended, we found that the subjects not only continued to exercise as frequently as they did before, but also continued to exercise in the unique path-dependent patterns that we had incentivized them with the dynamic incentive schemes.

*D. Heterogeneity in goal achievement*

Path dependence in goal achievement gives rise to distinct behavioral trajectories across experimental conditions. While mean levels of goal attainment are broadly comparable across treatment groups, the distribution of goal achievement every week reveals systematic heterogeneity. To facilitate comparison, Figure 7A-B illustrates the proportion of participants attaining their goal for varying numbers of days per week in each treatment condition. Results for the control condition are reported separately in the Appendix: Figure Sx to maintain visual clarity.

Participants assigned to the Gain and Loss conditions exhibited broadly similar patterns of goal achievement per week. In contrast, those in the Gain Streak condition were significantly more likely to achieve goals across all days of the week: On average, 32.8% achieved the goal on all 7 days in a week, exceeding proportions observed in the Gain (26.6%, $p < 0.001$), Loss (25.3%, $p < 0.001$), Loss Streak (29.3%, $p = 0.001$) and Control (10.7%, $p<0.001$) conditions.

Notably, the Loss Streak condition appeared to mitigate non achievement of goal per week. On average, only 7.9% of participants in this group failed to meet their goal on any day of a week,

a significantly lower fraction than in the Gain (12.7%, $p < 0.001$), Loss (13.4%, $p < 0.001$), Gain Streak (13.8%, $p < 0.001$) and Control (20.5%, $p<0.001$) conditions.

Strikingly, these behavioral divergences persisted beyond the intervention period. During the post-treatment phase, 31.7% of participants in the Gain Streak condition continued to meet their goal on all 7 days per week—significantly higher than Loss (24.7%, $p < 0.001$), Loss Streak (24.5%, $p < 0.001$) and Control (8.4%, $p < 0.001$) groups. Although the proportion of subjects who achieved the goal for all 7 days in the gain streak condition was directionally higher, it did not significantly differ from the Gain condition (29.1%, $p = 0.17$). Conversely, the Loss Streak condition continued to mitigate total disengagement, with only 14.8% of participants on average failing to meet their goal on any day of the week during post-intervention. This proportion remained significantly lower than in the Loss (18.4%, $p < 0.001$), Gain Streak (24.9%, $p < 0.001$), and Control (27.3%, $p < 0.017$) conditions. However, the difference in the proportion of subjects who achieved the target on zero days between the Loss Streak and Gain conditions was not statistically significant (16.9%, $p = 0.158$).

**Figure 7A. Proportion of subjects achieving the goal by number of days per week during treatment period**

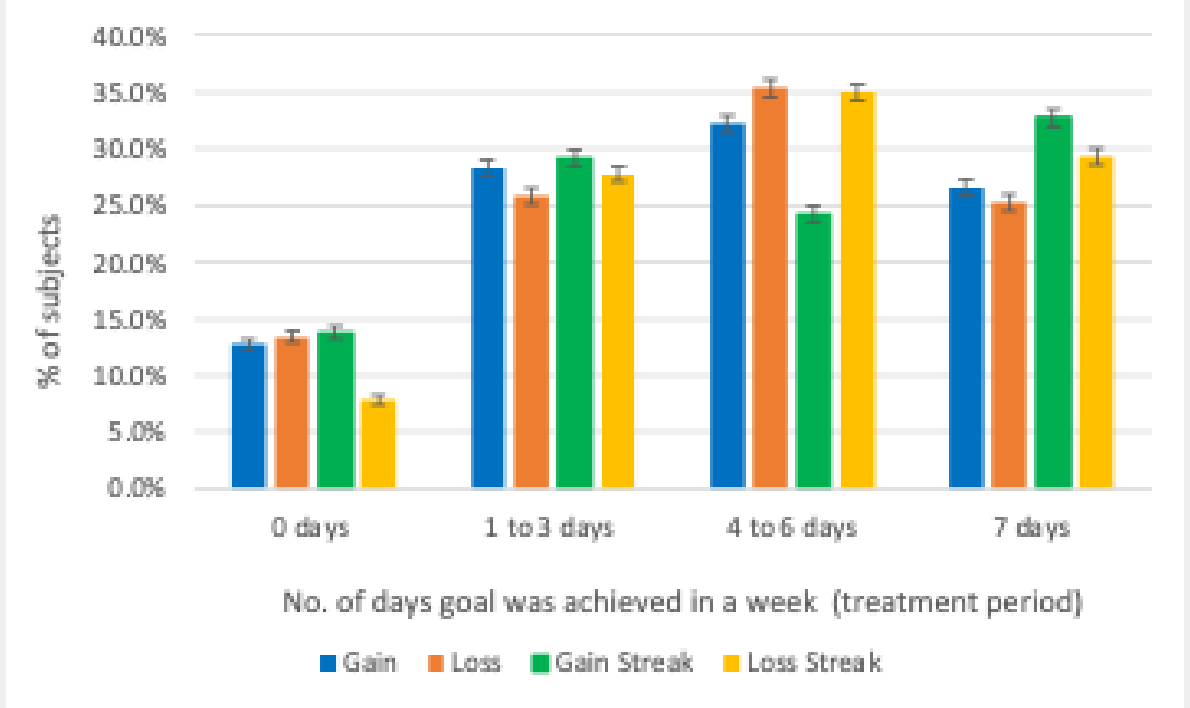
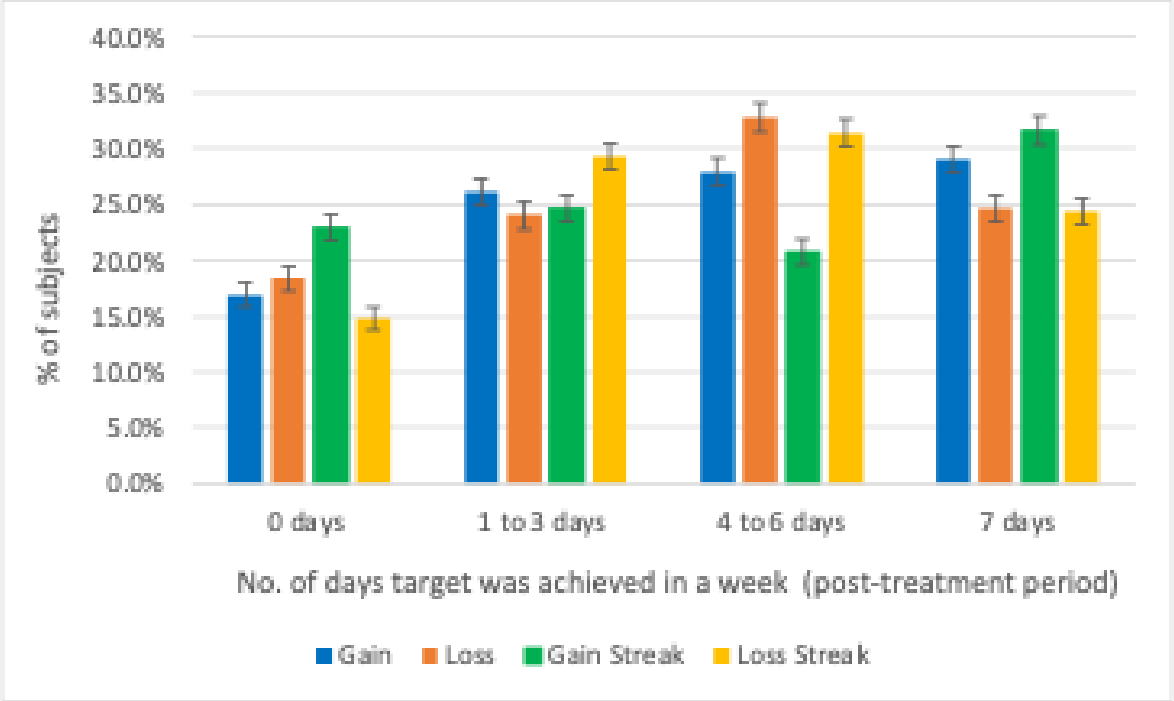


**Figure 7B. Proportion of subjects achieving the goal by number of days per week during post-treatment period**




*Effect of demographics on goal achievement*

We also compared the effect of our treatments on goal achievement separately for each gender. Among participants in the control condition, females increased their target achievement by 23.8% from baseline, which is higher than the 18.2% increase observed among males ($p<0.001$). All incentive treatments led to an increase in step count and goal achievement for males (compared to the control condition). However, among females, only the loss streak treatment resulted in a significant improvement in goal achievement (compared to the control) ($p<0.001$). Overall, these findings suggest that incentive responses differ by gender, with males responding more consistently across treatments and females showing a stronger response in control condition and to loss streak. The regression table is in the Appendix: Table S2.

When we focus on the effect of age, younger participants (participants less than 41 years) in the control group experienced an 11.9 percentage point increase in goal attainment rate, while the treatment group younger saw a 34.9 percentage point increase. In contrast, older

participants (participants 41 years or older) in the control and treatment groups experienced 27.8 and 39.5 percentage points increases in goal achievement rates, respectively. The incremental effect of incentives on goal-achievement rate was 23 percentage points for younger and 12 percentage points for older participants. Figures S1 in the Appendix illustrates the effect of treatment on different age group and gender.

## Discussion

Many attempts have been made in recent years—by academics and policymakers alike—to promote healthy habits with the use of monetary incentives. However, the incentive-based interventions have not been as effective as researchers had initially hoped, hindered by people's tendency to revert to their pre-intervention habits once the incentives are removed. In this paper, we have investigated whether more robust habit formation can be achieved through appropriate modifications to the intervention design. We gave daily goals that were personalized to the individual, coupled with real-time monitoring through a step-tracking device; we more than doubled the duration of the intervention period compared to previous studies in order to increase the repetition of "feedback and reward"; and introduced dynamic (gain streak and loss streak) incentive conditions to encourage behavioral momentum.

Interestingly, our randomized controlled trial (RCT) reveals that simply providing daily step goals—without any accompanying financial incentives, as implemented in our control condition—significantly increases step counts both during the intervention and in the post-intervention period. This effect appears to be driven by a high baseline willingness among participants to increase their physical activity, averaging **xx** on a 7-point scale. Prior research in psychology and marketing similarly suggests that even simple goals can effectively motivate behavioral change (CITE). However, our study contributes novel longitudinal evidence, demonstrating both the initiation and persistence of behavioral change using objective step count data.

We find almost perfect persistence—or "stickiness"—of elevated physical activity levels at the aggregate level in our treatment condition. This is highly surprising and novel to the literature. We conducted a novel test of habit formation by examining the participants' exercising patterns. The hypothesis is that, if exercising habits were formed, participants would not only continue

to exercise as much as they did before, but they would also continue to exercise in the same patterns as they did during the intervention. Importantly even after the incentives are removed, participants in the gain and loss streak continued to exercise in such unique path dependent pattern i.e., gain streak participants were more likely to meet the target if they met the target. This unique path dependent goal achievement also lead to distinct heterogeneity in number of days goal was achieved. We find that the gain streak treatment led to a significantly higher proportion of subjects achieving the target on all seven days of the week, compared to other conditions. In contrast, the loss streak significantly decreased the number of subjects who failed to meet the target on any day of the week, relative to other treatments. More importantly, this behavioral also sustained in the post-treatment.

Our findings suggest that gain and loss streak treatments can be strategically deployed based on the behavioral pattern desired. Gain streaks are particularly effective for inculcating daily behaviors—such as handwashing—that require consistent, uninterrupted repetition. In contrast, loss streaks are better suited for behaviors that involve periodic engagement, such as yoga, dance, or sports practice, where the goal is to maintain a minimum frequency (e.g., at least once per week). By aligning the incentive structure with the required behavioral cadence, these treatments can enhance habit formation. Importantly, our framework has broader applicability across domains such as energy conservation (Somasundaram et al., 2020), smoking cessation (Volpp et al., 2008), digital addiction (Allcott et al., 2022), and weight loss (Volpp et al., 2009), offering a flexible toolkit for sustained behavioral change.

We also find that gain streak condition was significantly more cost effective than the rest of the treatments because of the way the incentives were structured (see Appendix: Table S3 for a detailed analysis). First, we found that participants do not respond linearly to incentives. For example, consider the first day of the week, where simple gain and gain streak participants were offered SG$2 and SG$0.50 to exercise, respectively. We found that simple gain participants exercised 60.1% of the time, and gain streak participants exercised 57.4% of the time ($p = 0.0160$). That is, a SG$0.50 incentive induced almost as much exercise as a SG$2 incentive. On the second day of the week, the gain incentive remained fixed at SG$2 while the gain streak incentive was either SG$0.50 (if the participant did not meet the goal yesterday) or SG$1 (if she met the goal yesterday). Despite being offered significantly lower incentives, the goal attainment rate of the gain streak condition was 58.4%, almost as high as 59.0% from the

gain condition (p = 0.5480). The gain streak condition benefited from the cost-effectiveness of smaller daily incentives.

Another potential explanation is participants' forward-looking behavior. Compared to a SG$2 reward in the simple gain condition, a gain streak participant would respond more to an immediate reward of SG$2, because achieving the goal today increases tomorrow's reward. We find consistent evidence of forward-looking behavior. For example, the gain streak participants are significantly more responsive to a SG$2 reward than comparable participants in the simple gain condition on the 1st-6th days of the week (i.e., when today's goal attainment will increase tomorrow's reward), but the difference is not statistically significant on the last day of the week (i.e., today's goal attainment has no effect on tomorrow's reward)

Several important questions remain. In this study, we made several modifications all at once to maximize the chances of habit formation. As a result, we were unable to test how important each modification was. For example, would simply setting daily goals without a real-time monitor also be equally effective? What is the minimum duration of an intervention period to change people's habits, and how can it be determined *ex ante*? These are tough but important questions that will help policymakers design effective behavioral interventions. We leave a thorough exploration of these questions to future research.

## Methods

Recruitment occurred from October 2017 to January 2018 through a newspaper advertisement. We targeted adults between 25-60 years old who walked less than 10,000 steps a day (based on self-report), have body-mass index (BMI) greater than 22, and are either Singapore citizens or permanent residents.[12] They also needed to have a smartphone or a tablet compatible with Garmin fitness device. People who were taking part in another fitness or health study were not eligible to participate in the study. People who suffered from any major illnesses, such as hypertension, heart attack, heart failure, and stroke, were also not eligible for the study. We also excluded women who were breast-feeding, pregnant, or planning to be pregnant during

[12] The standard BMI cutoff for "overweight" category is 25. However, we used a lower threshold to reflect recent research that suggests that a lower threshold may be more appropriate for Asians. See https://www.moh.gov.sg/docs/librariesprovider4/guidelines/obesity-cpg_main.pdf

the study period. Finally, we only allowed one member per household to participate in our study, to minimize the chance of participants knowing about other treatments during the study.

Our study required four in-person visits. The first visit was for enrollment. During this visit, participants filled out the consent form and a demographic survey, and we gave each participant a wearable step-tracking device (Garmin Vivo 3). The second visit occurred at the end of week 2. During this visit, participants were randomly assigned to one of five conditions: control, gain, loss, gain streak, or loss streak condition. They were informed about their daily step goal, which was equal to 2,500 steps higher than their baseline average (rounded to the nearest 100 steps). To remind participants of their step goal, we gave them a physical card with their goal written on it, and input the goal into their Garmin app. We also informed them about the incentive structure of the condition that they were assigned to and took their body measurements. The third visit occurred at the end of the intervention period. During this visit, the participants completed another survey and had body measurements taken. We then paid them their incentive earnings. The fourth visit happened at the end of the post-intervention period. During this visit, participants completed an exit survey and had body measurements taken. In addition to incentive earnings, participants were paid for their in-person visits: SG$20 for the 2nd visit, SG$40 for the third visit, and SG$40 for the fourth visit.

## References

Barberis, Nicholas C. "Thirty years of prospect theory in economics: A review and assessment." *Journal of Economic Perspectives* 27.1 (2013): 173-96.

Carrera, M., Royer, H., Stehr, M., & Sydnor, J. (2018). Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members. *Journal of Health Economics*, *58*, 202-214.

Carrera, M., Royer, H., Stehr, M., & Sydnor, J. (2020). The structure of health incentives: Evidence from a field experiment. *Management Science*, *66*(5), 1890-1908.

Cawley, J., & Price, J. A. (2013). A case study of a workplace wellness program that offers financial incentives for weight loss. *Journal of Health Economics*, *32*(5), 794-803.

Charness, G & Gneezy, U. (2009). Incentives to exercise. *Econometrica,* 77(3): 909-931.

Chokshi, N. P., Adusumalli, S., Small, D. S., Morris, A., Feingold, J., Ha, Y. P., ... & Patel, M. S. (2018). "Loss-framed financial incentives and personalized goal-setting to increase physical activity among ischemic heart disease patients using wearable devices: the ACTIVE REWARD Randomized Trial." *Journal of the American Heart Association*, *7*(12), e009173.

Finkelstein, E. A., Linnan, L. A., Tate, D. F., & Birken, B. E. (2007). A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees. *Journal of Occupational and Environmental Medicine*, *49*(9), 981-989.

Finkelstein, E. A., Haaland, B. A., Bilger, M., Sahasranaman, A., Sloan, R. A., Nang, E. E. K., & Evenson, K. R. (2016). Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. *The Lancet Diabetes & Endocrinology*, *4*(12), 983-995.

John, L. K., Loewenstein, G., Troxel, A. B., Norton, L., Fassbender, J. E., & Volpp, K. G. (2011). Financial incentives for extended weight loss: a randomized, controlled trial. *Journal of General Internal Medicine*, *26*(6), 621-626.

Just, D. R., & Price, J. (2013). Using incentives to encourage healthy eating in children. *Journal of Human Resources*, *48*(4), 855-872.

Kahneman, D. & Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk". *Econometrica*. **47** (4): 263–291.

Loewenstein, G., Price, J., & Volpp, K. (2016). Habit formation in children: Evidence from incentives for healthy eating. *Journal of Health Economics*, *45*, 47-54.

Miller, J., Shenhav, A. & Ludvig, A. (2019). Habits without values. *Psychological Review*, 126(2): 292.

Mochon, D., Schwartz, J., Maroba, J., Patel, D., & Ariely, D. (2017). Gain without pain: The extended effects of a behavioral health intervention. *Management Science*, *63*(1), 58-72.

Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Heuer, J., ... & Wesby, L. (2016). Framing financial incentives to increase physical activity among overweight and obese adults: a randomized, controlled trial. *Annals of Internal Medicine*, *164*(6), 385-394.

Rohde, K. & Verbeke, W. (2017). We like to see you in the gym – A field experiment on financial incentives for short and long term gym attendance. *Journal of Economic Behavior & Organization*, 134, 388-407.

Royer, H., Stehr, M., & 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.

Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive–based approaches for weight loss: a randomized trial. *JAMA*, *300*(22), 2631-2637.

Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., ... & Corbett, E. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine*, *360:* 699-709.

# Appendix

**Figure S1. Maximum possible earnings per week, by treatment condition**

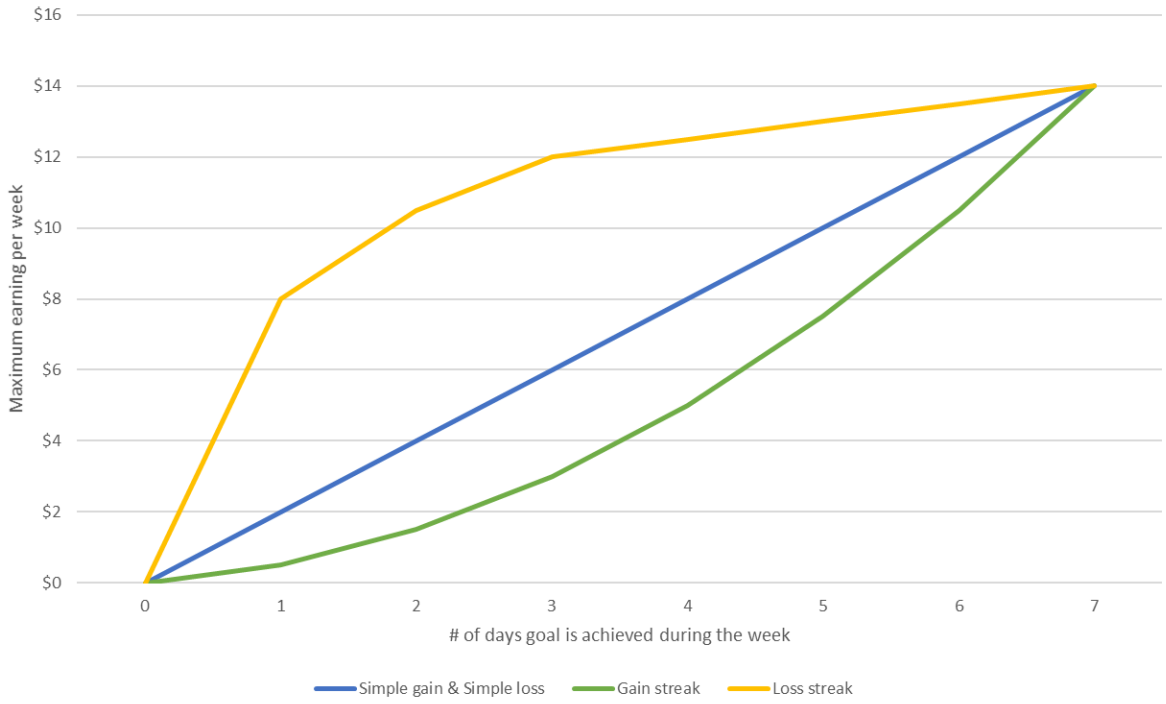



**Table S1. Summary statistics**

| | Control | Simple Gain | Simple Loss | Gain Streak | Loss Streak |
|---|---|---|---|---|---|
| Baseline # steps per day | 10,396.190 | 10,354.580 | 10,066.260 | 10,388.630 | 10,589.670 |
| *Demographics:* | | | | | |
| Male | 0.620 | 0.588 | 0.610 | 0.606 | 0.602 |
| Age | 40.989 | 42.333 | 40.971 | 39.692 | 41.852 |
| Low income (< SG$30,000) | 0.185 | 0.184 | 0.171 | 0.163 | 0.130 |
| High income (> SG$100,000) | 0.130 | 0.114 | 0.076 | 0.106 | 0.120 |
| High education (University or higher) | 0.674 | 0.614 | 0.667 | 0.663 | 0.676 |
| Married | 0.554 | 0.640 | 0.590 | 0.625 | 0.704 |

| | | | | | |
|---|---|---|---|---|---|
| # of children | 0.935 | 1.044 | 1.114 | 1.115 | 1.250 |
| # of members in household | 2.391 | 2.535 | 2.514 | 2.731 | 2.611 |
| *Baseline health indicators:* | | | | | |
| Baseline BMI | 27.070 | 27.146 | 27.078 | 27.148 | 26.761 |
| Baseline neck (cm) | 36.433 | 36.276 | 36.478 | 37.157 | 36.408 |
| Baseline waist (cm) | 86.618 | 87.812 | 87.492 | 86.921 | 87.007 |
| Baseline hip (cm) | 101.047 | 101.731 | 102.496 | 101.539 | 101.450 |
| Baseline systolic blood pressure | 126.054 | 125.268 | 127.324 | 124.034 | 124.838 |
| Baseline diastolic blood pressure | 77.288 | 74.624 | 77.848 | 74.688 | 75.264 |
| # Participants | 92 | 114 | 105 | 104 | 108 |

**Table S2: Effect of Treatment on Goal Achievement based on Gender**

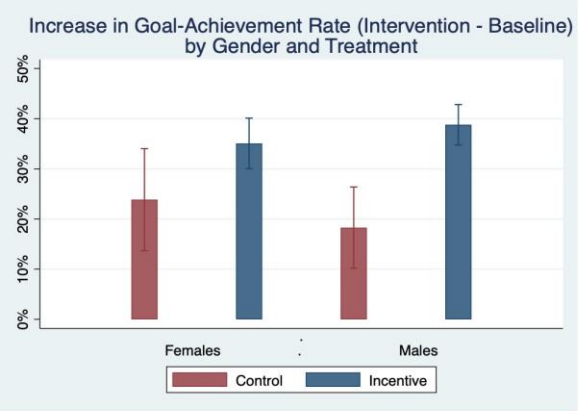
| | Dependent variable: | |
|---|---|---|
| | Goal achieved or not (1/0) | |
| | Male | Female |
| Intervention | 0.182*** | 0.238*** |
| | (0.040) | (0.049) |
| Post | 0.128*** | 0.227*** |
| | (0.040) | (0.054) |
| treatmentnameGain:Intervention | 0.198*** | 0.106 |
| | (0.055) | (0.071) |
| treatmentnameGain Streak:Intervention | 0.195*** | 0.058 |
| | (0.064) | (0.069) |
| treatmentnameLoss:Intervention | 0.183*** | 0.126* |
| | (0.054) | (0.072) |
| treatmentnameLoss Streak:Intervention | 0.249*** | 0.163** |
| | (0.056) | (0.072) |
| treatmentnameGain:Post | 0.273*** | 0.091 |
| | (0.058) | (0.077) |
| treatmentnameGain Streak:Post | 0.220*** | 0.074 |
| | (0.067) | (0.076) |
| treatmentnameLoss:Post | 0.217*** | 0.129* |
| | (0.058) | (0.078) |
| treatmentnameLoss Streak:Post | 0.241*** | 0.126* |
| | (0.058) | (0.076) |
| Observations | 109,148 | 71,285 |
| R2 | 0.034 | 0.027 |
| Adjusted R2 | 0.031 | 0.024 |
| F Statistic | 377.877*** (df = 10; 108822) | 194.647*** (df = 10; 71068) |

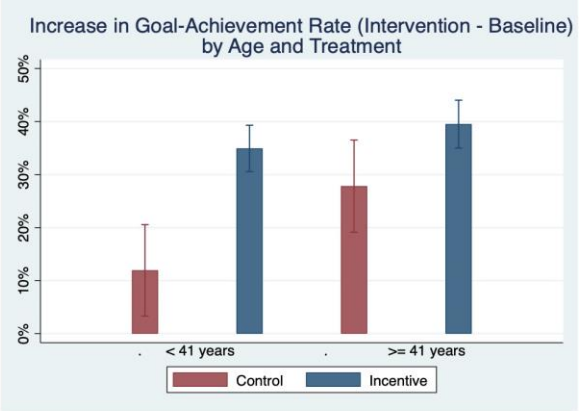
Note: *p<0.1; **p<0.05; ***p<0.01

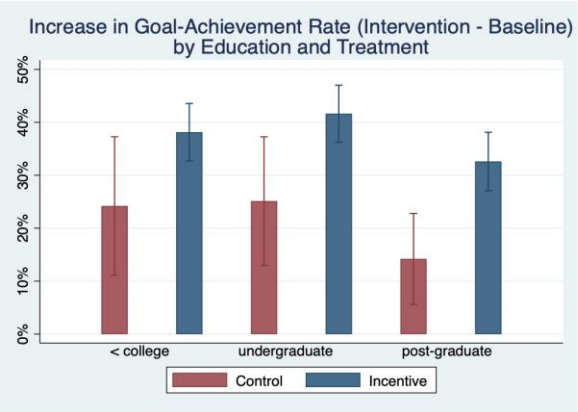
*Cost effectiveness*

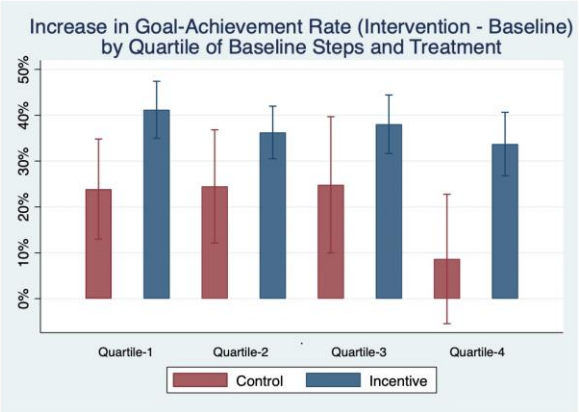
Table S3, column 1 shows the average exercise frequency per week observed during the baseline period.[1] Column 2 shows the long-run steady-state exercise frequency predicted by the post-intervention Markov models in Table 4. Column 3 shows the difference between columns 2 and 1. Column 4 shows the total experiment cost per participant, which includes SG$85 for the wearable step-tracking device, SG$100 in participation fee, and additional financial incentives. Column 5 shows the increase in weekly exercise frequency per SG$100 spent.

**Figure S2: Comparison of Increase in the Goal-Achievement Rate between Control and Incentivized Subjects by Gender (top left), Age (top right), Education (bottom left), and Baseline Step Count (bottom right).**




## Table S3. Cost effectiveness

| | (1) Baseline exercise frequency per week | (2) Long-run steady-state exercise frequency per week | (3) Increase in weekly exercise frequency | (4) Total cost per participant (SG$) | (5) Increase in weekly exercise frequency per SG$100 |
|---|---|---|---|---|---|
| Control | 1.467 | 2.690 | 1.223 | 185.00 | 0.661 |
| Simple gain | 1.535 | 4.169 | 2.634 | 479.05 | 0.550 |
| Simple loss | 1.557 | 4.021 | 2.353 | 480.64 | 0.490 |
| Gain streak | 1.668 | 4.037 | 2.480 | 415.56 | 0.597 |
| Loss streak | 1.488 | 4.067 | 2.579 | 612.12 | 0.421 |

Notes: Columns (1) – (3) report average exercise frequency per week. We define a participant to have exercised if she has met the daily step goal. Column (1) shows the average exercise frequency during the baseline period. Column (2) shows the long-run steady-state exercise frequency predicted by the post-intervention Markov model (Table 4). Column (3) shows the difference between Columns (2) and (1). Column (4) shows the total cost per participant, which includes SG$85 for Garmin Vivofit 3 (wearable step-tracking device), SG$100 in participation fee, and additional financial incentives. Column (5) is obtained by dividing Column (3) by (4), then multiplying by 100.