Journal of Economic Behavior and Organization 205 (2023) 560–573

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Intermittent incentives to encourage exercising in the long run ☆,☆☆,★

Ayala Arad [a,*], Uri Gneezy [b], Eli Mograbi [a]

[a] Coller School of Management, Tel Aviv University, Sderot Henry Gildred, Tel Aviv-Yafo, Israel
[b] Rady School of Management, UC San Diego, 9500 Gilman drive, La Jolla, CA 92093, USA

ARTICLE INFO

Article history:
Received 2 October 2021
Revised 11 November 2022
Accepted 13 November 2022
Available online 9 December 2022

Keywords:
Exercise
Field experiment
Incentives
Irregular reinforcement

ABSTRACT

We report the results of incentivizing students to exercise. We compare a no-incentive control to a per-visit payment and two intermittent incentive schemes: monetary rewards at increasing intervals and monetary rewards with unpredictable timing. Based on the psychology literature, we predicted that irregular incentives would facilitate the maintenance of reinforced behavior over the long run. In line with this prediction, we found that although all incentive schemes worked well during the incentivized period, only the two intermittent schemes were more effective than the control after incentives were removed and over an extended period of time. These results suggest an innovative way to incentivize individuals, which sets out to improve the long-term success of interventions that encourage the formation of habits.

1. Introduction

Can incentives be used to encourage the initiation and maintenance of good habits? We study the effect of incentives on exercising in the short and long run, with the goal of increasing the long-term impact. We develop novel intermittent incentive schemes that, unlike fixed incentive schemes, remain effective after incentives are removed. Increasing exercising has far-reaching consequences, with physical inactivity ranked as the fourth most important health risk factor worldwide. The estimated global annual cost of physical inactivity is $54 billion in direct healthcare expenditure and another $14 billion in productivity loss (WHO, 2018).

A possible reason for physical inactivity is that individuals simply do not want to exercise. Alternatively, people may want to exercise but fail to do so. This failure may have multiple causes, such as lack of self-control, which is often described as present-biased preferences (DellaVigna and Malmendier 2006; Laibson, 1997; O'Donoghue and Rabin, 1999). In these cases, incentives that increase utility in the present may be able to help overcome the failure to exercise (Charness and Gneezy, 2009). Moreover, incentivizing behavior in the short run might create exercising habits, as in Becker and Murphy's (1988)

☆ We acknowledge support from the US-Israel Binational Science Foundation (BSF) and the NSF. We thank Sean Joffe, Neta Klein, Aviv Halak, May Kenet, Guy Tikochinsky, and Itamar Amir for their research assistance, and the Tel Aviv University sports center, especially Pavel Kolosovsky, Shiri Eitan, and Yael Keilin, for their cooperation. We also thank Adi Shany, Sally Sadoff, Dotan Persitz, Yair Antler, and participants at the Coller Conference on Behavioral Economics, Advances with Field Experiments Conference, and the ESA European Conference for useful comments.
☆☆ This study is registered in the AEA RCT Registry, and the unique identifying number is AEARCTR-0005873.
★ The online appendix is available at https://www.tau.ac.il/~aradayal/appendix_intermittent_incentives.pdf.
* Corresponding autho.
E-mail address: aradayal@tauex.tau.ac.il (A. Arad).

https://doi.org/10.1016/j.jebo.2022.11.015
0167-2681/

model in which the stock of exercising behavior (how much one exercised in the past) may enhance the utility from exercising in the future. Thus, if individuals are incentivized to exercise in the short run, they may find exercising in the long run easier or more enjoyable and may continue to exercise even after the incentives are removed.

Findings from the research on incentives to exercise show that incentives are successful in motivating individuals to start exercising, but are less effective in overcoming the challenge of sustaining the effect in the long run, after the incentives are removed (Acland and Levy, 2015; Charness and Gneezy, 2009; Roemmich et al., 2012; Royer et al., 2015; see two recent reviews by Mitchell et al., 2019 [1], and Strohacker et al., 2014). The challenge of producing an enduring behavior change is not unique to exercising and appears in other health-related contexts, such as dietary choice, as noted in Belot et al. (2016) and Just and Price (2013). Although the focus of the current paper is on how to induce exercising in the long run, we hope our findings will prove useful for additional behavior change problems.

Our research involved a field experiment in which students who did not exercise regularly were incentivized to exercise in a university gym. All participants received a six-month free membership to the university gym. We examine their exercise frequency throughout the membership period as well as 12 and 18 months after the beginning of the experiment. We assigned participants into four groups.[2] Participants in the first group were not offered any incentives (*Control*), whereas those in the second group were offered 20 NIS per visit (*Per-visit*) for the first two months of the membership (henceforth, the *incentivized period*). In past studies, per-visit incentives encouraged exercising while in place, but after their removal, behavior returned to baseline levels (Hardman et al., 2011; Royer et al., 2015). We used this incentive scheme to ensure behavior in our setting is consistent with that observed in other studies, and thus we did not expect it to have long-run effects. The third group received frequent rewards at the beginning of the incentivized period, which were aimed at overcoming the high initial cost of starting to exercise. The incentives were then gradually became less frequent but larger, such that, on average, every visit was still rewarded with 20 NIS (*Increasing*). The fourth group was offered incentives on a random basis, such that not every visit was rewarded. Randomness was introduced by way of the number of visits required to receive the reward (1, 2, 3, or 4) (*Unexpected*). This treatment was designed to yield an expected reward of 20 NIS per visit in the incentivized period, as in the other incentivized treatments. Further details on the incentive schemes and comparison with past studies can be found in Section 2.3.

Although regular incentive schemes may crowd out the intrinsic motivation to exercise (Gneezy et al., 2011), intermittent incentives, such as those provided to the *Increasing* and *Unexpected* groups, may succeed—due to their irregular nature—in establishing the habit of exercising without creating an enduring association between it and the monetary incentive. This habit formation is achieved by not getting the participants accustomed to receiving a reward after every visit, with the goal of encouraging the maintenance of the exercise habit in the long run, after removal of incentives.

The idea of intermittent rewards builds on psychology literature. Ferster and Skinner (1957) introduced the concept of a schedule of reinforcement and demonstrated experimentally, using lab animals, that changing the reinforcement schedule affects the initiation and maintenance of the desired behavior. In light of the empirical evidence suggesting intermittent schemes lead to better maintenance of the reinforced behavior after the reinforcement is removed, Amsel (1958) developed what is known as frustration theory, whereby the use of intermittent incentives leads individuals to hold two contradictory expectations simultaneously: that of being rewarded and that of not being rewarded. Individuals continue to pursue the reinforcement because the magnitude of the anticipation of a reward is large enough. With repeated experience, they learn to cope with the frustration of sometimes not being rewarded for a proper response. Thus, the reinforced response also becomes conditioned in the situation of no reward, increasing the likelihood that the behavior that was reinforced will continue after the incentives are removed (Amsel, 1992; Pittenger, 2002; Papini, 2003; Domjan, 2010).

The phenomenon whereby intermittent rather than regularly scheduled incentives maintain a behavior for a more extended period is known as the partial-reinforcement extinction effect (PREE; Hochman and Erev, 2013; Humphreys, 1939; Pittenger, 2002). In a number of laboratory experiments, participants who received continuous reinforcement for some initial period (by being rewarded for every choice they made) were less likely to keep playing after “failures” in which they did not receive a reward, compared with participants who had received intermittent reinforcements in the initial period (Golz, 1992; Hogarth and Villeval, 2014; see Pittenger, 2002, for a review). Milkman et al. (2021) are a unique example of using a field experiment that includes an incentive scheme with an intermittent part. They found that the most successful intervention in the post-intervention month is one that provides small incentives for gym visits as well as for returning to the gym after a missed (planned) workout. Although this creative scheme may be cost effective in some contexts, we suspect higher incentives are needed to induce habit formation among people who do not exercise.

[1] This review reports on only four studies that found post-intervention effects. One study might have suffered from a selection problem—the treatment group with long-term effects was willing to sign commitment contracts, which requires a larger motivation to exercise (Royer, Stehr, and Sydnor 2015). In the second, participants were incentivized for three weeks, and the post-intervention effect lasted for only two weeks (Condliffe, Isgin, and Fitzgerald (2017). Using a lottery incentive scheme, Petry et al. (2013) found an effect three months after the intervention ended among 20 adults per group over the age of 55. However, they only tested one particular week, three months after the intervention, and reminded the participants before the beginning of that week, possibly causing a commitment effect for the incentivized group. Finally, Rohde and Verbeke (2017) examined gym attendance among participants who had paid for a membership and thus were likely to have a strong intrinsic motivation to exercise. Following a six-month intervention, they did not find a significant effect on the total number of visits during the first three months following the incentivized period. However, they did find less of a decrease in the number of visits (relative to the previous quarter) in the treatment group. This effect disappeared in the subsequent quarter.

[2] Originally in our research proposal, we planned to include five groups, yet due to insufficient funds, we decided to focus on four.

Our study focuses on encouraging exercising among people who do not exercise regularly and probably require a big boost to start doing so. Our two intermittent schemes, which have not been studied before, proved to be successful in the long run in our field setting. By contrast, the regular incentive scheme was more effective than the control only during the incentivized period. These findings demonstrate the importance of examining the effectiveness of incentives over the relevant horizon, because not all incentives that work while in place are also successful in the long run.

2. Experimental design

The participants were 213 students attending Tel Aviv University who were selected after a pre-screening process. They were offered the chance to participate in an experiment aimed at increasing their physical activity. Participants received a six-month membership (January–June 2018) to the Tel Aviv University sports center. Financial incentives for exercising were provided only for the first two months (January–February 2018), allowing us to observe whether the effects of the incentives persisted during the subsequent four months. We also examined whether participants continued to exercise at two six-month intervals, after the free membership had ended (12 and 18 months after the start of the experiment).

2.1. Recruitment

Past research (e.g., Charness and Gneezy, 2009) usually found the most significant effects among participants who did not previously exercise, which guided us in the recruitment. To this end, all university students were invited to answer a short three-minute lifestyle questionnaire that was distributed online (see the online appendix). To incentivize the students to answer the questionnaire, they were told that 5% of them would be randomly selected to receive 50 NIS (at the time of the experiment, 3.6 NIS=$1). The questionnaire included several questions about lifestyle, including filter questions used to determine who was eligible to participate in the experiment.

The students filling out the questionnaire did not know the details of the experiment and its incentives, and thus did not have an incentive to lie. We designed the selection criteria to identify individuals who (1) did not exercise at all or exercised only once a week but not in a gym and not swimming, (2) had a commute time of up to 120 minutes from their residence to the university (participants who live closer to the gym can more easily exercise there), and (3) wanted to exercise more. Initially, we intended our criteria to be stricter, with criterion (1) being not exercising at all, and criterion (2) being a commute time of up to 30 minutes. However, we did not have enough students who answered our stricter criteria. Of the 1,115 participants who completed the pre-study questionnaire, we invited 420 to participate in the experiment based on the lenient criteria.

We then sent the selected students invitations to participate. We told them that based on their answers to the lifestyle questionnaire, they were eligible to participate in a research project aimed at encouraging them to exercise in the university gym and that they would be given a free six-month membership (a value of 1,500 NIS). Of the 420 students invited to participate, 229 showed up to the introductory session. Of the 229 that showed up, 10 were removed and 6 dropped out, leaving us a total of 213 participants (see explanations below).

2.2. Introductory session

Participants attended a 90-minute introductory session. The 29 sessions were held between December 17 and December 28, 2017, in the Interactive Decision-Making Lab at Tel Aviv University. Each session was attended by participants from the same treatment in order to minimize the chance of exposure to participants from other treatments.

Before being informed about the incentives, participants were asked to sign a consent form to participate in the experiment; 223 agreed to sign the consent form (six participants dropped out at this stage). When signing the form, participants were also asked to give consent for us to access their future grade transcripts, their Israeli SAT scores, and their matriculation exam scores. After signing the consent form, participants received the experiment's instructions for the treatment to which they were assigned. The instructions were read aloud to them and fully explained. The instructions included the incentive they would be offered to exercise, how to use the mobile app that was designed for the experiment, how rewards would be distributed during the experiment, and so on. Participants also filled out a medical questionnaire, as required by Israeli law, to ensure exercising was safe for them.

Participant ID numbers were reported to the university sports center for registration purposes. In line with our inclusion criteria, from the 223 participants, 10 participants who were members of the sports center during the previous year (2017) were paid 100 NIS and removed from the study, leaving a total of 213 students.[3]

The four treatment groups each had between 50 and 55 participants. We tested whether the treatment groups were balanced according to age, gender, commute time, year of study, and prior exercising (i.e., whether they were previously involved in some non-gym physical activity once a week or not at all). Table 1, which includes all 213 participants, shows the means of these variables. We found the groups were balanced according to these variables except for prior exercising

[3] We removed one participant from *Control*, one participant from *Per-visit*, two participants from *Increasing*, and six participants from *Unexpected*.

Table 1
Balance in the assignment to treatments.

Treatment Variable	Control	Per-visit	Increasing	Unexpected
No. of Participants	50	54	55	54
Age	23.77	24.32	23.68	24.36
	(2.32)	(2.77)	(2.91)	(4.17)
Commute time (minutes)	36.46	38.13	38.4	35.81
	(28.97)	(26.27)	(27.58)	(26.21)
Gender (Male)	0.36	0.33	0.35	0.33
	(0.48)	(0.48)	(0.48)	(0.48)
Prior exercise	0.06	0.09	0.20*	0.11
	(0.24)	(0.29)	(0.40)	(0.31)
Year 1	0.5	0.43	0.41	0.51
	(0.51)	(0.49)	(0.49)	(0.50)
Year 2	0.26	0.35	0.31	0.26
	(0.44)	(0.48)	(0.47)	(0.44)
Year 3	0.16	0.11	0.16	0.14
	(0.37)	(0.31)	(0.37)	(0.35)
Year 4	0.08	0.11	0.11	0.07
	(0.27)	(0.31)	(0.31)	(0.26)

Notes: Group means of relevant characteristics in the four treatment groups. Standard deviations appear in parentheses. Asterisks indicate the significance of the differences between the means relative to the other groups jointly.
* $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$.

(in particular, the *Increasing* group consisted of a larger proportion of participants who did some sport prior to the study than did the other three groups jointly). We control for all these variables in our regression analysis.[4]

Participants answered the following psychological questionnaires in the introductory session: (1) the Propensity to Plan scale (Lynch et al., 2010), which measures an individual's tendency to plan, a trait that is typically necessary to exercise on a regular basis; (2) CFC—Consideration of Future Consequences (Strathman et al., 1994), which estimates the extent to which participants take into account future consequences, and in this case, those of the decision to exercise; (3) DOSPERT—Domain-Specific Risk-Taking (Blais and Weber 2006; Weber et al., 2002), which measures risk tolerance; and (4) a happiness questionnaire, based on a subset of questions from the Oxford Happiness Questionnaire (Hills and Argyle, 2002) that measures differences in levels of happiness between people who exercise and people who do not. Participants also answered additional questions about general lifestyle. While participants were answering the questionnaire, they were taken out one by one to a separate room where a nursing student measured their medical indicators (pulse, weight, and body fat percentage) using non-invasive devices—an Omron Body Composition Monitor BF511 for weight and fat percentage, and an Omron M3 device for pulse.

Table S1 in the appendix shows no major differences in the psychological and physiological measures between the different treatments. An ANOVA test indicates that the variables' means in the four treatment groups are not significantly different except for the risk tolerance measure (in particular, the *Control* group's mean score is significantly greater than that of the other groups jointly, $p < 0.10$). This suggests the aforementioned assignment to treatments induced fairly balanced groups. The measurement of the psychological and physiological indicators during the introductory session were repeated at the end of the six-month sports center membership.

2.3. Treatments

To facilitate comparison, we set the mean reward per visit to the gym at roughly 20 NIS for the three incentivized treatments.

We implemented the following four treatments:

1 *Control*: Participants received no monetary incentives to visit the gym.
2 *Per-visit*: Participants received 20 NIS for each visit to the gym.
3 *Increasing*: The number of visits required to receive the next reward and the size of the next reward increased after the payment of each reward. A reward of 20 NIS was paid after the first visit, a reward of 40 NIS after two additional visits (i.e., after the third visit), a reward of 60 NIS after three additional visits (i.e., after the sixth visit), and so on. Thus, the average reward per visit was 20 NIS, as in the *Per-visit* treatment, if the participant received a payment on his last visit in the incentivized period. (Because participants were not necessarily paid for their last few visits in the incentivized period, the average payment per visit may have been less than 20.) Participants were aware of this setup,

[4] In the regressions reported in the paper, we did not include Age as a control variable, because its variance in the sample is very small. Its inclusion does not affect the results.

and the information on the timing of the next reward and its size was available to them from the mobile app. Under this incentive structure, participants can more easily receive a reward early on, when the need to compensate for the high initial costs of exercising is bigger. Over time, the reward is paid less frequently. Yet, the salience of the high potential rewards coupled with overoptimism (Carrera et al., 2018; DellaVigna and Malmendier, 2006) may make this scheme attractive, while allowing participants to gradually get used to visiting the gym without a reward. In other words, the incentives are frequent enough to establish the exercise habit, but not to the extent that it becomes strongly associated with receiving a reward. This balance is accomplished by means of the reward's decreasing frequency, thus mitigating crowding-out effects.

To the best of our knowledge, our study is the first to use an incentive scheme in which participants must exercise more each time in order to receive the next reward. A recent study by Bachireddy et al. (2019) included an incentive scheme in which the size of the reward increased every two days regardless of the participant's behavior. In addition, their reward was paid out on a daily basis, based on an average of $0.0002 per step measured by a pedometer. Thus, the amount of exercising required to obtain a reward did not increase, as it does in our study. Participants in this group did not exercise more than those in other groups during and after the incentivized period, perhaps because their setup for the increasing-reward scheme is not likely to mitigate the crowding-out effect with exercising every day being rewarded.

1 *Unexpected*: Participants received 50 NIS after X visits to the gym, where X is a number between 1 and 4 determined randomly by lottery. After a participant in this group received a reward, a new random number was drawn, and a new count of visits began immediately. Participants knew only that a number between 1 and 4 had been drawn, but did not know which, and therefore could not know exactly when they would receive the reward. That is, the expected number of visits required to obtain the 50 NIS reward was 2.5, and hence, the expected reward per visit was 20 NIS. Thus, the average payment per visit in the *Unexpected* treatment was designed to be 20 NIS, as in the *Per-visit* and *Increasing* treatments. The likelihood of obtaining the reward increased each visit, until it was received: it was 0.25 in the first visit; then, conditional on not receiving the reward in the first visit, it was 0.33 in the second visit (the reward was obtained in one of the three remaining visits with equal probability); conditional on not receiving it in the first two visits, it was 0.5 in the third; finally, if it had not yet been received, it was received for sure in the fourth visit.

Previous studies such as Andrade et al. (2014); Patel et al. (2018); Wing et al. (1996) incentivized individuals by rewarding them with a lottery ticket each time they exercised. However, the lottery left participants with uncertainty as to whether and when they would receive a reward. By contrast, because X was bounded in our scheme, participants knew they would always get a reward by the fourth visit, thereby eliminating part of the uncertainty that existed in previous studies. Carrera et al. (2019) also used intermittent incentives to encourage participants to exercise. However, their incentives motivated the participants only while in place. Their design differs from ours in several aspects, which may drive the different results. Most notably, participants in their study knew in advance whether they would receive incentives for the coming week, which made the incentives expected.

Our goal was to test the effect of helping participants begin forming the habit of exercising, by providing them with certainty that they would eventually receive the monetary reward and by avoiding the frustration over not receiving a reward for an overly long period. On their first visit, these participants received 20 NIS to ensure them that they would receive rewards and to help establish the researchers' credibility among the students. After that visit, the first X was determined by lottery.

At the end of the experiment, we discovered that due to a technical error, the lotteries had drawn a random X of between 1 and 3 rather than 1 and 4. Participants are unlikely to have identified this error when they decided whether to visit the gym during the incentivized period. Despite the (unintentional) offer of a higher expected value, the results reported below show this incentive scheme had the weakest effect among the three treatments during the incentivized period.

In all treatments, no more than one entry to the gym per day was counted. Thus, the maximal number of visits eligible for a reward was 59, the number of possible days during the incentivized period (January–February 2018).

2.4. Sports center membership

As part of the membership, each participant received one personal training session, during which they received an exercise program suited to their needs. The goal of providing an individualized program was to enhance the exercise's effectiveness (Jeffery, 2012).

The experiment had an associated mobile app (a customized website for smartphones). A direct link to the app was uploaded to all of the participants' smartphones during the introductory session. The app contained information regarding the participant's exercise program and membership, including the number of visits, the number of visits in the previous week, the rewards that had already been received, the size and timing of their next reward, and so on. Each participant also received a text message on the last day of every week telling them how many times they had exercised in the previous week and whether they had reached their weekly goal (which was set at the recommended three visits). For example, a text might read as follows: "You exercised at the gym twice this week. Good job! You can do even better next week. The recommendation is to exercise at least 3 times a week."

Visits to the gym were recorded by chip swipes at the entrance to the sports center and to the gym. An employee of the sports center at the entrance verified that the chip used belonged to the individual that swiped it (a picture of the member pops up when the chip is swiped). This system applies to all members of the sports center.

When participants were eligible for a reward for a particular visit, they received a text message approximately 15 minutes after they swiped the chip. We set this delay so participants would still be at the gym if they were actually exercising, in order to prevent participants from swiping the chip to enter, receiving the reward, and leaving immediately. An example of such a message follows: "You are entitled to a reward of 20 shekels for exercising at the gym today. You can pick up the money at the gym reception desk. In order to receive the payment, please enter the app or the link in order to confirm receiving the money."

After receiving this message, participants had until the end of the day to pick up their reward from the gym reception desk. Even if, despite our precautions, some participants left immediately without exercising, note that the main interest of our study is their behavior in the non-incentivized period, when participants had no reason to swipe the card and leave the sports center.

During the four months following the incentivized period (March–June 2018), participants could still access the gym and use the app, and they continued to receive weekly text messages, but they did not receive any rewards. Of the 213 participants who started the experiment, eight canceled their participation at some point during the six months: three from *Control*, one from *Per-visit*, two from *Increasing*, and two from *Unexpected*. Thus, 205 participants remained in the study until its completion.[5]

2.5. Post-membership period

During the first two weeks of June 2018, all participants were invited to a concluding lab session similar to the introductory one held in December and received 100 NIS for attending. During the session, participants answered the same questions as in the original questionnaire, along with some additional ones (see the online appendix). One hundred seventy-one participants attended the concluding session (roughly 85% of the remaining participants).

Participants were sent two short follow-up questionnaires in which they were asked whether they had continued to exercise and where. The first online questionnaire was sent in January 2019, 12 months after the beginning of the incentivized period, and the second was sent in June 2019, 18 months after the beginning of the incentivized period and prior to the university's exam period. The first questionnaire had a response rate of roughly 95% (194 responses), whereas the second had a response rate of 92% (189 responses). Participants received 50 NIS for answering each questionnaire.

3. Results

The analysis follows the plan outlined in our proposal of the Binational Science Foundation (BSF) grant supporting this research, called "Intermittent Incentives for Encouraging Physical Activity."[6] As planned, we discuss the results according to time period.

Our main findings are as follows: during the incentivized period (months 1–2), participants in the three incentivized treatments visited the gym more often than participants in *Control*. During the non-incentivized period (months 3–6), participants in the two intermittent-schemes treatments exercised more than the *Control* group, whereas participants in the *Per-visit* group did not exercise more than the *Control* group. We also analyzed the 12- and 18-month follow-up questionnaires and found that only participants in the *Unexpected* treatment exercised more than participants in *Control*.[7] Thus, over time, after incentives were removed, the *Per-visit* incentive scheme became inferior. The *Unexpected*, although not performing as well as the other schemes during the incentivized period, became superior in the long run.

For the dependent variable for the number of visits during the membership period, no more than one entrance to the gym per day was counted, as in the payment of rewards. We omitted outliers according to the standard cutoff of mean ± 3 standard deviations of the dependent variable.[8] We estimate the effect of treatment relative to the *Control* group using a Tobit model. We chose that model in view of the large number of zero observations in our sample (left-censored data), which represent individuals who did not visit the gym at all. We repeated the analysis using an OLS model as a robustness check and report it in the online appendix. We present five specifications: one without any control variables and four that include subsets of the following control variables: gender, dummies for the year of study, commute time, and prior exercising.[9] For each specification, we present the incentivized treatments' coefficients, the *Control* group's mean, and the

[5] Why they decided to drop out is unclear. While they were a part of the experiment, like other participants, we counted their entrances. After they dropped out, we recorded their status as participants who did not exercise or attend the gym. The qualitative results hold regardless of whether we include these participants. They were not included in the two follow-ups, because they did not answer the questionnaires.

[6] www.267c5708-a14e-40fe-aa99-aa8d5d15b476.filesusr.com/ugd/08b008_4f13793501f04c67b60f497e54283dc3.pdf

[7] Eventually, the planned eight-month follow-up questionnaire was not carried out, due to budget constraints.

[8] In our main analysis of the membership period, we examined visits to the gym in different time periods (months 3–6, 3–4, and 5–6), in each of which different outliers emerges. We omitted only participants who were outliers in all these non-incentivized periods. In Tables S11*a* and S11*b* in the appendix, we re-analyze gym visits in the membership period using different approaches for dealing with outliers.

[9] As previously mentioned, we were planning to screen participants based on a commute time of less than 30 minutes and no physical activity prior to the study. Because this screening would have severely limited sample size, we decided to recruit participants who reported exercising once a week (though

Table 2
Visits to the gym during the membership period.

	Panel A: Gym visits in months 1–2					Panel B: Gym visits in months 3–6				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Constant	-2.83	-2.57	-2.73	-4.14*	-4.18*	-3.26	-3.33	-3.66	-2.22	-2.28
	(1.90)	(2.25)	(2.27)	(2.36)	(2.37)	(2.13)	(2.66)	(2.67)	(2.67)	(2.68)
Per-visit	10.21***	10.21***	10.02***	9.52***	9.43***	4.35	4.44	4.12	4.13	3.93
	(2.59)	(2.60)	(2.55)	(2.57)	(2.55)	(2.82)	(2.83)	(2.77)	(2.74)	(2.71)
Increasing	12.6***	12.6***	12.02***	12.08***	11.69***	6.14**	6.17**	5.14*	6.27**	5.34*
	(2.51)	(2.51)	(2.50)	(2.55)	(2.55)	(2.73)	(2.73)	(2.75)	(2.73)	(2.74)
Unexpected	7.70***	7.68***	7.41***	7.53***	7.34***	5.7**	5.77**	5.42*	5.67**	5.35*
	(2.49)	(2.5)	(2.51)	(2.42)	(2.44)	(2.79)	(2.80)	(2.81)	(2.75)	(2.76)
Control Mean			3.12 [5]					3.47 [5.74]		
Joint equality test (p-value)	0.12	0.11	0.14	0.15	0.17	0.78	0.8	0.88	0.7	0.82
Observations			211					211		

Notes: Tobit estimates. The dependent variable is the number of visits to the gym in a particular period. Each panel represents a different time period. In each period, five specifications are presented, each consisting of a different subset of control variables: (1) None; (2) Gender and Commute time; (3) Gender, Commute time, and Prior exercise; (4) Gender, Commute time, and Year of study; and (5) Gender, Commute time, Year of study, and Prior exercise. Robust standard errors appear in parentheses. Standard deviations appear in brackets.
* $p < 0.10$,; ** $p < 0.05$,; *** $p < 0.01$.;

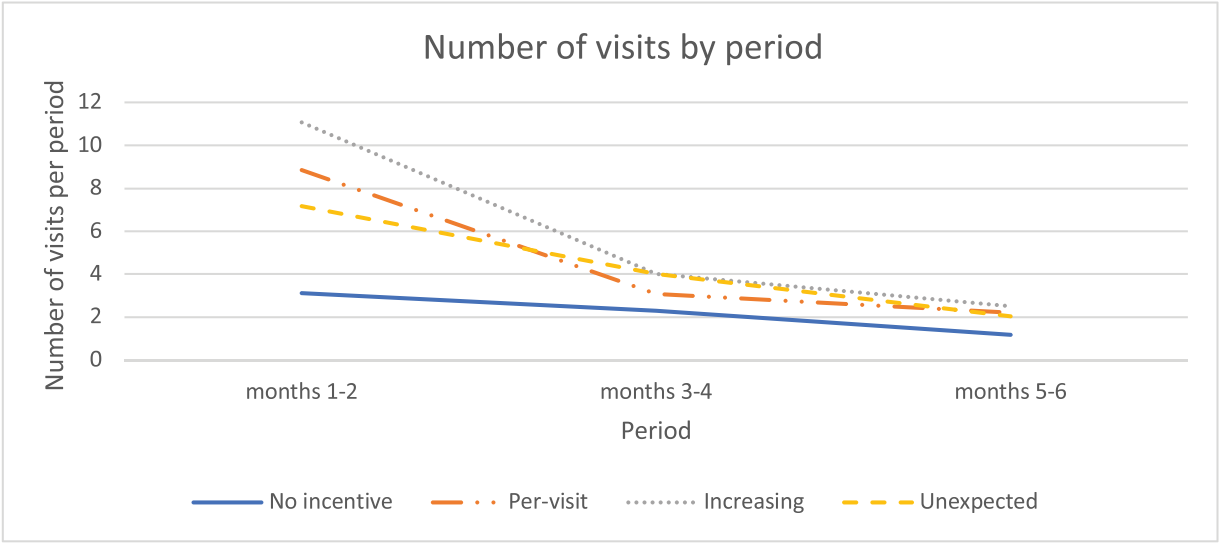

Fig. 1. Number of gym visits by period.

p-value obtained in a Wald test which examines the hypothesis that the coefficients of *Per-visit*, *Increasing* and *Unexpected* are all equal.

The focus of the paper is on determining the long-term effects on the groups that received intermittent incentives relative to the non-incentivized *Control* group. In what follows, we analyze the membership period, the follow-up, and the dynamic of exercising over time, ending with a number of robustness checks and an analysis of additional data collected in the experiment.

3.1. Membership period

Table 2 compares each of the incentivized treatments with the *Control* group in the incentivized and non-incentivized membership periods. Fig. 1 summarizes the dynamic of exercising during the membership period, and Supplementary Fig. S1 in the appendix depicts the distribution of the number of visits to the gym in these membership periods. The mean number of visits to the gym in the incentivized period (months 1–2) was 11.07 ($SD = 10.24$) in *Increasing*, 8.85 ($SD = 10.13$) in *Per-visit*, 7.17 ($SD = 8.96$) in *Unexpected*, and 3.12 ($SD = 5$) in *Control*. Table 2, Panel A, which relates to the incentivized

not in a gym) and who had a commute time of up to 120 minutes. Therefore, we included prior exercise (0 or 1) and commute time as control variables. Using additional control variables (e.g., age) do not seem to change our results.

Table 3
Visits to the gym during the membership period.

	Panel A: Gym visits in months 3–4					Panel B: Gym visits in months 5–6				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Constant	-1.95	-2.75*	-2.97*	-2.28	-2.32	-6.86***	-4.45**	-4.60**	-2.89	-2.81
	(1.36)	(1.66)	(1.66)	(1.66)	(1.66)	(1.94)	(2.26)	(2.28)	(2.30)	(2.33)
Per-visit	1.9	2.0	1.77	1.73	1.59	2.8	2.74	2.47	2.54	2.31
	(1.76)	(1.77)	(1.73)	(1.71)	(1.7)	(2.38)	(2.36)	(2.33)	(2.31)	(2.3)
Increasing	3.52**	3.55**	2.86*	3.55**	2.96*	3.78*	3.64	2.82	3.66	2.81
	(1.72)	(1.70)	(1.71)	(1.68)	(1.69)	(2.27)	(2.26)	(2.33)	(2.27)	(2.32)
Unexpected	3.45*	3.55**	3.31*	3.48**	3.27*	2.72	2.49	2.15	2.28	1.91
	(1.81)	(1.8)	(1.8)	(1.75)	(1.75)	(2.28)	(2.28)	(2.30)	(2.27)	(2.3)
Control Mean			2.29 [3.66]					1.18 [2.64]		
Joint equality test (p-value)	0.56	0.57	0.65	0.46	0.55	0.86	0.85	0.95	0.78	0.91
Observations			211					211		

Notes: Tobit estimates. The dependent variable is the number of visits to the gym in a particular period. Each panel represents a different time period. In each period, five specifications are presented, each consisting of a different subset of control variables: (1) None; (2) Gender and Commute time; (3) Gender, Commute time, and Prior exercise; (4) Gender, Commute time, and Year of study; and (5) Gender, Commute time, Year of study, and Prior exercise. Robust standard errors appear in parentheses. Standard deviations appear in brackets.

* $p < 0.10$,; ** $p < 0.05$,; *** $p < 0.01$.;

period, shows participants in the three incentivized treatments visited the gym significantly more than the *Control* group, with Tobit estimators (in the full specification) ranging from 7.34 to 11.69, depending on the incentivized treatment.

Our main interest is in behavior during the non-incentivized membership period (months 3–6). The mean number of visits in months 3–6 was 6.53 ($SD = 8.47$) in *Increasing*, 6.04 ($SD = 9.07$) in *Unexpected*, 5.30 ($SD = 8.70$) in *Per-visit*, and 3.47 ($SD = 5.74$) in *Control*. Panel B of Table 2 shows that during this period, the results for the Per-visit treatment were not significantly different from the *Control* group at the 10% level, whereas the number of visits to the gym in both the *Increasing* and *Unexpected* treatments were significantly larger than in the *Control* group (with similar estimators of the differences in the full specification: 5.34 and 5.35, respectively). The findings demonstrate the divergence between the short term, in which all the incentive schemes did better than *Control*, and the long term, in which the two intermittent groups visited the gym significantly more than *Control*, whereas the *Per-visit* group did not.[10]

Next, we separately examine the first and the second half of the non-incentivized membership period. The regression results are presented in Table 3, and the distribution of the number of visits is presented in Supplementary Fig. S2 in the appendix. In months 3–4, participants in the intermittent-incentives groups exercised significantly more than those in the *Control* group: the Tobit estimators are 2.96 ($SE = 1.69$) in *Increasing* and 3.27 ($SE = 1.75$) in *Unexpected*. By contrast, the *Per-visit* group did not exercise significantly more than the *Control* group. In months 5–6, the *Unexpected,* the *Increasing*, and the *Per-visit* did not show significant differences from the *Control* (except for the *Increasing* in a specification without control variables).

As can be seen from Fig. 1, the number of visits declined steeply after the incentivized period in all the incentivized groups, though to a smaller extent in *Unexpected*. The figure also suggests the success of the intermittent incentives in the non-incentivized membership period (months 3–6) is driven by both their effect in months 3–4 and their effect in months 5–6, with the former being larger.

We note that none of our incentivized treatments are worse than the no-incentive *Control*. That is, even if crowding out of intrinsic motivation takes place due to the use of monetary incentives, its negative effect is not larger than the positive effect of forming a new habit in the incentivized period. Furthermore, looking at the whole membership period (months 1–6), gym visits in the incentivized treatments are much more frequent than in the *Control* treatment: the mean number of visits was 17.6 ($SD = 16.59$) in *Increasing*, 14.15 ($SD = 17.35$) in *Per-visit*, 13.2 ($SD = 14.9$) in *Unexpected*, and 6.59 ($SD = 9.71$) in *Control*.

To better understand the dynamics, we measured how much an extra visit to the gym induced by our incentives in months 1–2 contributed to increased exercising in months 3–6. In this estimation, the treatment serves as an instrumental variable when estimating the effect of visits per month in the incentivized period (months 1–2) on visits per month in the non-incentivized period (months 3–6). We estimated the effect of an extra visit in the incentivized period in each of our treatments and found significant returns of 0.32 ($SE = 0.17$) extra visits per month in *Unexpected* and 0.17 ($SE = 0.08$) in *Increasing*. In the *Per-visit* treatment, the estimated return of 0.15 ($SE = 0.10$) is not significant at the 10% level (all the estimation results appear in Table S7 in the appendix). The returns per month during the two months after the intervention (months 3–4) are 0.43 ($SE = 0.23$) extra visits in *Unexpected* and 0.18 ($SE = 0.09$) in *Increasing* ($p < 0.10$ for both estimations), whereas the return of 0.12 ($SE = 0.13$) in the *Per-visit* treatment is not significant at the 10% level.

[10] Note the sample size provides us with 80% power in identifying a mean difference of about 4 gym visits in months 3-6 with 5% significance and of 3.5 visits with 10% significance. Identifying a significant ($p < 0.10$) mean difference of only 2 visits (with 80% power) would require a sample of about 650 participants. Our sample size also limits our ability to detect significant differences between the three treatments' coefficients.

Table 4
How many times do you exercise each week on average?

	Panel A: Weekly sessions in month 12					Panel B: Weekly sessions in month 18				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Constant	0.35	0.22	0.18	0.29	0.28	0.75**	0.63*	0.60*	0.66*	0.66*
	(0.33)	(0.36)	(0.36)	(0.4)	(0.4)	(0.31)	(0.33)	(0.33)	(0.38)	(0.38)
Per-visit	0.37	0.38	0.37	0.32	0.32	0.54	0.56	0.55	0.49	0.48
	(0.42)	(0.42)	(0.42)	(0.42)	(0.43)	(0.39)	(0.39)	(0.39)	(0.39)	(0.39)
Increasing	0.41	0.43	0.34	0.37	0.29	0.32	0.34	0.27	0.28	0.22
	(0.42)	(0.43)	(0.42)	(0.41)	(0.40)	(0.4)	(0.4)	(0.4)	(0.39)	(0.39)
Unexpected	0.78*	0.8*	0.77*	0.81*	0.78*	0.66*	0.68*	0.67*	0.67*	0.66*
	(0.43)	(0.43)	(0.43)	(0.42)	(0.42)	(0.4)	(0.4)	(0.39)	(0.39)	(0.39)
Control Mean			0.98 [1.24]					1.2 [1.29]		
Joint equality test (p-value)	0.54	0.52	0.49	0.39	0.37	0.63	0.61	0.52	0.51	0.44
Observations			192					187		

Notes: Tobit estimates. The dependent variable is the number of exercise sessions in a particular period. Each panel represents a different time period. In each period, five specifications are presented, each consisting of a different subset of control variables: (1) None, (2) Gender, (3) Gender and Prior exercise, (4) Gender and Year of study, (5) Gender, Year of study, and Prior exercise. Robust standard errors appear in parentheses. Standard deviation appears in brackets.
* $p < 0.10$.; ** $p < 0.05$, *** $p < 0.01$.;

Thus, our two intermittent incentives indeed produced an enduring effect more successfully than the *Per-visit* treatment, where the decay between the incentivized period and the non-incentivized period is the smallest in the *Unexpected* group. The return to exercise in the incentivized period in our study is comparable to the pooled effect of the successful incentive schemes in Milkman et al. (2021) mega study, which was 0.3 in the 10 weeks after the intervention.

In conclusion, the *Per-visit* scheme is effective only in the short term; the *Increasing* scheme is effective in both the short term and the long term; and the *Unexpected* scheme is the least effective during the short term, although its effect declines the least over time, a pattern that becomes important in the 12- and 18-month follow-ups.

3.2. Follow-up

We used the 12- and 18-month follow-up questionnaires to determine whether participants continued to exercise after the free six-month sports center membership had expired. Participants received 50 NIS to answer each questionnaire. Both questionnaires had a high response rate that was similar across groups though not identical (in the 12-month questionnaire, the response rate was 91% in Control, 94% in Increasing, and 96% in Per-visit and Unexpected; in the 18-month questionnaire, the response rate was 87% in Control, 92% in Unexpected, and 94% in Per-visit and Increasing). The main question was "How many times a week do you exercise on average?" and we used the responses as the dependent variable in the analysis. Although it is based on a self-report and is not as accurate as our measure in the membership period, we have no reason to expect different biases in self-reporting between the treatments. We chose not to limit the analysis to exercising in the sports center, because only about 15% of the students continued their membership after the six-month free membership period, which also meant we no longer had any reason to use commute time as a control variable. Furthermore, some of the participants graduated by the time they received the follow-up questionnaires and thus may have had more or less time to exercise. Therefore, controlling for the year of study (at the beginning of our experiment) is important.

Table 4 compares each of the incentivized treatments with the *Control* group in the two follow-up periods. Supplementary Fig. S3 in the appendix depicts the distribution of the number of exercise sessions in the two follow-up periods. Panel A of Table 4 shows that participants in the *Unexpected* treatment exercised significantly more than participants in the *Control* group one year after the beginning of the experiment: an effect of 0.78 ($SE = 0.42$) is estimated in the full specification. By contrast, the *Increasing* and *Per-visit* groups did not exercise significantly more than the *Control*.

Because self-selection into the follow-up survey may bias the results, we re-analyzed the data using a weighted regression. First, we used a logit model to calculate the probability that a participant responded to the 12- and 18-month surveys as a function of his or her individual characteristics and treatment, and then we used the inverse of these probabilities as the observations' weights in a weighted linear regression (see Table S5 in the appendix). The results support our conclusion that only participants in the *Unexpected* treatment exercised significantly more than participants in the *Control* group 12 months after the beginning of the experiment.[11]

For the 18-month follow-up questionnaire, Panel B of Table 4 suggests a significant, though weaker, effect of the *Unexpected* Treatment. However, in a weighted regression analysis we found no significant difference between the incentivized treatments and *Control* (see Table S5 in the appendix), suggesting the results in Panel B of Table 4 may be driven by se-

[11] As another robustness test, we assumed that participants who did not respond to the survey do not exercise at all (i.e., zero exercise sessions a week) and re-estimated the effect of our treatments. We found that if such an extreme selection takes place, our conclusions from the 12-month follow-up questionnaire hold (see Table S6 in the appendix).

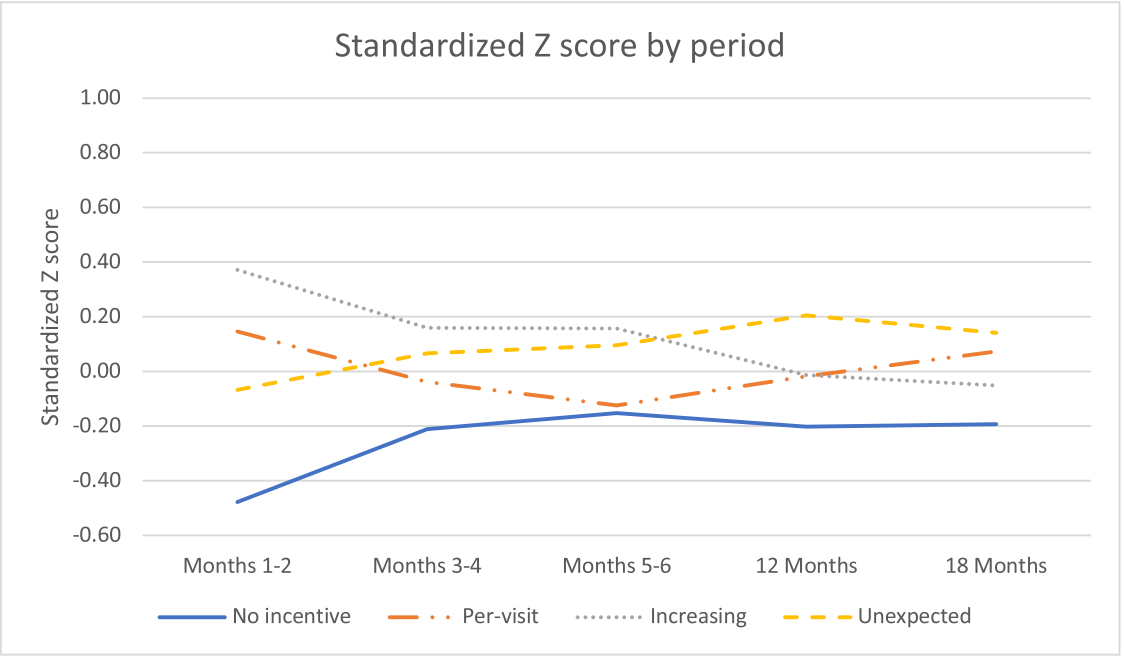

Fig. 2. Standardized *Z*-score by period for visits to the gym and follow-ups.

lection.[12] Yet, the findings of the long-term follow-ups are encouraging in view of the robust result that the exercise effect persisted one year after the beginning of the experiment (10 months after the end of the incentivized period).

Thus, whereas the effect of the incentives for participants in the *Per-visit* and *Increasing* groups diminished considerably over time, participants in the *Unexpected* scheme experience less of an effect during the incentivized period, but it diminished less subsequently. This difference between the incentivized groups also mitigates the concern regarding misreporting, which should not be different between groups. Nevertheless, these results should be interpreted cautiously because self-reports are not as accurate as our measure of visits to the gym, and self-selection into the follow-up survey might be related to unobserved characteristics.

3.3. *Persistence of incentive effects*

Fig. 2 presents the mean standardized *Z*-score for number of exercise sessions [$Z = (x - \textit{Sample Mean}) / \sigma$], for each group in each period (months 1–2, months 3–4, months 5–6, 12-month follow-up, and 18-month follow-up), and shows the difference in trends between the treatments. We used *Z*-scores because behavior was measured using different scales during the membership period and the follow-up period. For the membership period, we used the number of visits to the gym, whereas for the follow-up period, we used the answer to the question "How many times do you exercise each week on average?"

Fig. 2 indicates the Control group had the lowest *Z*-score during the incentivized period. After the incentives were removed, the *Control* group's *Z*-score increased and then remained stable. Participants in the *Per-visit* treatment exercised more than the *Control* and the *Unexpected* when incentives were present but had a sharp decrease in gym visits after the incentives were removed. Participants in the Increasing treatment visited the gym more than the other groups during the incentivized period and the non-incentivized membership period. However, over time, after the membership period, participants in the *Unexpected* group exercised more. The effect of the Unexpected treatment, which was the least effective among the incentivized groups in getting participants to start exercising, steadily improved relative to the other treatments. Thus, although it was less effective in getting the participants to start exercising in the short term, it led to greater exercise perseverance in the long run.

Table 5 presents each treatment's effect as a percentage of the *Control* group mean (i.e., the percentage increase) and thus provides another way to see the magnitude of the effects across the different time periods. The patterns that appear in Fig. 2 are evident in the table as well. In particular, whereas the large effect of the *Per-visit* scheme in the incentivized period fades away in later periods, a much smaller decrease occurs in the *Unexpected* group: its effect in the four months

[12] Interestingly, participants in all four groups exercised more after 18 months than after 12 months (see Supplementary Fig. S3 in the appendix). This improvement can be attributed to the different seasons, because individuals in Israel tend to exercise less in the winter than in the summer (the two follow-ups were conducted in January and June, respectively).

Table 5
Treatment effect as percentage of the *Control* mean.

	months 1-2	months 3-6	month 12	month 18
Per-visit	302%	113 %	33%	40%
Increasing	375%	154%	31%	18%
Unexpected	235%	154%	80%	55%

Notes: The treatments effect as a percentage of the *Control* group's mean. The measure is based on the Tobit estimates of the treatments' effects in the full specification in Tables 3, 4, and 5 and on the *Control* group's mean number of gym visits, which is reported in these tables for each period.

Table 6
Did the participant visit the gym at least once?

	Panel A: Months 1–2					Panel B: Months 3–6				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Constant	0.43***	0.40***	0.40***	0.37***	0.37***	0.41***	0.38***	0.36***	0.41***	0.40***
	(0.07)	(0.09)	(0.09)	(0.10)	(0.10)	(0.07)	(0.09)	(0.09)	(0.10)	(0.10)
Per-visit	0.36***	0.36***	0.36***	0.35***	0.35***	0.16	0.16	0.15	0.16	0.15
	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Increasing	0.37***	0.37***	0.36***	0.36***	0.35***	0.19*	0.19**	0.16	0.20**	0.17*
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Unexpected	0.28***	0.27***	0.27***	0.28***	0.27***	0.20**	0.21**	0.19*	0.20**	0.19*
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Control Mean			0.43 [0.5]					0.41 [0.5]		
Joint equality test (p-value)	0.46	0.49	0.51	0.55	0.56	0.89	0.89	0.9	0.88	0.93
Observations			211					211		

Notes: OLS estimates. The dependent variable is whether participants visited the gym at least once during a particular period. Each panel represents a different time period. In each period, five specifications are presented, each consisting of a different subset of control variables: (1) None; (2) Gender and Commute time; (3) Gender, Commute time, and Prior exercise; (4) Gender, Commute time, and Year of study; and (5) Gender, Commute time, Year of study, and Prior exercise. Robust standard errors appear in parentheses. Standard deviations appear in brackets.
* $p < 0.10$,; ** $p < 0.05$,; *** $p < 0.01$.;

after the incentivized period reflects a 154% increase relative to the *Control* group. Furthermore, the effect of the *Unexpected* scheme persists even 10 months after the incentivized period, with an 80% increase relative to the *Control.*

3.4. Additional analysis

This section presents the results of several robustness checks and an analysis of additional data collected in the experiment.

3.4.1. Exercised or not

A different way to determine whether the incentives were able to encourage exercise is to use the metric of whether participants exercised at least once per period. Thus, we use an OLS regression to examine the effect of the treatment on whether an individual exercised at least once during each period.[13] Although in the pre-study questionnaire, the participants indicated they wanted to exercise more, and although they had attended a one-hour instruction session, a proportion ranging from 20% to 60% of each group did not exercise at all during the incentivized period (Supplementary Fig. S1 in the appendix). The highest proportion (60%) occurred in the Control group, demonstrating that free membership was insufficient to get most of this group exercising. By contrast, the proportion of participants who did not exercise at all in the other three incentivized treatments is between 20% and 30%. Panel A of Table 6 shows that during the incentivized period, a larger proportion of participants across all incentivized treatments, ranging from 0.27 to 0.35 more than the *Control* group, visited the gym at least once.

Panel B of Table 6 presents the results of the analysis for the non-incentivized membership period. During this period, a proportion of 0.17–0.19 more participants exercised in the *Increasing* or *Unexpected* treatment than in the *Control* group, and no significant differences existed between the *Per-visit* group and the *Control* group at the 10% level. These results, like the previous ones, show a more rapid decline in exercising under the *Per-visit* incentive scheme, which demonstrates its inferiority in the long run, after incentives are removed.

[13] An analysis using a logistic regression is reported in Table S10 in the appendix.

3.4.2. Sports center entrances

Besides use of the gym, the sports center membership allows hanging out in the swimming pool, participating in various classes, such as yoga, and other activities. Although participants were only incentivized to attend the gym, they may also have taken advantage of the other activities at the sports center. Thus, we performed a similar analysis to our main analysis, with the number of entries only to the sports center without entering the gym, instead of the number of visits to the gym, as the dependent variable (see Table S8 and S9 in the appendix). Our results show no significant differences between all treatments in all membership periods, except for a difference in the incentivized period between the *Unexpected* treatment and the *Control* ($\beta_{Unexpected}$ = -0.82, SE = 0.49, p = 0.095). Thus, our study did not cause participants in any of the three incentivized treatments to exercise more than the *Control* group in the other facilities of the Sports Center. That is, no spillover effect occurred during the membership period.

3.4.3. Physiological and psychological indicators

Eighty-two percent of the participants attended the concluding session, with substantial differences across groups (74.5% in *Control*, 84.9% in *Per-visit*, 92.5% in *Increasing*, and 78.8% in *Unexpected(*. For those who attended the session, we calculated the changes in seven physiological and psychological indicators between the introductory session and the concluding session (body fat, weight, pulse, consideration of future consequences, propensity to plan, risk preferences, and happiness). To mitigate selection concerns, we used propensity score matching to obtain a matched sample (N=128) on which we performed our analysis. We then examined whether exercising induced by our incentives in the membership period affected the changes in the physiological and psychological measures, using the *incentivized treatments* jointly as an instrumental variable (see Table S12 in the appendix).

Examining each measure separately, we found that an increase of one visit to the gym (compared to the *Control*) during the membership period reduced the participant's body fat percentage by 0.22 percentage points (2SLS estimate: β_{visits} = -0.22, SE = 0.13, p = 0.096). Overall, the incentivized treatments (jointly) reduced the participant's body fat percentage by 1.70 percentage points (OLS estimate: $\beta_{treatments}$ = -1.70, SE = 0.94, p = 0.072). However, using Bonferroni correction for multiple hypotheses testing suggests that the reduction in body fat is not significant at the 10% level. We did not detect any significant effect of exercising on the other six measures.

3.4.4. Students' grades

Following Cappelen et al. (2022), we examined whether exercising in the membership period affected the students' grades. We received from Tel Aviv University the grades of 184 participants in two semesters of the 2017–2018 academic year.[14] The exam period of the first semester started three weeks after the beginning of the experiment and ended a few days after the end of incentivized period. Therefore, we examine whether grades in the first semester were affected by exercising in months 1–2. The exam period of the second semester started roughly when the six-month membership period ended, and we also examined whether exercising in the membership period affected the participants' grades in the second semester of 2017–2018.

We estimated the effect of exercising on grades by using the treatment as an instrumental variable, as in Cappelen et al. (2022). We examined the effect of each treatment separately as well as the effect of the three treatments jointly. We regressed the grades in each semester on the number of gym visits in the corresponding period, instrumented on the treatment status. In different specifications, we used different subsets of the following control variables: gender, commute time, year of study, prior exercising, field of study, Israeli SAT scores, matriculation exam scores, and additional psychological measures.

Although our treatments increased exercising in these periods relative to the *Control*, experiment-driven exercising in months 1–2 did not affect the students' grades in the first semester and exercising in months 1–6 did not affect their grades in the second semester (see Table S13 in the appendix).

3.4.5. Incentives' costs and benefits

Recall that in our study, we encouraged exercising also among the *Control* group, by providing them with a free membership to the gym, an app, and so on. Because the fixed costs are the same in all of our treatments, we focus on estimating the incentives' effects on exercising and the associated benefits relative to the *Control* group.

Using the OLS estimations of our treatments' effects, we calculated the cost per extra exercise session—the extra cost in each of our incentivized treatments relative to the *Control*,[15] divided by the estimated number of extra sessions induced by the incentives in a year (i.e., in the membership period and the following six months). Assuming the 12-months self-reporting on exercising reflects the average amount of exercising during the preceding six months, the cost per extra session is $3.82 in *Per-visit*, $3.49 in *Increasing,* and $1.72 in *Unexpected*. Making a different assumption that the estimated effect in months 3–6 remains the same in the following six months, the calculated cost per extra session is $4.27 in *Per-visit*, $3.57

[14] We received the grades of all undergraduate students at the 2017–2018 academic year, out of 204 participants who signed at the beginning of the experiment that they agreed to share their grades with us.

[15] For simplicity, we assumed a visit in the incentivized period was rewarded by 20 shekels on average in all the incentivized treatments. Thus, the average cost per person is 44.55 in the *Per-visit*, 55.33 in the *Increasing*, and 36.16 in *Unexpected*.

in *Increasing,* and $3.29 in *Unexpected*. Note that these calculations do not take into account additional exercise sessions (induced by our incentives) that occur after a year.

Carlsen et al. (2015) estimated the annual health-care expenditures of physically active versus inactive adults in the US and found they are between $920 and $1,437. The weekly benefits from being active (i.e., about five exercise sessions a week) is hence between $17.69 and $27.63, and a session's benefit is between $3.54 and $5.53, on average. Given these estimated benefits, using our intermittent incentives may be justified.

4. Conclusion

We study whether intermittent incentives can induce individuals to establish an exercising habit and maintain it after the incentives are removed. We examine the short- and long-run effects of two novel intermittent incentive schemes and a per-visit incentive scheme on exercising. Although all three incentives schemes were effective while in place, only the *Increasing* and *Unexpected* incentive schemes induced more exercising than the *Control* group after incentives were removed. The *Unexpected* treatment's effect lasted for 10 months after the end of the incentivized treatment. Together, the results indicate that whereas the *Per-visit* scheme works well in the short run, the intermittent schemes are superior in the long run. The results point to an innovative way to incentivize individuals, which can potentially improve the long-run success of interventions. The results are particularly important in designing incentives-based policies to encourage the creation of new habits.

Data availability

Data will be made available on request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2022.11.015.

References

Acland, Dan, Levy, Matthew, 2015. Naiveté, projection bias, and habit formation in gym attendance. Manag. Sci. 61, 146–160.
Amsel, Abram., 1958. The role of frustrative nonreward in noncontinuous reward situations. Psychol. Bull. 55, 102–119.
Amsel, Abram., 1992. Frustration Theory. An analysis of Dispositional Learning and Memory. Cambridge University Press, New York, NY.
Andrade, Leonardo F., Barry, Danielle, Litt, Mark D., Petry, Nancy M., 2014. Maintaining high activity levels in sedentary adults with a reinforcement-thinning schedule. J. Appl. Behav. Anal. 47, 523–536.
Bachireddy, Chethan, Joung, Andrew, John, Leslie K., Gino, Francesca, Tuckfield, Bradford, Foschini, Luca, Milkman, Katherine L., 2019. Effect of different financial incentive structures on promoting physical activity among adults: a randomized clinical trial. JAMA Netw. Open 2 (8), e199863.
Becker, Gary S., Murphy, Kevin M., 1988. A theory of rational addiction. J. Polit. Econ. 96, 675–700.
Belot, Michele, James, Jonathan, Nolen, Patrick, 2016. Incentives and children's dietary choices: a field experiment in primary schools. J. Health Econ. 50, 213–229.
Blais, Ann-Renee, Weber, Elke U., 2006. A domain-specific risk-taking (DOSPERT) Scale for adult populations. Judgm. Decis. Mak. 1, 33–47.
Cappelen, Alexander W., Charness, Gary, Ekström, Mathias, Gneezy, Uri, Tungodden, Bertil., 2022. Exercise Improves Academic Performance. NHH Dept. of Economics Discussion Paper No. 08/2017.
Carlson, Susan A., Fulton, Janet E., Pratt, Michalel, Yang, Zhou, Adams, Kathleen E., 2015. Inadequate physical activity and health care expenditures in the United States. Prog Cardiovasc Dis 57, 315–323.
Carrera, Mariana, Royer, Heather, Sther, Mark, Sydnor, Justin, 2018. Can financial incentives help people trying to establish new habits? Experimental evidence with new gym members. J. Health Econ. 58, 202–214.
Carrera, Mariana, Royer, Heather, Sther, Mark, Sydnor, Justin, 2019. The structure of health incentives: evidence from a field experiment. Manag. Sci. 66 (5), 1890–1908.
Charness, Gary, Gneezy, Uri, 2009. Incentives to exercise. Econometrica 77, 909–931.
Condliffe, Simon, Isgin, Ebru, Fitzgerald, Brynne, 2017. Get thee to the gym! A field experiment on improving exercise habits. J Behav Exp Econ 70, 23–32.
DellaVigna, Stefano, Malmendier, Ulrike, 2006. Paying not to go to the gym. Am. Econ. Rev. 96 (3), 694–719.
Domjan, Michael., 2010. The principles of learning and behavior, 6th ed. Wadsworth/Cengage Learning, Belmont CA.
Ferster, Charles S. and Burrhus F. Skinner. 1957. "Schedules of reinforcement." Appleton-Century-Crofts, New York, NY.
Golz, Sonia M., 1992. A sequential learning analysis of decisions in organization to escalate investments despite continuing costs or losses. J. Appl. Behav. Anal. 25, 561–574.
Hardman, Charlotte A., Horne, Pauline J., Lowe, C.Fergus, 2011. Effects of rewards, peer-modelling and pedometer targets on children's physical activity: a school based intervention study. Psychol Health 26 (1), 3–21.
Gneezy, Uri, Meier, Stephan, Rey-Biel, Pedro, 2011. When and why incentives (don't) work to modify behavior. J. Econ. Perspect. 25 (4), 191–210.
Hills, Peter, Argyle, Michael, 2002. The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being. Pers Individ Dif 33, 1073–1082.
Hochman, Guy, Erev, Ido, 2013. The partial-reinforcement extinction effect and the contingent-sampling hypothesis. Psychon. Bull. Rev. 20, 1336–1342.
Hogarth, Robin M., Villeval, Marie Claire, 2014. Ambiguous incentives and the persistence of effort: experimental evidence. J. Econ. Behav. Organ. 100, 1–19.
Humphreys, Lloyd G., 1939. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. J. Exp. Psychol. 25, 141–158.
Jeffery, Robert W., 2012. Financial incentives and weight control. Prev. Med. 55, S61–S67.
Just, David R., Price, Joseph, 2013. Using incentives to encourage healthy eating in children. J. Hum. Resour. 48 (4), 855–872.
Laibson, David., 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112, 443–477.
Lynch, John G., Netemeyer, Richard G., Spiller, Stephen A., Zammit, Alessandra, 2010. A generalizable scale of propensity to plane: the long and the short of planning for time and for money. J. Consum. Res. 37 (1), 108–128.
Milkman, Katherine L., Gromet, Dena, Ho, Hung, et al., 2021. Megastudies improve the impact of applied behavioural science. Nature 600, 478–483.

Mitchell, Marc S., Orstad, Stephanie L., Biswas, Aviroop, Oh, Paul I., Jay, Melanie, Pakosh, Maureen T., Faulkner, Guy, 2019. Financial incentives for physical activity in adults: systematic review and meta-analysis. Br. J. Sports Med. 54 (21), 1259–1268.
O'Donoghue, Ted, Rabin, Matthew, 1999. Doing it now or later. Am. Econ. Rev. 89, 103–124.
Rohde, Kirsten I.M., Verbeke, Willem, 2017. We like to see you in the gym – A field experiment on financial incentives for short and long term gym attendance. J. Econ. Behav. Organ. 134, 388–407.
Papini, Mauricio R., 2003. Comparative psychology of surprising nonreward. Brain Behav. Evol. 62, 83–95.
Patel, Mitesh S., Volpp, Kevin G., Rosin, Roy, L.Bellamy, Scarlett, Small, Dylan S., Heuer, Jack, Sproat, Susan, Hyson, Chris, Haff, Nancy, Lee, Samantha M., Wesby, Lisa, Hoffer, Karen, Shuttleworth, David, Taylor, Devon H., Hilbert, Victoria, Zhu, Jingsan, Yang, Lin, Wang, Xingmei, Asch, David A., 2018. A randomized, controlled trial of lottery-based financial incentives to increase physical activity among overweight and obese adults. Am. J. Health Promot. 32 (7), 1568–1575.
Petry, Nancy M., Andrade, Leonardo F., Barry, Danielle, Byrne, Shannon, 2013. A randomized study of reinforcing ambulatory exercise in older adults. Psychol. Aging 28 (4), 1164–1173.
Pittenger, David J., 2002. The two paradigms of persistence. Genet. Soc. Gen. Psychol. Monogr. 128, 237–268.
Roemmich, James N., Lobarinas, Christina L., Barkley, Jacob E., White, Tressa M., Paluch, Rocco, Epstein, Leonard H., 2012. Use of an open-loop system to increase physical activity. Pediatr. Exerc. Sci. 24, 384–398.
Royer, Heather, Stehr, Mark, Sydnor, Justin, 2015. Incentives, commitments, and habit formation in exercise: evidence from a field experiment with workers at a fortune 500 company. Am. Econ. J.: Appl. Econ. 7, 51–84.
Strohacker, Kelley, Galarraga, Omar, Williams, David M., 2014. The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. Ann. Behav. Med. 48, 92–99.
Strathman, Alan, Gleicher, Faith, Boninger, David S., Edwards, Scott, 1994. The consideration of future consequences: weighting immediate and distant outcomes of behavior. J. Pers. Soc. Psychol. 66 (4), 742–752.
Weber, Elke U., Blais, Ann-Renee, Betz, Nancy E., 2002. A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. J Behav Decis Mak 15, 263–290.
Wing, Rina R., Jeffery, Robert W., Pronk, Nicolaas, Wendy, Hellerstedt L., 1996. Effects of a personal trainer and financial incentives on exercise adherence in overweight women in a behavioral weight loss program. Obes. Res. 4, 457–462.
World Health Organization (WHO). 2018. "Global action plan on physical activity: more active people for a healthier world." Retrieved online at October 13 2019 from: https://www.who.int/publications-detail/global-action-plan-on-physical-activity-2018%E2%80%932030