The Unexpected Downside of Paying or Sending Messages to People to Make Them Walk: Comparing Tangible Rewards and Motivational Messages to Improve Physical Activity

8

MAURO CHERUBINI, GABRIELA VILLALOBOS-ZUÑIGA, and MARC-OLIVIER BOLDI, University of Lausanne
RICCARDO BONAZZI, HES-SO

People do not exercise as much and as regularly as they should. To support users in adopting healthy exercise routines, app designers integrate persuasive techniques in their apps. In this study, we focus on two of these techniques, i.e., offering *tangible rewards* and sending *motivational messages* to users. Past research has demonstrated the effects of these techniques in nudging recipients to increase their physical activity levels. However, the effect of these interventions on the *intrinsic motivation* of the participants has not yet been studied. We conducted a 10-month study involving 208 participants; this research consisted of a 3-month baseline (pre-phase), a 4-month experiment and a 3-month follow-up (post-phase). The participants were randomly assigned to one of the following three interventions: either they receive money ((i.) through a fixed incentive or (ii.) a lottery), or (iii.) informative messages. Their daily goal was to walk 10K steps. Through their smart phones, we recorded how many steps they walked every day. These interventions had no effect on the main outcome variable (i.e., the number of steps). However, the manipulations produced a detrimental effect on the intrinsic motivation of the participants, measured through a standardized questionnaire. This negative effect extended into the follow-up period. Our study reveals that tangible rewards and motivational messages decrease the intrinsic motivation of the participants, hence their connected physical activity. In our findings, we highlight the importance of intrinsic motivation in setting up healthy exercise routines that will be carried on autonomously by the participants after the period of the intervention. Finally, we present implications for the design of persuasive apps.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **HCI design and evaluation methods**;

Additional Key Words and Phrases: LMER, longitudinal study, persuasive technology, randomized controlled trial

ACM Reference format:
Mauro Cherubini, Gabriela Villalobos-Zuñiga, Marc-Olivier Boldi, and Riccardo Bonazzi. 2020. The Unexpected Downside of Paying or Sending Messages to People to Make Them Walk: Comparing Tangible Rewards and Motivational Messages to Improve Physical Activity. *ACM Trans. Comput.-Hum. Interact.* 27, 2, Article 8 (March 2020), 44 pages.
https://doi.org/10.1145/3365665

This work was supported by the UNIL HEC research fund and by the research fund of HES-SO, Switzerland.
Authors' addresses: M. Cherubini, G. Villalobos-Zuñiga, and M. Olivier Boldi, University of Lausanne, Faculty of Business and Economics (HEC), Lausanne, VD, Switzerland; emails: {mauro.cherubini, gabriela.villalobos, marc-olivier.boldi}@unil.ch; R. Bonazzi, HES-SO, Route du Rawil 47, Sion, VS, Switzerland; email: Riccardo.Bonazzi@hevs.ch.

1073-0516/2020/03-ART8 $15.00
https://doi.org/10.1145/3365665

1 INTRODUCTION

We all know that exercising is good for us [6]. Yet, many fail at giving it the importance it deserves [39, 72]. We often do not choose how to optimally invest our time because our brains are tricked by, among others, relativity of choices[1], emotions, and social norms [4]. However, once we understand when and where we make erroneous decisions, we can teach ourselves to think differently and embrace healthier behavior. Technology can play a major role in empowering users to overcome their natural limitations and to opt for healthy choices.

Designers have taken this challenge and today we can find thousands of apps on the market that aim at supporting behavior-changing programs. These apps have various purposes, for instance, helping the user to lose weight, to learn a new language, to quit smoking, or to stay hydrated. Many of these applications are based on behavior-changing techniques that are supposed to help users establish healthy routines or cease negative ones. We still lack, however, systematic design guidelines for these persuasive features that can help designers produce more effective apps. For instance, it is not clear which of these techniques is better than the other and for which kind of user.

This challenge can be described with a single question: *what motivates people*? Some people are money driven, whereas others are completely uninterested in material incentives. Some people are extremely sensitive to awareness campaigns, e.g., global warming, and others are unresponsive to them. Theories on human motivation describe motivation in a continuum that goes from extrinsic to intrinsic. [98]. *Extrinsic motivation* supports actions that are executed because there is an element external to the activity that moves the individual to perform the activity (e.g., reading a book because it will be evaluated in an exam next week). Whereas, a person has *intrinsic motivation* toward specific actions that they might find to be satisfying (e.g., the pleasure of reading a book because the content is appealing).

Given these two extremes, app designers developed the following two classes of interventions: The first class involves *tangible rewards* that are meant to encourage a person by offering benefits external to the activity [14]. In the second class, we find *motivational messages* that are supposed to grow the inner values that each person associates with performing an activity [19]. Although the motivational messages might scale easily to support large-scale deployment, the tangible rewards might require subsidies from private companies or governmental agencies.

Examples of applications belonging to the first group are SweatCoin[108], Helsana+[52], ExerciseRewards[33], or Healthy Wage[50]. These apps assign virtual points to the users who achieve specific goals (e.g., walking a certain amount of steps or exercising for a given amount of time). Then, these points can be converted in real money to buy goods or services. Other applications, such as Pact[46] or DietBet[115], propose to the users that they bet on whether they will achieve the goal. If users win the bet, they gain actual money. Other applications follow the other strategy described earlier; they send motivational messages to the users. Examples of apps in this second group include Runkeeper[94], Fitbit[35], and Food stand[38]. These apps send informative messages at a regular frequency to the users. These messages are designed to provide convincing evidence about why performing the activity could be beneficial and recommendations that can help the recipient set up routines to achieve the target behavior (e.g., exercise more often).

Unfortunately, research is still inconclusive about whether these interventions provide consistent adherence to healthy routines and whether they are equally supportive to all kinds of users. Concerning the financial incentives, researchers still debate the long-term benefits of providing tangible rewards to support changes in behavior.[7, 69]. Furthermore, it is still unclear how to

[1]Humans rarely choose in absolute terms. Rather we look at the relative advantage of one option over another.

properly design the interventions that involve financial payouts [43, 83]. Similarly, research on motivational messages has not yet contributed strong evidence that regularly providing this information to users leads to the consistent adoption of the target behavior [60, 112]. Unfortunately, as of this writing, little work has been conducted on the effects that these techniques have on the attitudes of people toward their desired outcomes (i.e., how motivated people are to perform a certain activity). Therefore we posed the following research question: *Will tangible rewards and motivational messages help or hinder people in attaining a more autonomous level of motivation to perform physical activity?*

We conducted a randomized controlled trial to compare the effectiveness of tangible rewards and motivational messages in motivating people to adopt healthy exercise routines. For an application domain, we choose that of *physical activity (PA)*, given that a large group of persuasive apps are typically built for this category of use. We compared the following three experimental conditions: two conditions where participants were offered money to increase their PA and one condition where participants were persuaded by motivational messages. The two conditions offering tangible incentives varied in the salience of the prizes being offered to participants. We provide evidence that shows the detrimental effects of the rewards and motivational messages on the users' motivation to perform PA. We discuss the implications of these results on the design of persuasive apps. Next, we frame the presented research within the larger body of work conducted in this domain.

2 RELATED WORK

To explain users behavior, scholars studying the effect of technological interventions on the increase of PA[2] refer to psychological theories of human motivation and to behavioral economics studies. Before reviewing studies that are specific to our research, we reviewed the seminal research on which these studies are built.

2.1 Behavioral Economics Studies and Behaviour Change

As briefly described in the introduction, people generally know that PA is good for their health. However, as revealed by surveys, a large portion of the population does not participate enough in these activities (see the Centers for Disease Control and Prevention report in the USA [39]). In the last few years, several studies have revealed systematic mistakes that people make when taking decisions and this could explain why people do not engage in PA as much as they should. These predictable "irrationalities" are often produced by cognitive biases. For instance, researchers found that people tend to favour *immediate* gratification over *delayed* gratification [75]. For example, as the benefits of exercising, for the most part, are perceived only after a certain amount of time (days, weeks, and months), the immediate returns of eating a meal might be preferred. An alternative explanation for why people avoid PA is offered by the *projection bias* [65]: the tendency to overestimate the normality of our beliefs in comparison to others' beliefs and our future beliefs. Basically, we tend to think that we will feel and act in the future the same as we do now. In the context of PA, this bias makes us prefer options that satisfy current desires or emotional states, instead of pursuing long-term goals.

Other scholars have tried to counter these cognitive biases by designing incentives and other interventions in order to change the perceived benefits of engaging in certain activities. For instance, Loewenstein et al. [64], study the design of different incentive schemes to encourage patients to take better care of themselves and find that small and frequent payments are more effective than discounts on monthly premiums or bonuses on their paycheck. Other studies conducted in this

[2]In the remainder of the article we will sometimes refer to physical activity with the acronym PA.

direction show that lotteries can be more effective than fixed payments, when nudging people to adhere to target behaviors [83, 85]. These results are explained by the *Prospect Theory* [58] that shows that people tend to overestimate small probabilities, and by the *Saliency Theory* [10] that states that true probabilities might be distorted by consistent payoffs. Both theories contribute to the explanation of why larger wins, which are improbable, might be more motivating than smaller wins that are more likely to occur.

These theories typically fall short in explaining what happens in the post-intervention periods. In this critical phase of these experiments, inconclusive or contradictory findings were reported [43]. Therefore, scholars sometimes refer to exogenous factors that would depress the effects recorded during the intervention phase in the post-experimental phase, when the intervention is removed. This is referred to as the *crowding-out effect* [13, 26, 70]. In this regard, other studies conducted in the domain of psychology provide additional data to better understand what happens after the intervention is removed.

2.2 Psychological Theories Explaining Behavior Change

The four theories most used to frame technological interventions in the PA domain are the following: Social Cognitive Theory (or SCT), Theory of Planned Behavior (or TPB), the Transtheoretical Model (or TTM), and Self-Determination Theory (or SDT) [12].

SCT [5] emphasizes the reciprocal causation relationship—people, behavior, environment—in which individuals are actors, as well as products of their environment. It states that behavioural changes occur due to a personal sense of control that is affected by various constructs; and *self-efficacy* is the most important construct. Self-efficacy indicates the belief that a person is capable of executing actions that reflect a sense of control over their environment. SCT posit that to measure *self-efficacy,* it is necessary to account for the following three dimensions: *strength* (a probabilistic judgment of how certain a person is of their ability to perform a specific task), *magnitude* (ordering the tasks by level of difficulty), and *generality* (the extent to which self-efficacy expectations about a particular situation or experience can be applied to other situations). Scholars who have applied SCT within the PA domain find that *self-efficacy* is a predictor of PA [31, 68, 91]. However, these studies only measured the dimension of *strength* but not the other dimensions, resulting in a misinterpretation of the utility of *self-efficacy* effects on PA. This suggests that the construct of self-efficacy might not be a suitable way to study and manipulate PA, which led us to consider other theories for this work.

TPB [2] explains the decision-making process individuals go through when setting up a new routine. The theory posits that *intention* is influenced by the following three constructs: *attitude*, *subjective norm,* and *perceived behavioral control.* Attitude indicates the person's evaluation toward the behavior. A subjective norm represents the belief that significant others want the person to engage in the behavior. Perceived behavioral control indicates the person's perception that the behavior is under their control. These three elements determine how strong the *intention* will be and, consequently, the likelihood of the individual to engage in the behavior. After the *intention* is conceived, a period of time occurs before the individual performs the behavior. Hausenblas et al. [49] suggest that this relation (intention-behavior) does not diminish over time, whereas Ajzen and Fishbein [3] recommend measuring it, as close as possible, from the beginning of the intended behavior. This ambiguity led us to consider that the suggested constructs were difficult to put in practice in the context of a behavior-changing intervention.

The TTM [88] is a practical approach, in six stages, to explaining behavior changes. Each of these six stages represents a temporal period characterized by a behavior pattern and a different degree of the readiness of the person to change. The first stage, *pre-contemplation*, indicates that the individual does not intend to change their behavior in the subsequent months. In the second stage, *contemplation*, the individual has the intention to change their behavior, however, does not act yet. In the next two stages, *preparation* and *action*, the individual is committed to engaging in the intended behavior and starts acting, however, discontinuously. In the fifth stage, *maintenance*, the individual has reached a moment of observable modification in their behavior and has achieved regular patterns. In the last stage, *termination*, the individual has full self-efficacy and is unlikely to return to an unhealthy behavior. The theory also describes the *processes of change*: activities people use to progress through the stages. Some of these include *consciousness raising*, *self-reevaluation,* and *self-liberation*. Despite having this precise process of progression, in a systematic review on PA about TTM-based studies [92], the authors find no evidence of a positive effect of stage-based interventions, as opposed to alternative interventions not based on these stages.

2.3 Self-Determination Theory

SDT is a meta-theory that is concerned with social conditions that facilitate or hinder human growth and well-being. The theory examines how specific contextual conditions "*either enhance or undermine the inherent human capacities for psychological growth, engagement, and wellness*" [101, p. 3]. SDT research studies factors that facilitate or hinder human motivation, ultimately leading the individual toward vitality or, alternatively, to unhappiness. SDT is an *organismic perspective*, meaning that it offers a unified approach to psychological growth, integrity, and wellness. One of the basic statements of the theory is that "*people manifest intrinsic tendencies to take interest in, deeply learn about, and gain mastery with respect to both their inner and outer worlds*" [101, p. 4]. These inclinations will be referred as *intrinsic motivation* in the remainder of this article.

Furthermore, the theory posits that these tendencies are natural yet conditional; they require contextual conditions to satisfy the following three basic psychological needs: *autonomy*, *competence*, and *relatedness*. Autonomy is the need to self-regulate one's experience and actions. This provides the individual a form of functioning associated with feelings of volition, congruency, and integration [21, 96]. Competence refers to the basic need of feeling effective and having mastery [27]. Finally, relatedness concerns the feeling of being socially connected and cared for by others [30].

Motivation is what moves people to action. SDT postulates that people have not only different amounts of motivation toward a certain activity but also—and more importantly—different *types* of motivation, specifically different orientations with underlying attitudes and goals that give rise to action [29]. The most basic distinction is between *intrinsic* and *extrinsic* motivation. The former refers to doing something because a person finds it inherently interesting or enjoyable, whereas the latter refers to doing something because it leads to a separate outcome. Furthermore, SDT proposes that there are several types of extrinsic motivation that differ in the degree of internalization (i.e., the degree to which the behavioral regulation is autonomous versus controlled). As described by Ryan and Deci "*behaviors can be externally regulated, meaning they are directly controlled by external and self-alien forces; or they can be controlled through introjection, in which case the person has taken but not fully accepted external controls*" [101, p. 14]. This source of the regulation of the autonomous behavior plays a very important role in moderating the basic need of autonomy and is often referred to as *perceived locus of causality* (PLOC, see [21]). Therefore, SDT organizes the different types along this *autonomy–control continuum* (see Figure 1). On the far left we find *amotivation*, the complete absence of intention to act. Next, we find *external regulation* that describes behaviors that are performed to obtain an externally imposed reward. Farther to the right, we find *introjected regulation* that describes behavior that are performed to avoid guilt. Then we find *identification* that describes a state where the person has accepted the regulation as their own (but not yet integrated). Integration occurs in the last two stages of the continuum; although in the *integration* stage, regulations have been evaluated and are considered congruent with the values, goals, and needs that belong to the self, in the last stage, these become fully assimilated to the self.

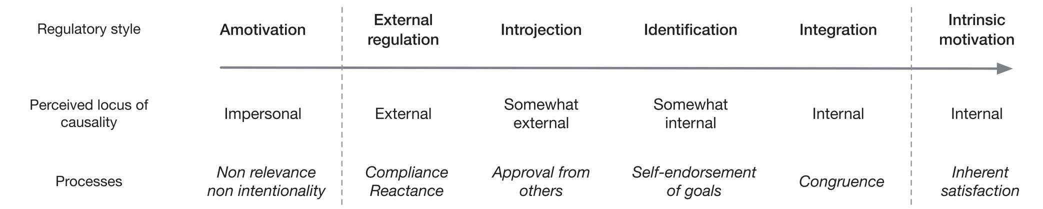

Fig. 1. Taxonomy of human motivation postulated by SDT (adapted from [98]).

According to SDT, extrinsic motivation types can encourage a person to behave a certain way in the short-term, but fail to maintain the behavior over time [29]. Furthermore, according to SDT when the needs are satisfied, the individual reaches higher levels of self-determined motivation, consequently moving toward the intrinsic motivation segment of the continuum. In the next section, we will review specific techniques that can be used to satisfy the basic needs described by the theory, and later specifically in the domain of sport and exercise.

2.4 Incentives and Motivational Messages in SDT

SDT research has extensively focused on how contextual conditions affect intrinsic motivation. During the 1970s and 1980s, there was a burst of experimental studies on rewards, punishments, feedback, and other extrinsic events. Essentially, these contextual episodes can support a person's sense of *autonomy* and *competence*; this support enhances intrinsic motivation, whereas episodes that affect the same basic needs negatively diminish intrinsic motivation.

From an SDT standpoint, incentives are a form of extrinsic and tangible *rewards*. SDT scholars demonstrated that, under well-specified conditions, rewards can yield detrimental effects on intrinsic motivation. In fact, externally administered rewards can be perceived as coercive and controlling, hence disrupting the basic need of autonomy. In an early study, Deci [22] demonstrate that participants who received money for solving puzzles (i.e., task-contingent reward) showed a decrease in their subsequent intrinsic motivation (measured as a free-choice persistence of the target behavior). In comparison, in a later study, researchers test the use of a monetary reward that is provided regardless of the successful completion of the activity (i.e., task non-contingent reward) and find that this kind of reward does not have any negative effect on intrinsic motivation [23]. Later, Deci and Ryan [28, 29] argue that offering an extrinsic reward (e.g., money) for an activity that was already intrinsically motivated can prompt people to experience an *external PLOC* [21] in their behaviour, hence to feel controlled. People exposed repeatedly to extrinsic rewards might come to see the activity as something they did only in order to receive the reward. Furthermore, using rewards to motivate people can signal that the activity is not worth doing for its own sake [101, p. 128]. A meta-analysis conducted by Ryan et al. [104] reports results of over 50 experiments that used *completion-contingent rewards*, specifically rewards that were offered when the target activity was completed. The analysis reveals that completion-contingent rewards undermine intrinsic motivation.

Concerning positive *feedback*, SDT researchers found that this can enhance peoples' sense of competence [28]. Especially when positive feedback is not expected, people are less likely to think that they completed the activity in order to get the feedback, thus not affecting the PLOC. However, other research shows that when feedback is experienced as an evaluation, pressure, or control, it prompts a shift in the PLOC, hence undermining intrinsic motivation [106]. Additionally, Ryan [95] demonstrates that simply providing positive feedback is not enough to motivate people if the participants do not also experience autonomy. A more recent study conducted by Hewett and Conway [53] finds that, overall, providing positive feedback is not detrimental to intrinsic motivation. However, if the feedback is very salient, it can be perceived as controlling hence hindering intrinsic motivation. By salient, it is implied here that the feedback must be *task-inherent* to enable recipients to gain a sense of their effectiveness.

In summary, external events can affect intrinsic motivation, depending on whether their functional significance is *controlling* (i.e., experienced as an external pressure toward a given goal) or *informational* (i.e., experienced as affirming autonomy and competence). Next, we look into studies that focus specifically on increasing PA.

2.5 Incentives and Motivational Messages Aiming at Increasing Physical Activity

Sport and outdoor recreation activities are most often intrinsically motivated because, by practicing them, people satisfy inherent psychological needs [100]. Whereas, exercising is often less enjoyable than playing a sport. Yet, people often persist at it perceiving the activity as instrumental to achieving extrinsic goals (e.g., looking attractive) [102]. Several researchers in the exercise domain [73, 82, 107, 110, 111, 116] find a predictive relation between autonomous types of motivation and PA. As we have seen in the previous section, these types of motivation can be assessed with self-reports, such as the *Intrinsic Motivation Inventory* (IMI [104]). Therefore, we can pose our first hypothesis: *People with a high intrinsic motivation will make more steps than participants with low intrinsic motivation* (*H1*).

Early studies on the use of rewards to motivate athletes showed that these can readily undermine intrinsic motivation [81]. Trophies, prizes, and other rewards can either enhance or diminish intrinsic motivation, depending on factors such as the nature of the contingency (i.e., whether delivery is based on performance, or activity completion) and the functional significance they foster (i.e., being perceived as informative or controlling). Unfortunately, often these rewards are used precisely to select the best players but yield negative consequences for the non-selected players. Recently, the use of incentives as an intervention strategy to increase PA has garnered renewed interest. A meta-review by Mitchell et al. [69] reports the results of several randomized controlled trials where several types of tangible rewards were tested [16, 17, 34]. Although most of the studies report positive effects of these incentives on PA, there is limited evidence to draw conclusions regarding long-term effects. An exception is the study of Jeffery et al. [56]; they recorded sustained exercise adherence for more than a year. Unfortunately, in their study they did not isolate the effect of the incentive over training. Furthermore, the most notable weakness of these studies is the lack of observation of the post-intervention behavior. An exception is the study of Charness and Gneezy [16] who find that the increase in gym attendance persisted for 16 weeks (with a 47 USD incentive per week) following a 5-week intervention. Another study that focuses on the post-intervention period conducted by Patel et al. [83] shows no effect of the intervention in the post-treatment phase; this indicates that more research is necessary to assess the effects of tangible rewards when the money are no longer offered to participants. In particular, there is a strong evidence to support the so-called *crowding out* effect [13, 26, 70]; the potential for new, external motivators to depress intrinsic motivation and harm post-intervention behavior. SDT states that it is precisely the effect of the intervention that harms intrinsic motivation. Rewards when salient and potent, can clearly motivate *immediate* behavior and still have detrimental effects on people's *subsequent* motivation [99]. Therefore, we pose our second hypothesis: *At the end of the experimental phase, the intrinsic motivation of the participants receiving a tangible incentive will be lower than the intrinsic motivation measured at the beginning of the experiment* [*H2a*]. If the detrimental effect of the manipulation would be validated in the experiment, then this would have in turn an effect on PA. Therefore we posed a related hypothesis focusing on the amount of activity: *During the experimental and post experimental phases participants receiving a tangible incentive will be less physically active than participants in other conditions* [*H3a*].

Furthermore, a body of research focuses on positive feedback provided to people under PA training. These works find that information given to trainees can promote autonomous motivation and provide greater vitality thus enhance intrinsic motivation [71]. Using a technological intervention, Duncan et al. [32] were able to enhance people's motivation with a series of verbal guided-imagery. Also, Carpentier and Mageau [15] study the effect of *change-oriented feedback* in which constructive, effectance-relevant messages were provided to improve motivation. Their findings show that such feedback predicts positive outcomes for athletes. A few studies focus on short message service-based interventions to improve PA (e.g., [37, 54, 59]). Unfortunately, these studies do not isolate the effect of messages alone, do not include follow-up periods, and lack a clear theoretical underpinning. An exception is provided by the study of Kinnafick et al. [60]; they study the effect of supportive text messages on PA. Their study, based on SDT, tests the effect of sending, on a weekly basis, supportive messages to individuals embarking on exercise programs. The study reveals an effect of the treatment on the levels of intrinsic motivation; it increases from pre- to post-intervention. Furthermore, there is some partial evidence of increase of PA (self-reported) between the treatment group and the control group. We therefore pose our research hypotheses, complementing *H2* and *H3* described above: *At the end of the experiment phase, the intrinsic motivation of the participants receiving motivational messages will be higher than the intrinsic motivation measured at the beginning of the experiment* [*H2b*] and *During the experimental and post-experimental phases, participants receiving motivational messages will be more physically active than participants participating under other conditions* [*H3b*].

In the rest of this article, we will use interchangeably the terms feedback, informative messages, and motivational messages. We acknowledge that the term feedback is usually used when information is provided in response to a person's performance, whereas informative and motivational messages might not bear any connection with the outcome of the task for which they are given. We will come back precisely to this point in Section 5.

3 METHODOLOGY

To answer our research questions, we chose a longitudinal study because we wanted to study the long-term effects of the tangible rewards and motivational messages. Also, we choose this research method because intra-subject variability is substantially less than inter-subject variability, consequently it is usually more powerful than a cross-sectional study and it provides a more sensible statistical test [51]. Therefore, we conducted a 10-month, randomized controlled trial that included a 3-month baseline (pre-phase), 4-month intervention (experiment), and 3-month follow-up (post-phase). This method enabled us to obtain objective measurements, and to mitigate the effects of possible biases due to the novelty effect of the technology involved in the intervention or the experimental-window-only observations.

Various research studies on human motivation were conducted using the above mentioned methodology (e.g., [47, 77, 90]). Studying human PA would require taking a holistic approach, as a person might perform a variety of activities (yoga, swimming, climbing, etc.) as part of their weekly routines. Unfortunately, reliably measuring the amount of PA that span all possible activities is a complex engineering problem for which a solution is yet to be found. In this experiment, we focused on walking (and running) for several reasons:

- It is a universal activity that is naturally performed by most people, even those who do not do sport. In fact, even going to work (or school) involves walking.
- It is a most basic activity that can be performed everywhere even outside of sport facilities.

— Software and hardware solutions for logging walking (or running) have greatly improved in the last few years and are now used to conduct scientific experiments.

In this section, we provide the details of the design of our experiment.

3.1 Sampling Procedures

A total of 461 potential participants enrolled online for the study and were assessed for eligibility. These people volunteered to be part of a subject pool (consisting of approximately 8K subjects) for behavioural experiments at the University of Lausanne. A specialized unit at our institution, called Labex managed the subject pool, took care of the randomization and enrollment processes, automated the transfers of financial incentives, and kept secure contact information of the participants of the study. We captured demographic data through a screener that also served to check whether respondents qualified for the study. Respondents were excluded from the experiment if they were participating (or had in the last 6 months) in a study involving PA, if during the expected duration of the study they planned to remain without internet connection for more than 7 consecutive days, or if they foresaw circumstances for which they were unable to move for more than 7 consecutive days (e.g., a planned medical intervention) because this would have produced substantial "holes" in the collected data. Also, we excluded participants if they did not have an `iPhone 5S` or a newer version. The screener was also further used to measure their IMI (see [95]) at the beginning of the experiment[3].

This inventory is a multidimensional measurement device intended to assess the participant's subjective experience related to a target activity. It has been used in several experiments related to intrinsic motivation and self-regulation (e.g., Deci et al. [25], Plant and Ryan [86], Ryan [95], Ryan et al. [97, 103]). The instrument is composed of the following seven sub-scale scores: the participants' interests/enjoyment, perceived competences, efforts, values/usefulness, pressure and tension felt, perceived choices, and relatedness while performing the activity. Recently McAuley et al. [67] examined the validity of the IMI and found strong support for its validity. After the eligibility selection, 179 respondents were excluded due to non-compliance with inclusion criteria, leaving 282 participants that were randomly assigned to each experimental condition. A CONSORT diagram showing participant progress through the study is shown in Figure 2. During the first weeks of the experiment, we discovered that older models of the iPhone that we originally considered compatible with the software needed for the study were, in fact, unable to track steps. Specifically, the iPhone 4S, 5 and 5C do not have the M7 Motion Sensor co-processor that tracks steps. Therefore, we had to exclude 53 participants, as they could not take part in the experiment. This number is listed in "Technical issues" in Figure 2. Also, during the analysis of the data, we noticed that 12 participants did not correctly log their steps. In fact, these users had around ∼40% of missing data. We contacted some of these participants and learned that, in some cases, they were not in the habit of bringing their mobile phones on campus and, in other cases, they owned multiple mobile phones and had installed the research app on a secondary device. Therefore, we decided to exclude these participants from the analysis (see Figure 2). In addition to participants that we had to exclude for technical reasons, there were also nine people who dropped out during the experiment phase and 23 who stopped participating during the post-phase, which results in a 14% attrition rate. The study was therefore conducted with a total of 220 participants, whereas the analysis was conducted with 208 participants. The sample size is discussed in Section 3.8.8 below.

[3]In the results, we will refer to this initial measure of the intrinsic motivation as IMI_1; it was recorded at week 14.

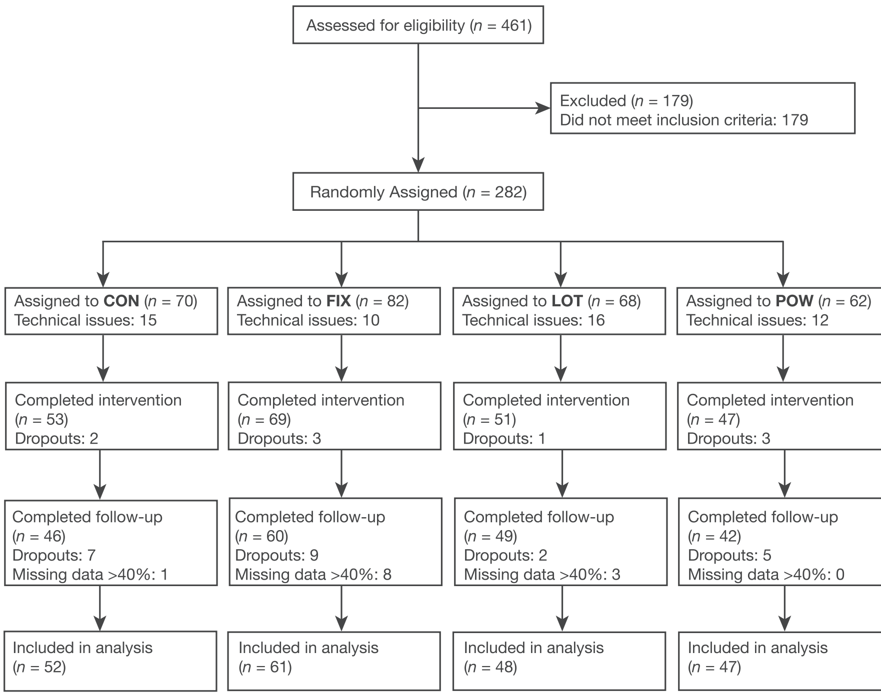

Fig. 2. CONSORT flowchart showing participants progress through the study.

3.2 Participants Characteristics

The participants sample included 220 students,[4] from the University of Lausanne, Switzerland, who were 18 years of age or older (M = 21.6, SD = 2.29), among whom 131 were women (59.5%), with an average IMI level of 4.8 (SD = 1.26)[5], see Table 1. To ensure a mix of technical and non-technical backgrounds, the students came from all seven faculties of the university. Through the enrollment questionnaire, we also measured additional baseline characteristics of the sample, such as the number of fitness activities they performed[6], whether they used a fitness tracker, whether they smoked and, finally, we measured a Socio-Economic Status (SES) indicator[7]. After randomly assigning the students to the experimental conditions, we checked an even distribution of the different characteristics across the conditions.

3.3 Experimental Design and Interventions

We used a between-subject design, with three conditions (fixed incentive condition (FIX), lottery condition (LOT), and motivational message condition (POW)) plus a control condition (CON). Participants were electronically randomly assigned to the control group or to one of the three intervention groups. The final number of participants in each group can be found in Table 1. The imbalance is due to exclusions, as explained in Section 3.1.

[4]Participants sampling is one of the limitations of our study. This is discussed in Section 5.2.
[5]IMI is expressed in a scale from 1 to 7. Low scores indicate a lack of intrinsic motivation.
[6]This is expressed on a scale from 1 to 7 and it was based on the amount of PA practiced by the respondent. Low numbers indicate a lack of fitness activities.
[7]This is a compound metric measured with five standardized items. For a full description of the instrument the reader can refer to Oesch [76].

Table 1. Characteristics of the Study Participants

	Control Group	FIX	LOT	POW	Total
Number of participants n (%)	52 (25)	61 (29.33)	48 (23.08)	47 (22.60)	208 (100)
Women n (%)	28 (13.46)	35 (16.83)	29 (13.94)	32 (15.38)	124 (59.62)
Age (years) mean (sd)	21.60 (2.11)	21.92 (2.35)	21.65 (2.18)	21.28 (2.52)	21.63 (2.29)
IMI[8] mean (sd)	4.62 (1.16)	4.90 (1.15)	5.11 (1.29)	4.79 (1.47)	4.85 (1.27)
Fitness Level mean (sd)	2.53 (0.86)	2.53 (0.71)	2.55 (0.71)	2.36 (0.73)	2.50 (0.75)
SES mean (sd)	2.55 (0.68)	2.51 (0.66)	2.53 (0.75)	2.61 (0.83)	2.55 (0.72)
Fitness Tracker n(%)	4(1.92)	11(5.29)	9(4.32)	8(3.85)	35(15.9)
Smoker n(%)	6(2.88)	10(4.81)	0(0)	9(4.32)	27(12.3)

Similarly to Patel et al. [83], FIX and LOT conditions were designed to provide tangible rewards. Participants of these conditions were assigned money as detailed below. Accrued incentives could be claimed at the end of the experiment phase. Similarly to Kinnafick et al. [60], the POW condition was designed to deliver motivational messages. All participants were required to install a research app on their mobile phone (described in Section 3.6). During the experiment phase, regardless of the condition in which they participated, all participants received a pop-up notification at 8 am on their smart phones, reminding them the goal for the day (see an example in Figure 3(a)). The experimental conditions differed in terms of the additional pop-up notifications that were sent to communicate to participants about the incentives specific to their treatments. Incentives and their related notifications were offered only during the 19 weeks of intervention. However, the monitoring of the PA continued for the entire 30 weeks.

Fixed Incentive Condition (or FIX): Participants in this group were assigned 0.60 CHF or Swiss franc (equivalent to 0.60 USD) every day they reached their daily goal (see Section 3.4). Participants who reached their daily goal received a notification the following day at 9:30 am; this reported the balance of their *digital wallet* (see an example in Figure 3(b)). This incentive was designed taking into account similar incentives in the Swiss market. For instance, Helsana+, a Swiss insurance group, rewards its clients for engaging in physical activities; they give them virtual points[9]. Later these virtual points can be redeemed for service or goods. Similarly, CSS, another insurance group, provides customers with a step-counter app. Customers reaching the daily target receive a reduction of 0.40 CHF on their monthly health insurance premiums[10]. From the SDT standpoint, rewards can detrimentally affect a person's intrinsic motivation. Also, externally administered rewards can be perceived as coercive and controlling hence hinder the basic need of autonomy.

Lottery Condition (or LOT): Participants in this group were assigned a weekly lottery ticket every time they reached their goal of daily steps. At the end of each week, a winning ticket was drawn

[8]IMI reported in the table was measured at week 0 and it is later referred to as IMI_1.
[9]See https://www.helsana.ch/microsite/plus/, last retrieved June 2018. These points have an equivalent value of 0.50 CHF.
[10]See https://www.css.ch/en/home/privatpersonen/kontakt_service/mycss/mystep.html, last retrieved June 2018.

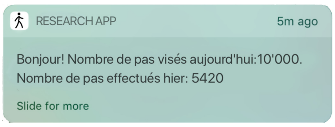

Fig. 3a. Daily Goal Notification Message (Good morning! Target steps for today: 10000. Yesterday you did: 5420)

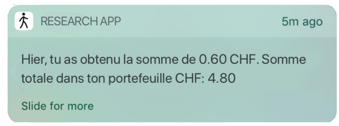

Fig. 3b. Wallet Status Notification Message (Yesterday, you have been assigned 0.60 CHF. Total CHF in your wallet: 4.80)

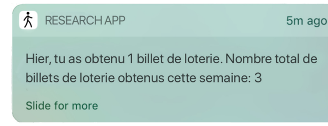

Fig. 3c. Lottery Notification Message (Yesterday, you have been assigned one lottery ticket. Total number of tickets this week: 3)

Fig. 3d. Lottery Price Notification Message (Congratulations, You won the lottery! Total CHF in your wallet: 20)

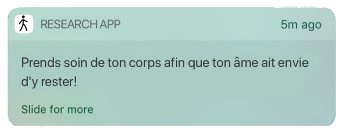

Fig. 3e. Example of Motivational Message Notification (Take care of your body, it's the only place your soul can live in.)

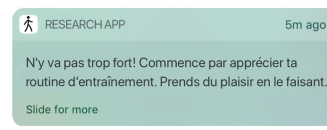

Fig. 3f. Example of Motivational Message Notification (Don't push yourself too hard! Start off by getting a feel for your exercise routine. Make the exercise itself a joy!)

from all the tickets issued that week. The owner of the ticket was credited 10 CHF on their *digital wallet*. Participants who were assigned a lottery ticket received a notification the following day at 10 am; this reported the total number of tickets for the week (see Figure 3(c)). Winners of the lottery were notified each Monday morning at 10:30 am (see Figure 3(d)). This lottery design was tested successfully in prior work [83, 85, 87]. Similarly to the FIX condition above, a lottery ticket can be considered a form of reward that hinders the basic need of autonomy.

Motivational Message Condition (or POW): Every day at 8:30 am, participants in this group received a motivational message pulled sequentially from a pre-arranged corpus (see two examples in Figure 3(e) and 3(f)). These messages were delivered as smart phone pop-up notifications. The method with which these messages were created is described in Section 3.5 below. These messages were delivered to all participants in the condition POW every morning, regardless of whether they had achieved their goal of daily steps. From an SDT perspective, these messages can help the users reflect on the reasons they want to engage in the specific activity, hence have the potential to support the basic need of autonomy.

Control Condition (or CON): Participants in the control group did not receive any other notifications, aside from the daily morning message announcing the goal for the day, described at the beginning of this section (see Figure 3(a)). Participants in this group did not receive either any additional incentives for achieving their goal of daily steps.

3.4 Daily Goal

The main task all participants were asked to complete was to walk 10K steps a day. This target reflects several deliberate design elements. This value has been proven by medical research to provide important health benefits such as lowering blood pressure, increasing exercise capacity and fostering an active life [55, 114]. This value was also used in prior studies aimed at increasing the daily number of steps (see for instance [11]). Furthermore, this value was also set from a pilot study involving a sample of 20 beta-testers, where we found that the average steps testers walked was 6K (with a standard deviation (SD) of ~3K steps), making it reasonable to request that they walk 4K more steps than their average number of steps (this corresponds to an additional

Table 2. The Stages of Change of the TTM with a Short Description (from de Vries et al. [19])

Stage of change	Description
Pre-contemplation (PC)	The individual is not willing to change in the foreseeable future (measured as the next 6 months). Individuals in this stage are mostly uninformed or demoralized.
Contemplation (C)	The individual is willing to change in the next 6 months. Individuals in this stage are aware of some of the pros of a change in behaviour, but are still more inclined to value the cons.
Preparation (P)	The individual is willing to change in the foreseeable future (measured as the next month) and has already taken some small steps towards change (in the past year). Individuals in this stage usually have some plan on how to tackle their lack of activity.
Action (A)	The individual has changed, but for or since no more than 6 months. Individuals in this stage have 'changed', but have not reached the duration that exemplifies a real change in behaviour.
Maintenance (M)	The individual has changed, longer than 6 months. Individuals in this stage have changed and are working not to relapse.

40 minutes of walking at regular walking speed). As FIX and LOT required a threshold to assign prizes to the participants, we set the same target for all conditions (i.e., walking 10K steps/day) in order to avoid possible co-founds that affect only some conditions and not others. Indeed, providing a specific goal to participants is in itself another motivational strategy that was shown to influence performance [63]. Goal-setting theory posits that goals should be, among other qualities, adaptable, challenging but feasible, and self-set by the individuals. In the context of this experiment, it is important to notice that (i) the goal of walking 10K steps a day was designated by the organizers rather than self-set, and (ii) it was not adaptable on the specific competences of each participant; basically it was the same for everybody. Using SDT-specific language, we find that setting a goal the way we did in the experiment did not satisfy the basic needs of autonomy and competence. Hence, we expected a negligible effect on the experimental design. Of course, it must be noted that setting an unobtainable goal might be demotivating. We will come back to this point in the results and discussion sections.

3.5 Motivational Messages Corpus

As we could not identify any publicly available corpus of motivational messages based on SDT, we looked for other sets that could be of use for the experiment.

We started from the set developed by de Vries et al. [19]. They devised a method to generate motivational messages to increase PA by using Mechanical Turk workers. To obtain motivational messages from the Mechanical Turk workers, the authors developed five scenarios, each about a different stage of change described by the *TTM* [88], then they asked ~500 workers to generate six motivational messages for each scenario. The resulting corpus contained ~3,000 messages. The messages of the original corpus were organized according to the stages of change (see Table 2). This model is typically employed in physical training when new activity routines are set up [42].

The first two authors of this article coded each message in the original corpus, according to how motivating they considered the message to be, on a scale from 1 ("Very demotivating") to 5 ("Very motivating") with a 3 as neutral. The two measures were then averaged. We used this information to rank the original messages, according to their effectiveness (as perceived by the first two authors).

Next, we selected messages for our corpus from the messages in the original corpus that were ranked highest in each stage of change. From these, we further excluded messages that were not relevant to walking/running.

We decided to include pre-contemplation messages in the corpus as these could have been relevant to participants who might have elected to participate in the study only for obtaining the economical incentive but that might have not been willing to change their routines around walking/running.

Then we organized the messages in chronological order so that *pre-contemplation* messages were listed at the beginning of the set, *contemplation* messages followed, and so forth. We did so in order that, during the first weeks, the participants would receive encouraging messages to establish a routine and that, later in the experiment, the messages would be aimed at supporting the participant in maintaining their routine. The original corpus also contained another code (orthogonal to the first), describing the processes of change, i.e., 10 processes usually experienced when successfully progressing through the stages of change (for the details, we defer the reader to the description reported in de Vries et al. [19]). When selecting the messages belonging to each stage of change, we ensured that collectively they covered all 10 different processes of change. This procedure provided us a diversified set of messages that followed the different stages of change. Incidentally, in the final set, we ended up selecting the 50 messages used by de Vries et al. in the second part of their study [19, p. 303].

Next, the messages were translated from English to French. The translation was curated by two independent translators who double checked the meaning of each message and the accuracy of the translations. The resulting selection contains 150 motivational messages that ensured full coverage of the experiment phase. A sample of the messages is reported in Table 15, in Appendix A. Before deployment, we considered the differences between the audience used by de Vries et al. to generate the original corpus and our target audience. Mechanical Turk workers were all based in the USA, their average age was 31 years (SD = 9 years), and the large majority had a college degree. Whereas, our participants were slightly younger and were all based in Switzerland. To make sure the selected messages could be used with our audience, we asked two students from the university to look through the selected messages and flag any content that did not make sense to them or that they would otherwise consider inappropriate. All messages in the corpus were validated by these two students.

3.6 Apparatus

There is a growing number of specialized devices that have been developed to sense human activity. These include bracelets[11], clips[12], or wearable sensors[13]. As consumer products, these devices are great, however, as research devices, they suffer serious limitations. For instance, users might forget to consistently wear them, or their batteries can run out of power if not systematically recharged. These points, among others, persuaded us to consider employing smart phones, without any additional hardware, as a tool for tracking steps. Smart phones can accurately predict walking activity [74] and are used to measure physical exercise [48, 62, 83]. They have the following advantages: They act as silent observers, letting participants carry on with their tasks without explicit reminders that their activity is being tracked, thus making the data capturing less intrusive. Furthermore, people are willing to carry their phones with them during the majority of their

[11] See for instance https://www.polar.com/en/products/lifestyle/loop, last retrieved June 2018.
[12] E.g., https://www.fitbit.com/zip, last retrieved June 2018.
[13] E.g., https://www.suunto.com/, last retrieved July 2018.

daily activities and are also aware of the battery level because they want to remain active in their social networks and be reachable via instant messaging and calls. The fact that most of daytime period users carry their phones diminishes the negative factors that usually come with the usage of wearable trackers. Finally, another methodological advantage of using smart phones as sensors is that they do not require participants to use additional hardware, thus increasing the *ecological validity* of the study.

Of the different hardware solutions, we focus on the `iPhone` because `Apple` has standardized both the hardware and the aplication program interface (API) that can be used to collect activity data, thus reducing the development costs required to build the experimental infrastructure for a study. Apple's `HealthKit` is a platform with a repository of PA data collected from the iPhone's internal accelerometer. The data collected is stored in an encrypted database called `Healthkit Store` from which step counts are retrieved[14]. Researchers have made use of this platform in the past to implement step tracking applications [41, 57, 118]. One of the advantages of using this service is that activity tracking always runs in the background and is activated by default on the iPhone (unless the user changes this preference in the phone's settings). As HealthKit also permits steps to be added manually, this could have led some participants to cheat. Fortunately, the HealthKit store enabled us to distinguish between readings coming from the sensor and manually entered steps[15]. Steps that were manually added to the HealthKit store through the iPhone `Health app` or 3rd party apps were ignored by the research app described in this article and not reflected in the daily stepcount.

To implement the different experimental conditions, we developed an app that ran on `iPhone 5s` or a later version, with an OS 8.0 or later version. The app provided three views that were accessed through the tab bar at the bottom of the screen (see Figure 4(a)). Right after installation, the app presented two requests for permissions from the user to ensure that it could display pop-up notifications (the main communication channel for receiving information about the incentives accrued in the study) and that it could retrieve step data from the Healthkit Stores. The main view of the app provided information on the user's current number of steps for the current day (see Figure 4(a)). When the daily goal was reached, the number of steps became bold and its color changed to blue (see Figure 4(b)). As HealthKit does not constantly refresh the step counts, we could not instantly update the number on our research app. To avoid misunderstandings with participants, the app indicated at the bottom of the screen the last date and time when the steps were refreshed. The second view displayed a contact e-mail to communicate with the research team, and the third view showed the version and copyright information about the app. We intentionally chose a minimalist design to keep the UI elements similar across experimental conditions, thus controlling for confounding effects of these elements. The purpose of the research app was to deliver the notifications and retrieve the activity logs from the participants' phones. Daily step counts were synced with a secure database on servers at our institution. Once the app was installed, participants were not required to reopen the app, although they could as often as they wished to check their daily progress. We also implemented a dashboard, as part of the overall architecture, through which we could monitor whether specific participants turned off notifications or stopped sending data; in this case, we could preemptively attempt to solve the emerging technical issues.

Before deploying the final version of the app, we ran a month-long pilot involving 20 beta testers. This study was useful for resolving bugs, for testing the delivery of the notifications and the accuracy of the step-count measurements.

[14]HealthKit also collects data from a variety of other sensors. However, in the context of this article we focus only on step count measurements.

[15]For the technical reader, entries were filtered using the `NSPredicate` logical condition.

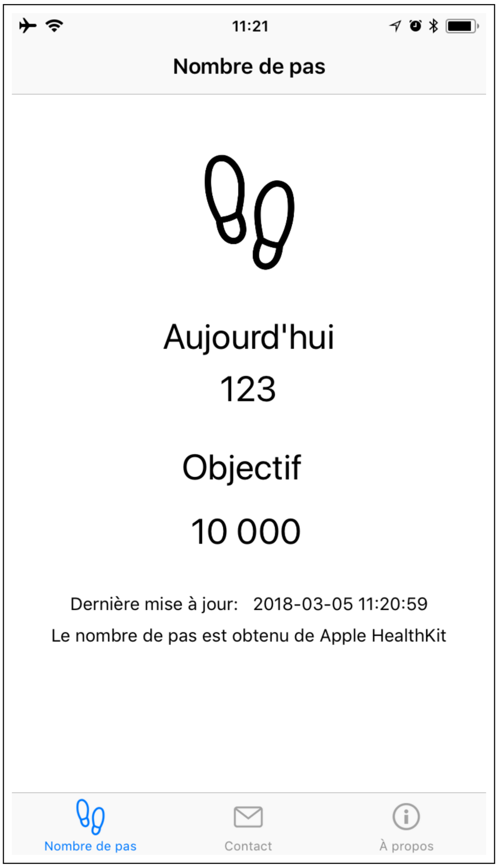

Fig. 4a. Home screen of the Research App (Number of steps; Today 123; Objective 10,000; Last update; The number of steps is obtained from Apple HealthKit).

Fig. 4b. Home screen when reaching the daily goal of 10K steps (Number of steps; Today 10,329; Objective 10,000; Last update; The number of steps is obtained from Apple HealthKit).

3.7 Procedure

All selected participants received an invitation e-mail, together with an informed consent. The consent form was adapted for the experimental condition the participant was assigned to. This document specified the conditions for participation and for winning a reward (i.e., simply making 10K steps for participants in FIX or winning the lottery for LOT), if any. All participants were requested to download from Apple App Store the research app described in the previous section. They were also asked to carry their mobile phone on them during their daily activities (for example in a arm band, pocket, or belt clip).

The study was designed to be run across various moments of the academic year (i.e., fall semester, winter holidays, and spring semester), which allowed various walking behavior observations. It began in early September 2017 (week 15) and finished in late March 2018 (week 44). Toward the end of the experiment phase (week 26), participants were asked to fill an intermediate questionnaire that presented the same standardized inventory used in the screener. The compound score was labeled IMI_2. The same instrument was again deployed at the end of the study (week 44) through which we collected the IMI_3 scores. At the onset of the experiment phase, the research app extracted and recorded the step count from 1st of June 2017 (week 1). The data of week 1–14 served as a baseline and is presented in the remainder of this article as the pre-experiment phase.

All participants were paid 10 CHF for participating in the experiment, specifically for installing and maintaining the research app on their phones for the duration of the study. In addition to this incentive, the participants who filled out the questionnaires through which we collected intrinsic motivation measurements were given the opportunity to participate in a lottery draw, with the possibility of winning a prize of 100 CHF (~100 USD). They were allowed to withdraw from the study at any point in time and without any consequences. The participants were identified through an ID, and their identity and contact information were separated from the steps data collected through the experiment. The University of Lausanne, HEC Institutional Review Board approved the study.

3.8 Statistical Analysis

Several analyses were performed. Their main characteristics are highlighted below. Data were analyzed with generalized mixed-effect regression models that are now widely used for the analysis of longitudinal data (see for example Fitzmaurice et al. [36]). For each outcome and model presented below, the results, reported in Section 4, include the unadjusted effects, means without controlling for the model factors, and the adjusted ones, controlling for the model factors.

3.8.1 The Outcomes. The first primary outcome was the participant IMI index measured at the onset of the study (IMI_1, weeks 0), in the experiment phase (IMI_2; weeks 12), and post-phase (IMI_3; weeks 30). The model for IMI index is presented in Section 3.8.3 below.

The second primary outcome was the average step count per day over each week during the experiment phase and post-phase. The measurements during the pre-phase were considered as secondary. The model for step counts is presented in Section 3.8.4 below.

Another secondary outcome is the binary indicator of whether a given participant in a given day reached the goal of 10K steps (we will refer to this as "participant-day"). This enables us to analyze the influence of the IMI score and conditions on the proportion of participant-day, where they reach the predefined goal, and to see if it is in line with the results for the step counts. The model for the goal achievement is presented in Section 3.8.5 below.

3.8.2 Missing Values. For each participant on each day (or participant-day), the observed daily number of steps could be missing if the participant turned off their smart phone for the entire day, left the smart phone at home, or if the application did not grant the permission to read the steps from HealthKit. In addition, the number of steps per participant-day that were larger than 54K or lower than 600 were considered unrealistic and replaced by a missing value code (i.e., `NA`). Although the upper-bound corresponded to running a marathon and rarely occurred (six times, for five people), the lower-bound was caused by a number of unspecified factors (e.g., placing the phone in a purse or overload of the CPU of the phone).

The measurement errors of PA by the accelerometers embedded in the mobile phones was also identified by prior research [84]; it demonstrated that, over the course of a several months-long study, these measurements errors have negligible effects.

Therefore, we conducted an analysis of the missing data by comparing each condition in each phase of the experiment. However, we did not conduct any statistical test on missing data because any test based on such a large number of observations would result significant, and this could mislead the reader. We thus conducted the analysis by simply computing the proportions. The results are reported in Table 3. In the pre-phase, we see that the proportion of missing participant-days is similar between groups CON, FIX, and LOT; whereas, in POW, this proportion is slightly lower. In the experiment phase, the proportions are overall similar between conditions and smaller

Table 3. Analysis of the Missing Data in Each Condition and Phase of the Experiment

	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Condition	SMP	NAs	Prop	SMP	NAs	Prop	SMP	NAs	Prop
CON	5044	910	0.18	6656	376	0.06	3848	710	0.18
FIX	5917	990	0.17	7808	506	0.06	4514	1206	0.27
LOT	4656	844	0.18	6144	411	0.07	3552	851	0.24
POW	4559	628	0.14	6016	431	0.07	3478	1123	0.32

SMP represents the number of observations. These numbers are computed from the 208 participants included in the full analysis reported in this article (cf. Section 3.1).

than those in the pre-phase. In the post-phase, the proportions increased to larger values, especially in the group POW. The differences between the pre-phase and the experiment phase are expected, as several participants had intentionally disabled the automatic capturing of their steps, hence creating more missing data (similarly across all conditions). Similarly, the differences between the experiment and the post-phase are due to the participants who dropped out from the experiment, which resulted slightly more frequent in the POW condition.

3.8.3 *IMI Model.* The adjusting factors are the participant effect (random intercept), the condition (categorical: CON, FIX, POW, and LOT), the phase (categorical: experiment and post-experiment), and their interaction. The research hypotheses are assessed using multiple comparisons between conditions at each phase, and between phases at each condition.

3.8.4 *Step Counts Model.* Because the trend in time (week effect) is nonlinear, the step counts were analyzed in two different stages. First, a smoothing spline was fitted to estimate the weekly average step counts. The smoothing parameter was selected by cross-validation. Second, the spline was subtracted from the average step counts before the regression analysis was run. The adjusting factors are the participant effect (random intercept), the IMI_1 level (numerical), the condition (categorical: CON, FIX, POW, and LOT), and the interaction between IMI_1 and the condition. The research hypotheses were assessed using multiple comparisons between conditions at several IMI_1 values (i.e., 2, 3.5, 4.5, 6). Also, the IMI_1 effect (the slope) was assessed at each condition.

3.8.5 *Goal Model.* A generalized linear model with binomial family was fitted. The adjusting factors were the same as for the step-count model. The estimates are given in *odds ratios* (for the unfamiliar reader, see for example Szumilas [109]).

3.8.6 *Statistical Software.* All the analyses were performed using the `R` packages `lme4` and `lmerTest` [8]. The mixed-effect models were fitted using the `REML`. All the multiple comparisons were obtained using the R function `lsmeans` [61]. Dunnett's correction was applied for comparison to the reference level. The smoothing spline was fitted using the `R` function `smooth.spline`.

3.8.7 *Limitations of the Statistical Analysis.* For the step counts, separating the analysis into two stages can bring bias in the results because the adjusted analysis does not incorporate the uncertainty of the smoothing spline estimates. We chose this for clarity of exposure and assumed that this bias was small. The results were checked using a generalized additive mixed model (R function `gamm`) [117]) that can incorporate both analyses at once, but whose presentation of results is more complex.

For the adjusted analysis, using the weekly averages of the daily counts, instead of the daily counts directly, can bias the analysis in the sense that the number of observations (i.e., full data) for building the averages are not taken into account. This choice was made for ease of presentation.

Results were checked on the full data, using the day of the week (Monday through Sunday) as an additional factor. The models lead to the same conclusions, but some cases could not be treated because of numerical limits (non-convergence).

For the mixed-effect models, all the inferential results (p-values and confidence intervals) assumed the normality of the residuals. In addition, missing at random is assumed.

3.8.8 Power Analysis. Similarly to [83], we made an approximate power analysis for the step counts, using a two-sided paired t-test based calculation with a Bonferroni adjustment of the α level to 0.017 (CON versus FIX, POW, and LOT). With a power of 80% and the participant design in each condition (see Table 1), an effect size of 0.60 can be detected. With an estimated pooled within-group SD of 3,306, a true contrast of 1,984 steps between two conditions can be detected.

For the IMI analysis, the Bonferroni adjustment implied an α level of 0.025 (IMI_1 versus IMI_2 and IMI_3). With a power of 80% and 112 responses (see Table 6), an effect size of 0.295 can be detected. With an estimated pooled within-group SD of 1.23, a true contrast in IMI of 0.36 can be detected.

3.9 Physical Activity and Outdoor Weather Conditions

It is reasonable to consider the weather conditions as a predictor for the PA of the participants. Indeed, we attempted to use them in the models to explain the step counts and the daily-goal achievements. In particular, we considered the daily mean temperature (in Celsius)[16]. We retrieved the data from the Federal Office of Meteorology and Climatology in Switzerland[17].

Although the data carries some predictive power, this was not useful to estimate the differences in physical activities between groups in the experiment. This can be explained by the following two observations: (1) all the participants are exposed to the same weather conditions; (2) the weather conditions are highly collinear with the time trend component of the LMER analysis. In order to keep the models simple, we did not include weather conditions in the models described above.

4 RESULTS

The participants' baseline characteristic means and SDs were similar across the four experimental groups, which implies that the groups were well balanced after randomization and exclusions (Table 1).

During the pre-experiment phase, the average daily step-count for all participants was 7299.2 steps (95% CI 7215.6 to 7382.8). During the the experiment phase, the same average was 7060.2 steps (95% CI 7003.8 to 7116.5). Finally, in the post-experiment phase, this average was 6604.3 steps (95% CI 6521.7 to 6686.9). See Tables 4 for the details. Figure 5 shows the average daily step-counts per week and the non-linear trend fitted as described in Section 3.8.4. The figure shows a global downward trend during the experiment phase with a drop around weeks 30–32 (end of December 2017, beginning of January 2018). The overall trend can be attributed to the fact that the experiment was held during different times of the academic year. The pre-phase began during the exams of the spring semester, then it overlapped the vacation time for most students. The experiment started with the beginning of the fall semester and culminated at Christmas and the winter-exam period. We observed that students typically tend to exercise less during the Christmas break and around the exam periods. During the experiment, the mean proportion of days that the participants achieved the 10K steps goal was 0.204 (95% CI 0.199 to 0.209), see details in Table 5. These statistics confirm that the 10K-step goal was indeed a reasonable target that could be attained by the majority of participants in the sample, with roughly 30 additional minutes of walking/running

[16]We also considered the total daily precipitations (in millimeters); however, as the set contained mostly zeroes we decided to exclude these from the analysis.
[17]See https://gate.meteoswiss.ch/idaweb, last retrieved March 2019.

Table 4. Unadjusted Daily Average Number of Steps

	Control Group	FIX	LOT	POW	Total
Pre-					
mean	7383.3	6916.1	7193.8	7793.1	7299.2
sd	5592.8	5053.8	5517.6	5996.2	5531.7
95% CI	(7212.8-7553.8)	(6775.0-7057.2)	(7018.6-7369.0)	(7605.6-7980.5)	(7215.6-7382.8)
Experiment					
mean	7392.8	6707.8	7055.7	7151.5	7060.2
sd	4731.2	4110.0	4759.0	4575.2	4535.7
95% CI	(7275.8-7509.8)	(6613.5-6802.1)	(6932.5-7178.9)	(7031.5-7271.5)	(7003.8-7116.5)
Post-					
mean	6696.3	6317.1	6602.3	6887.4	6604.3
sd	4617.8	4340.4	4582.1	4535.6	4518.3
95% CI	(6534.8-6857.9)	(6169.2-6465.1)	(6429.5-6775.1)	(6704.1-7070.6)	(6521.7-6686.9)

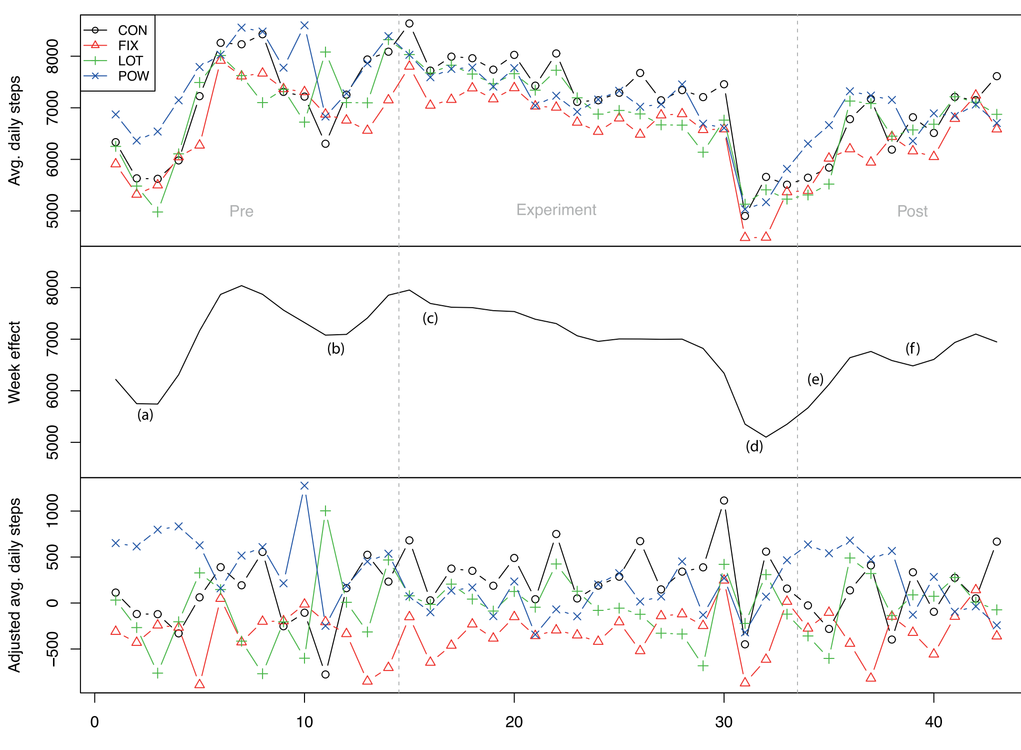

Fig. 5. Average daily step-counts per week. Top: unadjusted numbers. Middle: weekly non-linear trend. The minimas correspond to specific periods of the academic calendar: (a) summer exam session; (b) summer retake session; (c) beginning of the fall courses; (d) end of courses and holidays season; (e) winter exam session; and (f) beginning of the spring courses. Bottom: numbers after adjustment for the weekly trend.

Table 5. Unadjusted Proportions of Participant-days Reaching 10,000 Steps

	Control Group	FIX	LOT	POW	Total
Pre-					
mean	0.231	0.206	0.234	0.252	0.231
sd	0.425	0.405	0.424	0.434	0.421
95% CI	(0.223-0.249)	(0.195-0.218)	(0.221-0.248)	(0.238-0.265)	(0.224-0.237)
Experiment					
mean	0.231	0.189	0.205	0.194	0.204
sd	0.422	0.391	0.403	0.395	0.403
95% CI	(0.221-0.242)	(0.180-0.198)	(0.194-0.215)	(0.184-0.204)	(0.199-0.209)
Post-					
mean	0.171	0.155	0.173	0.184	0.170
sd	0.377	0.362	0.378	0.388	0.375
95% CI	(0.158-0.184)	(0.143-0.168)	(0.158-0.187)	(0.169-0.200)	(0.163-0.176)

Table 6. Unadjusted IMI Measured at Week 0 (IMI_1), Week 12 (IMI_2), and Week 30 (IMI_3)

	Control Group	FIX	LOT	POW	Total
IMI_1					
n	52	61	48	47	208
mean (sd)	4.62 (1.16)	4.90 (1.15)	5.11 (1.29)	4.79 (1.47)	4.85 (1.27)
95% CI	(4.30-4.94)	(4.61-5.19)	(4.75-5.48)	(4.36-5.21)	(4.68-5.03)
IMI_2					
n	46	55	41	43	185
mean (sd)	4.96 (1.16)	4.57 (1.15)	4.92 (1.22)	4.48 (1.40)	4.73 (1.24)
95% CI	(4.62-5.30)	(4.26-4.88)	(4.55-5.30)	(4.06-4.90)	(4.55-4.90)
IMI_3					
n	35	28	20	18	101
mean (sd)	5.13 (1.41)	4.89 (0.99)	4.64 (1.18)	5.15 (1.32)	4.97 (1.24)
95% CI	(4.65-5.60)	(4.52-5.27)	(4.11-5.17)	(4.52-5.78)	(4.73-5.21)

per day[18]. Below, we assess our hypotheses by using the adjusted estimates from the models and inferential results. All the statistical significances are meant with an alpha level of 0.05.

4.1 H1. Participants with a high IMI_1 score will make more steps than participants with a low IMI_1 score

The results of the step counts model described in Section 3.8.4 are reported in Table 7. The model shows a linear increase in the step counts with IMI_1 in each phase for the different experimental conditions: for the control group, in the pre-experiment phase step-counts increased of 1129.7 steps for each unit of IMI_1 (95% CI, 620.3 to 1639.0; $p < .01$), in the experiment phase, step counts increased by 798.1 steps for each unit of IMI_1 (95% CI, 304.0 to 1292.2; $p < .01$), and in the post-experimental phase step counts increased by 1118.8 steps for each unit of IMI_1 (95% CI, 620.7 to 1616.9; $p < .01$).

During the pre-phase, the linear increase described for the control condition was also found in the other conditions, which indicates that the IMI_1 score predicts the amount of PA a participant

[18]See http://www.wellocracy.com/2013/10/how-many-steps/, last retrieved March 2019.

Table 7. Adjusted Increase in Daily Step Counts for one IMI_1 Level by Experimental Condition and Measurement Phase

	Increase in daily step count for 1 point change in the IMI score								
	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Condition	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% C.I.	Adj. P
CON	1129.7	(620.3 ; 1639.0)	**<0.01**	798.1	(304.0 ; 1292.2)	**<0.01**	1118.8	(620.7 ; 1616.9)	**<0.01**
FIX	497.1	(39.1 ; 955.0)	**0.03**	437.9	(−18.6 ; 894.5)	0.06	445.6	(−22.6 ; 913.9)	0.06
LOT	466.0	(3.48 ; 935.4)	**0.05**	265.9	(−196.9 ; 728.7)	0.26	436.8	(−38.3 ; 911.9)	0.07
POW	502.7	(93.0 ; 912.4)	**0.02**	372.2	(−36.9 ; 781.2)	0.07	571.8	(126.4 ; 1017.2)	**0.01**

Table 8. Adjusted Slope of the IMI_1 in the Goal Model by Experimental Condition and Measurement Phase. The unit is the odds ratios

	Increase in odds ratio for 1 point change in the IMI score								
	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Condition	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% C.I.	Adj. P
CON	1.73	(1.37 ; 2.17)	**<0.01**	1.64	(1.27 ; 2.12)	**<0.01**	1.78	(1.36 ; 2.33)	**<0.01**
FIX	1.27	(1.04 ; 1.55)	**0.02**	1.22	(0.97 ; 1.54)	0.10	1.26	(0.99 ; 1.61)	0.06
LOT	1.22	(0.99 ; 1.49)	0.06	1.17	(0.92 ; 1.48)	0.20	1.32	(1.02 ; 1.70)	**0.03**
POW	1.24	(1.04 ; 1.48)	**0.02**	1.22	(0.98 ; 1.50)	0.07	1.36	(1.08 ; 1.73)	**0.01**

will do. In the experimental phase, this linear increase remained mildly positive, whereas, for the control group, this remained significant, for the other conditions this was not statistically significant[19]. Similar results were found in the post-experiment phase. Although for FIX and LOT, the linear increase of the step counts with IMI_1 remained non-significant, we found a significant relationship in the POW condition where the step count increased by 571.8 steps for each unit of IMI_1 (95% CI, 126.4 to 1017.2; $p < .01$).

If we look at the goal model described in Section 3.8.5, we find the same results described above (reported in Table 8). Therefore, H1 is *partially validated*: Although the relation between steps and IMI_1 score was significant throughout the different phases of the experiment for the control group, the same relation was significant for the other conditions in the pre-phase, then it became non-significant during the experiment and, after the experiment, returned to being significant but only for the POW condition. This finding reveals that the association between the IMI_1 score and the amount of steps was stable in the control group, whereas it was perturbed by the intervention. Therefore, we looked more specifically at the influence of the experimental manipulation on the intrinsic motivation with H2.

4.2 H2a. During the experimental phase, the IMI score of participants in FIX and LOT will decrease

Adjusted differences in the IMI scores recorded at week 26 and at week 44, against the reference measured at week 14, are displayed in Table 9. IMI scores of the participants in the control group increased by 0.44 IMI points at the end of the post-phase, when compared to the scores registered at the beginning of the experiment (95% CI, .03 to .85, $p = .03$). Observing the trends on the IMI scores, we note that, while these increased during the experiment and post-experiment phase for participants in the control condition, they decreased for participants in the FIX and LOT conditions.

[19]In the remainder of this article, we will use term *significant* to mean *statistically significant.*

Table 9. Adjusted Contrasts Between IMI Levels in Each Condition

	Change of IMI score during the Experiment and Post– phases					
	IMI_2 - IMI_1			IMI_3 - IMI_1		
Condition	Estimate	95% CI	Adjusted P	Estimate	95% C.I.	Adjusted P
CON	0.30	(−0.06 ; 0.67)	0.12	0.44	(0.03 ; 0.85)	**0.03**
FIX	−0.33	(−0.67 ; 0.01)	0.06	−0.01	(−0.45 ; 0.44)	>0.99
LOT	−0.26	(−0.65 ; 0.13)	0.24	−0.37	(−0.90 ; 0.15)	0.20
POW	−0.29	(−0.67 ; 0.10)	0.18	0.12	(−0.43 ; 0.66)	0.84

Measurements at week 26 (IMI_2) and 44 (IMI_3), againstthe reference measured at week 14 (IMI_1).

Table 10. Difference in Differences Analysis for H2. Fixed Effect Coefficients of the Regression of the IMI in Function of the Phase and of the Condition

	Estimate	Std. Error	df	t value	Pr(>|t|)	CI (2.5%)	CI (97.5%)
Intercept	4.74	0.15	398.05	31.69	0.00	4.45	5.03
condition							
FIX-CON	0.18	0.21	398.05	0.88	0.38	−0.22	0.59
LOT-CON	0.39	0.21	398.05	1.84	0.07	−0.02	0.81
POW-CON	0.11	0.22	398.05	0.52	0.60	−0.31	0.54
phase							
IMI_2-IMI_1	0.30	0.16	319.58	1.85	0.07	−0.02	0.62
IMI_3-IMI_1	0.44	0.18	326.78	2.40	**0.02**	0.08	0.80
interactions							
(FIX-CON):(IMI_2-IMI_1)	−0.63	0.22	316.24	−2.83	**0.01**	−1.07	−0.20
(LOT-CON):(IMI_2-IMI_1)	−0.56	0.24	323.23	−2.36	**0.02**	−1.03	−0.10
(POW-CON):(IMI_2-IMI_1)	−0.59	0.24	320.02	−2.47	**0.01**	−1.05	−0.13
(FIX-CON):(IMI_3-IMI_1)	−0.45	0.27	328.41	−1.65	0.10	−0.98	0.08
(LOT-CON):(IMI_3-IMI_1)	−0.82	0.30	334.74	−2.74	**0.01**	−1.39	−0.24
(POW-CON):(IMI_3-IMI_1)	−0.33	0.31	334.22	−1.07	0.29	−0.92	0.27

Reference levels are CON group and IMI_1, respectively. The subject effect is controlled with a random intercept (not shown below).

To mitigate the effects of extraneous factors (e.g., seasonal effects) or selection bias, we conducted a difference-in-difference (DID) analysis. DID is a common analysis technique used for example in Abadie [1] or Blundell et al. [9].

We compared the average change over time in the IMI score in FIX and LOT groups, compared to the average change over time of the IMI score for the CON group. This analysis is valid under the assumption that individuals in any treatment group (FIX, LOT, or POW) would behave the same way as participants in the control group. In particular, we assume that in average the IMI scores in any treatment group would change over time by the same amounts of the IMI scores of the control group. Though we do not have data to test it, this is a reasonable assumption to make as we do not have reasons to believe that extraneous factors could have unevenly affected the individuals in different experimental conditions. Therefore, the estimates of the DID study are obtained from the interaction terms in the underlying regression used previously in the within group study. These coefficients and their p-values are reported in Table 10[20]. The difference of IMI scores between the FIX and CON groups decreased by 0.63 points between the beginning and

[20] Scripts for reproducing our analyses, as well as full R outputs are available on https://osf.io/m28xe/.

the end of the experiment[21] (95% CI, −1.07 to −0.20, $p < .01$). Similarly, that difference between the LOT and CON groups decreased by 0.56 points between the beginning and the end of the experiment (95% CI, −1.03 to −0.10, $p = .02$). When looking at the differences between the pre- and post-experiment phases, the decrease of the IMI scores between the FIX and CON groups is no longer significant (95% CI, −0.98 to −0.08, $p = .10$, ns), while that decrease of the difference between the LOT and CON group, estimated to −0.82, is still significant (95% CI, −1.39 to −0.24, $p < .01$). In summary, these results *enable us to validate* H2a. The negative trends on the IMI scores in FIX and LOT are indicative of a *detrimental effect* of the tangible rewards over the intrinsic motivation of the participants. Next, we looked at the effect of the motivational messages on the IMI scores.

4.3 H2b. During the Experimental Phase, the IMI Score of Participants in POW will Increase

For this analysis we refer to the same DID study reported in Section 4.2 and in Table 10 above. Contrary to our expectations, the difference in the IMI scores between the POW and CON groups decreased by 0.59 points between the pre- and experiment phases (95% CI, −1.05 to −0.13, $p < .01$). That decrease between the pre- and post-experiment phase is no longer significant (95% CI, −0.92 to −0.27, $p = .29$, ns). These *results do not enable us to validate* H2b. Not only we do not see the expected increase of the IMI scores, but we observe a negative trend similar to those obtained in the FIX and LOT intervention. These negative trends are indicative of a detrimental effect of the motivational messages, which we did not anticipate. The motivational messages sent in POW did not increase the intrinsic motivation of the participants. Instead, these messages had a negative effect on the IMI scores. Therefore, we checked the step counts to see whether these trends produced visible effects on the overall PA.

4.4 H3a. During the Experimental and Post-experimental Phase, Participants in FIX and LOT Walked Less than Participants in the Control Condition

Adjusted differences in the mean daily step-counts are displayed in Table 11. The analysis of the goal study is reported in Table 12 and confirms the results of the steps study[22]. The results reported in these two tables were obtained using models where IMI_1 is a numerical factor. This accounts for the effect of IMI_1 on the number of steps for Table 11 and on the probability of reaching their goal for Table 12. As this effect varies from one group to another, an interaction between IMI_1 and condition was included. Hence, to compare the conditions, because of the presence of interactions, we fixed the value of IMI_1 (because the difference between two conditions varies with the value of IMI_1). We then selected $IMI_1 = 2, 3.5, 4.5$, and 6 in an ad-hoc way, to globally cover the range of possible IMI_1 values.

The experimental conditions were not significantly different at the various IMI_1 levels. We now turn to the description of trends estimated by the model. These results should be taken with caution, as the lack of significance does not allow us to imply causal relations between the variables and should be cross-validated in future research. During the experiment, for FIX we observe a

[21] The unfamiliar reader might wonder whether a difference of half-point in the IMI score is clinically significant. In a similar study, Tsigilis [113] found that the IMI score significantly predicted the performance obtained on a 20 m shuttle field test. They found that a difference of 0.50 in the IMI score resulted in the variation of the VO_2 max, or maximal oxygen consumption, between 30.5 and 33.5 (mL/Kg · minute) for a population of students 19 years old, which results in a difference between a poor and fair fitness level [45].

[22] As an odds ratio below 1 indicates a decrease in the proportion, we see that these trends follow the results of the step counts.

Table 11. Adjusted Contrasts of the Step Counts Between Conditions for Several IMI_1 Levels

	Change in daily step count for fixed IMI_1 levels								
	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
IMI_1 = 2.0									
FIX - CON	1167.3	(−1284.3 ; 3618.9)	0.53	170.8	(−2243.7 ; 2585.3)	0.99	1299.3	(−1162.4 ; 3760.9)	0.45
LOT - CON	1393.5	(−1191.6 ; 3978.6)	0.43	909.2	(−1621.1 ; 3439.4)	0.71	1446.9	(−1117.2 ; 4010.9)	0.40
POW - CON	2131.9	(−211.8 ; 4475.6)	0.08	767.6	(−1532.1 ; 3067.4)	0.75	1786.8	(−602.9 ; 4176.5)	0.19
IMI_1 = 3.5									
FIX - CON	218.5	(−1176.1 ; 1613.0)	0.94	−369.5	(−1746.8 ; 1007.8)	0.83	289.6	(−1122.4 ; 1701.5)	0.90
LOT - CON	398.0	(−1120.4 ; 1916.4)	0.84	110.8	(−1378.7 ; 1600.4)	0.99	423.9	(−1085.1 ; 1932.8)	0.81
POW - CON	1191.5	(−180.4 ; 2563.4)	0.11	128.7	(−1221.8 ; 1479.1)	0.98	966.4	(−441.4 ; 2374.2)	0.25
IMI_1 = 4.5									
FIX - CON	−414.1	(−1383.9 ; 555.6)	0.60	−729.7	(−1686.2 ; 226.8)	0.18	−383.5	(−1374.9 ; 607.8)	0.66
LOT - CON	−265.7	(−1329.6 ; 798.3)	0.85	−421.4	(−1465.6 ; 622.8)	0.64	−258.1	(−1323.7 ; 807.5)	0.86
POW - CON	564.5	(−449.6 ; 1578.7)	0.41	−297.3	(−1299.5 ; 704.9)	0.80	419.4	(−638.8 ; 1477.6)	0.65
IMI_1 = 6.0									
FIX - CON	−1363.0	(−2774.7 ; 49.8)	0.06	−1270.0	(−2648.8 ; 108.7)	0.08	−1393.2	(−2813.3 ; 26.8)	0.06
LOT - CON	−1261.2	(−2671.5 ; 149.2)	0.09	−1219.7	(−2593.8 ; 154.3)	0.10	−1281.1	(−2698.3 ; 136.1)	0.09
POW - CON	−375.9	(−1817.6 ; 1065.9)	0.84	−936.2	(−2354.4 ; 482.0)	0.28	−401.0	(−1905.1 ; 1103.0)	0.83

Table 12. Adjusted Contrasts Between Conditions in the Goal Model, for Several IMI_1 Levels

	Change in odds ratio for fixed IMI_1 levels								
	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
IMI_1 = 2.0									
FIX - CON	1.95	(0.66 ; 5.80)	0.33	1.80	(0.52 ; 6.30)	0.54	2.40	(0.64 ; 9.08)	0.28
LOT - CON	2.40	(0.76 ; 7.57)	0.18	1.96	(0.53 ; 7.32)	0.48	2.01	(0.50 ; 8.10)	0.49
POW - CON	2.87	(1.01 ; 8.13)	**0.05**	1.73	(0.52 ; 5.73)	0.56	2.62	(0.72 ; 9.55)	0.20
IMI_1 = 3.5									
FIX - CON	1.23	(0.66 ; 2.29)	0.74	1.15	(0.56 ; 2.36)	0.91	1.43	(0.66 ; 3.08)	0.54
LOT - CON	1.42	(0.72 ; 2.78)	0.47	1.18	(0.54 ; 2.56)	0.90	1.28	(0.56 ; 2.91)	0.79
POW - CON	1.74	(0.95 ; 3.19)	0.08	1.10	(0.54 ; 2.22)	0.96	1.76	(0.82 ; 3.76)	0.20
IMI_1 = 4.5									
FIX - CON	0.91	(0.59 ; 1.38)	0.87	0.86	(0.52 ; 1.40)	0.78	1.01	(0.60 ; 1.72)	> 0.99
LOT - CON	1.00	(0.63 ; 1.59)	> 0.99	0.84	(0.49 ; 1.44)	0.75	0.95	(0.54 ; 1.68)	0.98
POW - CON	1.25	(0.80 ; 1.93)	0.49	0.81	(0.48 ; 1.37)	0.65	1.35	(0.77 ; 2.35)	0.45
IMI_1 = 6.0									
FIX - CON	0.57	(0.31 ; 1.04)	0.08	0.55	(0.27 ; 1.11)	0.12	0.60	(0.29 ; 1.25)	0.25
LOT - CON	0.59	(0.32 ; 1.08)	0.10	0.50	(0.25 ; 1.02)	0.06	0.60	(0.29 ; 1.25)	0.25
POW - CON	0.76	(0.41 ; 1.40)	0.56	0.52	(0.25 ; 1.07)	0.09	0.90	(0.42 ; 1.96)	0.96

The unit is the odds ratios.

negative trend in the step counts for $IMI_1 \geq 3.5$, and for LOT we observe a similar negative trend for $IMI_1 \geq 4.5$. However, these differences were visible also in the pre-experiment phase, hence we were not able to derive any conclusion on the effects of the experimental manipulation. Therefore, these *results do not enable us to validate* H3a: The tangible rewards did not change the amount of steps of the participants during and after the experiment.

Finally, we look at whether motivational messages have an effect on the main outcome.

4.5 H3b. During the Experiment and Post-experiment Phases, the Number of Steps of the Participants in POW will be Higher than Those of Participants in the Other Conditions

We find that all the differences between the POW group and the control group, in terms of daily step counts, are not significant (see Table 11). Similar results can be observed in the goal study, reported in Table 12. Therefore, we turned to the description of trends, as per H3a above. During the experiment, we observed a negative trend for participants with $IMI_1 \geq 4.5$. However, these trends are not indicative of any effect on the PA of the participants during and after the experiment, as the same trends were present in the Pre-phase.

Therefore, these *results do not enable us to validate* H3b, as they contradict the effect we expected for motivational messages.

Instead of encouraging participants to walk more, motivational messages had no effect on the number of steps of the participants.

We then wondered whether the discrepancies we observed between our experimental results and the effects predicted by the theory could have been artifacts of our experimental design.

Therefore, we look at the following two aspects of our experiment that could have influenced the results: (1) the choice of 10K steps with regard to the levels of PA before the beginning the experiment; and (2) the incentive mechanism of the experimental conditions with regard to the goal of reaching the 10K steps. Finally, given the results obtained in the IMI_1 study, reported in Section 4.2 and 4.3 and Table 9, we conduct an additional analysis to evaluate the effect of the experiment on the *most* and *least* motivated people in the groups of participants.

4.6 Additional Subgroup Analysis for the Steps Study

In addition to testing the hypotheses derived from the theory, we explored whether the interventions worked for participants for whom the 10K steps was a reasonable challenge better than for those for whom this was a target too difficult to achieve. Therefore, we performed a post-hoc split of participants, based on their level of PA as recorded in the 3 weeks that preceded the beginning of the experiment (i.e., week 12 to 14). We chose this particular time window to limit the seasonal effects of summer during the baseline. We then organized the participants in three groups, based on the following definitions:

- **Challenging**: those for whom the target of 10K steps was above the mean plus 1.5 times the SD of the steps calculated on 3 weeks before the beginning of the experiment.
- **Feasible**: those for whom the target fell between the mean and the mean plus the SD.
- **Achieved**: those for whom the target was below the mean.

We then computed the frequencies in each PA activity group by experimental condition. The results are available in Table 16 in Appendix B. The results are reported in Table 17 in Appendix B. We did not observe any significant difference between the experimental groups and the control group during the experiment and post-phase. We now look at the analysis of incentive mechanism, with regard to the the goal of reaching 10K steps.

Table 13. Adjusted Contrasts Between Odds Ratio of Exceeding 11K Steps Among the Day-participants Exceeding 10K Steps

	Odds ratio of exceeding 11K steps once the daily goal has been achieved								
	Pre-phase (wk 1–14)			Experiment (wk 15–33)			Post-phase (wk 34–44)		
Contrast	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
CON - FIX	1.16	(0.88 ; 1.52)	0.42	0.89	(0.72 ; 1.08)	0.35	1.01	(0.71 ; 1.42	1
LOT - FIX	0.95	(0.71 ; 1.26)	0.92	0.79	(0.64 ; 0.99)	**0.03**	0.97	(0.67 ; 1.39)	0.98
POW - FIX	0.85	(0.64 ; 1.13)	0.41	0.71	(0.57 ; 0.89)	**<0.01**	0.96	(0.66 ; 1.38)	0.97

4.7 Additional Subgroup Analysis for the Goal Study

In addition to the aforementioned analyses suggested by the literature, we looked at whether setting a specific goal at 10K steps was perceived as an implicit contract by participants and provoked a *ceiling effect* on the step counts. This could have been particularly true for participants in the FIX condition, for whom attaining this specific goal would lead to the most immediate form of reward. Extrinsic incentives might cause participants to stop striving once their goal is reached, or might lead them to stop trying if they cannot even come close to attaining it. Therefore, we computed in each phase, in each group, and for each level of IMI_1, the number of day-participants exceeding 11k steps among the day-participants exceeding 10k steps. We then ran a logistic regression on this indicator with factor condition and phase, including the interaction, and controlling for the IMI_1. The odd ratio comparisons were done using a Dunnet's correction of the group FIX against any other group. The results of this analysis are reported in Table 13. We found that indeed, during the experiment only, the FIX group had an odd ratio smaller than the POW group (0.71, 95% CI 0.57 to 0.89, $p < .01$) and the LOT group (0.79, 95% CI 0.64 to 0.89, $p < .05$). The other contrast FIX - CON is not significantly different. In the pre- and post-phases no significant differences emerge. This result confirms that setting up a static goal, combined with offering a monetary incentive, can limit the chances that participants of this intervention will exceed this limit. We now look at the effects of the interventions on the most and least motivated people among the participants.

4.8 Additional Subgroup Analysis for the IMI Study

The analysis reported in Section 4.2 suggests that the experimental manipulations had a negative effect on the intrinsic motivation of the participants. To study further the trends reported in Table 9, we looked more closely at the least and most motivated participants in the sample.

We conducted a sub-group analysis of the IMI model among the 23 participants with IMI_1 smaller or equal to 3. For this analysis, we chose this particular threshold because selecting and IMI smaller or equal to 2 would have left us with too few participants. Hence, we conducted the analysis with 5 participants in CON, 6 in FIX, 6 in LOT, and 6 in POW. The results are shown in Table 14. The analysis revealed a significant effect on the FIX condition in the experimental phase; the IMI score increased by 1.15 units (95% CI, .02 to 2.33, $p = .05$). The estimates of the model for the other conditions and in the post-phase are not significant.

Concerning the most motivated participants, an analysis of the sub-group of the IMI model was carried out among the 63 participants with IMI_1 greater or equal to 6 (specifically for 13 participants in CON, 17 in FIX, 19 in LOT, and 14 in POW). The results are shown in Table 14. The analysis reveals a significant effect on the FIX condition in the experimental phase: The IMI score decreased by 1.48 units (95% CI, −2.08 to −0.89, $p < .01$). The estimate for the FIX condition during the post-phase was not significant (95% CI, .06 to −2.11, $p = .07$, ns). Furthermore, the analysis reveals significant effects for the LOT condition; the IMI score decreased by 0.99 units in the experiment-phase (95% CI, −1.55 to −0.43, $p < .01$) and −1.25 in the post-phase (95% CI,

Table 14. Subgroups Analysis

	Change of IMI score for fixed levels of IMI_1					
	IMI_2 - IMI_1			IMI_3 - IMI_1		
Condition	Estimate	95% CI	Adj. P	Estimate	95% CI	Adj. P
$IMI_1 \leq 3$						
CON	0.88	(−0.33 ; 2.10)	0.18	1.11	(−0.10; 2.33)	0.07
FIX	1.15	(−0.02 ; 2.33)	**0.05**	1.21	(−0.07 ; 2.48)	0.06
LOT	0.61	(−0.67 ; 1.89)	0.44	0.58	(−0.84 ; 2.01)	0.54
POW	0.36	(−0.75 ; 1.47)	0.66	−0.40	(−2.68 ; 1.87)	0.87
$IMI_1 \geq 6$						
CON	−0.51	(−1.30 ; 0.28)	0.26	−0.30	(−0.33 ; 2.10)	0.63
FIX	−1.48	(−2.08 ; −0.89)	**<0.01**	−0.84	(−0.33 ; 2.10)	0.07
LOT	−0.99	(−1.55 ; −0.43)	**<0.01**	−1.25	(−0.33 ; 2.10)	**<0.01**
POW	−1.42	(−2.06 ; −0.78)	**<0.01**	−1.04	(−0.33 ; 2.10)	**0.02**

Adjusted contrasts between IMI levels measured at different phase in each condition.

−2.00 to −0.50, $p < .01$). Similarly, we find significant effects on the POW condition: The IMI score decreased by −1.42 units in the experiment phase (95% CI, −2.06 to −0.78, $p < .01$) and −1.04 in the post-phase (95% CI, −1.95 to −0.13, $p < .05$). Finally, the estimates for the CON condition were not significant.

These results show that the experimental manipulations have an effect on the intrinsic motivation of the participants. The least motivated participants in the FIX condition increased their IMI score during the experiment. The most motivated participants are negatively affected by the experimental manipulations, as their IMI score decrease during the experiment and continue decreasing for those in the LOT and POW condition in the Post-phase.

The results should be taken with caution, as this sub-group analysis was decided a posteriori. In particular, the IMI level threshold of 6 and 3 were selected based on the results of the step counts and goal analysis. There was no a priori justification for these thresholds, thus these results should be confirmed in future research.

For a final step, to find possible qualitative accounts for the quantitative trends observed in the data, we looked at the results of the questionnaires delivered at week 26 and week 44.

4.9 Participants' Perception of the Experimental Manipulations

As we explained in the methodology section, we deployed two short questionnaires at week 26 and at week 44 of the study. The goal was to collect qualitative feedback on the application, and about the design of the persuasive features of the research app. Specifically, we asked the following questions:

(1) *Que pensez-vous de la récompense qui a été donnée lorsque l'objectif quotidien a été atteint?* [What do you think about the reward that was given when your daily goal was reached?]
(2) *Que pensez-vous des messages de motivation envoyés chaque matin?* [What do you think about the motivational messages that were sent every morning?]

The first question was asked exclusively to participants in the conditions FIX and LOT, and the second was asked only to respondents in the condition POW. Two researchers analyzed the answers provided to the questions above, and they looked specifically for recurring comments that could account for the perceptions that participants had about the interventions. We summarize the main findings in the subsections below.

4.9.1 Perceptions of the Tangible Rewards. The participants with a low IMI felt generally positive about the tangible rewards that were offered during the experiment phase: e.g., *"Oui je marchais et courrais lègérement plus"* [Yes I walked and ran a little more] (Id: 292, IMI: 2.57, Condition: FIX); and *"Ça m'a un peu motivé mais pas au point de toujours aller faire du sport car je n'ai pas le temps."* [It did motivate me a bit, but not to the point of doing sport every day because I do not have the time] (Id: 297, IMI: 2, Condition: FIX). However, the participants with a high IMI did not perceive the reward to be sufficient to motivate them to increase their PA level: e.g.,*"une plus grande récompense m'aurait encouragé davantage"*, (Id: 244, IMI: 4.71, Condition: FIX) [a bigger reward would have encouraged me more.]; *"La récompense fait plaisir mais me pousse pas à faire davantage d'effort"* [The reward is fun but not pushing me to make more of an effort.] (Id: 241, IMI: 3.75, Condition: FIX); and *"Au début j'étais très motivée, mais les motivations financières n'étant pas assez élevées, j'ai arrêté de faire des efforts."* [At first I was very motivated, but the financial rewards were not high enough, I stopped making an effort."] (Id: 287, IMI: 4.86, Condition: FIX). Finally, participants with an IMI equal or greater than 6 expressed rather negative sentiments about the rewards: e.g., *"Je n'ai pas fait cette étude pour gagner de l'argent. Je fais assez activité physique pour moi-même"* [I did not participate in this study to make money. I do enough physical activity for my own benefit] (Id: 176, IMI: 6.77, Condition: LOT); and *"Je ne modifiais pas particulièrement mon activité physique en fonction de cette "récompense", cela dépend plutot du temps libre que j'ai."* [I did not particularly modify my physical activity according to this "reward", it depends rather on the free time that I have.] (Id: 252, IMI: 7, Condition: FIX). In summary, we found that participants with a low IMI felt generally positive about the rewards, while those with an higher IMI had rather negative comments.

4.9.2 Perceptions of the Motivational Messages. Similarly to what we saw in the other conditions, the participants assigned to the condition POW and with low IMI reported being influenced by the messages and perceived them as encouraging: e.g., *"Oui je suis influencée par les commentaires que l'on reçoit."* [Yes I am influenced by the comments we receive.] (Id: 148, IMI: 3.44, Condition: POW); and *"Parfois ça envoie des messages encouragants"* [Sometimes it sends encouraging messages.] (Id: 120, IMI: 2.86, Condition: POW).

On the contrary, participants with high IMI felt rather negative about the usefulness of the messages: e.g., *"Commentaires pas spécialement motivant, j'ai arrêté de les lire après quelques jours."* [Not particularly motivating comments, I stopped reading them after a few days.] (Id: 131, IMI: 5.86, Condition: POW); and *"Je n'aime pas beaucoup les commentaires qu'envoit l'application du genre "vos amis vous préfèrent lorsque vous faite du sport" etc.* [I do not like very much the comments that come with the application like "your friends prefer you when you play sports" etc.] (Id: 309, IMI: 6.58, Condition: POW).

Interestingly, some participants reported that the messages were not personalized enough hence were not useful for them: e.g., *"Les messages de motivation ne me touche pas... ce n'est pas assez personnalisé je trouve."* [the motivational messages do not affect me ... they are not personalized enough, I think.] (Id: 154, IMI: 4, Condition: POW).

Also, some participants indicated that they received too many notifications and consider them annoying: e.g., *"Les notifications journalières son très pénibles"*["Daily notifications are very painful"] (Id: 196, IMI: 5.86, Condition: POW); and *"Certaines notifications etaient un peu trop agressive et/ou intrusive"*["Some notifications were a little too aggressive and/or intrusive"] (Id: 151, IMI: 4.57, Condition: POW).

Other participants expressed being more attentive to the messages at the beginning of the intervention: *"Elle m'a fait changer au tout début, quand je voyais les notifications. C'est vrai que je faisais plus attention. Par contre, les notification sont devenus habituelles et a un moment donné je faisais* *plus attention."* [It made me change at first, when I saw the notifications, it's true that I was paying more attention. On the other hand, the notifications became habitual and at a certain moment I did not pay attention.] (Id: 122, IMI: 5.14, Condition: POW); and *"Oui, au début, j'ai vraiment fait l'effort de prendre moins les transports en commun et de marcher plus dans ma vie de tous les jours. Puis les messages sont devenus agaçants."* [Yes, at first, I really made the effort to take less public transportation and walk more in my everyday life, then the messages became annoying.] (Id: 155, IMI: 6.57, Condition: POW).

5 DISCUSSION

One of the main outcomes of this work is that we learned that tangible rewards do not help establish lasting healthy routines. During the course of the experiment, we observed that there was not any significant difference in the amounts of steps walked, whether we offered participants money or not to perform this activity. The estimates of the step model are not significant for each contrast and for each phase of the study. Hence, we cannot validate H3a and cannot confirm the predictions of the SDT theory [101] or those of empirical studies [81, 83] for which a reduction of PA in FIX and LOT was expected. However, these results do not contradict the theory either, as it is evident that the monetary incentives we tested in this experiment did not increase the level of PA of the participants.

A second—and more important—outcome of this research is that although PA did not substantially change, the intrinsic motivation of the participants indeed changed as a consequence of the experimental manipulations. At the end of the experiment phase, the DID analysis indicate that the IMI scores of participants in the control group increased, whereas those of the participants in FIX and LOT lowered. In the post-phase, these negative trends recovered a bit but remained negative (see Table 10 and the analysis reported in Section 4.2). The subgroup analysis reported in Section 4.8 further reveals that though for the least motivated participants the rewards increased the IMI scores at the end of the experiment, for the most motivated participants it was just the opposite. In other words, an economic incentive can motivate people who are not yet interested in walking or running, but it can have the opposite effect on people who are already have the habit of walking (or running).

The extent of these negative effects largely depends on the *intrinsic motivation* of the person receiving the rewards; the more the person is already self-motivated about performing the activity, the higher the chance is that rewards will have a detrimental effect.

Offering money to a person who is already motivated to perform a certain activity can confuse the person about their real motives. This finding is in line to previous research conducted by Orji et al. who found that monetary rewards do not work as a behavior-change strategy as "*people tend to view the rewards and the values they get from them as the only benefit of adopting a healthy behavior" [80, p. 29]*. Hence, failing to establish lasting benefits. This is also signaled by the need of some participants to re-state their original interest in walking or running (see qualitative examples reported in Section 4.9.1). This finding was also predicted by SDT; the theory connected the detrimental effects of the rewards to the basic need of *autonomy* [99]. However, little research to date has provided an empirical account of the theory that demonstrates the effect of the intervention on the intrinsic motivation. Although most studies infer the effect on intrinsic motivation by using a *free-choice* paradigm (i.e., the frequency with which the participants engage with the target activity after the intervention has been removed), we also demonstrated the negative effects through a direct measurement of the IMI score.

It is important to also highlight that the findings we present apparently contradict those that report positive effects of rewards in increasing PA, (e.g., as discussed in the literature review section [16, 17, 34], among others). We argue that this contradiction is simply apparent because SDT also posits that when tangible rewards are salient to the activity for which they are given, they can indeed propel action. We speculate that if we had chosen a larger economic incentive, we could have produced a visible increase of PA, as posited by SDT. However, the theory also cautions (and these findings reveal) that though these interventions might have positive effects in the short term, they might—in the long term—harm the intrinsic motivation of the participants. In our findings, this is clearly evident, as the participants' IMI scores reduced during the experiment and post-phase. Therefore, our findings provide empirical evidence of this "downside" of tangible rewards, by showing that paying people to do more PA can harm the inner resources (for self-determined action) people might already possess. Furthermore, even if salient rewards work with people who are not yet motivated, these interventions will not encourage people to progress in the motivation continuum posited by the SDT, hence people will not ultimately become autonomous in performing the target behavior. Therefore, this type of intervention will always be dependent on the delivery of the rewards, as we expect that as soon as the incentive is suspended, the person will fall back into their previous behavior.

Another important finding is that we do not find any difference between the tangible rewards provided by the fix incentive and the lottery condition. During the experiment, people who were submitted to the LOT condition did not behave differently than those in the FIX condition; we observe the same negative trends on PA overall. We also observe the same detrimental effect on the IMI scores during the experiment. We designed the LOT condition to offer an incentive more salient than the FIX condition. Therefore, we expected the reward with higher salience to yield a larger effect, as predicted by the salience theory [10]. However, obtaining the reward of the LOT intervention required two steps, i.e., gaining a ticket and winning the lottery. This further delayed the gratification of the participants (see the work of O'Donoghue and Rabin [75]), which can explain the presented results.

Comparing these results with those obtained by Patel et al. [83], we noticed several differences in the experimental design. (1) The maximum prize for winning the lottery was 50 USD, an incentive five-times larger than the incentive we offered; (2) the daily goal was 7K steps, instead of the 10K in our study; and (3) the participants of their study included obese people, which could result in lower IMI scores at the onset of the experimental phase of the study. In particular, this last point can account for the differences in efficacy of the lottery treatment in our study and in the study conducted by Patel et al.; if we consider the subgroup analysis for the IMI study reported in Section 4.8, we notice that the trends on the IMI scores are all positive for the the least motivated people in the sample. A salient reward can be effective in motivating participants who are not intrinsically motivated to perform the activity. However, this positive trend might eventually stall, because once participants raise their intrinsic score, this mechanism will not be sufficient to make them progress towards autonomous behaviour (Figure 1). This comparison between the two studies and its conclusion is speculative and should be confirmed by future research.

The third intervention that we tested in this experiment was related to sending daily motivational messages (condition POW). We find that this intervention does not increase the PA of the participants. Similar to what we describe above for the FIX and LOT conditions, we do not observe any differences in the control group, in terms of the number of walked steps. However, we observe a detrimental effect on the IMI scores measured at the end of the experiment phase. These findings contradict our expectations that were based on SDT and prior empirical research on motivational messages [60]. This difference, however, has to be interpreted by looking at how we translated the theory into design. In fact, SDT specifies that in order for feedback to provoke positive effects on the recipients' sense of competence, it has to provide information that can help recipients reflect on their own performance [101, p. 154]. If we consider the corpus of messages more carefully (Appendix A), we observe that the messages did not provide information related to the specific execution of the task by the participants (i.e., how much a participant was walking or running in regard to the goal of the 10K steps). Instead, these texts communicated generic information about why running or walking is beneficial for health and provided tips on how to setup healthy exercise routines. This information did not really provide useful information for the participants to assess progress from one week to the next, or to understand whether—aside from meeting the daily goal—the overall PA was increasing or decreasing. Therefore, the motivational messages that were delivered in this study fell very short of providing information that participants could use to reflect on their performances, hence the lack of effect on the basic need of competence. The corpus used in this experiment was modeled on the TTM [88]. The messages followed a specific chronological order that tailored the intervention to the stages of change, i.e., progressing stages of changing behavior. However, the same message was sent to all participants in the POW condition, regardless of the specific status of each participant. In a later work, the same researchers, who created the original corpus, conducted additional research and found that these motivational messages could be more effective if their delivery is tailored on the personality and gender of the recipient [20], or even the designer of the message, i.e., an expert or a peer [18]. This is one aspect that should be considered in future research involving motivational messages. A second aspect to consider is that repeatedly receiving the messages every day was experienced by participants as an external pressure about the daily goal hence was perceived as controlling. This aspect of the study could explain the negative effect that we observed on the IMI scores and is documented by the participants' comments reported in Section 4.9.2 where they express being annoyed by the repeated notifications. The frequency with which we sent the motivational messages to the participants could also explain a difference in the results of Kinnafick et al. [60]. These authors could not observe any significant difference of PA (however, they observed a positive trend). Yet, they measured an increase of the IMI scores at the end of the experimental phase. In their experimental design, they delivered the messages only twice a week. We can speculate that this low frequency of delivery could have been perceived by participants as less controlling, compared to the frequency of our messages (i.e., a message every day). Furthermore, our results reveal that, similarly to what we observed for the tangible rewards, the negative effect of sending these motivational messages affects particularly the most motivated participants in the sample. In summary, these results reveal that the most motivated participants are extremely susceptible to the interventions that we tested in our experiment.

5.1 Implications for Design

The most direct implication of this research is that "one solution fits all" does not work for persuasive design. Depending on the characteristics of the user, interventions can have detrimental effects. Therefore, we argue that interventions should be personalized in order to maximize their efficacy and to avoid negative outcomes.

Previous research demonstrated that profiling the personality of the participants [79] or the persuasive strategy that works best for the participants [78] can improve the overall effectiveness of the intervention.

The results presented in this work support the argument of designing personalized interventions in the following two ways: (1) in terms of establishing the level of the intrinsic motivation of the participant at the onset of the intervention; and (2) in terms of designing optimal challenges for the participant. Concerning point (1), in this study, we found that the level of intrinsic motivation toward a given activity is a predictor of the effect of the intervention, and we argue that this metric could be used to choose the most effective persuasive techniques. In particular, for users who have an IMI score lower than 4.5, tangible rewards could be considered a persuasive technique for encouraging a person's internalization process. However, after 3–5 weeks, this technique should be replaced by a different technique able to continue the introjection of the PLOC. Extending the intervention further might hinder subsequent improvements. Furthermore, for users with an IMI score higher than 4.5, using tangible rewards is altogether not recommended, as this technique might simply lead to outcomes that are contrary to the desired results. Unfortunately, at this stage, we are unable to advise which persuasive techniques should be used instead (or in combination with), as this goes beyond the scope of this article.

One of the specific advantages of using the IMI score to personalize the user experience would be related to the design of practical instruments that capture the metric. Building personality profiles of participants usually involves long surveys[23], whereas capturing the IMI can typically be done with 6 questions, thus simplifying the data-capturing aspect.

Concerning point (2), we found that imposing the same goal (i.e., 10K steps/day) to all participants led to a ceiling effect on the outcomes that some participants could have achieved. Therefore, this study suggests that participants should be offered personalized optimal challenges that could increase (or decrease), depending on their short-term achievements. For instance, if in a given week a person does 8K steps, a reasonable goal for the following week could be an increase between 5% and 10% (i.e., 8.4K to 8.8K steps). These adjustable targets could also take care of the seasonal effects we observed in the study (e.g., lower PA due to exam seasons and holidays). The concept of *optimal challenge* was already discussed within the SDT [24] and demonstrated empirically in game immersion [89, 93]. Providing adaptable targets is also one of the key properties of effective interventions, as posited by the goal-setting theory [63].

A second relevant implication of this research deals with the use of tangible rewards to provoke behavior changes that last. Although many scholars still debate the usefulness of these rewards and how to best design the incentive schemes [43, 83], our results highlight the importance of looking at the impact of these interventions in the long run. Providing well-designed and salient rewards can often boost compliance to the target behavior. However, this can create negative effects in the long-term perspective. Therefore as designers, we should reflect on what kind of changes we want to provoke. We argue that having a short-sighted approach to the problem might result in creating technology that acts as a *prosthesis* for humans but that will not ultimately empower users in *exploiting their full potential*. In other words, design solutions must not create dependency on the technology itself. Tangible rewards should be reconsidered in persuasive design, not only because in many cases they produce negative results even in the short term (i.e., for highly motivated people, as discussed above) but most importantly because even in the cases where they seem to work in the short-term they will discourage the users in their future endeavours. From this perspective, persuasive app designers should seek solutions that *empower* users rather than those that create dependence. Unfortunately, the two interventions tested in this study do not posses this characteristic.

Finally, this study also shed light on the design of motivational messages. We learned that providing informative messages does not work as a persuasive mechanism. Sending feedback at a high frequency might also be perceived negatively by users and demotivate them in their goals of establishing a healthy exercise routine. Feedback should be used by app designers to *satisfy the basic need of competence*. However, for feedback to work, it has to provide useful information that can help the users reflect on their own performance and gain a sense of their effectiveness. An example of this task-inherent feedback could be *Well done! Today you walked 12345 steps. This week you improved your step activity by 3%, compared to last week. Keeping up this level of activity for an additional three weeks will result on an overall improvement of your muscular tone*. This example

[23]One of the standard to capture the personality profile of a person is the BigFive questionnaire that involves 50 questions in its shortest form [44].

demonstrates the following features of feedback as required by SDT: *(a.) personal,* it provides content that is specific to the participant; *(b.) contextual,* it provides task-inherent information that can help the user connect their performance of the activity with its outcomes; and *(c.) goal-oriented,* it provides the next challenge to push their work further by being phrased in a way that is specific to the level of the user. We provide implications about how motivational messages should be delivered. Using the results of the study, we argue that the delivery of feedback should not follow a regular schedule, as this could have negative consequences for the user's basic need of autonomy. Instead, to avoid repetition and the related feeling of being controlled, it would be more beneficial to *deliver the feedback opportunistically* when the user performs a spontaneous activity, perhaps at a time or place where this did not occur in the past.

5.2 Limitations

Our study has several limitations. First, the participants were all university students, which limits the generalization of our findings. The physical activity of the students is largely influenced by the academic calendar that defines examination and class periods, and summer/winter holiday seasons. Class sessions demand long sitting-hours, whereas holidays can lead to an increase in physical activity. Second, the study was limited to one country and region. Swiss students living in the French-speaking region of Switzerland have exercise opportunities and transportation habits different than students of the same age living in denser metropolitan areas, or in locations with more severe winters or more constrained recreational facilities. Third, the study was limited to one type of physical activity, which does not allow us to measure the complete spectrum of physical activities that the participants were involved into (e.g., swimming and cycling). The interventions we tested in this experiment might have had an effect in promoting the overall physical activity, but this might have not been necessarily reflected by the number of steps of the participants. Given that the design of the interventions was focused on the number of steps, we consider this last point unlikely but not impossible. In this article, we are unable to provide any data on what types of PA participants might have been involved in (other than walking and running) and whether their overall PA changed during the experiment. Fourth, participants' physical activities were not tracked if they did not carry their smart phones. These last two points might account for under-reported levels of PA by the participants in this study. Additionally, the study was limited to iPhone users, thus excluding other market segments[24]. Furthermore, we did not test a combination of tangible rewards and motivational messages in the same intervention. Finally, we acknowledge that the procedure we used to create the corpus of motivational messages is not perfect, and that in order to generate a more effective corpus, it could be ameliorated in future research. We suggest the following two ideas for improvement: (1) the effectiveness of each message can be achieved through Mechanical Turks; and (2) as reported in the discussion, effective feedback should be personalized, contextual, and goal-oriented.

6 CONCLUSIONS

In this article, we shed light on the elements of design of persuasive apps in their ability to provoke change in behavior. We have provided evidence that contributes to the body of knowledge concerning the use of tangible rewards and motivational messages in order to increase physical activity. To the best of our knowledge, this is the first study comparing, in the same longitudinal study, these two persuasive techniques. Also, this is one of the first studies focusing specifically on the detrimental effects of these interventions on the intrinsic motivation of participants in the post-intervention phase. Although most of the prior work measured experimental effects on the

[24]It has to be reported that the iPhone has one of the largest market share in Switzerland. See https://www.nzz.ch/wirtschaft/unternehmen/apple-bleibt-platzhirsch-1.18563593, last retrieved June 2018.

main outcome variable (e.g., steps) during the intervention, we have reported and discussed the negative effects that these interventions can have in the long-term on the intrinsic motivation of the participants of the intervention. The specific contribution of this work is to add to the current discussion by suggesting that in order to determine whether an intervention is effective, we should look at both the main outcome measurement of PA and, more importantly, at the intrinsic motivation. Indeed, the intrinsic motivation has often been neglected, yet it determines the development of people's inner resources that sustain their self-determined actions.

Furthermore, we argue that only by embracing a holistic approach to these behavioral interventions can we significantly advance the design of persuasive technology. By this, we mean there is a need for a systematic experimentation and comparison of different designs over a longitudinal segment of time spanning several months. Furthermore, we argue that careful testing of the post-intervention phase of these interventions might reveal unexpected side effects.

An additional benefit of designing this kind of study is that we can expect to eventually reconcile the apparent differences between studies conducted in the domain of behavioral economics and psychology (e.g., the differences of results around rewards, see for instance the discussion reported by Gneezy and Rustichini [43]).

For future work, we plan to replicate this study with a different population. In particular, it would be valuable to study whether the same effects would be found with participants with an active job or with older participants who are retired. Furthermore, we plan to conduct additional systematic evaluations of design elements that might contribute to satisfying the basic needs identified by the SDT. Future studies could include other standardized measurements that would help us to comprehend the perception of autonomy, competence and relatedness (e.g., Locus of Causality for Exercise Scale [66], Physical Self-Perception Profile [40] and Social Support for Exercise Questionnaire [105]). Another idea that emerged in this work was related to the concept of an optimal challenge. Future work should examine whether exposing participants to adaptable goals could increase relevance and decrease the feeling of being controlled perceived by some when feedback was delivered.

APPENDIX

A EXCERPT OF MOTIVATIONAL MESSAGE CORPUS

Table 15. Excerpt from Motivational Message Corpus

Stage	Translated version in French	Original message in English	Day
PC	Marcher régulièrement prévient le risque de complications cardiaques et d'AVC. Une marche de 10min./jour est un bon moyen pour réduire la pression artérielle.	If you walk regularly, you are less likely to suffer a stroke and other cardiovascular problems. Taking 10-minute walks every day is an effective way to lower your blood pressure	1
PC	Planifier ta marche ou ta course à l'avance t'aidera à aller jusqu'au bout.	Planning in advance when you will go for a walk or run will help you to go through with it.	2
PC	Marcher régulièrement aide tes poumons à devenir plus sains. S'entraîner accroit tes cycles respiratoires. Cela dégage tes poumons et les rend plus forts.	Walking regularly helps your lungs to become more healthy. When exercising you increase your respiration cycles. This keeps your lungs clear and makes them stronger.	3

(Continued)

Table 15. Continued

Stage	Translated version in French	Original message in English	Day
PC	Tes amis remarqueront les changements si tu t'entraines davantage.	Your friends will notice the changes if you exercise more.	11
PC	Lance toi des défis par ètapes et non pas par mois. Monte 500 marches aujourd'hui, 501 demain et une de plus le jour suivant.	Challenge yourself in steps not months. Today walk 500 steps, tomorrow walk 501, and the next day add another step.	13
PC	Chaque minute de marche rapide par jour diminue le risque d'obésité de 5%. Marcher rèduit de moitié l'effet des gènes d'obésité.	Every minute of brisk walking throughout the day lowers the risk of obesity by 5%. Walking cuts the effects of obesity-promoting genes by half.	14
C	Tu progesseras lentement. Tu ne seras pas soudainement plus souple ou en meilleure santé, mais tu sentiras peu à peu l'effet de 15 minutes de marche par jour.	This process takes patience. You are not going to be more flexible or in better health over night. But I promise from adding like 15 minute walk every day you will see results, you will feel them.	15
C	Apporte un plus à ta vie! En procrastinant, ta santé risque de se déteriorer. Après 3 semaines d'activité physique, point besoin de médicaments.	Start soon on something you want to make a part of your life! While I procrastinated my health deteriorated. After three weeks of walks I was no longer relying on pills.	16
C	Attrape un de tes ami · e· es et sortez marcher ou courir ensemble!	Grab one of your friends and go for a walk / run together.	17
C	Prendre les escaliers ou se garer un peu plus loin est un bon moyen d'exercer son coeur sans faire trop d'effort.	Taking the stairs or parking a little farther out is a great way to squeeze in a few minutes of heart healthy exercise without working too hard!	18
C	Sache qu'un travail assidu paie toujours: tu récolteras les fruits de ton travail!	Just know that hard work pays off, and the seeds you plant now will reap a harvest.	27
C	L'activité physique est bon pour ton cerveau et marcher améliore ta mémoire. Marcher 40 minutes 3 fois par semaine augmente de 2% tes capacités cognitives.	Exercise is good for the brain but walking in specific is good for boosting your memory. Walking for 40 minutes three times a week could increase the volume of the hippocampus by 2%!	28
P	Le plus difficile, c'est le début. Mais c'est fait. à partir de maintenant ça ne peut être que plus facile.	Starting to exercise is often the hardest part, and you've already managed that! It'll only get easier from here on.	29
P	Essaie de te fixer des objectifs personnels. Garde à l'esprit qu'ils doivent être réalistes et atteignables.	Try to set some personal goals. Keep in mind that they have to be realistic and reachable.	30
P	Pense à tout ce que tu as accompli la semaine passée et essaye de faire le lien avec tes objectifs.	Think about what you have achieved last week and try to connect your goals to that.	31
P	Voilà 4 semaines que tu t'entraînes régulièrement. Bravo!	You have been exercising regularly for 4 weeks now. Good job!	32

(Continued)

Table 15. Continued

Stage	Translated version in French	Original message in English	Day
P	Tu te sentiras plus en forme en rentrant du travail si tu as marché un peu plus durant la journée. Quiconque peut le faire: quelques minutes de plus suffisent!	You will feel more energetic when you get home from work if you walked a few extra minutes during the day. Anyone can do that: a few extra minutes!	59
P	Après une longue période d'activité physique régulière, tu verras que ta silhouette s'améliore.	After a longer time of regular exercising you will see how your body changes in a positive way.	60
A	La vie se construit au jour le jour. Chacun de ces jours compte.	Lives are built on day at a time. Each of those days matters.	64
A	Ceux qui s'entraînent sontgénéralement plus heureux que lesautres!	People who exercise are generally happier than people who do not!	65
A	Te sens-tu capable de faire plusd'efforts?	Do you feel like going harder/faster/moreintense?	66
A	Cela peut paraître difficile à présent, mais plus tu es actif sur la durée, plus cela te semblera facile et naturel	It may seem hard now, but the more active you are over time, the easier and more natural it comes.	67
A	Que dirais-tu de te joindre à tes amis· e · es qui s'entraînent déjà? Cela améliorerait non seulement ta santé, mais aussi tes relations sociales.	How about joining your friends that are already exercising? It will not only improve your health, but also your social relationships.	102
A	Savais-tu déjà que tu brûles plus de calories en marche rapide qu'en courant?	Did you already know that you burn more calories with a strong walk than with jogging?	103
M	N'abandonne jamais!	Never give up.	106
M	Prends du temps pour toi et pour t'entraîner. Cela devrait être un priorité pour toi.	Making time for yourself and to workout should be a priority in your life. It's important to make time for you!	107
M	S'entraîner n'est pas que du sérieux. Fais-le en t'amusant et tu voudras le faire plus souvent !	Exercise isn't all about being serious, have some fun with it and you'll want to do it more often!	108
M	Entraîne-toi avec tes amis pour pouvoir courir 5 km!	Train with your friends for a 5k!	109
M	Vas en ville ou au centre commercial. Change tes habitudes. Parfois on n'a pas envie d'être dans la nature et le lèche-vitrine nous attire plus.	Head for town or for the mall. Sometimes nature just isn't calling. You may be more entertained window shopping.	139
M	Ecoute de la musique en marchant. Crée un playlist sur ton natel. Parfois, c'est encore plus motivant et réjouissant d'aller marcher pour écouter la musique.	Listen to Music While You Walk. I always keep a walking playlist on my phone. Sometimes I look forward to listening to my music as much as I look forward to the walk.	140

The first column reports the stage of change as coded in the original corpus. The last column reports the day of deployment of the message during the course of the experiment.

B TABLES OF SUBGROUP ANALYSIS OF STEP MODEL

Table 16. Frequencies of Participants in Each Subgroup

	CON	FIX	LOT	POW
achieved	11	8	11	8
feasible	23	24	15	20
challenging	18	26	19	18

Table 17. Adjusted Contrasts Between Conditions in the Steps Model, for Subgroups of PA at Baseline

	Experiment (wk 15–33)			Post-phase (wk 34–44)		
	Estimate	95% C.I.	Adj. P	Estimate	95% C.I.	Adj. P
achieved						
FIX - CON	−1156.6	(−3316.7 ; 1003.5)	0.44	−1590.2	(−3941.2 ; 760.76)	0.26
LOT - CON	−30.8	(−2013.1 ; 1951.5)	>0.99	−178.7	(−2367.8 ; 2010.3)	0.98
POW - CON	811.3	(−1348.5 ; 2971.1)	0.68	2201.0	(−189.6 ; 4591.6)	0.08
feasible						
FIX - CON	−1176.2	(−2532.8 ; 180.5)	0.11	−460.8	(−2011.4 ; 1089.8)	0.79
LOT - CON	−898.7	(−2441.8 ; 644.3)	0.38	−429.7	(−2124.8 ; 1265.4)	0.85
POW - CON	−829.9	(−2251.24 ; 591.4)	0.38	−169.5	(−1760.8 ; 1421.8)	0.97
challenging						
FIX - CON	640.7	(−784.7 ; 2066.2)	0.57	742.8	(−841.3 ; 2326.9)	0.54
LOT - CON	210.1	(−1319.2 ; 1739.4)	0.96	200.1	(−1507.6 ; 1907.8)	0.97
POW - CON	244.9	(−1305.5 ; 1795.2)	0.94	332.5	(−1511.2 ; 2176.2)	0.92

ACKNOWLEDGMENTS

We would like to thank several colleagues that contributed directly and indirectly to the development of the ideas reported in this article: Valérie Chavez, José Mata, Luís Santos-Pinto, Kashyap Todi, and Roelof de Vries. A special thank goes to Katja Schwab-Weis for helping in coordinating communications with the participants of this study. Finally, we would like to thank Holly Cogliati-Bauereis for her precious comments. This research was financed by the HEC Research fund 2017.

REFERENCES

[1] Alberto Abadie. 2005. Semiparametric difference-in-differences estimators. *The Review of Economic Studies* 72, 1 (2005), 1–19. DOI: https://doi.org/10.1111/0034-6527.00321

[2] Icek Ajzen. 1985. From intentions to actions: A theory of planned behavior. In *Action Control*. Springer, Salmon Tower Building New York City, 11–39. DOI: https://doi.org/10.1016/0091-7435(87)90022-3

[3] Icek Ajzen and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behaviour*. Prentice-Hall, Upper Saddle River, NJ. 278 pages.

[4] Dan Ariely. 2010. *Predictably Irrational : The Hidden Forces that Shape Our Decisions*. Harper Perennial, HarperCollins Publishers, 195 Broadway, New York, NY, 10007. 349 pages.

[5] Albert Bandura. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

[6] Charlotte A. Barnes and Patricia M. Schoenborn. 2012. *Trends in Adults Receiving a Recommendation for Exercise or Other Physical Activity From a Physician or Other Health Professional*. Technical Report. NCHS data brief, no 86, Hyattsville, MD: National Center for Health Statistics. Retrieved from https://www.cdc.gov/nchs/data/databriefs/db86.pdf.

[7] Jeroen C. M. Barte and G. C. Wanda Wendel-Vos. 2017. A systematic review of financial incentives for physical activity: The effects on physical activity and related outcomes. *Behavioral Medicine* 43, 2 (2017), 79–90. DOI: https://doi.org/10.1080/08964289.2015.1074880

[8] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

[9] Richard Blundell, Monica Costa Dias, Costas Meghir, and John van Reenen. 2004. Evaluating the employment impact of a mandatory job search program. *Journal of the European Economic Association* 2, 4 (2004), 569–606. DOI: https://doi.org/10.1162/1542476041423368

[10] Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. 2012. Salience theory of choice under risk. *The Quarterly Journal of Economics* 127, 3 (2012), 1243–1285. Retrieved from https://scholar.harvard.edu/files/shleifer/files/salience_theory_of_choice_under_risk_qje.pdf.

[11] Dena M. Bravata, Crystal Smith-Spangler, Vandana Sundaram, Allison L. Gienger, Nancy Lin, Robyn Lewis, Christopher D. Stave, Ingram Olkin, and John R. Sirard. 2007. Using pedometers to increase physical activity and improve health: A systematic review. *JAMA* 298, 19 (2007), 2296–2304. DOI: https://doi.org/10.1001/jama.298.19.2296

[12] Duncan S. Buchan, Stewart Ollis, Non E. Thomas, and Julien S. Baker. 2012. Physical activity behaviour: An overview of current and emergent theoretical practices. *Journal of Obesity* 2012 (2012), 1–11. DOI: https://doi.org/10.1155/2012/546459

[13] Rachel J. Burns, Angela S. Donovan, Ronald T. Ackermann, Emily A. Finch, Alexander J. Rothman, and Robert W. Jeffery. 2012. A theoretically grounded systematic review of material incentives for weight loss: Implications for interventions. *Annals of Behavioral Medicine* 44, 3 (2012), 375–388. DOI: http://dx.doi.org/10.1007/s12160-012-9403-4

[14] Rachel J. Burns and Alexander J. Rothman. 2018. Comparing types of financial incentives to promote walking : An experimental test. *Applied Psychology: Health and Well-Being* 10, 2 (2018), 1–22. DOI: https://doi.org/10.1111/aphw.12126

[15] Joëlle Carpentier and Geneviève A. Mageau. 2013. When change-oriented feedback enhances motivation, well-being and performance: A look at autonomy-supportive feedback in sport. *Psychology of Sport and Exercise* 14, 3 (2013), 423–435. DOI: https://doi.org/10.1016/j.psychsport.2013.01.003

[16] Gary Charness and Uri Gneezy. 2009. Incentives to exercise. *Econometrica* 77, 3 (2009), 909–931. DOI: https://doi.org/10.3982/ECTA7416

[17] Ahmad Daryanto, Ko de Ruyter, Martin Wetzels, and Paul G. Patterson. 2010. Service firms and customer loyalty programs: A regulatory fit perspective of reward preferences in a health club setting. *Journal of the Academy of Marketing Science* 38, 5 (2010), 604–616. DOI: https://doi.org/10.1007/s11747-009-0165-x

[18] Roelof A. J. De Vries, Cristina Zaga, Franciszka Bayer, Constance H. C. Drossaert, Khiet P. Truong, and Vanessa Evers. 2017. Experts get me started, peers keep me going: Comparing crowd- versus expert-designed motivational text messages for exercise behavior change. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, 155–162. DOI: https://doi.org/10.1145/3154862.3154875

[19] Roelof A. J. de Vries, Khiet P. Truong, Sigrid Kwint, Constance H. C. Drossaert, and Vanessa Evers. 2016. Crowd-designed motivation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM Press, New York, New York, 297–308. DOI: https://doi.org/10.1145/2858036.2858229

[20] Roelof A. J. de Vries, Khiet P. Truong, Cristina Zaga, Jamy Li, and Vanessa Evers. 2017. A word of advice: How to tailor motivational text messages based on behavior change theory to personality and gender. *Personal and Ubiquitous Computing* 21, 4 (2017), 675–687. DOI: https://doi.org/10.1007/s00779-017-1025-1

[21] R. DeCharms. 1968. *Personal Causation: the Internal Affective Determinants of Behavior*. Academic Press, Cambridge, Massachusetts. Retrieved from https://books.google.ch/books?id=GrV9AAAAMAAJ.

[22] Edward L. Deci. 1971. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* 18, 1 (1971), 105.

[23] Edward L. Deci. 1972. The effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance* 8, 2 (1972), 217–229.

[24] E. L. Deci. 1975. *Intrinsic Motivation*. Plenum Publishing Company Limited, 233 Spring Street, New York, NY 10013-1522. 324 pages. Retrieved from https://books.google.it/books?id=gbV9AAAAMAAJ.

[25] E. L. Deci, H. Eghrari, B. C. Patrick, and D. R. Leone. 1994. Facilitating internalization: the self-determination theory perspective. *Journal of Personality* 62, 1 (1994), 119–142. DOI: https://doi.org/10.1111/j.1467-6494.1994.tb00797.x

[26] Edward L. Deci, Richard Koestner, and Richard M. Ryan. 1999. The undermining effect is a reality after all-extrinsic rewards, task interest, and self-determination: Reply to Eisenberger, Pierce, and Cameron (1999) and Lepper, Henderlong, and Gingras (1999). *Psychological Bulletin* 125, 6 (1999), 692–700. DOI: https://doi.org/10.1037/0033-2909.125.6.692

[27] Edward L. Deci and Arlen C. Moller. 2005. The concept of competence: A starting place for understanding intrinsic motivation and self-determined extrinsic motivation. In *Handbook of Competence and Motivation*. Guilford Publications, New York, NY, 579–597.

[28] Edward L. Deci and Richard M. Ryan. 1980. The empirical exploration of intrinsic motivational processes. In *Advances in Experimental Social Psychology*. Vol. 13. Elsevier, 125 London Wall, 125 London Wall, London EC2Y 5AJ, UK, 39–80. DOI: https://doi.org/10.1016/S0065-2601(08)60130-6

[29] Edward L. Deci and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-determination in Human Behavior*. Springer Science & Business Media, Salmon Tower Building New York City.

[30] Edward L. Deci and Richard M. Ryan. 2014. Autonomy and need satisfaction in close relationships: Relationships motivation theory. In *Human Motivation and Interpersonal Relationships: Theory, Research, and Applications*. Springer Science + Business Media, Salmon Tower Building New York City, 53–73. DOI: https://doi.org/10.1007/978-94-017-8542-6_3

[31] Rod K. Dishman, Robert W. Motl, Ruth Saunders, Gwen Felton, Dianne S. Ward, Marsha Dowda, and Russell R. Pate. 2004. Self-efficacy partially mediates the effect of a school-based physical-activity intervention among adolescent girls. *Preventive Medicine* 38, 5 (2004), 628–636. DOI: https://doi.org/10.1016/j.ypmed.2003.12.007

[32] Lindsay R. Duncan, Craig R. Hall, Philip M. Wilson, and Wendy M. Rodgers. 2012. The use of a mental imagery intervention to enhance integrated regulation for exercise among women commencing an exercise program. *Motivation and Emotion* 36, 4 (2012), 452–464. DOI: https://doi.org/10.1007/s11031-011-9271-4

[33] ExerciseRewards. 2019. Exercise reards: it pays to be active. Retrieved from https://www.exerciserewards.com/.

[34] Eric A. Finkelstein, Derek S. Brown, David R. Brown, and David M. Buchner. 2008. A randomized study of financial incentives to increase physical activity among sedentary older adults. *Preventive Medicine* 47, 2 (2008), 182–187. DOI: https://doi.org/10.1016/j.ypmed.2008.05.002

[35] Fitbit. 2019. Fitbit. Retrieved from https://www.fitbit.com/.

[36] G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs. 2008. *Longitudinal Data Analysis*. CRC Press, 6000 Broken Sound Pkwy NW #300, Boca Raton, FL 33487. Retrieved from https://books.google.ch/books?id=zVBjCvQCoGQC.

[37] Brianna S. Fjeldsoe, Yvette D. Miller, and Alison L. Marshall. 2010. MobileMums: A randomized controlled trial of an SMS-based physical activity intervention. *Annals of Behavioral Medicine* 39, 2 (2010), 101–111. DOI: https://doi.org/10.1007/s12160-010-9170-z

[38] FoodStand. 2019. FoodStand: Ditch dieting. Build healthy habits instead. Retrieved from https://www.thefoodstand.com.

[39] Centers for Disease Control and Prevention. 2014. *The State Indicator Report on Physical Activity*. Technical Report. Centers for Disease Control and Prevention, Department of Health and Human Services, Atlanta, GA. 28 pages. Retrieved from https://www.cdc.gov/physicalactivity/resources/reports.html.

[40] Kenneth R. Fox and Charles B. Corbin. 1989. The physical self-perception profile: Devlopment and preliminary validation. *Journal of Sport and Exercise Psychology* 11, 4 (1989), 408–430. DOI: https://doi.org/10.1123/jsep.11.4.408

[41] Julian Frank. 2017. *A Personalized Support Tool for the Training of Mindful Walking: The Mobile "Mindfulwalk" Application*. Master's thesis. Ulm University. Retrieved from http://dbis.eprints.uni-ulm.de/1583/.

[42] Karly S. Geller, Rob W. Motl, Caroline C. Horwath, Kristin K. Wertin, and Rodney K. Dishman. 2011. A research agenda to examine the efficacy and relevance of the Transtheoretical Model for physical activity behavior. *Psychology of Sport and Exercise* 12, 1 (2011), 7–12. DOI: https://doi.org/10.1016/J.PSYCHSPORT.2010.04.004

[43] Uri Gneezy and Aldo Rustichini. 2000. Pay enough or don't pay at all. *The Quarterly Journal of Economics* 115, 3 (2000), 791–810. DOI: https://doi.org/10.1162/003355300554917

[44] Lewis R. Goldberg. 1992. The development of markers for the big-five factor structure. *Psychological Assessment* 4, 1 (1992), 26. DOI: https://doi.org/10.1037/1040-3590.4.1.26

[45] A. Guyton and J. E. Hall. 2015. *Textbook of Medical Physiology* (13th ed.). Saunders, New York, NY. 48 pages.

[46] GymPact. 2019. GymPact: Incentivize your exercise. Retrieved from http://static.gym-pact.com.

[47] Michael D. Hanus and Jesse Fox. 2015. Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education* 80 (2015), 152–161. DOI: https://doi.org/10.1016/j.compedu.2014.08.019

[48] Tim Harries, Parisa Eslambolchilar, Ruth Rettie, Chris Stride, Simon Walton, and Hugo C. Van Woerden. 2016. Effectiveness of a smartphone app in increasing physical activity amongst male adults: A randomised controlled trial. *BMC Public Health* 16, 1 (2016), 1–10. DOI: https://doi.org/10.1186/s12889-016-3593-9

[49] Heather A. Hausenblas, Albert V. Carron, and Diane E. Mack. 1997. Application of the theories of reasoned action and planned behavior to exercise behavior: A meta-analysis. *Journal of Sport and Exercise Psychology* 19, 1 (1997), 36–51. DOI: https://doi.org/10.1123/jsep.19.1.36

[50] HealthyWage. 2019. HealthyWage: weight loss challenges with cash prices. Retrieved from https://www.healthywage.com/.

[51] Donald Hedeker and Robert D. Gibbons. 2006. *Longitudinal Data Analysis*. Vol. 451. John Wiley & Sons, Hoboken, New Jersey.

[52] Helsana+. 2019. Helsana+ Programme de bonus. Retrieved from https://www.helsana.ch.
[53] Rebecca Hewett and Neil Conway. 2016. The undermining effect revisited: The salience of everyday verbal rewards and self-determined motivation. *Journal of Organizational Behavior* 37, 3 (2016), 436–455. DOI: https://doi.org/10.1002/job.2051
[54] Robert Hurling, Michael Catt, Marco De Boni, Bruce William Fairley, Tina Hurst, Peter Murray, Alannah Richardson, and Jaspreet Singh Sodhi. 2007. Using internet and mobile phone technology to deliver an automated physical activity program: Randomized controlled trial. *Journal of Medical Internet Research* 9, 2 (2007), 88–93. DOI: https://doi.org/10.2196%2Fjmir.9.2.e7
[55] Masataka Iwane, Mikio Arita, Shigehiro Tomimoto, Osamu Satani, Masanobu Matsumoto, Kazuhisa Miyashita, and Ichiro Nishio. 2000. Walking 10,000 steps/day or more reduces blood pressure and sympathetic nerve activity in mild essential hypertension. *Hypertension Research* 23, 6 (2000), 573–580. DOI: https://doi.org/10.1291/hypres.23.573
[56] Robert W. Jeffery, Rena R. Wing, Carolyn Thorson, and Lisa R. Burton. 1998. Use of personal trainers and financial incentives to increase exercise in a behavioral weight-loss program. *Journal of Consulting and Clinical Psychology* 66, 5 (1998), 777. DOI: https://doi.org/10.1037/0022-006X.66.5.777
[57] Hayeon Jeong, Heepyung Kim, Rihun Kim, Uichin Lee, and Yong Jeong. 2017. Smartwatch wearing behavior analysis. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–31. DOI: https://doi.org/10.1145/3131892
[58] Daniel Kahneman and Amos Tversky. 1979. On the interpretation of intuitive probability: A reply to jonathan cohen. *Cognition* 7, 4 (1979), 409–411. DOI: https://doi.org/10.1016/0010-0277(79)90024-6
[59] Bang Hyun Kim and Karen Glanz. 2013. Text messaging to motivate walking in older african americans: A randomized controlled trial. *American Journal of Preventive Medicine* 44, 1 (2013), 71–75. DOI: https://doi.org/10.1016/j.amepre.2012.09.050
[60] Florence Emilie Kinnafick, Cecilie Thøgersen-Ntoumani, and Joan Duda. 2016. The effect of need supportive text messages on motivation and physical activity behaviour. *Journal of Behavioral Medicine* 39, 4 (2016), 574–586. DOI: https://doi.org/10.1007/s10865-016-9722-1
[61] Russell V. Lenth. 2016. Least-squares means: The R package `lsmeans`. *Journal of Statistical Software* 69, 1 (2016), 1–33. DOI: https://doi.org/10.18637/jss.v069.i01
[62] Chung Tse Liu and Chia Tai Chan. 2016. Exercise performance measurement with smartphone embedded sensor for well-being management. *International Journal of Environmental Research and Public Health* 13, 10 (2016), 1–13. DOI: https://doi.org/10.3390/ijerph13101001
[63] Edwin A. Locke and Gary P. Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* 57, 9 (2002), 705–717. DOI: https://doi.org/10.1037/0003-066X.57.9.705
[64] George Loewenstein, David A. Asch, and Kevin G. Volpp. 2013. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs* 32, 7 (2013), 1244–1250. DOI: https://doi.org/10.1377/hlthaff.2012.1163
[65] George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003. Projection bias in predicting future utility. *The Quarterly Journal of Economics* 118, 4 (2003), 1209–1248. DOI: http://www.jstor.org/stable/25053938
[66] David Markland. 1999. Self-determination moderates the effects of perceived competence on intrinsic motivation in an exercise setting. *Journal of Sport and Exercise Psychology* 21, 4 (1999), 351–361. DOI: https://doi.org/10.1123/jsep.21.4.351
[67] Edward McAuley, Terry Duncan, and Vance V. Tammen. 1989. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport* 60, 1 (1989), 48–58. DOI: https://doi.org/10.1080/02701367.1989.10607413
[68] Edward McAuley, Gerald J. Jerome, David X. Marquez, Steriani Elavsky, and Bryan Blissmer. 2003. Exercise self-efficacy in older adults: Social, affective, and behavioral influences. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 25, 1 (2003), 1–7. DOI: https://doi.org/10.1207/S15324796ABM2501_01
[69] Marc S. Mitchell, Jack M. Goodman, David A. Alter, Leslie K. John, Paul I. Oh, Maureen T. Pakosh, and Guy E. Faulkner. 2013. Financial incentives for exercise adherence in adults: Systematic review and meta-analysis. *American Journal of Preventive Medicine* 45, 5 (2013), 658–667. DOI: https://doi.org/10.1016/j.amepre.2013.06.017
[70] Arlen C. Moller, H. McFadden, Donald Hedeker, and Bonnie Spring. 2012. Financial motivation undermines maintenance in an intensive diet and activity intervention. *Journal of Obesity* 2012 (2012), 810–827. DOI: https://doi.org/10.1155/2012/740519
[71] Frederiki C. Moustaka, Symeon P. Vlachopoulos, Chris Kabitsis, and Yannis Theodorakis. 2012. Effects of an autonomy-supportive exercise instructing style on exercise motivation, psychological well-being, and exercise attendance in middle-age women. *Journal of Physical Activity and Health* 9, 1 (2012), 138–150. DOI: https://doi.org/10.1123/jpah.9.1.138
[72] Dariush Mozaffarian, Emelia J. Benjamin, Alan S. Go, Donna K. Arnett, Michael J. Blaha, Mary Cushman, Sarah De Ferranti, Jean-Pierre Després, Heather J. Fullerton, Virginia J. Howard, et al. 2015. Heart disease and stroke

statistics–2015 update: A report from the american heart association. *Circulation* 131, 4 (2015), e29–e322. DOI: https://doi.org/10.1161/CIR.0000000000000152

[73] Elaine Mullan, David Markland, and David K. Ingledew. 1997. A graded conceptualisation of self-determination in the regulation of exercise behaviour: Development of a measure using confirmatory factor analytic procedures. *Personality and Individual Differences* 23, 5 (1997), 745–752.

[74] Meaghan Nolan, J. Ross Mitchell, and Patricia K. Doyle-Baker. 2014. Validity of the apple iphone®/ipod touch®as an accelerometer-based physical activity monitor: A proof-of-concept study. *Journal of Physical Activity and Health* 11, 4 (2014), 759–769. DOI: https://doi.org/10.1123/jpah.2011-0336

[75] Ted O'Donoghue and Matthew Rabin. 2000. The economics of immediate gratification. *Journal of Behavioral Decision Making* 13, 2 (2000), 233.

[76] Daniel Oesch. 2006. Coming to grips with a changing class structure: An analysis of employment stratification in britain, germany, sweden and switzerland. *International Sociology* 21, 2 (2006), 263–288. DOI: https://doi.org/10.1177/0268580906061379

[77] W. L. Quint Oga-Baldwin, Yoshiyuki Nakata, Philip Parker, and Richard M. Ryan. 2017. Motivating young language learners: A longitudinal model of self-determined motivation in elementary school foreign language classes. *Contemporary Educational Psychology* 49 (2017), 140–150. DOI: https://doi.org/10.1016/J.CEDPSYCH.2017.01.010

[78] Rita Orji, Regan L. Mandryk, and Julita Vassileva. 2017. Improving the efficacy of games for change using personalization models. *ACM Transactions on Computer-Human Interaction* 24, 5 (2017), 1–22. DOI: https://doi.org/10.1145/3119929

[79] Rita Orji, Lennart E. Nacke, and Chrysanne Di Marco. 2017. Towards personality-driven persuasive health games and gamified systems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, ACM, New York, NY, 1015–1027. DOI: https://doi.org/10.1145/3025453.3025577

[80] Rita Orji, Julita Vassileva, and Regan L. Mandryk. 2014. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction* 24, 5 (2014), 453–498. DOI: https://doi.org/10.1007/s11257-014-9149-8

[81] Terry D. Orlick and Richard Mosher. 1978. Extrinsic awards and participant motivation in a sport related task. *International Journal of Sport Psychology* 9, 1 (1978), 27–39.

[82] Katherine B. Owen, Jordan Smith, David R. Lubans, Johan Y. Y. Ng, and Chris Lonsdale. 2014. Self-determined motivation and physical activity in children and adolescents: A systematic review and meta-analysis. *Preventive Medicine* 67 (2014), 270–279. DOI: https://doi.org/10.1016/j.ypmed.2014.07.033

[83] Mitesh S. Patel, David A. Asch, Roy Rosin, Dylan S. Small, Scarlett L. Bellamy, Jack Heuer, Susan Sproat, Chris Hyson, Nancy Haff, Samantha M. Lee, Lisa Wesby, Karen Hoffer, David Shuttleworth, Devon H. Taylor, Victoria Hilbert, Jingsan Zhu, Lin Yang, Xingmei Wang, and Kevin G. Volpp. 2016. Framing financial incentives to increase physical activity among overweight and obese adults. *Annals of Internal Medicine* 164, 6 (2016), 385. DOI: https://doi.org/10.7326/M15-1635

[84] Mitesh S. Patel, Luca Foschini, Kurtzman Gregory W., et al. 2017. Using wearable devices and smartphones to track physical activity: Initial activation, sustained use, and step counts across sociodemographic characteristics in a national sample. *Annals of Internal Medicine* 167, 10 (2017), 755–757. DOI: http://dx.doi.org/10.7326/M17-1495

[85] Mitesh S. Patel, Kevin G. Volpp, Roy Rosin, Scarlett L. Bellamy, Dylan S. Small, Jack Heuer, Susan Sproat, Chris Hyson, Nancy Haff, Samantha M. Lee, Lisa Wesby, Karen Hoffer, David Shuttleworth, Devon H. Taylor, Victoria Hilbert, Jingsan Zhu, Lin Yang, Xingmei Wang, and David A. Asch. 2018. A Randomized, Controlled Trial of Lottery-Based Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults. *American Journal of Health Promotion* 32, 7 (2018), 1568-1575. DOI: https://doi.org/10.1177/0890117118758932

[86] Robert W. Plant and Richard M. Ryan. 1985. Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality* 53, 3 (1985), 435–449. DOI: https://doi.org/10.1111/j.1467-6494.1985.tb00375.x

[87] Stephen R. Porter and Michael E. Whitcomb. 2003. The impact of lottery incentives on student survery response rates. *Public Opinion Quarterly* 44, 4 (2003), 389–407. DOI: https://doi.org/10.1086/378964

[88] James O. Prochaska and Carlo Diclemente. 1983. Stages and processes of self- change of smoking - toward an integrative model of change. *Journal of Consulting and Clinical Psychology* 51, 3 (1983), 390–395. DOI: https://doi.org/10.1037//0022-006X.51.3.390

[89] Hua Qin, Pei-Luen Patrick Rau, and Gavriel Salvendy. 2010. Effects of different scenarios of game difficulty on player immersion. *Interacting with Computers* 22, 3 (2010), 230–239. DOI: https://doi.org/10.1016/j.intcom.2009.12.004

[90] Michael Reinboth and Joan L. Duda. 2006. Perceived motivational climate, need satisfaction and indices of well-being in team sports: A longitudinal perspective. *Psychology of Sport and Exercise* 7, 3 (2006), 269–286. DOI: https://doi.org/10.1016/j.psychsport.2005.06.002

[91] Barbara Resnick and Ann Marie Spellbring. 2000. Understanding what motivates older adults to exercise. *Journal of Gerontological Nursing* 26, 3 (2000), 34–42. DOI: https://doi.org/10.3928/0098-9134-20000301-08
[92] R. P. Riemsma, J. Pattenden, Christopher Bridle, A. J. Sowden, L. Mather, I. S. Watt, A. Walker, et al. 2002. A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change. *Health Technology Assessment* 24, 6 (2002), 1–231. DOI: https://doi.org/10.3310/hta6240
[93] Scott Rigby and Richard M. Ryan. 2011. *Glued to Games: How Video Games Draw us in and Hold us Spellbound.* Praeger/ABC-CLIO, Santa Barbara, CA. xiii, 186–xiii, 186 pages.
[94] Runkeeper. 2019. Runkeeper: Everyone, every run. Retrieved from https://runkeeper.com/.
[95] Richard M. Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology* 43, 3 (1982), 450–461. DOI: https://doi.org/10.1037/0022-3514.43.3.450
[96] Richard M. Ryan. 1993. Agency and organization: Intrinsic motivation, autonomy, and the self in psychological development. In *Proceedings of the Nebraska Symposium on Motivation, 1992: Developmental Perspectives on Motivation.* University of Nebraska Press, Lincoln, NE, 1–56.
[97] Richard M. Ryan, James P. Connell, and Robert W. Plant. 1990. Emotions in nondirected text learning. *Learning and Individual Differences* 2, 1 (1990) 1–17. DOI: https://doi.org/10.1016/1041-6080(90)90014-8
[98] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist* 55, 1 (2000), 68–78. DOI: https://doi.org/10.1037/0003-066X.55.1.68
[99] Richard M. Ryan and Edward L. Deci. 2000. When rewards compete with nature. In *Intrinsic and Extrinsic Motivation.* Elsevier, 125 London Wall, 125 London Wall, London EC2Y 5AJ, UK, 13–54. DOI: https://doi.org/10.1016/B978-012619070-0/50024-6
[100] Richard M. Ryan and Edward L. Deci. 2007. *Active Human Nature: Self-determination Theory and the Promotion and Maintenance of Sport, Exercise, and Health.* Human Kinetics Publishers, 107 Bradford Road Stanningley, Leeds LS28 6AT, United Kingdom. 1–19 pages.
[101] Richard M. Ryan and Edward L. Deci. 2017. *Self-Determination Theory Basic Psychological Needs in Motivation, Development and Wellness.* THE GUILFORD PRESS, 370 Seventh Avenue, Suite 1200, New York, NY 10001. 770 pages. Retrieved from https://www.guilford.com/books/Self-Determination-Theory/Ryan-Deci/9781462528769.
[102] Richard M. Ryan and Christina Frederick. 1997. On energy, personality, and health: Subjective vitality as a dynamic reflection of well-being. *Journal of Personality* 65, 3 (1997), 529–565. DOI: https://doi.org/10.1111/j.1467-6494.1997.tb00326.x
[103] Richard M. Ryan, Richard Koestner, and Edward L. Deci. 1991. Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotion* 15, 3 (1991), 185–205. DOI: https://doi.org/10.1007/BF00995170
[104] Richard M. Ryan, Valerie Mims, and Richard Koestner. 1983. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology* 45, 4 (1983), 736. DOI: https://doi.org/10.1037/0022-3514.45.4.736
[105] James F. Sallis, Robin M. Grossman, Robin B. Pinski, Thomas L. Patterson, and Philip R. Nader. 1987. The development of scales to measure social support for diet and exercise behaviors. *Preventive Medicine* 16, 6 (1987), 825–836. DOI: https://doi.org/10.1016/0091-7435(87)90022-3
[106] Ronald E. Smith and Irwin G. Sarason. 1975. Social anxiety and the evaluation of negative interpersonal feedback. *Journal of Consulting and Clinical Psychology* 43, 3 (1975), 429. DOI: https://doi.org/10.1037/h0076855
[107] Martyn Standage, Simon J. Sebire, and Tom Loney. 2008. Does exercise motivation predict engagement in objectively assessed bouts of moderate-intensity exercise?: A self-determination theory perspective. *Journal of Sport and Exercise Psychology* 30, 4 (2008), 337–352. DOI: https://doi.org/10.1123/jsep.30.4.337
[108] SweatCoin. 2019. SweatCoin: Sweat for coins. Retrieved from https://sweatco.in.
[109] M. Szumilas. 2010. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 19, 3 (2010), 227–229.
[110] Pedro J. Teixeira, Eliana V. Carraça, David Markland, Marlene N. Silva, and Richard M. Ryan. 2012. Exercise, physical activity, and self-determination theory: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity* 9, 1 (2012), 78. DOI: https://doi.org/10.1186/1479-5868-9-78
[111] Cecilie Thøgersen-Ntoumani and Nikos Ntoumanis. 2006. The role of self-determined motivation in the understanding of exercise-related behaviours, cognitions and physical self-evaluations. *Journal of Sports Sciences* 24, 4 (2006), 393–404. DOI: https://doi.org/10.1080/02640410500131670
[112] Debbe Thompson, Dora Cantu, Riddhi Bhatt, Tom Baranowski, Wendy Rodgers, Russell Jago, Barbara Anderson, Yan Liu, Jason A Mendoza, Ramsey Tapia, and Richard Buday. 2014. Texting to increase physical activity among teenagers (TXT Me!): Rationale, design, and methods proposal. *JMIR Research Protocols* 3, 1 (2014), e14. DOI: https://doi.org/10.2196/resprot.3074

[113] N. Tsigilis. 2005. The influence of intrinsic motivation on an endurance field test. *The Journal of Sports Medicine and Physical Fitness* 45, 2 (2005), 213–6. DOI: http://www.ncbi.nlm.nih.gov/pubmed/16355083
[114] Catrine Tudor-Locke and David R. Bassett. 2004. How many steps/day are enough? *Sports Medicine* 34, 1 (2004), 1–8. DOI: https://doi.org/10.2165/00007256-200434010-00001
[115] WayBetter. 2019. DietBet. Retrieved from https://www.dietbet.com/.
[116] Philip M. Wilson and Wendy M. Rodgers. 2004. The relationship between perceived autonomy support, exercise regulations and behavioral intentions in women. *Psychology of Sport and Exercise* 5, 3 (2004), 229–242. DOI: https://doi.org/10.1016/S1469-0292(03)00003-7
[117] S. N. Wood. 2017. *Generalized Additive Models: An Introduction with R, Second Edition.* CRC Press, Boca Raton, Florida.
[118] Tongtong Xu, Ao Guo, and Jianhua Ma. 2017. Analysis of temporal features in data streams from multiple wearable devices. In *Proceedings of the 3rd IEEE International Conference on Cybernetics.* IEEE, IEEE Operations Center 445 Hoes Lane Piscataway, NJ 08854-4141, 1–6. DOI: https://doi.org/10.1109/CYBConf.2017.7985758

Received July 2018; revised June 2019; accepted September 2019