Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Gain-Loss Incentives and Physical Activity: The Role of Choice and Wearable Health Tools

Idris Adjerid, George Loewenstein, Rachael Purta, Aaron Striegel

http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 68, No. 4, April 2022, pp. 2642–2667
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Gain-Loss Incentives and Physical Activity: The Role of Choice and Wearable Health Tools

Idris Adjerid,[a] George Loewenstein,[b] Rachael Purta,[c] Aaron Striegel[c]

[a] Business Information Technology, Pamplin College of Business, Virginia Tech, Blacksburg, Virginia 24061; [b] Social and Decision Sciences, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213; [c] Computer Science and Engineering, College of Engineering, University of Notre Dame, South Bend, Indiana 46556

Contact: iadjerid@vt.edu, https://orcid.org/0000-0002-2786-1244 (IA); gl20@andrew.cmu.edu (GL); rachaelpurta@gmail.com (RP); striegel@nd.edu (AS)

Received: April 2, 2019
Revised: June 28, 2020; December 7, 2020
Accepted: January 22, 2021
Published Online in Articles in Advance: May 28, 2021

https://doi.org/10.1287/mnsc.2021.4004

Abstract. Economic incentives are a promising approach for improving health behavior but have been limited by their short-lived benefits. In this manuscript, we examine whether coupling economic incentives with motivational tools provided by health wearables can address this limitation and drive longer-term changes in health behavior. We focus on "gain-loss" incentive schemes that offer both an economic reward for goal attainment and a penalty for failure to meet a goal. In an experiment conducted on individuals wearing Fitbit wearables, we find that gain-loss incentives can drive increases in physical activity but are limited by the element of choice. Specifically, we find modest and short-lived increases in physical activity for those provided the choice of gain-loss incentives. The subpar benefits for this group seem to emerge because those who benefit most from these schemes do not opt into them when they are voluntary. In contrast, we find significant and persistent increases in physical activity for those assigned (oftentimes against their preference) to the same gain-loss incentives. These individuals recorded ~2,000 additional steps daily during the incentive period, and benefits persisted for six months after incentives ended. Critically, the persistent gains to this group were driven by individuals who also utilized the wearable's goal-setting tool. Our results suggest that a novel approach toward motivating sustained changes in health behavior couples aggressive incentive schemes that jolt individuals out of their comfort zone in the short term with motivational tools built into health wearables that help individuals sustain healthy behavior after economic incentives end.

History: Accepted by Kartik Hosanagar, information systems.
Funding: This work was supported by the National Health, Lung, and Blood Institute of the National Institutes of Health [Grant 5R01HL117757-04].
Supplemental Material: Data and the online appendices are available at https://doi.org/10.1287/mnsc.2021.4004.

Keywords: healthcare • field experiment • health wearable • behavioral economics • technology health tools

1. Introduction

The World Health Organization estimates that one in four adults worldwide and a staggering 80% of the world's adolescent population are not sufficiently active.[1] This lack of physical activity contributes to most of the leading causes of death in the United States (Booth and Chakravarthy 2002, Booth et al. 2017). Recognizing that behavioral factors are a critical determinant of this situation, research in behavioral economics has explored incentive-based interventions to increase motivation and help individuals overcome systematic barriers to improving their health behavior (Charness and Gneezy 2009, Gneezy et al. 2011). Although these approaches have shown promise, health gains from economic incentives have not typically persisted after incentives end (Volpp et al. 2008, Gneezy et al. 2011, Mitchell et al. 2013). Given that it is typically infeasible to offer economic incentives in perpetuity, this is a critical limitation of incentive-based interventions for health.

This limitation provides the impetus for the current research. Specifically, scholars have conjectured that leveraging emerging health technologies alongside incentive-based interventions has the potential to yield more sustained changes in health behavior (relative to incentive-only interventions). For example, Rogers et al. (2014, p. 2066) suggest that effective interventions require "novel strategies for maintaining people's engagement and attention at key moments when drop-off or discontinuation is a risk." They suggest that one way to achieve this is by "leveraging advances in wireless technologies to make dynamic feedback easier, cheaper, and timelier." Relatedly, Mitchell et al. (2013, p. 666) argue that advances in health wearables and mobile health may be "leveraged to more promptly assess and reward behaviors on a population scale, further reducing the need for prohibitively costly incentives." However, to our knowledge, research along these lines has been scant. We address this knowledge gap by testing a novel intervention that combines economic incentives tied to physical activity with motivational tools enabled by popular health wearables (activity tracker, smart watches, etc.).

The first aspect of our intervention is the economic incentives. We drew inspiration for the design of these incentives from promising approaches that provide individuals the choice to impose negative consequences on themselves if they fail to meet health goals in the longer term (Volpp et al. 2008, Giné et al. 2010, John et al. 2011, Royer et al. 2015). For example, Royer et al. (2015) offer individuals "deposit contracts" whereby they place their own money into an account and are only allowed to keep it if they continue to attend the gym regularly for six months after initial incentives end. We offer a variant of these incentive approaches that, like prior work, includes penalties for failures of goal attainment but, unlike most prior work, also includes economic rewards for goal attainment (effectively coupling a carrot with the stick). We henceforth refer to our modified incentive scheme as "gain-loss incentives." The second aspect of our intervention is the use of motivational tools built into health wearables. Specifically, we focus on goal-setting tools that allow individuals to set specific physical activity targets, continuously observe progress toward these targets, and receive positive reinforcement when the target is achieved. To our knowledge, the ability to set physical activity targets is common to all major health wearable platforms and is one of the most salient and commonly used health tools for adopters of wearable devices (Mercer et al. 2016).

With these two aspects of our intervention in mind, our first research objective is to examine whether the combination of gain-loss economic incentive and wearable health tools leads to meaningful and sustained changes in health behavior. Ex ante, we conjectured that exposure to the more aggressive gain-loss incentives would improve health behavior (relative to standard gain-only incentive approaches) in the short run. Moreover, we conjectured that wearable goal-setting tools would help individuals avoid reverting to previous levels of physical activity after gain-loss incentives ended. With such wearable tools remaining in use, instead of having an abrupt end to positive reinforcement when payments stop, individuals will continue to be reminded, through their health wearable, of the higher activity goal set for them, observe their prior success in increasing their physical activity, and continue to receive positive daily reinforcement from the wearable device and associated mobile application when they meet the goal after incentives end.

Our second objective is to evaluate whether the effects of gain-loss incentives and wearable health tools are heterogeneous for individuals who prefer the standard gain-only incentive over the gain-loss one. Although more aggressive incentive schemes (e.g., gain-loss incentive) have proven useful for those who choose them, a challenge with these approaches is that a small and nonrepresentative sector of the population typically participates (Giné et al. 2010; Gneezy et al. 2011; Halpern et al. 2015, 2016; Royer et al. 2015).[2] This concern is especially pronounced in the context of programs encouraging physical exercise, as prior work suggests that economic incentives are most effective for those who were previously sedentary (Charness and Gneezy 2009, Mitchell et al. 2013). Thus, we conjecture that exposure to gain-loss incentives will be particularly beneficial for those who would not choose them on their own. Moreover, prior work suggests that goal setting and tracking provided by wearables is particularly valuable when individuals are exposed to gain-loss incentives that impose ambitious health goals that they would not undertake on their own (Becker 1978, Bandura and Cervone 1983, Erez and Zidon 1984, Locke and Latham 1990). Specifically, individuals exposed to gain-loss incentives (potentially against their preference) may require the additional support of wearable health tools to achieve health goals they wouldn't otherwise be motivated to meet or would likely perceive to be too difficult.

To meet our research objectives, we conducted a field experiment over a 29-week (~seven month) period during which we had access to daily data on steps walked, as captured by Fitbit Charge HR activity trackers. In the key part of the experiment, individuals participated in a week-long "Step Challenge" in which they attempted to meet a new step goal for seven days. During this period, participants in the treatment conditions were offered additional incentives to meet their new health goal. The experimental design involved first eliciting the preference of participants for either a standard gain-only incentive or the more aggressive gain-loss incentive and then randomly assigning them to either the gain-only or gain-loss incentive scheme (regardless of the scheme they preferred) or giving them their choice. The elicitation of preferences for gain-only versus gain-loss incentives prior to either randomizing individuals into one of the schemes or providing them with their choice of incentive schemes is a unique aspect of our design that allows us to compare the impact on behavior of giving people a choice of incentive schemes as compared with assigning them (randomly) to either their preferred or nonpreferred incentive scheme.

In parallel to the economic incentives offered, we provided some randomly selected participants with a prompt encouraging them to update the step target in the Fitbit application to match the new step goal we provided them. The random prompt encourages individuals to update their Fitbit step target and provides exogenous variation in the use of this wearable health tool. This variation allows us to examine heterogeneity in the effects of incentives depending on wearable health tool use.

We find that, during the challenge week, participants who were either provided their choice or were assigned to the standard gain-only incentive scheme saw modest benefits and walked 770 and 955 steps (respectively) more a day compared with participants not provided an economic incentive. In contrast, individuals assigned to the gain-loss scheme walked 2,068 more steps per day compared with the no incentive group. These additional steps for individuals assigned to the gain-loss incentives translate into one additional mile of physical activity daily during the challenge week.[3] More striking is that the effect for those assigned to the gain-loss incentive persisted well after the incentives ended. Participants who were assigned to this scheme continued to outpace those without an initial economic incentive during the 24 weeks (~six months) after the challenge ended. In contrast, participants who were assigned to the standard gain-only incentives, or were provided a choice, did not outperform the no-incentive group during the same period. We find that the use of the wearable health tool explains this large and persistent effect for those assigned to gain-loss incentives. Specifically, those who were both assigned to the gain-loss incentives and randomly prompted to match the step target on the Fitbit wearable with their new step goal saw considerably more benefit during the incentive period. Moreover, the persistence of physical activity in the long term for those assigned to gain-loss incentives was driven by individuals who were also prompted to update their Fitbit step target.

Taking into account whether an individual prefers the gain-loss incentives further informs these results. We find that the performance of those who were assigned to the gain-loss incentives emerges for two reasons. First, participants who prefer the standard gain-only incentive scheme perform poorly when they can choose between incentive schemes but perform significantly better when forced into the gain-loss incentive scheme. In contrast, those who prefer the gain-loss incentives perform well regardless of whether they are provided this incentive scheme (by choice or random assignment) and even if they are provided the standard gain-only incentives. Thus, this subgroup does not benefit from being provided a choice. These latter results point to an interesting phenomenon whereby those who are willing to take on more motivational incentive schemes may actually not need them to perform well. Second, wearable health tools were particularly useful for those who preferred the standard gain-only incentives but who were then thrust into gain-loss incentives. Interestingly, we don't identify consistent benefits of wearable health tools for those assigned to the standard gain-only incentives or those provided their choice.

We validated the heterogenous impacts of gain-loss incentives for those who prefer the gain-only versus gain-loss schemes in a second experiment on Amazon Mechanical Turk (AMT). The design of this experiment mirrored the field experiment but presented participants with a series of numerical puzzles and set a performance goal in terms of the number of puzzles to solve correctly. Again, we find evidence that when provided a choice between incentive schemes, people who preferred gain-only incentives but were assigned to the gain-loss incentives saw the largest improvements to performance. In contrast, those who preferred the gain-loss incentives performed well independent of the incentive scheme they were provided.

Our interpretation of these results is that when people are provided with the option of working under gain-loss incentives, the wrong people—those who do not in fact benefit from them—choose them over conventional gain-only incentives. Those who would have benefited more from the gain-loss scheme are more likely to eschew gain-loss incentives in favor of standard gain-only incentives. One explanation for the disproportionate benefit of wearable health tools for those who were assigned into the gain-loss incentives is that individuals in this condition are pushed to achieve health goals that they would otherwise perceive to be infeasible or difficult. Thus, they were more motivated to effectively use the health tool to achieve health goals; their success in the challenge week could have been particularly motivating, changing their beliefs about what they could do. Overall, these results suggest that one effective strategy for encouraging sustained health behavior combines mandatory gain-loss incentives to jolt complacent individuals out of their comfort zone with motivational tools built into wearable devices that help individuals sustain good behavior after incentives end. These results extend research on health wearables and incentive design for health and have implications for entities (e.g., employers, policy makers) interested in promoting sustained changes in health behavior.

Our findings contribute to the behavioral economics literature on the tension surrounding the benefits of offering individuals the choice to opt-in to punishing mechanisms for regulating their own behavior (Giné et al. 2010, Gneezy et al. 2011, Royer et al. 2015). Specifically, our results suggest that providing individuals the option to opt-in to more aggressive incentive schemes may not always be optimal, as those who choose gain-loss economic incentives may be exactly the people who least benefit from them. Our results help to explain why recent work finds limited benefits from individuals voluntarily subjecting themselves to economically punishing incentive schemes (Larkin and Leider 2012, Halpern et al. 2015, Chaudhry and Klinowski 2016). Our findings are particularly distinct from the extant literature finding benefits of offering choice of incentives in the workplace (Mellizo et al. 2014, Kaur et al. 2015) and highlight how contextual factors (e.g., less extrinsic motivators in the physical activity context compared with the workplace) can lead to different effects. Moreover, we contribute to the behavioral economics literature by exploring longer-term effects of behavioral interventions in the field—a point of significant focus in the literature (Levitt and List 2007, 2009; Gneezy et al. 2011; Chen and Konstan 2015).

Second, our paper contributes to the literature evaluating the impact of mobile health interventions on health outcomes (Dugas et al. 2018, Sun et al. 2019) and the literature evaluating the impact of wearables on health outcomes (Finkelstein et al. 2016, Jakicic et al. 2016, Piwek et al. 2016). Importantly, the literature focused on wearables has found limited health benefits from these devices (Finkelstein et al. 2016, Jakicic et al. 2016). Patel et al. (2015) suggest that one explanation for lackluter benefit is that individuals using these devices are not motivated enough to engage with them. To our knowledge, our manuscript is one of the first to consider how the design of economic incentives alongside the use of wearable health tools impacts health. By doing so, we inform this literature and suggest that carefully designed economic incentives can fill the initial motivation gap that limits the value of wearables and, by doing so, can help unlock health benefits from them. Specifically, our findings show that gain-loss incentives can be beneficial for those who would not typically choose them, if coupled with interventions that effectively increase engagement with health wearables.

Finally, our paper contributes to a burgeoning literature evaluating the intersection of experimental and behavioral economics on the one hand and information systems research on the other. Scholars argue that this intersection of research streams holds significant promise (Goes 2013, Gupta et al. 2018). Recent works draw from experimental economics to evaluate the dynamics of digital piracy (Hashim et al. 2017, Hashim et al. 2018), the intersection of behavioral economics and privacy (Adjerid et al. 2019), and the impact of economic incentives on user-generated content (Burtch et al. 2017, Khern-am-nuai et al. 2018). We extend this literature by highlighting complementarities between motivational technology tools and incentive-based interventions.

2. Conceptual Background: Incentives and Health Wearables

In this section, we review the relevant findings in the behavioral economics, healthcare, and information systems literature and articulate the research gaps and theoretical tensions that motivate our research objectives.

2.1. Health Wearables and Economic Incentives

The first stream of work with relevance to this manuscript is from behavioral economics and focuses on the impact of economic incentives on health behavior (Volpp et al. 2008, Charness and Gneezy 2009, Giné et al. 2010, Mitchell et al. 2013, Royer et al. 2015). Generally, this body of work demonstrates the significant potential of incentive-based interventions to increase motivation and help individuals overcome systematic barriers to improving their health behavior (Giné et al. 2010, Royer et al. 2015). However, this literature finds that the benefits of incentives in healthcare are not always evenly distributed: for example, Charness and Gneezy (2009) found that the effect of economic incentives on gym attendance is driven wholly by those who had not been attending previously. Moreover, economic incentives are not intended as a long-term health intervention. Rather, the hope is that incentives kickstart the intended health behavior and, after individuals experience the positive aspects of a healthy lifestyle, their motivation will increase enough to help them continue their improved habits even without the extrinsic motivation that incentives provide (Gneezy et al. 2011). This hope has not materialized, as in several studies the health gains from economic incentives have not persisted after incentives end (Volpp et al. 2008, Gneezy et al. 2011, Mitchell et al. 2013).

To improve the consistency and longevity of health benefit from economic incentives, recent work has evaluated the impact of incentive schemes that provide individuals the option to impose negative consequences on themselves if they fail to achieve desired goals (Volpp et al. 2008, Giné et al. 2010, Royer et al. 2015). The premise behind these approaches is that individuals will sustain good health behavior for longer periods if they "have skin in the game" and place their own income at stake. This premise is substantiated by the well-documented phenomenon of "loss aversion" whereby individuals are more motivated to change behavior by a potential loss compared with an equal-sized potential gain (Tversky and Kahneman 1991). In this paper, we consider the impact of gain-loss incentive schemes that include *both* an additional reward for meeting health goals and an economic punishment for failure to meet the goal.

The second stream of work with relevance to our manuscript relates to the impacts of wearables on health. Modern wearables are multisensor, with the ability to measure heart rate, steps, sleep, energy expenditure, floors climbed, and more. In addition to health tracking, wearables have associated digital platforms that offer several tools aimed at improving health behaviors. These include the ability to set physical activity targets, observe progress toward these targets, and to compete with friends who are also on the platform. Distinct from other mobile health interventions (e.g., health apps), the addition of the wearable device allows for relevant information and feedback to be regularly communicated to and easily accessed by the user throughout the day (even while engaged in the physical activity itself). For example, the wearable device can provide tactile feedback (e.g., vibrations) alongside wearable screen notifications when a daily step target is met.

Fueled by the promise of health gains (Swan 2013, Lupton 2016), health wearables have enjoyed the fastest rates of adoption of any consumer wearable category (PwC 2016). However, to our knowledge, the information systems literature on health wearables is still developing and rigorous evaluations of the impact of wearable health tools on health outcomes are sparse. One recent work by James et al. (2019) does not evaluate health outcomes but considers the likelihood of use (and by extension) benefit from wearable device functionality conditional on the type of user. They find evidence that those high in intrinsic motivation would be attracted to social interaction features (e.g., wearable leaderboards), whereas those low in intrinsic motivation benefit most from an exercise control feature (step targets) of wearable devices. In addition to the information systems literature, a growing stream of work in healthcare and economics has started to evaluate the promised benefits of health wearables (Finkelstein et al. 2016, Jakicic et al. 2016, Piwek et al. 2016, Handel and Kolstad 2017). Notably, these works have failed to uncover consistent evidence of anticipated health gains. In a year-long randomized control trial, Jakicic et al. (2016) found no additional health benefit for participants provided with a wearable device. Handel and Kolstad (2017) conducted a large-scale experiment in which they manipulated access to tools that utilize wearable data to monitor performance and provide plans to improve health. They found that, among wearable users, those who had access to these tools did not exhibit significant increases in physical activity. Finkelstein et al. (2016) found that activity trackers resulted in no meaningful gains to steps walked compared with the control group. In their review of the literature, Piwek et al. (2016, p.2) suggest that "current empirical evidence is not supportive" of benefits from health wearables.

These streams of research highlight that, although both gain-loss economic incentives and health wearables hold promise, they also have important limitations and have shown uneven benefits. We conjecture that combining wearables with these novel incentive schemes can result in more significant and persistent changes in healthful behavior. Specifically, complementarities between gain-loss economic incentives and health tools enabled by health wearables can emerge for several reasons. Most critically, wearable health tools can help address the issue of fleeting benefits from economic incentives by continuing to reinforce good health behavior when the motivation from economic incentives wanes (Mitchell et al. 2013, Rogers et al. 2014). For example, Sullivan and Lachman (2017) suggest that health wearables have the potential to increase feelings of control over one's own exercise behaviors, leading to more sustained behavior change. Moreover, prior work finds that incentives can have a pronounced effect when individuals have additional support toward meeting a desired health goal. For example, Babcock and Hartman (2010) elicited individuals' social networks prior to implementing incentives for exercise and found that study participants benefited more from incentives when they had social support from others in their network. Because wearable health tools can provide a regular form of support for meeting health goals tied to incentives, they may similarly increase the efficacy of economic incentives.

In addition, economic incentives could be one of the missing pieces toward unlocking the value of wearables. Specifically, Patel et al. (2015) argue that one reason for lackluster benefits of wearables is that these devices are facilitators, not drivers, of health behaviors and that they can facilitate health gains only if individuals are already willing to engage with them. Given that economic incentives have been shown to improve health behavior in the short term, they may be a promising mechanism for providing an initial spark of engagement that then allows individuals to benefit more from wearable health tools. The psychological and behavioral momentum literature substantiates this conjecture and suggests that economic incentives that help individuals achieve health goals in the short term can provide psychological momentum that helps them persist in meeting health goals after incentives end (Nevin et al. 1983, Briki and Markman 2018). The health tools built into wearables can help maintain this momentum by reinforcing goal adherence and counteracting frictions (laziness, demotivation, etc.) that reduce an individual's positive psychological momentum.

To our knowledge, evaluation of complementarities between novel economic incentives and technology health tools has been scant. Finkelstein et al. (2016) address this intersection partially by evaluating the combination of a cash incentive and an activity tracker. They find that individuals provided this combination walk more steps compared with the control group but that this effect diminishes after incentives end. Their work is, however, limited because, lacking experimental conditions with cash incentives alone and activity trackers alone, they are not able to determine if there is any additional benefit from the combination of a tracker and economic incentives. This gap in the literature, and the potential for significant complementarities between incentives and health wearable technologies, motivates our first research objective focused on examining whether the combination of gain-loss economic incentive and wearable health tools leads to meaningful and sustained changes in health behavior.

2.2. Gain-Loss Incentives, Health Wearables, and the Dilemma of Choice

The benefits of combining aggressive incentives with health wearable health tools are not likely to accrue homogenously. Although aggressive incentive schemes that introduce a potential economic loss for low performers have proven effective for those who choose them (Giné et al. 2010, Royer et al. 2015), the element of choice in these schemes introduces a dilemma. On the one hand, choice makes these approaches easier to implement; forcing individuals into economically punishing incentive schemes is often infeasible or fails to attract willing participants (Halpern et al. 2015, 2016). In addition, those who choose their own incentive schemes may perform better because they have deliberately bought in to them (Charness et al. 2012, Bartling et al. 2014). This result has been validated in workplace contexts, where researchers find evidence that workers who have a say in their compensation outperform those who do not (Mellizo et al. 2014, Kaur et al. 2015). On the other hand, for these approaches to be broadly useful individuals must recognize their self-control problems and be willing to opt-in to economically punishing incentive schemes as a mechanism for regulating their own long-term health behavior. Even when a sizable fraction of individuals chooses to use them, prior work suggests that gains may be the result of initial individual differences rather than the scheme itself (Larkin and Leider 2012, Halpern et al. 2016).

Introducing gain-loss incentives to those who would not select them can alleviate some of these concerns. For example, one approach is to provide these incentives on a nonvoluntary basis and expose individuals to potentially beneficial incentive schemes they would not choose for themselves. In particular, incentive schemes with an economic punishment may more effectively motivate a departure from the status quo for individuals who would not voluntarily choose them (e.g., those with higher levels of loss aversion) and provide a proof of concept for them that they can meet an ambitious health goal. This conjecture is again consistent with theories of psychological momentum that posit that the positive psychological momentum gained from short-term goal attainment is highest for those for whom goal attainment is most difficult (Nevin et al. 1983, Briki and Markman 2018).

Of course, nonvoluntary incentive schemes introduce their own set of challenges. Primarily, it's not clear that exposing individuals to gain-loss incentive schemes they would not have chosen themselves is beneficial. For example, the benefits of these incentives could be limited because individuals have not bought into them (Charness et al. 2012, Bartling et al. 2014); gain-loss incentives that punish poor behavior could further demotivate individuals by making failures to meet health goals more salient. This tension highlights an important open question related to the design of more aggressive incentive-based interventions: Should incentive schemes that include economic punishments always be choice-based or can providing gain-loss incentives on a nonvoluntary basis be beneficial?

It is possible, in fact, that complementarities between aggressive incentives (e.g., gain-loss incentives) and wearable health tools may be most pronounced for individuals who would not choose them on their own. If individuals who avoid aggressive incentives do so because of an elevated degree of loss aversion (Tversky and Kahneman 1991), this subgroup may be more willing to fully engage and utilize wearable health tools when assigned to aggressive incentives (i.e., in order to avoid the associated loss). If individuals who avoid aggressive incentives do so because they perceive the goal to be difficult or are not motivated to meet the goal (or both), they may be most likely to need health tools that support their efforts. This conjecture is substantiated by the literature on goal setting and motivation, which finds that realistic but challenging goals generate the largest performance gains (Locke and Latham 1990). However, prior work also finds that performance gains from challenging goals can level off if commitment to the goal lapses (Erez and Zidon 1984). Fortunately, this literature also finds that measurement, continuous feedback, and reinforcement of the goal is critical to maintaining commitment to challenging goals and their associated performance gains (Becker 1978, Strang et al. 1978, Bandura and Cervone 1983). Wearable health tools that offer the ability to designate a health target and then observe progress toward that target on an accessible wearable device provide an excellent source of this type of feedback.

These theoretical tensions and potential for heterogenous benefit from combining health wearables and gain-loss incentives motive our second research objective: to evaluate whether the effects of gain-loss incentives and wearable health tools differ between individuals who prefer conventional gain-only incentives and those who prefer more aggressive gain-loss incentives.

3. Fitbit Field Experiment

The first experiment was in the field and focuses on the impact on physical activity of coupling gain-loss incentives with wearable health tools. The evidence on the health impacts of physical activity points to numerous benefits, including the prevention of chronic diseases (e.g., cardiovascular disease, diabetes, cancer, hypertension, obesity, and osteoporosis), reduction in the risk of premature death, and improved mental health and mood (Paluska and Schwenk 2000, Bauman 2004, Penedo and Dahn 2005). Importantly, research finds that even moderate levels of physical activity can be enough to see these promised benefits. The American College of Sports Medicine and the American Heart Association recommend a minimum of 30 minutes of brisk walking daily for adults aged 18–65 and suggest that this can substitute for more vigorous activity (Haskell et al. 2007). Morris and Hardman (1997) note that a brisk walk is the most natural activity and is the only sustained aerobic exercise that is accessible to everyone except for the seriously disabled. Lee and Buchner (2008) argue that regular physical activity has the potential to have a large public health impact because of its accessibility and its documented health benefits. Given the significant health value of any form of physical activity, we focus our intervention on an individual's total daily steps.

3.1. Experiment Population and Data

We drew participants in the field experiment from a population of students at a North American university who had been provided Fitbit Charge HR armbands as part of a long-running study (referred to henceforth as the "StudentHealth" study). At the time of our field experiment, this population included 402 active participants. The participants in this study were compensated monthly for wearing an activity tracker and allowing researchers access to the data from their device. Most individuals in the StudentHealth population had been using health wearables for a year prior to the experiment and were acclimated to the device and its functions. Individuals were also periodically surveyed in order to collect data on their demographic, health, and psychological features.

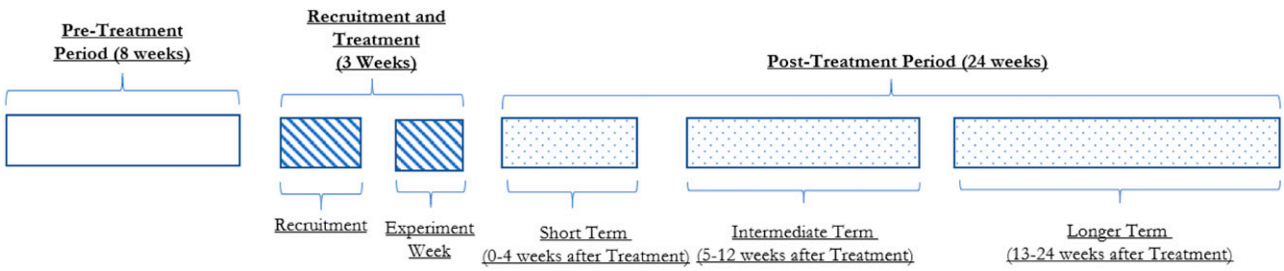
Participants in the StudentHealth population were compensated based on how frequently they wore their Fitbit device (i.e., compliance). During the field experiment period, participants were provided their full monthly compensation ($20) if they wore their Fitbit more than 40% of the time in each month. At the time of our field experiment, 341 participants were maintaining a 40% or higher compliance rate. Overall, participants in the field experiment maintained a high compliance rate during the 29 weeks of our study and wore their device, on average, 71% of each day. To maintain their compliance and ensure payment, individuals were responsible for syncing their device to Fitbit's servers in a timely manner. This could be done either by using the dongle provided with the Fitbit and the syncing program on their computer or through the Fitbit app. The Fitbit Charge HR is advertised to hold up to seven days of data, so syncing less frequently than every seven days could result in lost data and lower compliance. Once students synced their data, it became accessible to registered apps, with students having provided permission for the data to be accessed. To collect the data, we ran a nightly set of scripts that gathered the latest synced data through the Fitbit Application Programming Interface (API) and stored it in our secure database.[4] Our focus in this experiment is on participants' physical activity measured by the number of steps walked each day.

In terms of data quality, Fitbit is not entirely open about the methods used to calculate the various measurements its devices generate. However, researchers have evaluated their validity. The extant literature generally finds steps to be a reliable, valid measure across all tested Fitbit devices, with mostly underestimation in the cases where error was higher (Evenson et al. 2015, Wahl et al. 2017). Many of these studies have specifically evaluated the Charge HR, the Fitbit model used in our study, in not only laboratory but free-living conditions and with diverse populations. They find that the Charge HR provides step measurements with relatively low error, less than 10% in most cases (Fokkema et al. 2017, Wahl et al. 2017, Bai et al. 2018). The accuracy of Charge HR's heart rate measure is less clear from the literature, as some studies have found the Charge HR equivalent to criterion measures (Stahl et al. 2016, Bai et al. 2018) and some have not (Wang et al. 2017). Our study primarily uses step count data, with heart rate data only being used to determine compliance. Adopting the procedure from prior studies (e.g., Purta et al. 2016), if the participants registered a heart rate, they were considered as wearing the Fitbit Charge HR (i.e., compliant) but not otherwise. We also utilized the survey data to extract control variables for individuals in our study.

3.2. Experimental Design and Prestudy

We conducted the field experiment over a 29-week (~seven month) period. In the pretreatment period (eight weeks), we established a physical activity baseline for participants. We then informed participants of an upcoming week-long "Step Challenge" that would require them to try and meet a new step goal for seven days, randomized them into the experimental manipulations, and then conducted the actual step challenge for those who agreed to participate (recruitment and random assignment into conditions lasted for two weeks, whereas the step challenge lasted for one week). We compensated individuals $10 for participating in the challenge and, depending on the experimental condition, compensated them an additional amount if they met their new step goal. After the end of the treatment period, we continued to observe physical activity for participants during a 24-week posttreatment period. We parsed the posttreatment period into three nonoverlapping segments. The short term consisted of the four weeks immediately following the end of the challenge. The intermediate term consisted of the eight weeks following the end of the short term. The long term consisted of the 12 weeks following the intermediate time period (see Figure 1 for an overview of our experimental implementation). Granular data on physical activity prior to the

Figure 1. (Color online) Overview of Field Experiment Implementation



experimental period allows us to evaluate pretreatment trends and the extended period of postexperiment data allows us to evaluate long-term effects of our interventions.

We first manipulate the exposure to individuals of gain-loss versus conventional gain-only incentives. In the no-incentive group, we enrolled participants in the step challenge and provided them a new step goal to meet but did not provide them any additional incentive for meeting this new goal. We presented the remaining participants with two options for additional compensation (labeled as "Option 1" and "Option 2") and asked them to choose which option they preferred. Option 1 was the conventional gain-only incentive scheme that provided participants an additional amount every day they met their goal but no penalty if they did not meet their goal. Option 2 was the more aggressive gain-loss incentive scheme, which rewarded them a slightly higher amount (relative to the gain-only scheme) if they met their goal but also penalized them if they failed to meet to their goal. We then assigned participants to either the gain-only (*gain-only assigned*) or gain-loss incentive scheme (*gain-loss assigned*). The final group of participants was provided the option they preferred (*choice*). The *choice* group was oversampled relative to groups assigned to one incentive or the other because individuals in this group further sort into the gain-only or gain-loss incentives. Although the endogenous sorting of these individuals precludes us from analyzing these subgroups directly, oversampling allows us to have similarly sized groups when analyzing incentive effects conditional on an incentive preference.[5] Random assignment into these groups allows us to evaluate the effect of providing gain-loss incentives on physical activity, whereas eliciting preferences for the different incentive schemes beforehand allows us to identify heterogeneity in the effect of more aggressive economic incentives conditional on incentive preference.

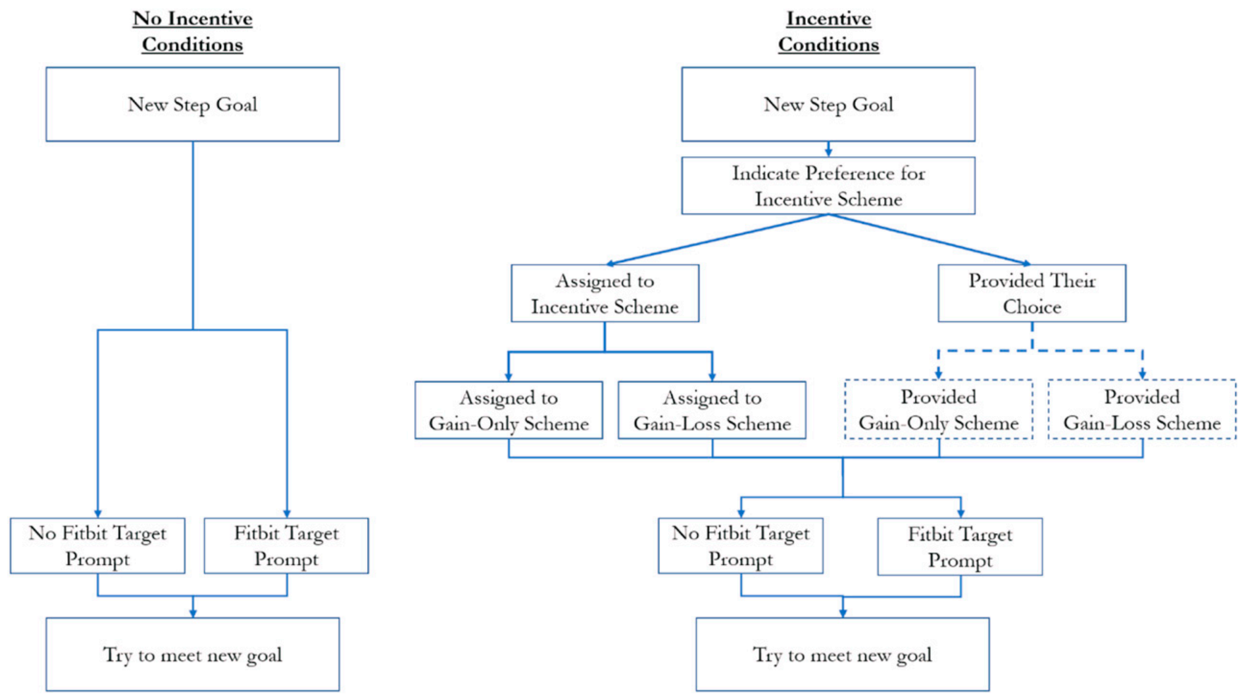
We also evaluate how Fitbit HR health tools moderate the impact of gain-loss economic incentives. We focus on the Fitbit health tools that allow users to designate a daily step target and track progress toward this target (the default step target for all Fitbit users is 10,000 steps per day). Figure 2 provides an example dashboard where the user has set a step target of 8,000 steps and the star icon indicates the days that they met this target. This type of health tool is available from all leading health wearables, and research suggests that most who own fitness trackers are primarily concerned with monitoring their steps (Mercer et al. 2016). An

Figure 2. (Color online) Example Fitbit Dashboard and Step Targets

Source. fitbit.com.

Management Science, 2022, vol. 68, no. 4, pp. 2642–2667, © 2021 INFORMS

Figure 3. (Color online) Overview of Experimental Design



Notes. Choice condition is further split into groups who were given their choice of the gain-only scheme and those given their choice of the gain-loss scheme (dashed lines). These subgroups are not exogenously assigned in our experiment.

important feature of this health tool is that it is highly accessible to users: the user's current steps are continuously available to them on the wearable device (as opposed to only on the Fitbit app on their mobile phone) and the wearable device vibrates and gives a congratulatory message when the designated step target is met. We preferred not to use heavy-handed experimental interventions that, for example, forced participants to utilize the health tool as part of the experiment, in order to avoid inorganic use and experimenter demand effects. Rather, we manipulated whether the communications to participants prompted them to update their Fitbit target to match the new, more ambitious goal set by us. Specifically, we provided a random subset (including from those not provided economic incentives) of participants the following message: "Take a minute now to change your step goal in the Fitbit App." We provided participants this message one day before the start of the challenge after we reminded them of the new step goal we had assigned them as part of the challenge (see Online Appendix A-2). This randomly assigned prompt is intended as a subtle way to exogenously shift use of the Fitbit health tools and allows us to examine how the use of this functionality moderates the impact of different incentive schemes. The experimental flow and key manipulations are provided in Figure 3.

Prior to running our field experiment, we sought to calibrate two key parameters. First, we wanted to identify what participants in our study would consider a reasonable but challenging new step goal. Second, we wanted to identify how the additional potential reward provided in the gain-loss incentive scheme would impact participants' propensity to choose it relative to the gain-only incentive scheme. The primary purpose of this calibration was to avoid severe skew in preference by participants toward one incentive scheme over the other. Specifically, if the gain-loss incentive scheme was either too lucrative or if the goal was perceived to be easy to achieve (or both), participants might only choose the gain-loss incentive scheme (and vice versa if the gain-loss option was not attractive or the goal was too hard to be achievable).

To avoid these issues, we ran a prestudy in which we recruited 87 undergraduate students (at the same university) who were not part of the StudentHealth population but reported using an activity tracker. Participants in the prestudy were asked to indicate what they would consider to be a reasonable new goal based on their current average step count as well as their preference for the gain-only versus gain-loss incentive scheme, varying the level of reward associated with the gain-loss incentive scheme. Specifically, we presented participants with nine scenarios asking them to choose between a gain-only incentive scheme (labeled "Option 1") that rewarded them \$1.00 daily for meeting their goal and a gain-loss incentive scheme (labeled "Option 2") that would reward them some amount daily for meeting their goal (the reward started at \$1.00 and was incremented by \$0.25 all the way up to \$3.00) but had a \$1.00 penalty if they missed their goal. At a \$1.25 reward, a slight majority of participants (52%) indicated they would choose the gain-loss scheme; at a reward of \$1.50 a day, 71% of participants indicated a preference for the gain-loss scheme. Using these results, we set the new goal for the field experiment as a 20% increase on participants' average step counts from the prior month and (if in the treatment conditions) offered participants a choice between a gain-only option that awarded \$1.00 for meeting a step goal (and no penalty) and a gain-loss scheme that awarded participants \$1.25 for meeting their goal but took away \$1.00 if they did not meet it. For more details on the prestudy, see Online Appendix A-1.

3.3. Experimental Procedure

We informed individuals in the StudentHealth population of an upcoming step challenge via text messages sent to their mobile phones over a two-week period, starting at the beginning of the fall semester. Individuals could opt-in to the step challenge at any point during this two-week period. Although there was minor variation between the recruitment text messages (see Online Appendix A-2), each message informed individuals that they had been invited to participate in a "Step Challenge," told them that they would be compensated \$10 for participating, and provided them a link to click if they wished to participate.

We sent individuals who clicked the link to an instruction page that informed them that the purpose of the step challenge was to have participants strive to achieve a new step goal and that the challenge would run for seven days. The instructions then provided participants with their average step count for the most recent four-week period for which we had complete data and gave them a new step goal which was a 20% higher than this average.[6] The four weeks of data used to calibrate the new goal preceded the pretrend period to allow participants' Fitbit data to be fully up to date. Using this slightly earlier time frame to calibrate new goals ensured that all participant data were up to date and fully synced for goal calibration. We dynamically populated participants' prior step counts and their new goal into the survey and were transparent about how their new goal was determined. For example, we would inform a participant that 11,000 was their average step count in a recent month and that their new step goal for the challenge week would be 13,200, a 20% increase over their prior monthly average.[7] We then informed participants that performance in the challenge was about "consistency" and would be based on whether they met their step goal on each day of the challenge. We made this point explicit to make clear that participants needed to try and meet their step goal on each day of the challenge (as opposed to simply achieving an average step count that exceeded their prior average).

We provided participants in the incentive conditions with an additional payment opportunity, and described both the gain-only and gain-loss incentive schemes to them. After reminding them of their new step goal, we asked participants to indicate their preference of incentive scheme. At this point, participants had no expectation that they would not be provided the incentive scheme they preferred. We provided participants in the "choice" condition their chosen incentive scheme. We randomized the remaining participants into one of the two incentive schemes, citing the rationale that we wanted to be "fair to all participants." We then informed all participants that they had been enrolled in the step challenge and that we would send them a text message the day before the challenge started.[8] The step challenge took place at the beginning of the third week of the semester, to allow students to enter a stable routine and avoid initial changes in health behavior associated with the start of a new semester. We did not contact the participants again until the day preceding the start of the challenge, when they all received a text message informing them that the challenge was starting the following day. It was at this point that some participants received the randomized prompt to update their Fitbit set target. We also provided participants a mid-challenge update as well as a text message indicating the end of the challenge and summarizing their performance and compensation.

3.4. Estimation Approach

The main dependent variable in our analysis is a measure of the number of steps walked by a participant on a given day of the challenge. Because steps are captured as a repeated measure (i.e., participants had observations of steps walked every day of the challenge), we use a panel random effects regression as our primary estimation approach, with heteroscedasticity-robust standard errors.

$$\{\mathrm{Steps}_{it}\} = \beta^{*}\mathbf{Treatment}_{i} + \alpha^{*}\mathbf{Demographics}_{i} + \pi^{*}\mathbf{Personality}_{i} + \delta_{t} + \theta_{i} + u_{ij}$$

Our dependent variable (Steps_{ij}) is an effectively continuous measure of steps walked by a participant i $\{1 \ldots N\}$ in day t $\{1 \ldots J\}$. $\mathrm{Treatment}_i$ is a vector of binary indicators for the treatment condition assigned to a participant; $\mathrm{Demographics}_i$ and $\mathrm{Personality}_i$ are vectors of controls capturing heterogeneity in individual demographics (e.g., age, gender, body mass index (BMI)[9])

Table 1. Summary Statistics

Variable	Description	Mean (1)	S.D. (2)
Steps	The number of steps recorded on a given day	9,779	6,030
New goal	The new step goal assigned to an individual based on prior activity	12,453	4,493
Goal met	Binary indicator for meeting the step goal assigned to them	0.29	0.45
Comply percentage	The percentage of the day that individuals wore their Fitbit	84.90	25.18
Male	Whether the individual is a male	0.47	0.50
BMI	The body mass index of the individual	22.91	2.99
Age	The age at the beginning of the experiment	18.61	0.40
[a]*Extraversion*	Big Five personality trait–extraversion (scale 1–5)	3.21	0.78
[a]*Agreeableness*	Big Five personality trait–agreeableness (scale 1–5)	3.74	0.55
[a]*Conscientiousness*	Big Five personality trait–conscientiousness (scale 1–5)	3.68	0.57
[a]*Neuroticism*	Big Five personality trait–neuroticism (scale 1–5)	2.90	0.70
[a]*Openness*	Big Five personality trait–openness (scale 1–5)	3.34	0.56

Note. S.D., standard deviation.
[a]BFI questionnaires based on John et al. (1991).

and personality types; δ_t is day fixed effects; θ_i is the participant-specific random effect; and u_{ij} is the error term. Estimates on randomly assigned treatments (Treatment_i) are unbiased because they are assumed to be uncorrelated with unobserved individual differences and the error term (we test this assumption extensively in Section 4.1 and Online Appendix A-4). The primary benefit of this estimation approach is that our model corrects standard errors for nonindependence of multiple observations of steps walked from a single participant (Zeger and Liang 1986) and allows us to more robustly account for time trends in our data.

We also analyze the impact of our interventions on whether individuals met the goal we set for them using a linear probability model (as well as the logit and probit models) and an identical specification. However, we focus on steps as our main outcome because it has more direct interpretability for health and effective randomization accounts for the relative nature of the new goal we set for participants across conditions. More so, goal attainment does not have as much relevance in the posttreatment period where individuals are not directly incentivized to meet the goal. Finally, some individuals significantly exceeded the new goal we set for them, whereas others, though striving to achieve it, fell short. This important variation would be lost if we focused primarily on the binary measure of goal attainment.

4. Results

Three hundred and ten individuals (M_{Age} = 19, 47% male) participated in the step challenge and had steps observed during the challenge period. One participant had observations for only the first day of the challenge period and was dropped from the study, leaving 309 usable responses. We had a 90% response rate for participants with a 40% compliance rate or higher (309/341) and a 76% response rate for all active participants (309/402). Table 1 provides descriptions of the variables in our analysis and summary statistics for the individuals who participated in the field experiment.

The average step count in the pretreatment period is just below 10,000 steps, and the average new goal that was ultimately set for this group was 12,453 steps a day.[10] Prior to the experiment starting, individuals were not aware that they would be assigned this new goal nor were they aware of the prospect of the experiment starting. They also rarely exceeded the step count of the new goal we would set for them (0.21). Our demographic is young with little variation in age (expected given our sample), evenly split between male and female, and with an average BMI of 22.91 (the healthy range is 18.5–24.9).[11]

In Table 2, we show the distribution of individuals across the interventions in our experiment. We observe a roughly equal split of individuals across the various interventions, with 99 (44 + 55) individuals randomly assigned to an incentive scheme and 100 who were provided their choice. Individuals in our experiment were also exposed to the Fitbit target prompt manipulation at similar rates. Of the participants offered the choice of bonus incentives (i.e., noncontrol conditions), 26% indicated a preference for the

Table 2. Breakdown of Sample by Experimental Conditions

	Control	Choice	Gain-only assigned	Gain-loss assigned	Total	No target prompt	Target prompt	Total
Sample size	766	675	304	383	2,128	1,095	1,033	2,128
Individuals	110	100	44	55	309	160	149	309

Table 3. Pretrend Between Incentive Manipulations—Eight Weeks Prior to Experiment

Variables	(1) Steps	(2) Goal met	(3) Compliance	(4) Steps	(5) Goal met	(6) Compliance
Choice	−136.7	0.00204	0.192			
	(442.6)	(0.0195)	(2.986)			
Gain-only assigned	8.527	0.0152	−2.861			
	(630.6)	(0.0317)	(4.107)			
Gain-loss assigned	595.8	0.0274	1.364			
	(573.3)	(0.0265)	(3.368)			
Fitbit target prompt				−403.8	−0.0319*	2.247
				(383.7)	(0.0174)	(2.440)
Constant	7,747***	0.172**	58.33**	8,010***	0.195***	57.18***
	(477.3)	(0.0273)	(2.956)	(427.5)	(0.0277)	(2.618)
Observations	14,420	14,420	16,638	14,420	14,420	16,638
Individuals	287	287	287	287	287	287

Notes. Robust standard errors in parentheses. Time fixed effects included in all columns.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

gain-loss incentives. The proportion of participants choosing the gain-loss incentives is lower than in the hypothetical prestudy but is still twice as large as in previous studies that evaluated commitment-based incentive schemes (e.g., Royer et al. 2015). For those offered their choice, 23% preferred the gain-loss incentives and 77% preferred the gain-only incentives. The stronger-than-expected preference for the gain-only incentives does not impact the estimation of our exogenously assigned group but has some implications for analysis conditioning on incentive preferences (discussed in detail in Section 4.4).

4.1. Balance Checks

Before we delve into the effects of our treatments, we evaluate whether there are pretreatment differences between experimental conditions. First, we extensively examine differences between individual demographics, BMI, and personality traits of individuals in our sample. We conduct pair-wise comparisons for all these variables for all experimental manipulations (incentives and Fitbit target prompt) and find evidence of balance between experimental conditions (see Online Appendix A-4). Nevertheless, we control for these variables in our analysis and observe consistent results.

We also examine differences in health behavior in the eight-week pretreatment period to evaluate whether there are significant differences in steps walked (*steps*) and how often participants met their new goal during that period (*goal met*), although they had not actually been assigned the goal at that time.[12] In addition, we evaluate whether participants across conditions differed in terms of the percentage of time they wore their Fitbit during this period (*compliance*). We did not observe any meaningful differences between experimental conditions during the preperiod for all these measures (Table 3). Specifically, no conditions are significantly different from the no-incentive group in the pretreatment period and all pair-wise comparisons are insignificant with most comparisons yielding large p-values. For example, with regard to our key variable of interest (steps), there is not a significant difference between gain-loss and choice (p = 0.21), between choice and gain-only assigned (p = 0.82), or between gain-loss assigned and gain-only assigned (p = 0.43).

We make similar comparisons for the Fitbit target prompt and again don't find significant differences in steps (column (4)), which is our key dependent variable, or in compliance levels (column (6)). We do identify a small but significant ($p < 0.1$) difference in the propensity of meeting the new goal we would create for them as part of our experiment in the pretreatment period (column (5)). Similar to individual characteristics, we control for pretreatment levels of activity and find that this has no bearing on our results (see Online Appendix A-4). Overall, we find balance on observed characteristics and pretreatment activity levels; any differences that exist are within the threshold of random chance and do not impact our findings.

4.2. Main Effects: Incentives

Model-free results suggest that all treatments had some positive effect on steps walked relative to the control condition during the step challenge (see Figure 4) However, we also see notable differences between conditions. Specifically, those assigned to the choice condition seem to perform slightly worse than those provided the gain-only incentives despite having the option of the presumably more motivational gain-loss incentives. Those assigned gain-loss incentives (many of which went against their preference) have treatment effects that are nearly twice that of other treatment groups.

Regression analysis confirms our summary results and is reported in Table 4.[13] On average, we find that assigning individuals to gain-loss incentives

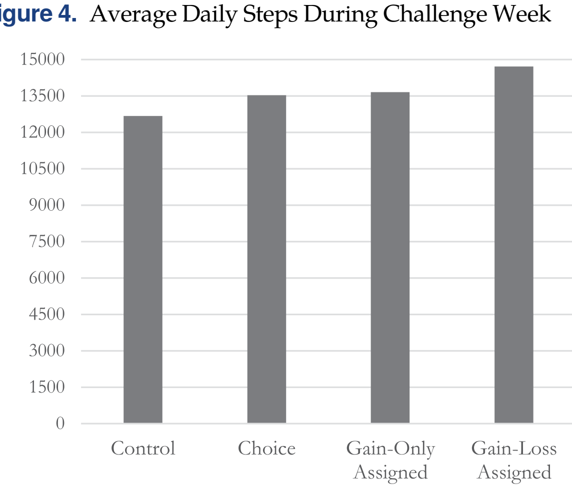
Table 4. Main Effect of Incentive Schemes on Steps

	Challenge week				Short term	Intermediate term	Long term
Variables	Steps (1)	Steps (2)	Goal met (3)	Goal met (4)	Steps (5)	Steps (6)	Steps (7)
Choice	770.0*	971.8**	0.111**	0.127***	895.7**	577.5	306.5
	(454.0)	(440.1)	(0.0463)	(0.0452)	(407.0)	(377.4)	(351.5)
G-O assigned	955.0*	1,114**	0.103*	0.126**	493.2	3.854	204.4
	(525.9)	(495.9)	(0.0589)	(0.0591)	(481.1)	(489.9)	(438.1)
G-L assigned	2,068***	2,116***	0.185***	0.183***	1,711***	1,398***	1,324***
	(685.1)	(684.5)	(0.0505)	(0.0499)	(555.9)	(533.0)	(513.3)
Constant	12,351***	−12,649	0.541***	−1.509	−5,232	−5,349	−15.79
	(363.6)	(9,583)	(0.0357)	(0.985)	(8,129)	(7,945)	(7,626)
Observations	2,128	2,114	2,128	2,114	8,117	13,740	19,870
Individuals	309	307	309	307	303	293	275
Demographic controls	NO	YES	NO	YES	YES	YES	YES
Personality controls	NO	YES	NO	YES	YES	YES	YES

Notes. Robust standard errors in parentheses. Time fixed effects included in all columns. G-O, gain only; G-L, gain loss.
***$p < 0.01$; **$p < 0.05$, *$p < 0.1$.

had the strongest effect on steps walked. Participants in this condition walked 2,068 ($p < 0.01$) more steps than their counterparts in the control condition (column (1)). Participants who were either provided a choice or were simply assigned to the gain-only incentives walked 770 and 955 steps (respectively) more than their counterparts in the control condition; these effects, however, were only significant at the 10% level (column (1)). These effects are consistent when controlling for demographic factors and personality characteristics (column (2)) except that the effect on gain-only assigned is more precise.[14] We also find consistent effects when examining the effect of our treatments on the probability of meeting the new goal we provided participants (*goal met*) during the challenge week (column (3)), and these results again are consistent when including individual demographic and personality controls (column (4)). These results are consistent when we utilize the logit and probit estimation approaches (see Online Appendix A-5).

Figure 4. Average Daily Steps During Challenge Week



4.2.1. Persistence of Effects. Evaluating the effect of these incentives after the end of the challenge week reveals heterogeneity in their resilience (Table 4). Those assigned to the gain-only incentive scheme had the least resilience posttreatment; this group quickly reverted back to control levels of steps walked (column (5)) and remained that way throughout the post period (columns (6) and (7)). The effect for those offered a choice seems to have some posttreatment resiliency, but this effect was only significant at the 10% level in the short term (column (5)) and was insignificant in the intermediate and longer term (columns (6) and (7)). Participants assigned to the gain-loss incentives, in contrast, did display posttreatment resilience and outperformed their counterparts in the control condition during the short (column (5)), intermediate (column (6)), and longer term (column (7)). These results are unchanged if controls are excluded.

4.2.2. Compensation. We also evaluated any differences in daily bonus compensation during the step challenge between conditions. We treat those assigned to the gain-only incentives as our baseline group, because the control group had no additional compensation potential. We find no significant differences between treatment conditions: participants in all treatment conditions were compensated \$4–\$5 in addition to the base compensation of \$10 given to all participants (see Online Appendix A-5).[15] This result suggests that participants who preferred gain-only incentives but were randomized into the gain-loss ones, avoided an economic loss by improving their performance in the challenge.

Table 5. Compensation for Participants Preferring Gain-Only Incentives

Condition (participants)	(1) Daily compensation (actual)	(2) Daily compensation (under gain-loss incentive scheme)	(3) Reward (%)	(4) Penalty (%)
Gain-loss assigned (55)	\$0.47	\$0.47	0.66	0.34
Choice (100)	\$0.51	\$0.21[a]	0.54	0.46[a]
Gain-only assigned (44)	\$0.57	\$0.31[a]	0.58	0.42[a]

[a]These would have been the values if participants had been provided the gain-loss incentives.

Table 5 illustrates this point and shows that daily incentive compensation for those who preferred gain-only incentives was comparable across treatment conditions (column (1)). We focused on those who prefer the gain-only incentives because they were put at risk for financial loss by being assigned to the gain-loss incentives. However, we show that those provided gain-only incentives would have been significantly worse off in the gain-loss scheme given the same performance (column (2)). Specifically, individuals provided their choice of gain-only incentives would have only earned an additional \$0.21 daily (as opposed to \$0.54) if assigned to the gain-loss incentives, assuming they had the same level of performance (see column (2)). Similarly, those given a choice of incentives and who opted for gain-only incentives would have only earned an additional \$0.31 daily (as opposed to \$0.58) if they had been assigned to the gain-loss incentives and had walked the same amount as they did (column (2)). The participants who preferred gain-only incentives but were assigned gain-loss ones avoided this loss by earning the higher reward more often and incurring fewer penalties (columns (3) and (4)).

4.3. Effect of the Fitbit Health Tool

Next, we evaluated whether the health tools provided by the Fitbit app impact performance in the step challenge. We focused on the tool that allows users to designate daily step targets, to track progress toward this target, and to receive feedback when it is met. To identify the effect of this wearable health tool, we introduced a randomly assigned prompt encouraging participants to update their Fitbit step target to match the new goal we set for them (*FitbitTargetPrompt*); see Online Appendix A-3. We found that 79/309 participants (~26%) updated their Fitbit step target to match the new goal provided by us and that the prompt was highly predictive of changes to Fitbit step targets ($\text{StepTargetUpdated}_{\text{Prompt=1}}$ = 47%, $\text{StepTargetUpdated}_{\text{Prompt=0}}$ = 6%, $t(307) = 9.42$, $p < 0.01$). We note that all participants had a new step goal in the experiment that was 20% higher than their prior step average, and this manipulation only relates to changing the step target on the wearable to match this new goal. We extend our analysis by including an indicator for the randomized prompt as well as interactions between the prompt that encourages Fitbit users to update their step target and our various incentive schemes (Table 6). For ease of exposition, all analyses include demographic and personality controls (results are consistent without these controls).

By evaluating the effect of our randomized prompt (as opposed to actual updates to Fitbit step targets), we identify a conservative estimate of the effect of changes to Fitbit step targets (an intention-to-treat analysis—Gupta 2011). We find that updating the Fitbit step target is an important moderator of the relationship between being assigned to a gain-loss incentive scheme and changes in daily steps walked, particularly in the long run. Generally, those who were both assigned to the gain-loss incentives and received the randomized prompt walked more steps throughout our period of observation, relative to those assigned to the gain-loss incentives but not provided the prompt (columns (1)–(4)).[16] These effects were robust well after the incentive period ended ($\beta_{\text{FitbitTargetPrompt*Gain-LossAssigned}}$ = 2,549; $p < 0.05$, column (4)). We also observe that the target prompt seems to trend negative for those in the control condition and this effect is significant in the long term. This result points to potentially demotivational effects of asking individuals to update their step target without an initial push (in the form of economic incentives) to change their behavior. Because individuals rarely undo these changes, the higher step goal could have a demotivational effect by reinforcing the lack of goal attainment for individuals.

Our results are consistent when looking at actual changes to Fitbit step targets. If we utilize actual changes to Fitbit step targets (as opposed to the randomized prompt) or use our randomized treatment as an instrument for changes in Fitbit step targets, we find even stronger effects of wearable health tools for those assigned to the gain-loss incentives against their preference. Specifically, we find large and significant effects on steps for them during both the study period and in the long term when they update their Fitbit step target but not otherwise (see Online Appendix A-6). These findings suggest that, with the right set of tools employed for increasing engagement (in this case, the prompt to update the steps goal on the app), wearable health tools may be important for sustaining physical activity when assigning individuals to gain-loss incentives.

Table 6. Effect of Incentives by Target Prompt

Variables	Step challenge week (1) Steps	Short term (2) Steps	Intermediate term (3) Steps	Long term (4) Steps
Choice	841.5	849.9	235.2	−429.8
	(634.9)	(592.1)	(533.4)	(463.5)
Gain-only assigned	961.6	343.5	−742.6	−275.9
	(721.7)	(743.8)	(721.6)	(598.5)
Gain-loss assigned	1,014	716.0	540.4	102.5
	(755.2)	(641.6)	(589.2)	(476.3)
Target prompt	−640.9	−559.5	−788.2	−1,407***
	(610.2)	(535.3)	(512.6)	(438.5)
*Choice*TargetPrompt*	208.6	55.70	704.1	1,529**
	(874.6)	(813.0)	(726.2)	(674.0)
*G-O assign * TargetPrompt*	284.7	291.5	1,508	947.4
	(939.6)	(930.8)	(949.4)	(836.6)
*G-L assign * TargetPrompt*	2,333*	2,119*	1,818*	2,549**
	(1,401)	(1,129)	(1,101)	(1,050)
Constant	−11,592	−4,232	−3,835	2,111
	(9,454)	(8,123)	(7,798)	(7,113)
Observations	2,114	8,117	13,740	19,870
Individuals	307	303	293	275
Demographic controls	YES	YES	YES	YES
Personality controls	YES	YES	YES	YES

Notes. Robust standard errors in parentheses. Time fixed effects included in all columns. G-O, gain only; G-L, gain loss.
*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

4.4. Heterogenous Effects by Incentive Preference

We also consider the effect of our manipulations while accounting for whether individuals indicated a preference for the gain-only or gain-loss incentive scheme. In the field experiment, we are not able to disentangle the preference for either the gain-loss or gain-only incentives for those in the control condition because we did not elicit from them their preference between the two schemes. Thus, we focus on differences between those in the treatment conditions who indicated a preference for a given incentive scheme (see Figure 5); the level of the control group is still provided for reference. One important caveat for this analysis is that relatively few individuals preferred the gain-loss incentive scheme compared with the gain-only incentives (52 versus 147). Although the panel data results in a reasonable sample size, the analysis conditional on individuals

Figure 5. (Color online) Average Daily Steps by Condition and Incentive Preference

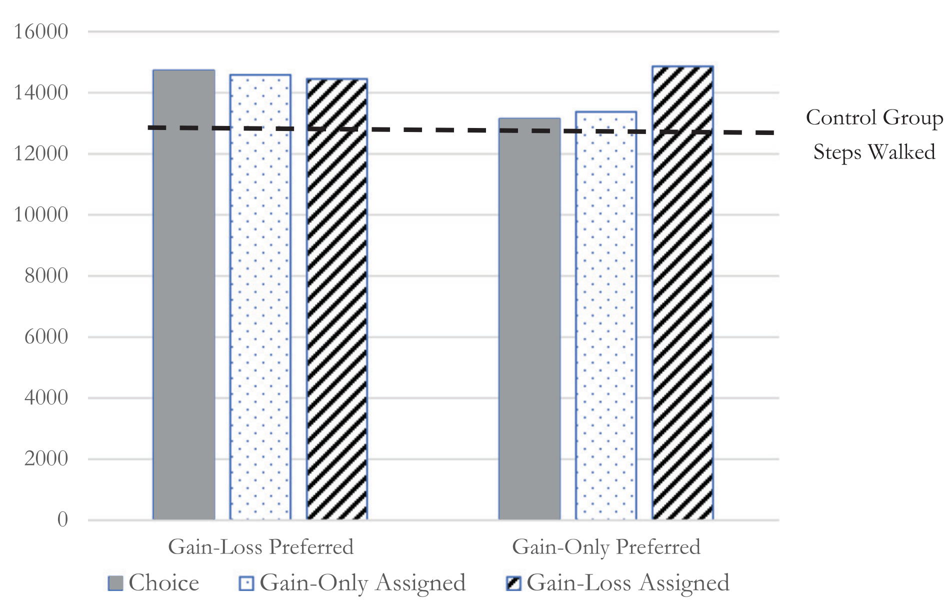


Table 7. Effect by Incentive Preference

	Gain only preferred				Gain loss preferred			
Variables	(1) Challenge week	(2) Short term	(3) Intermediate term	(4) Long term	(5) Challenge week	(6) Short term	(7) Intermediate term	(8) Long term
Choice	−297.2	342.2	827.2	90.51	884.9	1,714*	135.1	−118.2
	(536.2)	(565.2)	(554.1)	(519.9)	(1,057)	(729.7)	(1,180)	(1,077)
Gain-loss assigned	1,390*	1,718*	1,929**	1,546*	744.4	1,354	544.8	592.5
	(853.2)	(730.3)	(736.9)	(716.2)	(1,261)	(924.5)	(1,253)	(1,160)
Constant	−358.2	9,921	14,229	19,056	−4,308	13,934	2,762	1,248
	(16,025)	(14,318)	(12,638)	(11,727)	(16,480)	(12,302)	(13,746)	(12,915)
Observations	998	3,893	6,724	9,455	357	1,342	2,184	3,248
Individuals	147	144	141	131	51	51	48	46
Demographic controls	YES	YES	YES	YES	YES	YES	YES	YES
Personality controls	YES	YES	YES	YES	YES	YES	YES	YES

Notes. All columns utilizes daily steps as robust standard errors in parentheses. Time fixed effects included in all columns; Daily steps is the outcome variable in all columns.
**$p < 0.05$; *$p < 0.1$.

preferring gain-loss incentives is driven by relatively few individuals. In contrast, the analysis conditional on individuals preferring the gain-only incentives has both a reasonable sample size and a reasonable number of individuals in the analysis. In Online Appendix A-7, we assess the balance on observables as well as the robustness of results after conditioning on a given incentive preference. The results are consistent with what we present below.

We find that participants who preferred the gain-loss scheme performed at a high level regardless of which condition they were in. In contrast, individuals who preferred the gain-only incentive scheme seemed to exhibit significant differences in their step counts depending on the treatment condition. Specifically, those who were assigned to the gain-only incentives or were provided their choice of the gain-only incentives performed worse than other treated group and did not differ from the control group. In contrast, those who preferred the gain-only incentives and were then assigned to the gain-loss incentives did significantly better.

These results are confirmed in our regression analysis (Table 7). Because we did not elicit incentive preferences for the control group, we conduct our analysis with those provided the gain-only incentive scheme as our baseline group. Consistent with our summary results, we find that participants who prefer gain-only incentives performed significantly better when assigned to the gain-loss scheme, relative to when they were provided their choice ($p < 0.05$) or assigned to the gain-only incentives ($p < 0.1$), as shown in column (1). In contrast, differences in treatment effect were not significant for those who preferred the gain-loss scheme (column (5)). In other words, those who preferred gain-loss incentives did equally well independent of the treatment they were provided.

4.4.1. Persistence of Effects.

Evaluating these effects in the posttreatment period reveals similar trends. We find that assigning participants who preferred gain-only incentives to the gain-loss ones resulted in large and significant differences in steps walked during the short, intermediate, and longer term (Table 7, columns (2)–(4). The lack of differences in treatment effect for those who preferred the gain-loss scheme persisted in the long run (Table 7, columns (6)–(8)). An important takeaway from these results is that the longer-term benefits reaped by those assigned to the gain-loss scheme are driven by those who would have elected into the alternate incentive scheme, given a choice. Evaluating the effect of our treatments while accounting for an individual's preference suggests that, provided a choice between schemes, the wrong people opted into the more challenging gain-loss incentive scheme. Those who would have benefited from the gain-loss scheme were more likely to choose the gain-only scheme, and it was those who did not need it who opted in.

4.4.2. Differences Between Individuals Preferring Gain-Only vs. Gain-Loss Incentives.

To further inform these results, we evaluate whether there are relevant differences between those who preferred the gain-only versus gain-loss incentive scheme. We find that participants who preferred gain-loss incentives had lower absolute goals to meet (10,472 versus 13,129, $p < 0.01$) and were meeting these goals more frequently in the two weeks prior to the start of the challenge (0.37 versus 0.27, $p < 0.01$). These results are consistent with the conjecture that individuals would avoid the gain-loss incentives if they were less confident that they could meet the goal we set for them. Because we are evaluating treatment effects conditioning on an individual's incentive preference, these differences should not significantly impact our results. Nonetheless, we

Table 8. Personality Traits by Incentive Preference

Panel A

Personality traits	Gain-only preferred (1)	Gain-loss preferred (2)	*p*-value (3)
Extraversion	3.20	3.12	0.49
Agreeableness	3.69	3.80	0.22
Conscientiousness	3.65	3.59	0.56
Neuroticism	3.01	2.79	0.06*
Openness	3.31	3.49	0.05**

Panel B

Variables	(1) *Openness*	(2) *Neuroticism*	(3) *Conscientiousness*
Gain-loss preferred	0.174*	−0.220**	−0.0582
	(0.0976)	(0.112)	(0.0885)
Constant	3.143	5.721**	0.497
	(2.334)	(2.653)	(2.000)
Observations	197	197	197
Dem controls	YES	YES	YES

Notes. Robust standard errors in parentheses. Time fixed effects included in all columns.
$^{**}p < 0.05$; $^{*}p < 0.1$.

extend our analysis to include a control for frequency of meeting the new goal before the challenge starts and find consistent results (Online Appendix A-7).

We also conjectured that the avoidance of the gain-loss incentives is more likely for those with elevated degrees of loss aversion. If people avoid gain-loss incentives because they are acutely loss-averse—they want to avoid losses—then we should expect those who do, against their own preferences, get assigned to such incentives to respond to them most strongly, in order to avoid experiencing the losses. We can actually test this idea, crudely, because we have data on personality traits that, prior literature suggests, are predictive of loss aversion. Specifically, prior work has shown that the openness trait is associated with more willingness to accept risks and lower levels of loss aversion (Dohmen et al. 2010) whereas neuroticism (and in one study, conscientiousness) is correlated with higher levels of loss aversion and lower risk taking (Borghans et al. 2008, Anderson et al. 2011, Sahinidis et al. 2020). With these works in mind, we evaluate differences in demographic variables and personality traits between those who prefer the gain-only versus gain-loss incentives. Specifically, those who select the gain-only incentives tended to be higher on the neuroticism personality trait ($p = 0.07$) and lower on the openness trait ($p = 0.05$)—see Table 8, Panel A. There was no significant difference on the conscientious trait. These results are confirmed utilizing regression analysis with controls for demographic variables (Table 8, Panel B, columns (1)–(3)). These combinations of results provide suggestive evidence that individuals who prefer the gain-only incentives do so because of diminished confidence that they can meet the new goal and likely higher levels of loss aversion.

4.4.3. Effects by Incentive Preference and Wearable Health Tool Usage. Finally, we evaluate the effect of wearable health tools specifically for individuals who preferred the gain-only incentives (Table 9). Recall that because the control group did not indicate an incentive preference, the analysis conditioning on an incentive preference utilizes those who were assigned to the gain-only incentives as the baseline group. We specifically examine the conjecture that those who prefer the gain-only incentives and then are assigned to gain-loss schemes benefit from the use of wearable health tools. We find that, during the challenge period, individuals who preferred the safe incentives but were assigned to the gain-loss incentives reaped twice as much benefit from the gain-loss incentives if they also received the prompt to utilize the Fitbit health tool (2,330 versus 783 additional steps walked)—columns (1)–(2). However, the coefficients were not significant for both groups (potentially because this subsample analysis utilizes fewer observations than prior analysis). Evaluating these effects in the longer term (which had a longer period of data collection and thus more observations), we see that those who preferred the gain-only incentives persisted in their physical activity (3,044 steps, $p < 0.1$) if they were both assigned to the gain-loss incentives and provided the prompt to utilize the wearable health tool (column (3)) but failed to do so otherwise (column (4)). Finally, we don't find significant impacts for the choice condition in any of the columns; this result likely emerges because these individuals were providing the same incentives as the gain-only group (our baseline). These findings suggest that motivational health tools offered by wearables may be important for sustaining physical activity, particularly when assigning individuals to gain-loss incentives against their preference.

4.5. Extensions and Robustness Checks

We extend our analysis in several ways (see Online Appendix A-8 for more details and results). First, we evaluate whether failure to achieve a goal (and by extension incurring the penalty under the gain-loss scheme) impacts physical activity for those assigned to the gain-loss incentives. We find that incurring the penalty in one day seems to motivate increased activity in the following day (although this result was not significant, $p = 0.18$). This result points to the motivational potential of the penalty aspect of the gain-loss incentive scheme. In addition, we evaluate the robustness of our main results. Specifically, we evaluate our effects with the addition of a control for daily $Compliance_{it}$ (measured as the percentage of each day

Table 9. Gain-Only Preferred and Fitbit Target Prompt

Variables	(1) Challenge period–target prompt	(2) Challenge period–no target prompt	(3) Long term–target prompt	(4) Long term–no target prompt
Choice	−760.3	90.96	421.9	34.90
	(869.2)	(761.3)	(762.6)	(727.1)
Gain-loss assigned	2,330	782.8	3,044*	667.4
	(1,716)	(988.2)	(1,396)	(743.2)
Constant	1,025	−4,062	21,463	18,488
	(26,584)	(20,809)	(20,901)	(15,581)
Observations	446	552	4,291	5,164
Individuals	65	82	59	72
Demographic controls	YES	YES	YES	YES
Personality controls	YES	YES	YES	YES

Notes. Robust standard errors in parentheses. Time fixed effects in all columns. Daily steps is the outcome variable in all columns. Comparison group for all columns is the *Gain-only assigned* group.
*$p < 0.1$.

t that participant i wore their Fitbit device) to account for differences in the wearing of the Fitbit device by participants. In addition, we evaluate our effects utilizing a Poisson regression approach. Finally, we also evaluate the effect of our manipulations on the log of steps, which addresses potential nonnormality in our data. We find consistent results across these robustness checks.

4.6. Discussion

The results of our field experiment suggest that allowing individuals to opt-in to gain-loss incentive schemes may result in the exclusion of those who would benefit most from such schemes. Our results also point to promising complementarities between economic incentives and wearable health tools built into increasingly popular health activity trackers. These complementarities may be pronounced when individuals are exposed to gain-loss incentive schemes that they would not select for themselves.

We offer several explanations for our observed results. If individuals who avoid gain-loss incentives do so because of an elevated degree of loss aversion (Tversky and Kahneman 1991), this subgroup may be more willing to exert effort to meet their new goal (i.e., in order to avoid the associated loss) and better utilize wearable health tools to do so. Moreover, individuals who avoid gain-loss incentives because they perceive the goal to be difficult or are not motivated to meet the goal (or both) may have received the largest psychological boost from goal attainment during the challenge period, which then provided them more lasting motivation after incentives end. Finally, these individuals seem to benefit most from the wearable health tool. This result is consistent with goal-setting literature that finds that when individuals pursue challenging goals, measurement, continuous feedback, and reinforcement of the goal is critical to performance gains (Becker 1978, Strang et al. 1978, Bandura and Cervone 1983).

There are some concerns with our field experiment. First, we do not have data to assess whether individuals were "gaming" the challenge week incentives (e.g., asking friends to wear their Fitbit devices). The Fitbit device has no measures, that we know of, of whether the same person is wearing the device throughout the day. However, it is unlikely that gaming the incentive is driving our main effects, which show the persistence of treatment effects after compensation ends. Second, we did not inform participants in the control condition about the various bonus options available to those in the treatment condition, so we cannot ascertain the preference for either incentive scheme for individuals in the control condition. This precludes us from evaluating the effects of incentive schemes against the control group conditional on incentive preference. Thus, we cannot speak to, for example, how those who prefer the gain-loss incentives and were provided this selection compare with those who prefer the same scheme but were not provided any incentive at all. However, this is not a core line of inquiry in our manuscript. In addition, although the pilot study suggested that approximately 50% of subjects would express a preference for the gain-loss incentives, a substantially smaller fraction ended up doing so. This dissonance in preferences could be due to the longer-term nature of the commitments associated with the field experiment that may not have been salient in the hypothetical choice provided in the hypothetical prestudy. Nonetheless, the smaller fraction of people who chose the gain-loss incentives could have been partly responsible for the effects in the field experiment.

To address these concerns, we conducted another experiment on Amazon Mechanical Turk involving a cognitively (as opposed to physically) taxing task. Responding to the limitations of the field study just noted, we elicited preferences for the standard gain-only versus aggressive gain-loss scheme from all participants (including those in the control condition) and offered a higher reward for choosing the gain-loss incentives, so as to elicit more opt-in to the aggressive incentive scheme than occurred in the field experiment. This internet study yielded results similar to those of the field study, reinforcing our confidence in its findings. This experiment is presented in full detail in Section 5.

5. Laboratory Experiment

In a second experiment, we sought to tweak some design aspects of the field experiment and validate our understanding of how various incentive schemes, as well as the choice to opt-in to these schemes, impacts individuals' propensities to meet a new goal. As part of the experiment, participants were asked to solve a series of numerical puzzles. The puzzle task was adapted from Shu et al. (2012), involving a 3×3 matrix in which each cell had a three-digit number, including two decimal points (see Figure 6(a)). To solve the puzzle, participants had to identify the combination of numbers in the matrix that add up to 10 (see Figure 6(b)).

All participants were provided the same goal to reach in terms of the number of puzzles to solve correctly, calibrated based on average solution rates from an earlier study. Although not about physical activity, this experiment has important overlaps with our field experiment. First, it allows us to validate the results from the field experiment and evaluate how various incentive schemes impact an individual's willingness to meet a new goal on an effortful task. Second, and like physical activity, participants have heterogenous preferences and aptitudes for the task. Third, the puzzles used in the experiment were pretested to be solvable by the average AMT participant and, in that way, mirrored the accessibility of daily steps evaluated in the field experiment.

This experiment involved the same four incentive conditions randomly assigned between subjects. In the control condition, participants were not provided with any additional incentive to meet the puzzle goals; however, distinct from the field experiment, they were asked to indicate their preference for the gain-only and gain-loss incentive schemes through hypothetical choice. The treatment conditions were identical to the field experiment and included conditions in which participants were either assigned to gain-only incentives (*gain-only assigned*) or gain-loss incentives (*gain-loss assigned*), regardless of the incentive option they preferred, and a choice condition in which participants were given the incentive scheme they preferred. Another important distinction from the field experiment is that we increased the relative magnitude of the gain in the gain-loss incentives to increase uptake and achieve better balance of individuals across experimental groups.

Figure 6. (a) Example Puzzle; (b) Instructional Text

(a)

7.30	5.20	0.17
5.52	7.40	8.19
1.77	8.23	9.15

Example Puzzle

(b)

"Your task involves searching for pairs of numbers in a limited amount of time. For each grid, find the two numbers that sum to 10."

Instructional Text

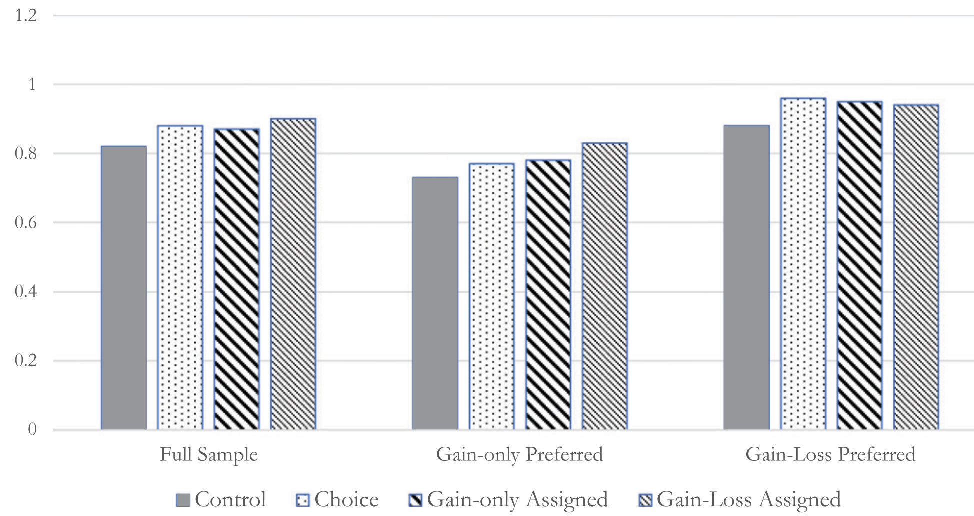
5.1. Experimental Procedure

Participants for the experiment were recruited from AMT to complete a study that was advertised to them as an effort "to better understand how people make decisions." All participants in the study were paid an initial \$0.50 upon accepting the request to participate in the study.

We first collected basic demographic information from participants (gender, age, race, and employment status) and introduced the task to them. To avoid confounds and preconceived notions associated with the term *puzzles*, we described the puzzles to participants as either "search tasks" or "grids" throughout the study. We then gave participants detailed instructions on how to properly solve the puzzle and enter their responses into the survey. We informed participants that they would not be penalized for incorrect responses and that they would not be able to return to a puzzle once an answer had been submitted. After the training, we provided participants five practice puzzles to solve. These puzzles were intended to help participants fine-tune their understanding of the rules and to correct any errors prior to engaging with the main task.

After they completed the practice round, we informed participants they would be awarded an additional bonus payment of \$0.50 upon completion of the study. We provided the \$0.50 bonus payment to endow participants with an amount (beyond their main payment) that they could risk if they chose or were assigned to the gain-loss incentives. They were also told that they would have five minutes to complete as many grids as they could and would be paid \$0.02 for each grid that they solved. We provided participants the \$0.02 payment to introduce a motive to persist in solving the puzzles, beyond simply the bonus for achieving the goal we assigned to them. We then provided participants with a grid goal to try and reach in the study. We informed participants that the goal provided for them was 10% more than what prior participants had solved on average (this number was calibrated based on a prestudy we ran). We had considered asking participants to indicate a goal for the task but worried that AMT participants may

Figure 7. (Color online) Goal Met by Condition and Incentive Preference



provide a goal strategically (i.e., anticipate that they may be asked to meet that goal in the future and provide a low goal). Also, participants on AMT may not be sufficiently familiar with the grid solving task to appropriately gauge what they considered a challenging goal. Thus, we simply chose a 10% increase from the average across all subjects of the prestudy baseline.

We described both the gain-only and gain-loss incentive schemes to all participants. After reminding them of the grid goal, we asked participants to indicate their preference for one of the incentives. Participants in the control condition made this choice as a hypothetical—that is, unincentivized. Participants in the treatment condition were asked to indicate an actual preference for one of the incentives. After they provided their preference, we informed participants in the choice treatment condition that "not all participants were given their choice of incentives" and that they were amongst those who were given their choice. Participants in the gain-loss assigned and gain-only assigned conditions were told they would be randomly assigned to one of the two incentive schemes. We asked these participants to select heads or tails and informed them that the computer would also randomly flip a coin. If the coins matched, we assigned them to the gain-only incentive option; if the coins did not match, we assigned them to the gain-loss incentive. Participants then progressed to solving the grids. After each grid attempt, participants were told whether they got the correct answer, reminded of their total score, and asked whether they wished to continue solving grids or to quit. Participants kept attempting grids until they finished all 30 grids available in the study or quit. Finally, participants answered a series of exit questions about solving the grids (how enjoyable the grid task was, confidence in reaching the goal, etc.).

5.2. Prestudies

Prior to running the main experiment, we conducted two prestudies to calibrate key parameters (detailed results and discussion of the prestudies are provided in Online Appendix A-9). In the first prestudy, we sought to establish a baseline for grids solved and identify a goal to provide to participants. Utilizing a largely identical protocol to the main experiment with baseline incentives to solve the grids (initial \$0.50 payment and \$0.02 per grid) but without a new goal or additional incentives options, we found that participants attempted 14.4 grids on average, got 13.7 of those grids correct, and spent 27 seconds on each grid. Overall, the results of the prestudy suggest that the grids are solvable but still cognitively taxing enough to deter some individuals from continuing with the task. In a second prestudy, we provided participants with a grid goal that was 10% greater than the average solved in the first prestudy (15 grids) and asked them to indicate (in hypothetical terms) their preference for gain-only versus gain-loss incentives. Our aim in this prestudy was to identify a reward amount that would result in uptake of the gain-loss incentive scheme such that each incentive scheme was close to equally preferred. Specifically, we presented participants with six scenarios, asking them to choose between a gain-only incentive scheme (labeled "Option 1") that rewarded them \$0.25 for meeting the grid goal and a gain-loss incentive scheme (labeled "Option 2") that would reward them a higher amount for meeting their goal (the reward started at \$0.50 and was incremented to \$0.80) but had a penalty if they missed their goal (the penalty started at −\$0.30 and was incremented to −\$0.40 for the highest reward option). Based on the results, we opted for a gain-only scheme that rewarded participants \$0.25 if they met their goal and no penalty

Table 10. Laboratory Experiment Main Results

	Goal met			Time on task			Compensation
Variables	(1) Whole sample	(2) Gain-only preferred	(3) Gain-loss preferred	(4) Whole sample	(5) Gain-only preferred	(6) Gain-loss preferred	(7) Total compensation
Choice	0.0568*	0.0345	0.0859**	32.12	53.30	22.78	0.207***
	(0.0330)	(0.0611)	(0.0326)	(36.72)	(52.50)	(50.01)	(0.0291)
Gain-only assigned	0.0530*	0.0412	0.0709*	23.26	38.58	17.41	–
	(0.0323)	(0.0593)	(0.0334)	(34.68)	(49.11)	(47.43)	–
Gain-loss assigned	0.0818***	0.0986*	0.0680**	33.14	96.7*	−9.981	0.344***
	(0.0311)	(0.0582)	(0.0333)	(36.07)	(51.38)	(49.32)	(0.0310)
Constant	0.818***	0.735***	0.875***	644.4***	584.8***	685.0***	1.620***
	(0.0248)	(0.0448)	(0.0277)	(25.89)	(34.09)	(36.51)	(0.0144)
Observations	998	420	578	998	420	578	756

Note. Robust standard errors in parentheses.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

if they failed to meet it and a gain-loss scheme that rewarded them \$0.65 if they met their goal but also penalized them −\$0.40 if they did not meet the goal.

5.3. Results and Discussion

Nine hundred ninety-eight individuals (M_{age}=36, 38% male) participated in our AMT experiment. A slight majority of participants (57%) preferred the gain-loss incentive scheme to the gain-only incentive scheme. Overall, 87% of the sample was able to meet the grid goal. The breakdown of individuals across the experimental conditions and incentive preference is provided in Online Appendix A-9.

Model free results suggest that all treatment conditions resulted in some increase in the probability of meeting the grid goal provided to participants, relative to the no incentive control (see Figure 7).

The effect of incentives was, however, most pronounced in the condition in which all participants were assigned to the gain-loss incentives. Moreover, similar to the field study, the effect of the treatment differed depending on participants' preference for the gain-only versus gain-loss incentives. Participants who preferred the gain-only incentives met the grid goal less often, relative to those who preferred the gain-loss scheme. This is indicative of strategic sorting, where those who have more aptitude for the task and expect to do better are more likely to self-select into gain-loss incentives. Most importantly, those who preferred the gain-only incentives did particularly well when their preference was ignored and they were assigned to gain-loss incentives. In contrast, those who preferred the gain-loss incentive scheme see nearly equal performance across all of the treatments relative to the control condition.

We confirm these results in regression analysis using a linear probability model and robust gain-only errors (see Table 10, column (1)). Not conditioning on individuals' preferences, all conditions had a significant effect on the probability of meeting the goal, with assignment to the gain-loss incentives having the strongest effect (8%, $p < 0.01$). However, effect sizes were similar for other treatment groups; the differences between treatment groups were not statistically different. Conditioning on an individual's preference for the gain-only versus gain-loss incentive scheme, however, again reveals a more nuanced result. Those who preferred the gain-only incentives only saw a significant increase in the probability of meeting their goal when they were assigned to the gain-loss scheme but not otherwise (column (2)). In contrast, those who preferred gain-loss incentives had a higher probability of meeting their goal, independent of the treatment they were in (column (3)). These results are consistent with the field experiment, even after significantly increasing the number of participants preferring the gain-loss incentives. This suggests that the imbalance in preferences in the field experiment favoring the gain-only incentives doesn't fundamentally impact our conclusions.

To evaluate whether participants exhibited more effort and persisted longer because of the incentives, we also examined the effect of our treatments on the time participants spent working on the grids. Although we find that individuals in the treatment conditions spent slightly more time on the task, these differences were not significant in the overall sample (column (4)). Conditioning on an individual's preference for the gain-only versus gain-loss incentives, we find that individuals with a preference for the gain-only incentives persisted more on the task (96.7 seconds, $p < 0.1$) when assigned to the gain-loss incentives but not otherwise (column (5)). In contrast, those who preferred the gain-loss incentives did not spend any additional time on the task in any of the conditions (column (6)). This result suggests that persistence in the task was not impacted by our treatments for those who preferred the gain-loss incentives. We acknowledge, however, that other dimensions of effort that we do not measure (e.g., focus on the task) may have been impacted by our treatments.

Finally, we evaluate differences in compensation between participants and find that participants earned more when provided their choice of incentives or assigned to the gain-loss incentives (relative to assignment to the gain-only incentives). This result highlights that, although a higher reward with gain-loss incentives can elicit higher opt-in, it may also be costlier from the perspective of entities administering the incentives (e.g., employers, government bodies). This additional cost is particularly problematic for the choice condition, as participants did not differ in performance from the more cost-effective gain-only approach.

The effect of our treatments on the number of grids attempted and total solved were directionally consistent but statistically insignificant in our analysis (results excluded for clarity of presentation). This result suggests that participants in the AMT sample were primarily motivated by the economic incentives tied to meeting the goal and exerted little additional effort on solving grids beyond the target we set for them. Also, this result suggests that our incentives largely influenced those at the margin who only needed to solve one or two additional grids to meet the goal.

Consistent with results in the field experiment, we find that diverse incentive schemes can promote desired changes in behavior but also find that these benefits depend on the incentive scheme as well as on an individual's incentive preference. Surprisingly, however, maximally incentivizing an individual does not entail giving them the incentives they desire. Almost the opposite, assignment to gain-only incentives as well as schemes that allow individuals to choose their incentive scheme have more modest effects on goal adherence relative to assignment to gain-loss schemes. These results are consistent with the field experiment results despite the change of experimental context, solicitation of incentive preferences for the control group, and significantly increasing the uptake of the gain-loss incentives. This consistency in results suggests that the dynamics observed in our field experiment are likely to persist even with more attractive (and thus more popular) aggressive incentives.

6. General Discussion: Implications for Research and Practice

Our results show that although both the standard gain-only and the more aggressive gain-loss incentive schemes can impact behavior, schemes that offer individuals a choice may underperform because those who would benefit the most from gain-loss incentives may be the least likely to choose them. In parallel, those who prefer more aggressive gain-loss schemes perform well with any of the incentive schemes. The results of our second experiment demonstrate that these dynamics persist in a different context and when increasing the reward associated with gain-loss incentives. In addition, our results suggest that longer-term adherence to improved physical activity as a result of incentive-based interventions is most likely when individuals are assigned to gain-loss incentives, against their own preferences, and are provided with prompts that increase engagement with the health tools available in modern-day and increasingly popular health wearables. Moreover, these results can apply beyond healthcare, as any context that requires persistence and regularity of effort and where performance can be digitally measured could benefit from a combination of economic incentives and motivational tools enabled by digital data collection capabilities.

These studies have limitations. One potential limitation is that the student population in the field experiment may not be representative of the general population (e.g., they are more physically active). Thus, extrapolating these results to less healthy populations may not be appropriate because less healthy individuals could have difficulty modifying their behavior in response to gain-loss incentive approaches. Further research is needed to explore how these results could differ for less active populations and consider how these strategies may still be utilized for them. For example, the goal increase could be diminished to account for more sedentary populations (e.g., a 5%–10% increase instead of a 20% increase).

Another concern is that we were not able to offer the choice of incentive scheme multiple times to participants. As a result, we cannot study possible learning effects by which individuals recognize their self-control issues and make more self-interested choices with more experience. Kaur et al. (2015) find some evidence of learning: individuals who exhibited self-control issues, in their experiment, increasingly opted into gain-loss incentives in a workplace context. However, these individuals were exposed to the incentive schemes daily over an extended period (their experiment lasted 13 months) and were able to immediately observe their financial outcomes because of their choices. Finally, we examine one health tool out of potentially many health tools available in health wearables (e.g., leaderboards) and cannot speak to whether our results would extend to other technology health tools.

These limitations notwithstanding, our results have significant implications for research and point to several additional directions for future work. For instance, our findings illustrate the value of exploring further the dynamics around choice-based incentive approaches, particularly when negative consequences are introduced as part of these schemes. Related to this point, future research could examine the mechanism behind the results in our manuscript. One possible explanation for the different reactions to incentive schemes for individuals who preferred the gain-loss versus gain-only incentives in our study is that those who preferred the gain-loss scheme were already motivated to change their physical behavior. Thus, they performed well above the goal provided to them independent of the incentive scheme during the challenge period and did not engage or effectively use the wearable health tool. However, as their motivation waned after the treatment period, so did their level of physical activity. In contrast, those who preferred the conventional gain-only incentives were less confident that they could meet their new goal or were unmotivated to increase their physical activity (or both). Thus, when they were provided a standard gain-only incentive scheme (either by choice or through random assignment) that did not include a punishment for failing to change their behavior, they failed to increase their level of physical activity. When they were forced into an incentive scheme that introduced a punishment, however, it motivated a departure from their status quo during the challenge period. Moreover, they recognized their need for support in meeting this new goal and effectively utilized the tools provided by the wearable device to meet this goal. Finally, in the follow-up period, the positive psychological boost from achieving their goal in the challenge period coupled with effective use of this wearable health tool resulted in a more prolonged commitment to elevated levels of physical activity. Future research evaluating this potential explanation (and others) for our results could provide additional insights into how to best incorporate choice, economic rewards, and economic penalties into the design of incentive schemes.

Our work also highlights the significant potential benefit of incorporating increasingly popular health wearables and the health tools they provide into behavioral economics research in health. These wearables allow for the continuous measurement of physical activity (and other health outcomes) for extended periods and can provide better insight into the long-term effect of various types of interventions. Moreover, in the future, wearables may be available at a larger scale, be more accurate, provide more granular data than what has been previously available, and be introduced to difference contexts. In addition to the measurement capabilities provided by health wearables, many of the technology platforms associated with them introduce innovative health tools (often incorporating insights from behavioral research) intended to motivate increased physical activity. Our results provide evidence that coupling the ability to set step targets via these devices with economic incentives can lead to significantly more value, particularly in the longer term. However, there are still many other features of health wearables we have not explored. For example, the Fitbit platform allows individuals to join leaderboards that introduce a competitive element to physical activity. Although these leaderboards are intended to elicit increased physical activity, they could as easily be demotivators (particularly for some subset of users). Future evaluation of these health tools and their intersection with economic incentives may yield interesting research insights and speak to the value of health wearables more generally.

Our research also has significant implications for organizations interested in utilizing economic incentives or health wearables to improve health outcomes. Most notably, our results highlight that either intervention alone may be limited in its efficacy. Organizations should, therefore, consider providing the two in conjunction to maximize value. More specifically, when possible, entities interested in these incentive-based interventions should consider ways to introduce more gain-loss incentive schemes in a nonvoluntary fashion. In parallel, organizations should consider ways to ensure that participants utilize, alongside gain-loss economic incentives, health tools that health wearables enable. We recognize, however, that some incentive schemes are infeasible without the element of choice (e.g., deposit contracts require individuals to put their own income or wealth at risk). However, there may be ways to do this while also limiting pushback from individuals. For example, an organization could provide a meaningful additional reward for meeting health goals (as opposed to only introducing a penalty).

The acceptance of nonvoluntary and gain-loss incentive schemes can also be facilitated by how they are designed. Our experiments offer some insights into to how to introduce nonvoluntary gain-loss incentives in a real-world setting. Specifically, in our experiment, we endowed individuals with an initial payment and then instituted (for some of them) a nonvoluntary gain-loss incentive scheme. Importantly, the loss did not exceed the initial amount endowed to them. Nonvoluntary gain-loss incentive schemes in the real world can be structured in a similar fashion. For instance, many employers offer health insurance premium reductions to employees who take preventative health measures (e.g., get a flu shot, sign up for prenatal care), provide economic incentives for filling out health profiles, or subsidize gym access or the purchase of an activity tracker. To shift to a gain-loss incentives scheme, instead of providing a one-off subsidy up front to those employees interested in participating (which is typically an additional benefit on top of an employee's base earnings), a firm can provide an additional reward to employees who meet related health goals but reduce the subsidies or render them moot (e.g., the initial subsidy is retracted) if employees do not meet these goals. For example, a firm could, without offering a choice, give all employees a $100 annual credit that they can claim toward a healthful activity (gym membership, health wearable purchase, etc.). Analogous to standard gain-only incentive schemes, this credit can increase to $150 if individuals meet certain health goals (e.g., daily activity). Alternatively, the firm could, again without offering a choice, offer an initial credit of $100. However, this credit can increase to $200 at the end of the year if individuals meet certain health goals or diminish to $50 if they don't. In both cases, the employee is better off by being in the program. This approach results in all individuals being exposed to the gain-loss scheme while also reducing the ethical and moral ramifications of the nonvoluntary nature of the program.

One additional way to leverage the insights from our work is to keep gain-loss incentives optional but identify approaches to nudge hesitant individuals toward more aggressive incentive schemes. These approaches could include having gain-loss schemes selected by default, providing education increasing awareness of the potential benefits, and highlighting positive adoption of these schemes by peers. More so, organizations could first elicit individual preferences and then try to convince those who select conventional gain-only incentive schemes to switch.

Overall, our results point to significant potential value from interventions that couple well-thought-out economic incentives and various health tools available from an evolving set of consumer health wearables. As these technologies advance and become more pervasive, the potential ways to complement them with economic incentives will likely only increase.

Acknowledgments

The authors thank their reviewers for constructive and helpful comments and suggestions, the associate editor for exceptional effort and guidance throughout the review process, and the department editor for support and feedback on the manuscript. The authors are also grateful for the useful feedback from participants at research seminars at the University of Rochester, University of Florida, and Indiana University. They also thank the Center for Network and Data Science at the University of Notre Dame and the National Institutes of Health for their support of this research. The statements made and views expressed in this paper are solely the responsibility of the authors. All errors are the authors' own.

Endnotes

[1] See http://www.who.int/mediacentre/factsheets/fs385/en/, accessed September 23, 2018.

[2] In one study, only 11% of participants opted into "deposit contracts" where they place their own money into an account and only receive it back if they cease smoking (Giné et al. 2010), and only 12% opted into a similar scheme in the context of gym attendance (Royer et al. 2015).

[3] See https://www.uwyo.edu/wintherockies_edur/win%20steps/coordinator%20info/step%20conversions.pdf.

[4] If students did not sync, or if they changed their password from the one assigned to them, then we were unable to update their data until the issue was corrected.

[5] In the prestudy described later in this section, we test different parameterization of the incentive schemes to try and achieve parity between those who have a preference for the gain-only versus gain-loss incentive scheme.

[6] We excluded the two highest and two lowest observations of daily steps for each participant in the four weeks used for calibration to avoid skew in the goal introduced by extreme days of (in)activity by participants.

[7] We used the most recent month of activity to set this baseline. For most of our sample (287 individuals), this was the four weeks prior to the start of recruitment. For individuals who lost or broke their Fitbit or had trouble syncing their data in the weeks prior to the experiment, this period could be earlier (e.g., two to three months prior to the start of the experiment).

[8] Participants were not informed of the date of the challenge to avoid any strategic behavior by participants.

[9] BMI may not always be reflective of individuals who are at a healthy weight in the sample and is only a proxy for healthy weight attainment.

[10] Recall that the step goal was based on the four weeks prior to the recruitment period, whereas the entire pretreatment period is eight weeks.

[11] See https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmicalc.htm.

[12] There are fewer individuals in the pretreatment period because students lost or broke their Fitbit during the summer months and their devices would be replaced once they returned to campus.

[13] There are few individuals with missing steps data during the experiment week (e.g., because they did not wear their Fitbit device). Hence, our observations do not equal the number of participants multiplied by the number of days.

[14] The addition of individual controls results in a negative coefficient on the constant term. However, adjusting for average age, personality characteristics, and other controls yields comparable mean steps for the control group.

[15] Note that only those in the treatment conditions had an opportunity for bonus payment, hence the smaller sample size.

[16] The p-value of the coefficient in column one is 0.12. The coefficient is significant at the 10% level if controls are excluded.

References

Adjerid I, Acquisti A, Loewenstein G (2019) Choice architecture, framing, and cascaded privacy choices. *Management Sci.* 65(5): 2267–2290.

Anderson J, Burks S, DeYoung C, Rustichini A (2011) Toward the integration of personality theory and decision theory in the explanation of economic behavior. IZA Discussion Paper No. 6750, Institute of Labor Economics, Bonn, Germany.

Babcock PS, Hartman JL (2010) Networks and workouts: Treatment size and status specific peer effects in a randomized field experiment. NBER Working Paper No. 16581, National Bureau of Economic Research, Cambridge, MA.

Bai Y, Hibbing P, Mantis C, Welk GJ (2018) Comparative evaluation of heart rate-based monitors: Apple Watch vs. Fitbit Charge HR. *J. Sports Sci.* 36(15):1734–1741.

Bandura A, Cervone D (1983) Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *J. Personality Soc. Psych.* 45(5):1017.

Bartling B, Fehr E, Herz H (2014) The intrinsic value of decision rights. *Econometrica* 82(6):2005–2039.

Bauman AE (2004) Updating the evidence that physical activity is good for health: An epidemiological review 2000–2003. *J. Sci. Medicine Sport* 7(1):6–19.

Becker LJ (1978) Joint effect of feedback and goal setting on performance: A field study of residential energy conservation. *J. Appl. Psych.* 63(4):428.

Booth FW, Chakravarthy MV (2002) Cost and consequences of sedentary living: New battleground for an old enemy. *President's Council Physical Fitness Sports Res. Digest* 3(16):1–10.

Booth FW, Roberts CK, Thyfault JP, Ruegsegger GN, Toedebusch RG (2017) Role of inactivity in chronic diseases: Evolutionary insight and pathophysiological mechanisms. *Physiological Rev.* 97(4):1351–1402.

Borghans L, Duckworth AL, Heckman JJ, Ter Weel B (2008) The economics and psychology of personality traits. *J. Human Resources* 43(4):972–1059.

Briki W, Markman KD (2018) Psychological momentum: The phenomenology of goal pursuit. *Soc. Personality Psych. Compass* 12 (9):e12412.

Burtch G, Hong Y, Bapna R, Griskevicius V (2017) Stimulating online reviews by combining financial incentives and social norms. *Management Sci.* 64(5):2065–2082.

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

Charness G, Cobo-Reyes R, Jiménez N, Lacomba JA, Lagos F (2012) The hidden advantage of delegation: Pareto improvements in a gift exchange game. *Amer. Econom. Rev.* 102(5):2358–2379.

Chaudhry SJ, Klinowski D (2016) Enhancing autonomy to motivate effort: An experiment on the delegation of contract choice. *Experiments in Organizational Economics*, vol. 19 (Emerald Publishing Ltd., Bingley, UK), 141–157.

Chen Y, Konstan J (2015) Online field experiments: A selective survey of methods. *J. Econom. Sci. Assoc.* 1(1):29–42.

Dohmen T, Falk A, Huffman D, Sunde U (2010) Are risk aversion and impatience related to cognitive ability? *Amer. Econom. Rev.* 100(3):1238–1260.

Dugas M, Crowley K, Gao GG, Xu T, Agarwal R, Kruglanski AW, Steinle N (2018) Individual differences in regulatory mode moderate the effectiveness of a pilot mHealth trial for diabetes management among older veterans. *PLoS One* 13(3):e0192807.

Erez M, Zidon I (1984) Effect of goal acceptance on the relationship of goal difficulty to performance. *J. Appl. Psych.* 69(1):69.

Evenson KR, Goto MM, Furberg RD (2015) Systematic review of the validity and reliability of consumer-wearable activity trackers. *Internat. J. Behav. Nutrition Physical Activity* 12(1):159.

Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EEK, Evenson KR (2016) Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): A randomised controlled trial. *Lancet Diabetes Endocrinology* 4(12):983–995.

Fishbach A, Dhar R (2005) Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *J. Consumer Res.* 32 (3):370–377.

Fokkema T, Kooiman TJ, Krijnen WP, Van CDS, De MG (2017) Reliability and validity of ten consumer activity trackers depend on walking speed. *Medicine Sci. Sports Exercise* 49(4):793–800.

Giné X, Karlan D, Zinman J (2010) Put your money where your butt is: A commitment contract for smoking cessation. *Amer. Econom. J. Appl. Econom.* 2(4):213–235.

Gneezy U, Meier S, Rey-Biel P (2011) When and why incentives (don't) work to modify behavior. *J. Econom. Perspect.* 25(4):191–209.

Goes PB (2013) Editor's comments: Information systems research and behavioral economics. *MIS Quart.* 37(3):3–8.

Gupta SK (2011) Intention-to-treat concept: A review. *Perspect. Clinical Res.* 2(3):109.

Gupta A, Kannan K, Sanyal P (2018) Economic experiments in information systems. *MIS Quart*. 42(2):595–606.

Halpern SD, French B, Small DS, Sausgiver K, Harhay MO, Audran-McGovern J, Loewenstein G, Asch DA, Volpp KG (2016) Heterogeneity in the effects of reward- and deposit-based financial incentives on smoking cessation. *Amer. J. Respiratory Critical Care Medicine* 194(8):981–988.

Halpern SD, French B, Small DS, Sausgiver K, Harhay MO, Audran-McGovern J, Loewenstein G, Brennan TA, Asch D, Volpp KG (2015) A randomized trial of four financial incentive programs for smoking cessation. *New England J. Med.* 372(22):2108–2117.

Handel B, Kolstad J (2017) Wearable technologies and health behaviors: New data and new methods to understand population health. *Amer. Econom. Rev.* 107(5):481–485.

Hashim MJ, Kannan KN, Maximiano S (2017) Information feedback, targeting, and coordination: An experimental study. *Inform. Systems Res.* 28(2):289–308.

Hashim MJ, Kannan KN, Wegener DT (2018) Central role of moral obligations in determining intentions to engage in digital piracy. *J. Management Inform. Systems* 35(3):934–963.

Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, Macera CA, et al. (2007) Physical activity and public health: Updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation* 116(9):1081.

Jakicic J, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, Rickman AD, Wahed AS, Belle SH (2016) Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: The IDEA randomized clinical trial. *JAMA* 316(11): 1161–1171.

James TL, Wallace L, Deane JK (2019) Using organismic integration theory to explore the associations between users' exercise motivations and fitness technology feature set use. *MIS Quart.* 43(1): 287–312.

John OP, Donahue EM, Kentle RL (1991) Big five inventory. *J. Personality and Soc. Psych.*

John LK, Loewenstein G, Troxel AB, Norton L, Fassbender JE, Volpp KG (2011) Financial incentives for extended weight loss: A randomized, controlled trial. *J. General Internal Medicine* 26(6): 621–626.

Kaur S, Kremer M, Mullainathan S (2015) Self-control at work. *J. Political Econom.* 123(6):1227–1277.

Khern-am-nuai W, Kannan K, Ghasemkhani H (2018) Extrinsic vs. intrinsic rewards for contributing reviews in an online platform. *Inform. Systems Res.* 29(4):871–892.

Larkin I, Leider S (2012) Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. *Amer. Econom. J. Microeconom.* 4(2):184–214.

Lee IM, Buchner DM (2008) The importance of walking to public health. *Medicine Sci. Sports Exercise* 40(7):S512–S518.

Levitt SD, List JA (2007) What do laboratory experiments measuring social preferences reveal about the real world? *J. Econom. Perspect.* 21(2):153–174.

Levitt SD, List JA (2009) Field experiments in economics: The past, the present, and the future. *Eur. Econom. Rev.* 53(1):1–18.

Locke EA, Latham GP (1990) Work motivation and satisfaction: Light at the end of the tunnel. *Psych. Sci.* 1(4):240–246.

Lupton D (2016) *The Quantified Self* (Polity Press, Cambridge, UK).

Mellizo P, Carpenter J, Matthews PH (2014) Workplace democracy in the laboratory. *Indust. Relations J.* 45(4):313–328.

Mercer K, Li M, Giangregorio L, Burns C, Grindrod K (2016) Behavior change techniques present in wearable activity trackers: A critical analysis. *JMIR Mhealth Uhealth* 4(2):e40.

Mitchell MS, Goodman JM, Alter DA, John LK, Oh PI, Pakosh MT, Faulkner GE (2013) Financial incentives for exercise adherence in adults: Systematic review and meta-analysis. *Amer. J. Preventative Medicine* 45(5):658–667.

Morris JN, Hardman AE (1997) Walking to health. *Sports Medicine* 23(5):306–332.

Nevin JA, Mandell C, Atak JR (1983) The analysis of behavioral momentum. *J. Experiment. Anal. Behav.* 39(1):49–59.

Paluska SA, Schwenk TL (2000) Physical activity and mental health. *Sports Medicine* 29(3):167–180.

Patel MS, Asch DA, Volpp KG (2015) Wearable devices as facilitators, not drivers, of health behavior change. *JAMA* 313(5):459–460.

Penedo FJ, Dahn JR (2005) Exercise and well-being: A review of mental and physical health benefits associated with physical activity. *Current Opinion Psychiatry* 18(2):189–193.

Piwek L, Ellis DA, Andrews S, Joinson A (2016) The rise of consumer health wearables: Promises and barriers. *PLoS Medicine* 13(2):e1001953.

Purta R, Mattingly S, Song L, Lizardo O, Hachen D, Poellabauer C, Striegel A (2016) Experiences measuring sleep and physical activity patterns across a large college cohort with Fitbits. *Proc. 2016 ACM Internat. Sympos. Wearable Comput.* (ACM, New York), 28–35.

PwC (2016) The wearable Life 2.0: Connected living in a wearable world. Accessed April 27, 2021, https://www.pwc.com/ee/et/publications/pub/pwc-cis-wearables.pdf.

Rogers T, Milkman KL, Volpp KG (2014) Commitment devices: Using initiatives to change behavior. *JAMA* 311(20):2065–2066.

Royer H, Stehr M, Sydnor J (2015) Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a fortune-500 company. *Amer. Econom. J. Appl. Econom.* 7(3):51–84.

Sahinidis AG, Tsaknis PA, Gkika E, Stavroulakis D (2020) The influence of the big five personality traits and risk aversion on entrepreneurial intention. *Strategic Innovative Marketing and Tourism* (Springer, Cham, Switzerland), 215–224.

Shu LL, Mazar N, Gino F, Ariely D, Bazerman MH (2012) Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison with signing at the end. *Proc. Natl. Acad. Sci. USA* 109(38):15197–15200.

Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM (2016) How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport Exercise Medicie* 2(1):e000106.

Strang HR, Lawrence EC, Fowler PC (1978) Effects of assigned goal level and knowledge of results on arithmetic computation: A laboratory study. *J. Appl. Psych.* 63(4):446.

Sullivan AN, Lachman ME (2017) Behavior change with fitness technology in sedentary adults: A review of the evidence for increasing physical activity. *Frontiers Public Health* 4:289.

Sun T, Gao G, Jin G (2019) Mobile messaging for offline group formation in prosocial activities: A large field experiment. *Management Sci.* 65(6):2717–2736.

Swan M (2013) The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data* 1(2):85–99.

Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Quart. J. Econom.* 106(4):1039–1061.

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

Wahl Y, Düking P, Droszez A, Wahl P, Mester J (2017) Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. *Frontiers Physiology* 8:725.

Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, Gillinov M (2017) Accuracy of wrist-worn heart rate monitors. *JAMA Cardiology* 2(1):104–106.

Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 73(1):121–130.