



University of Pennsylvania
ScholarlyCommons

Operations, Information and Decisions Papers

Wharton Faculty Research

2-2014


Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling

Katherine L. Milkman
University of Pennsylvania

Julia A. Minson

Kevin. G. Volpp
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/oid_papers

 Part of the [Exercise Science Commons](#), [Other Education Commons](#), and the [Sports Management Commons](#)

Recommended Citation

Milkman, K. L., Minson, J. A., & Volpp, K. G. (2014). Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling. *Management Science*, 60 (2), 283-299. <http://dx.doi.org/10.1287/mnsc.2013.1784>

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/oid_papers/150
For more information, please contact repository@pobox.upenn.edu.

Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling

Abstract

We introduce and evaluate the effectiveness of temptation bundling—a method for simultaneously tackling two types of self-control problems by harnessing consumption complementarities. We describe a field experiment measuring the impact of bundling instantly gratifying but guilt-inducing “want” experiences (enjoying page-turner audiobooks) with valuable “should” behaviors providing delayed rewards (exercising). We explore whether such bundles increase should behaviors and whether people would pay to create these restrictive bundles. Participants were randomly assigned to a full treatment condition with gym-only access to tempting audio novels, an intermediate treatment involving *encouragement* to restrict audiobook enjoyment to the gym, or a control condition. Initially, full and intermediate treatment participants visited the gym 51% and 29% more frequently, respectively, than control participants, but treatment effects declined over time (particularly following Thanksgiving). After the study, 61% of participants opted to pay to have gym-only access to iPods containing tempting audiobooks, suggesting demand for this commitment device.

Keywords

commitment devices, temptation bundling, self-control, field experiment, exercise

Disciplines

Exercise Science | Other Education | Sports Management

Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling

Katherine L. Milkman

Julia A. Minson

Kevin G.M. Volpp

The Wharton School, University of Pennsylvania

Last Revised: June 6, 2013

Abstract: We introduce and evaluate the effectiveness of temptation bundling – a method for simultaneously tackling two types of self-control problems by harnessing consumption complementarities. We describe a field experiment measuring the impact of bundling instantly-gratifying but guilt-inducing “want” experiences (enjoying page-turner audiobooks) with valuable “should” behaviors providing delayed rewards (exercising). We explore whether such bundles increase *should* behaviors and whether people would pay to create these restrictive bundles. Participants were randomly assigned to a full treatment condition with gym-only access to tempting audio novels, an intermediate treatment involving *encouragement* to restrict audiobook enjoyment to the gym, or a control condition. Initially, full and intermediate treatment participants visited the gym 51% and 29% more frequently, respectively, than control participants, but treatment effects declined over time (particularly following Thanksgiving). After the study, 61% of participants opted to pay to have gym-only access to iPods containing tempting audiobooks, suggesting demand for this commitment device.

Keywords: commitment devices; temptation bundling; self-control; field experiment

Acknowledgements: *We thank the Pottruck Health & Fitness Center and Wharton Behavioral Lab for their assistance with this project. We are particularly grateful to Elizabeth Herrick, Amy Wagner, Kaity Moore and Young Lee as well as our Research Assistants Samantha Lee, Benjamin Kirby, Daniel Milner and Alexander Rogala. We also thank Max Bazerman, Barbara Mellers, Klaus Wertenbroch, Uri Simonsohn, and participants at the SCP Winter 2013 Conference as well as the CMU-Penn 2010 and 2012 Roybal Center Retreats for their insightful feedback on this research. Finally, we thank the Wharton Dean’s Research Fund, the Wharton Behavioral Lab, the Penn-CMU Roybal Center on Behavioral Economics and Health (NIA 1P30AG034546-01), and the NBER Roybal Center for Behavior Change in Health and Saving (NIH P30AG034532) for funding support.*

1. Introduction

With 68% of adult Americans overweight or obese as of 2008 (Flegal, Carroll, Ogden, & Curtin, 2010) and 112,000 deaths in the United States per year attributable to obesity (Flegal, Graubard, Williamson, & Gail, 2007), promoting weight loss is an urgent public health priority. Further, in light of skyrocketing healthcare costs caused in part by obesity, programs designed to encourage weight loss are of tremendous interest to most organizations (Finkelstein, Fiebelkorn, & Wang, 2005; Finkelstein, daCosta DiBonaventura, Burgess, & Hale, 2010). Despite the many benefits that exercise provides, including promoting weight loss (Andersen, 2010), only 50% of Americans exercise sufficiently, a percentage that has been steadily declining ("U.S. Physical Activity Statistics," 2007).

Recent research has highlighted the possibility that public policy interventions built on an understanding of the psychology surrounding the challenges associated with increasing physical activity may be particularly effective. For example, groundbreaking research conducted in the last several years has shown that incentivizing exercise is not only an effective way to increase physical activity, but that incentivizing *repeated* gym attendance can produce long-lasting exercise habits that remain after incentives are removed (Charness and Gneezy, 2009; Acland and Levy, 2011). This work underscores the value of taking both economics and psychology into account when seeking to increase exercise rates.

If low exercise rates are in part the result of self-control problems, as much past research suggests (see for example Della Vigna and Malmendier (2006)), interventions that use psychological tools and insights to tackle this obstacle may be particularly valuable and cost-effective. Limited willpower has been shown to play an important role in decisions made by individuals that affect weight gain, such as healthy eating and exercise: people intend to exercise and diet tomorrow but frequently lack the necessary willpower to act on those good intentions today (Della Vigna & Malmendier, 2006; Milkman, Rogers, & Bazerman, 2009; Read & van Leeuwen, 1998; Royer, Stehr, & Sydnor, 2012). On the other hand, limited willpower makes it difficult for individuals to resist engaging in many highly tempting behaviors involving indulgences that induce regret after-the-fact (for a review, see Milkman, Rogers, & Bazerman (2008)).

We propose that valuable, healthy behaviors could be increased while guilt and wasted time from indulgent behaviors are simultaneously decreased through the use of a previously unstudied intervention: “temptation bundling.” Temptation bundling involves the coupling of instantly gratifying “want” activities (e.g., watching the next episode of a habit-forming television show, checking Facebook, receiving a pedicure, eating an indulgent meal) with engagement in a “should” behavior that provides long-term benefits but requires the exertion of willpower (e.g., exercising at the gym, completing a paper review, spending time with a difficult relative). For example, imagine only allowing yourself to: enjoy the next episode of your favorite TV show while exercising, receive a pedicure while completing an

overdue manuscript review, or indulge in the burger you crave when spending time with your cranky uncle. Temptation bundling can solve two problems at once by increasing the desire of those with self-control problems to engage in beneficial behaviors requiring willpower, and reducing the likelihood that people will engage in indulgent activities that they will later regret. Temptation bundling may be particularly effective because it exploits complementarities that often exist between *wants* and *shoulds* to create added value. The simultaneous engagement in *wants* and *shoulds* can reduce the guilt associated with indulgences and offer a distraction from the unpleasantness of many beneficial activities.

We theorize that people with limited willpower who are aware of their self-control problems (“sophisticates” (O’Donoghue & Rabin, 1999)) would gain from and value an opportunity to force their future selves to engage in beneficial *should* activities while simultaneously preventing those future selves from indulging in pleasurable but ultimately guilt-inducing *want* activities. We test a previously unexplored method for enforcing these preferences by creating a temptation bundling program that bundles a highly tempting activity (listening to low-brow, page-turner audio novels) with an activity that requires exerting self-control (exercising).¹ By bundling access to a hedonic experience with exercise, exercise is made “tempting” and increasingly appealing, while the squandering of time and resources on a potentially regret-inducing indulgent activity is prevented. In short, the inverted shapes of the utility streams obtained from engaging in *want* and *should* behaviors are strategically combined by temptation bundling. This insures that those who discount the future heavily will engage in *shoulds* and will limit their engagement in *wants* to moments when the downstream negative consequences (e.g., guilt and wasted time) are minimized if not eliminated (see Figure 1).

In the present investigation we focus on two questions pertaining to the value of “temptation bundling.” First, our field experiment examines whether temptation bundling programs have the potential to induce behavior change, setting aside the question of whether individuals would be “sophisticated” enough about their self-control problems to voluntarily seek out such programs (O’Donoghue & Rabin, 1999). Answering this first question allows us to establish whether the temptation bundling idea has value. We do this by measuring the effectiveness of temptation bundling as a means of increasing exercise frequency. We also examine whether individuals are able to effectively self-impose a suggested temptation bundling rule. Previous research on mental accounting and goal setting indicates that without external referees, people often (though not always: see Burger, Charness and Lynham, 2011) have the capacity to adhere to pre-determined rules designed to mitigate self-control problems (Abeler & Marklein,

¹ Note that exercise does not fit the definition of a *should* for all individuals, nor does listening to lowbrow audio novels universally fit the definition of a *want*. However, we follow past *want/should* research that defines goods and experiences as *wants* and *shoulds* based on the attitudes of most individuals (see for example Milkman, Rogers, & Bazerman, 2009).

2008; Camerer, Babcock, Loewenstein, & Thaler, 1997; Cheema & Soman, 2008; Heath, Larrick, & Wu, 1999; Milkman & Beshears, 2009; Shefrin & Thaler, 1988; R. Thaler, 1985; R. H. Thaler, 1990, 1999; R. H. Thaler & Shefrin, 1981). By including an intermediate intervention in our experiment to test the effectiveness of suggested temptation bundling, we are able to disentangle the effectiveness of merely giving people the insight regarding the potential value of this technique from that of creating a structured environment that limits their ability to behave myopically.

The second question we address is whether individuals are willing to restrict their own behavior in order to garner the benefits of temptation bundling. We investigate this critical question after first establishing the power of temptation bundling to change behavior in our field experiment. Measuring willingness to pay for temptation bundling devices is important because non-zero willingness to pay would be crucial for widespread adoption of temptation bundling without government or employer incentives, suggesting a potential market in which for-profit entities could package and sell these types of commitment devices. This second question is further of considerable theoretical interest, as evidence of a non-zero willingness to pay would add to the mounting literature contradicting the neoclassical economic models of behavior whereby rational actors see no value in restricting their future selves.

Indeed, past research has demonstrated that people value mechanisms that prevent their future selves from making unwise decisions such as procrastinating (Ariely & Wertenbroch, 2002), under-saving (Ashraf, Karlan, & Yin, 2006; Beshears, Choi, Laibson, Madrian, & Sakong, 2011), smoking cigarettes (Giné, Karlan, & Zinman, 2010), failing to achieve workplace goals (Kaur, Kremer, & Mullainathan, 2010), and giving into repeated temptations in the laboratory (Houser, Schunk, Winter, & Xiao, 2010). Study participants have expressed a willingness to use “commitment devices” such as placing money in restrictive accounts that prevent premature savings withdrawals (Ashraf et al., 2006; Beshears et al., 2011); self-imposing deadlines with associated late penalties to prevent procrastination on coursework (Ariely & Wertenbroch, 2002); and placing money on the line for forfeiture if they fail to quit smoking, exercise or lose weight (Halpern, Asch, & Volpp, 2012; John, Loewenstein, & Volpp, 2012; Royer et al., 2012; Volpp et al., 2008).

Conceptually, temptation bundling devices are a previously unstudied form of commitment device with several distinct features. First, temptation bundling devices do not require monetary transfers (or any other form of punishment) between the user and an outside entity. Rather than imposing a cost on individuals who break commitments to exert self-control, a temptation bundling device makes the activity whose pursuit requires willpower more alluring. Second, temptation bundling may be particularly effective if complementarities exist between a temptation item and the healthy behavior it is bundled with. For example, complementarities may exist between exercising and listening to tempting audio novels such that exercising *while* listening to fiction may create more net utility than that created from engaging

in both activities separately. Third, to the extent that the tempting activities bundled with *should* behaviors are habit-forming, they may be particularly powerful motivators (Solomon & Corbit, 1974). In other words, individuals may be particularly eager to return to the gym to hear the next chapter of their novel or to view the next episode of a television show after a cliffhanger. Finally, previous psychology research has suggested that engaging in healthy behaviors like exercise depletes willpower, which is a finite resource (Baumeister, Bratslavsky, Muraven, & Tice, 1998; M. Muraven, Tice, & Baumeister, 1998; Mark Muraven & Baumeister, 2000), while engaging in indulgences has the opposite, repleting effect.² Temptation bundling may be particularly effective if giving into a temptation increases an individual's available willpower, making the net impact of exercise on willpower less depleting (and potentially even positive).

2. Research Overview

To investigate our first research question – whether temptation bundling can create value – we conducted a three-condition randomized, controlled trial in collaboration with a large University fitness facility. Participants were randomly assigned to one of three experimental conditions and their frequency of gym attendance was measured. In the *full treatment condition*, participants were given access to an iPod containing four “want” audio novels of their choice that they could only listen to at the gym. In the *intermediate treatment condition*, participants were also given access to four “want” audio novels of their choice, but these novels were loaded onto their personal iPods, which they could access at any time. These participants were encouraged to try self-imposing a rule whereby they only allowed themselves to enjoy audio novels while exercising. Finally, in the *control condition*, participants were given a \$25 Barnes and Noble gift certificate at the start of the study (valued equivalently to the loan of four audio novels, see 2.2 Procedures).

After approximately 9 weeks, the study concluded and treatment condition participants returned their loaned audio novels and/or iPods. At this time, to investigate our second research question – whether temptation bundling devices are attractive to people – we assessed individuals' willingness to pay

² Muraven and Baumeister (2000) argue that self-control acts like a muscle that can be both exhausted by repeated use and restored through rest (or in other words, by giving into temptation rather than resisting it). In one study supporting this theory, subjects who were forced to exert self-control by consuming a *should* item rather than a *want* item subsequently exhibited less persistence when working on unsolvable puzzles than others who were allowed to consume *wants* instead of *shoulds* (Baumeister et al., 1998). Furthermore, past experimental research has shown that enjoying a *want* stimulus (such as watching a comedy video) restores the willpower depleted by an initial act of self-regulation to baseline levels and significantly more successfully than engaging with a neutral stimulus (Tice et al., 2007). Specifically, after enjoying a *want* film, experimental participants across three studies (Tice et al., 2007) exerted more willpower by persisting for longer on a frustrating ball-rolling task (one study), an uncomfortable handgrip task (a second study), and a puzzle-solving task (a third study).

for this type of commitment device. Specifically, through an incentive-compatible elicitation method, all participants reported the amount they were willing to pay for one month of restricted, gym-only access to an iPod loaded with a single tempting audio novel of their choice, which they would otherwise be able to take home and access at any time free of charge.

Comparing average weekly gym attendance frequencies across experimental conditions, we find that attendance rates increased meaningfully and significantly with access to the temptation bundling program, suggesting that temptation bundling creates value; but, the effectiveness of the intervention wears off over time (particularly following Thanksgiving). In addition, the majority of participants stated a non-zero reservation price for one month of access to a temptation bundling device (restricting their access to an iPod they would otherwise be entitled to use freely) at the end of the intervention period, suggesting a market for this type of commitment device.

This study is the first to describe and evaluate temptation bundling devices – a novel tool engineered based on past self-control research to harness complementarities and cost-effectively increase desirable, *should* behaviors while simultaneously decreasing undesirable, *want* behaviors. We find evidence that temptation bundling is indeed a cost-effective means of increasing exercise. Further, our participants' willingness to restrictively bundle their access to tempting experiences (requiring willpower to resist) with healthy behaviors (requiring willpower to carry out) is also of considerable theoretical interest. Our research adds to a growing body of evidence that many people are aware of the limitations on their willpower and are actively seeking new avenues for overcoming those limitations.

3. Experimental Design and Procedures

3.1 Study Participants

The participants in this study were 226 students, faculty and staff at a large, Northeastern U.S. University who responded to recruitment advertisements in September 2011 describing an opportunity to participate in a 10-week exercise study in exchange for \$100. Requirements for participation included having a University ID card, belonging to the university gym, owning an iPod (Shuffle, Nano, Classic or iTouch) with at least 1 GB of available storage space, being between 18 and 75 years old, and typically exercising on a stationary aerobic machine. All 226 participants in the study were randomly assigned to one of three groups: full treatment ($n = 75$), intermediate treatment ($n = 75$) or control ($n = 76$). Thirty-four percent of the study participants were male and 86% were undergraduates. The average self-reported minutes spent per week exercising in our sample prior to the study's launch was 100.2 (s.d. = 66.7, min = 0, max = 240), and the average number of visits made to the gym by our participants during the first week of classes (before the study's launch) was 1.5 (s.d. = 1.6, min = 0, max = 7) with 40% of participants never visiting the gym during the first week of classes. Participants' self-reported weight and height

allowed us to calculate their BMIs: the average BMI in our study was 23.3 (s.d. = 3.9, min = 18.7, max = 52.6).

3.2 Procedures

In order to sign up for the study, participants visited a website and filled out a short pre-screening online survey to: confirm their eligibility for the study, indicate their availability to attend intake sessions, consent to participate and have their gym attendance tracked, and self-report their gender, weight, height and average minutes spent exercising weekly at the university gym. Figure 2 presents a diagram of the flow of study participants. Those selected for study participation were first randomly assigned to a condition (stratified³ by self-reported gender, BMI, and weekly gym attendance) and then asked to complete a second online survey (see Electronic Companion for all survey materials) prior to visiting the University's behavioral lab for a 1.5 hour study "intake visit."

The second online survey differed across conditions. Participants in the control condition provided information about their typical time per week spent: (a) exercising on different types of machines, (b) exercising in different locations, and (c) listening to their iPod, as well as barriers to exercising as frequently as desired. Participants in the two treatment conditions also answered these questions but in addition, they selected six novels from a list of 82 deemed highly tempting and engrossing in pre-tests (see details below). These participants were told: "In this research study, you will be provided with an opportunity to listen to audio novels while exercising. Your goal is to select a set of novels to listen to that will be so engrossing and addictive that the prospect of listening to the next chapter will help motivate you to return to the gym each day to exercise."⁴ We provided participants with a webpage containing detailed descriptions and cover art for each of the 82 novels and required participants to select and rank six novels in order of preference.

Next, we required all participants to visit the University Behavioral Lab for a 1.5-hour study intake visit. Of the 226 participants who were randomly assigned to a condition, 215 attended an intake

³ We first sorted individuals into bins based on their self-reported gender (male or female), weekly minutes spent at the gym (grouped into the following nine bins: <15 min.; 15 – 45 min.; 46 – 75 min.; >76 – 105 min.; 106 – 135 min.; 136 – 165 min.; 166 – 195 min.; 196 – 225 min.; >225 min.), and BMI (grouped into the following 13 bins: BMI = 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28-30, 31-46, >46). Within each of these bins, individuals were randomly assigned to treatment conditions. This type of stratified random assignment algorithm (or randomization within blocks) ensures balanced samples across experimental conditions along the dimensions of stratification and typically decreases variance in estimated treatment effects (List, Sadoff and Wagner, 2010).

⁴ We worked with undergraduate research assistants to develop non-academic language that would resonate with participants and encourage them to select stimuli that met the criteria of theoretical interest in this research. Because the word "addictive," as used in lay parlance, is a vivid means of conveying the idea that the novels participants picked should be both instantly gratifying and attention-grabbing, we selected this term. Our research assistants did not raise concerns that the term "addictive" could trigger thoughts about the clinical definition of the term, but this of course cannot be ruled out.

session. These visits took place from September 26 to October 3, 2011, approximately one to two weeks after participants completed their second online survey. During these intake visits, we confirmed required iPod ownership and collected biometric data (body fat, pulse rate, weight, waist size, BMI) and online survey data. In addition, a research assistant took all participants across all conditions to the gym to exercise on an aerobic machine for 30 minutes and complete a short oral survey about their enjoyment of the workout.

In the control condition, the online survey that participants completed during their study intake visit reminded them of the importance of exercising in order to improve their health. Control participants received a \$25 gift card to Barnes and Noble prior to their 30 minute workout at the gym – a gift selected because pre-tests revealed that undergraduates perceived it to be of equal value to receiving 4 audio novels on an iPod for a ten week loan.

In the intermediate treatment condition, the online survey used identical language to the control condition to remind participants of the importance of exercising. Next, we introduced participants to the idea of using audio novels in temptation bundles with exercise. We explained the idea as follows:

“In order to make each workout you engage in at [University gym’s name] gym more tempting (so that after a long day, you will actually find yourself craving a workout rather than dreading one), we recommend that you try making a rule for yourself: **only let yourself enjoy these novels while exercising.**

The hope is that you will frequently find yourself longing to find out what happens next in your novel, and this will lead to trips to the gym to satisfy that curiosity. Before long, we hope you will find yourself exercising more regularly as a result of your audio-novel addiction.”

After completing a comprehension check to ensure they understood the program and the terms of the 10-week audio novel loan, participants loaded the four audio novels they had rated most highly onto their personal iPods. Finally, we required participants to listen to the first 30 minutes of one of these audio novels during their aerobic workout at the University gym and to answer questions about their enjoyment of the novel and the workout.

The procedures in the full treatment condition were nearly identical to those in the intermediate treatment condition. However, participants’ audio novel selections were loaded onto an iPod Shuffle that was loaned to them for the 10-week study but stored in a monitored locker at the University Gym. Thus in this condition, the idea of using audio novels in temptation bundles with exercise was both suggested and enforced. We explained the program to participants as follows:

“In order to make each workout you engage in at [University gym’s name] gym more tempting (so that after a long day, you will actually find yourself craving a workout rather than dreading one), **we will only let you enjoy these novels while exercising at [University gym’s name].** The iPod we are loaning you during this study will be required to remain in a locker at [University gym’s name] at all times when you are not checked into the gym facility, and study administrators will check the locker regularly to ensure compliance with the study protocols.

The hope is that you will frequently find yourself longing to find out what happens next in your novel, and this will lead to trips to the gym to satisfy that curiosity. Before long, we hope you will find yourself exercising more regularly as a result of your audio-novel addiction.”

After completing a comprehension check to ensure they understood the program and the terms of the 10 week iPod loan and after agreeing not to discuss the study with others, participants loaded the four audio novels they had rated most highly onto their loaned iPods. Finally, we required participants to listen to the first 30 minutes of one of these audio novels during their required intake visit aerobic workout at the University Gym and to answer questions about their enjoyment of the novel and the workout. Our research assistant also provided participants with the combination that would grant them access to a monitored iPod locker at the entrance of the University Gym (see Appendix Figure A1) and instructed them on how to store their iPod there and how to sign it out during exercise.

Participants in all arms of the study received weekly emails with a link to an online survey asking a series of questions about workout patterns and enjoyment as well as audio novel listening patterns in the treatment conditions. Despite frequent reminders, weekly survey completion rates were low, so these data were not analyzed.

It is important to highlight that participants across all experimental conditions in our study received the same, strong encouragement to exercise at the outset of our intervention period (as well as emailed surveys monitoring and therefore encouraging exercise throughout the study), along with equal incentive payments. The only difference between conditions was the provision of tools (temptation bundling apparatus) to participants in the full and intermediate treatment groups that were designed to assist with these participants’ efforts to exercise more regularly.

After a minimum of 9 weeks (exact study duration depended on participants’ intake and outtake dates), participants completed a 1-hour outtake visit at the University’s Behavioral Lab. Of the 215 participants who completed an intake visit, 214 returned for an outtake visit between December 5 and December 12, 2011. During these outtake visits, participants completed an online survey about their exercise experience throughout the study, and biometric data were again collected. At the end of the study, all participants across all conditions were truthfully told that they would have an approximately 10% chance of winning a 1 GB iPod Shuffle loaded with one audio novel of their choice (again from the list of 82 tempting novels). This iPod and its contents would be theirs to take home and use as they pleased. They were then told about a temptation bundling program that they could purchase if interested:

“For the first month of the next semester, we would hold the iPod shuffle that we gave you in a locked and monitored cabinet at the check-in counter of the Penn gym you prefer to visit. You would only be allowed to access your new iPod shuffle when in the gym and would be required to return it upon exiting. The benefit of this program is that your restricted access to the iPod will make it more tempting to go to the gym (so that after a long day, you will actually find yourself craving a workout rather than dreading one). The idea is that you will frequently find yourself eager to find out what

happens next in your novel, and this will lead to trips to the gym to satisfy that curiosity. Before long, you may find yourself exercising more regularly as a result of your audio-novel addiction. In addition, you won't waste time listening to your new audio-novel outside of the gym.”

In short, this program would restrict participants’ access to an iPod that they would otherwise be able to take home and use freely. Thus, the program is value-destroying under traditional, rational actor assumptions, as it reduces an individual’s access to one of her possessions. However, the program is value-enhancing if individuals anticipate benefits from temptation bundling.

After reading this program description, participants were first asked to indicate whether this program sounded appealing (a “yes”/“no” question). Next, participants’ reservation willingness to pay for the program was elicited using the incentive-compatible Becker-DeGroot-Marschak (BDM) method (Becker, DeGroot, & Marschak, 1964). This method was explained in great detail, and all participants were required to correctly complete a mathematical, four-question comprehension check ensuring they understood the BDM procedure before reporting their reservation willingness to pay for the temptation bundling program. If a participant provided any incorrect answers on this comprehension check, she would be required to re-read the description of the BDM method and successfully answer five mathematical questions from a second, different comprehension check (see pages 45-49 of *Electronic Companion*). If any question was answered incorrectly on this second comprehension check, a participant would exit the survey and this portion of the study. Of the 214 participants who attended an outtake visit, 212 completed this survey and 211 successfully completed the BDM comprehension check(s) and provided reservation prices between the minimum and maximum allowable (\$0 - \$100) for the one-month temptation bundling program.⁵ Finally, participants completed a series of questions designed to measure inter-temporal discount rates and locus of control.

3.3 Stimuli Selection

In order to develop a list of novels for use in this study, 233 undergraduates were recruited to complete a short survey at the University Behavioral Lab (see *Electronic Companion*). Participants were asked to list “the five novels you have read in the last several years that you found it the most difficult to put down once you had begun reading. Specifically, we are looking for the names of five ‘addictive’ fiction books.” From the set of suggested novels as well as a list of the top 20 bestselling novels each year from 1990-2011, a research assistant selected a set of 82 novels that were both frequently suggested, highly rated on “addictiveness,” and available for purchase on audio CD (see Appendix for complete list). The audio novels in our study lasted an average of 11.7 hours (minimum = 6 hours; maximum = 27

⁵ Two participants managed to skip the final survey in the study without detection.

hours). The most popular novels in our study were selections from The Hunger Games trilogy. Other popular selections included novels from the Da Vinci Code trilogy, the Twilight series, The Help, and The Devil Wears Prada.

3.4 Statistical Analysis

3.4.1 Effectiveness of Temptation Bundling Intervention

The primary outcome of interest in our study of the effectiveness of temptation bundling is gym attendance. In order to enter any University Gym facility, individuals are required to swipe their University ID card. Electronic entrance records for the University gyms for all of the participants in our study were provided from September 1, 2011 through December 9, 2011.

We evaluate the impact of our interventions on an intent-to-treat basis by calculating weekly gym attendance frequencies for participants in all conditions of our experiment and comparing the two treatment conditions to the control condition. Our analyses rely on an Ordinary Least Squares (OLS) regression with data clustered by participant to predict total gym visits by a given individual during each week of the study. This is the same modeling approach that was taken by both Acland and Levy (2011) and Royer et al. (2012) in their analyses of the effects of different interventions designed to increase gym attendance. This clustered or multi-level approach allows us to account for the fact that we obtained multiple periods of observations of gym attendance per participant, while adjusting standard errors to account for the fact that data points provided by the same participant are not independent. This approach thus maximizes statistical power while ensuring that non-independent observations do not artificially inflate significance levels. Anticipating that the effectiveness of our intervention might vary over time (or decay) also necessitated this modeling approach, because it allows for exploration of interactions between variables at different levels of analysis such as interactions between week of the study (a within-participant variable) and experimental condition (a between-participant variable). We report robust standard errors to alleviate concerns about heteroscedasticity in the data.

All regression analyses include binary indicators for a participant's experimental condition as primary predictors. All regression analyses also control for each participant's number of gym visits during the first week of the University's fall 2011 term when the study had not yet begun (linear and squared) as well as each participant's self-reported average minutes spent weekly exercising at the University Gym in the online pre-screening survey (linear and squared) to control for dramatic pre-treatment individual differences in gym usage. A control is also included for weeks since a given participant began the study (ranging from 0-8 since all participants completed 9 full weeks of the study prior to an outtake visit).

3.4.2 Willingness to Pay for Temptation Bundling Devices

The primary outcome of interest in our exploration of demand for temptation bundling devices is reservation willingness to pay for the month-long temptation bundling device offered to study participants at the beginning of the upcoming spring semester. Our analyses rely on one-sample mean comparison t-tests to evaluate whether willingness to pay is greater than zero and to evaluate whether a significant portion of our study population values our temptation bundling device.

4. Results

4.1 Effectiveness of Temptation Bundling

All 226 participants who completed our pre-screening survey and met study requirements (see 3.1 *Study Participants*) were randomly assigned to our full treatment, intermediate treatment, or control group. We found no significant differences in individual characteristics measured prior to the intervention across conditions (see Table 1).

Table 2 presents average, week-by-week gym attendance frequencies during our intervention period by experimental condition, with the first week for a participant commencing on the date of his or her study intake visit.⁶ Table 2 also presents the average total gym visits over the course of our study by condition. Figure 3 plots week-by-week attendance across conditions after subtracting participants' baseline, pre-treatment gym attendance (in order to provide a standard point of reference across groups).⁷ As Table 2 and Figure 3 both highlight, our treatment conditions directionally produce the hypothesized effects during our entire study period. However, these effects are driven by the seven weeks of the study prior to the University's Thanksgiving break (when the University Gym was closed) when participants in the treatment condition exercised consistently more than participants in the control condition. During this seven-week period, the average total number of gym visits by participants in the full treatment group was 7.8, while it was 6.5 in the intermediate treatment group, and 6.1 in the control group. Further, the average percentage of participants who visited the gym at least once in a given week of the study during the study's first seven weeks was 51% in the full treatment group compared to 44% in the intermediate treatment group and 42% in the control group (for week-by-week frequencies of gym visitors, see Appendix Table A2). During this pre-Thanksgiving period, a regression (Table 3, Model 1) to predict an individual's gym visits per week, with observations clustered at the participant level, indicates a

⁶ There are no significant differences in the dates of intake visits across conditions. For participants who never completed an intake visit, for purposes of intent-to-treat analyses, the intervention is assumed to begin at the date and time of the first available intake session (September 26, 2011 at 9 am EST).

⁷ Note that because gym attendance declines precipitously following the beginning of the fall semester, baseline gym attendance levels (measured at the outset of the fall semester) are higher than attendance levels during our intervention across all experimental groups.

difference between the full treatment condition and the control condition of 0.31 gym visits per week ($p = 0.026$) and a difference between the intermediate treatment condition and the control condition of 0.14 gym visits per week ($p > 0.10$), controlling for pre-study exercise frequency and (self-reported) duration as well as weeks since the study's start. Uncontrolled, nonparametric ranksum tests examining participants' net visits to the gym during this seven week period minus their gym visits during the first week of the fall term (prior to the start of our study) by condition produce similar results (H_0 : full treatment = control, $p = 0.075$; H_0 : intermediate treatment = control, $p > 0.10$)

The declining efficacy of the full treatment over time and particularly following Thanksgiving illustrated in Table 2 and Figure 3 (and reflected in the results presented in Table 3, Models 1 and 2, which show the treatment effect during the pre-Thanksgiving weeks of the intervention versus the full intervention period, respectively) is consistent with past research on gym attendance. Specifically, previous research on gym attendance suggests that habits can be formed around exercise, but they wear off over time and especially precipitously during holiday breaks (Acland & Levy, 2011). To assess the significance of this trend, we tested a regression model with our primary predictors, adding a control for the number of weeks since a participant's intake visit (ranging from 0-8) and an interaction between this variable and each of our treatment indicators (Table 3, Model 3). We find that in the first week after a participant's intake visit (week zero), the regression-adjusted difference between the full treatment condition and the control condition is 0.48 visits per week ($p = 0.004$) – a 51% increase over the regression-estimated 0.94 visits for the control group. The difference between the intermediate treatment condition and the control condition is 0.27 visits per week ($p = 0.092$) reflecting a 29% increase in attendance. However, the full treatment effect is significantly attenuated over time, decreasing by 0.07 gym visits per week over the full nine weeks of the study ($p = 0.005$), and the intermediate treatment's marginally significant benefit is marginally significantly attenuated over time, decreasing by 0.05 gym visits per week over the nine week study ($p = 0.069$).^{8,9}

4.2 Treatment Effect Interactions

⁸ All of the results we report become larger in magnitude and more statistically significant if only the participants who attended an intake visit ($N = 215$), and thus experienced our full intervention, are included in our analyses (see Table 3, Model 4).

⁹ We observe no significant differences between experimental groups in any of the biometric variables measured during intake and outtake visits (body fat, pulse rate, weight, waist size, BMI), as shown in Appendix Table A1. Although disappointing, it should also be noted that measures were not collected by medical professionals leading to errors in: use of BMI equipment, waist size measures, and pulse rate measures. Further, given the net magnitude of the effects of our intervention on exercise and our sample size, the lack of a measurable impact on health outcomes is not surprising.

Importantly, we find that the benefits of the full treatment condition vary depending on participants' self-reported enjoyment of the workout they completed during their study intake visit. Specifically, a one-standard deviation increase in self-reported intake visit workout enjoyment corresponds to an increase in the initial treatment effect of 0.21 gym visits per week ($p = 0.041$) (Table 3, Models 6 and 7). Further, when we compare initial workout enjoyment ratings across conditions, we observe a marginally significant increase in the number of people claiming their workout was enjoyable as opposed to negative or neutral when it involved an audio novel (in the treatment conditions) than when it did not (in the control condition; $t(224) = 1.84$; $p = 0.067$). These findings are in line with our theory whereby the act of exercising is itself less unpleasant when experienced in combination with an audio novel. They further support the idea that the more pleasant participants find exercising with audio novels, the more benefits they experience from temptation bundling.

Notably, anyone in our treatment conditions who did not enjoy exercising with an audio novel during our required, initial study workout would be free to exercise unencumbered by novels throughout our study. Thus, unless enrollment in our program created psychic costs, our intervention would not be expected to *reduce* workout enjoyment in the period following this initial workout even for those who disliked the temptation bundle we provided - it would rather only be expected to *benefit* those who found exercise more enjoyable when it was bundled with an audio novel.

Although not anticipated *ex ante*, we discovered one additional variable that interacted significantly with our treatment effect. The number of intake visit timeslots for which a participant indicated availability in our pre-screening survey was negatively correlated with intervention effectiveness. Availability is a proxy for how busy a participant is, and this variable significantly interacts with the effectiveness of our intervention. We find that temptation bundling particularly benefits our busiest participants: a one standard deviation decrease in availability corresponds to an increase in the initial treatment effect of 0.21 gym visits per week ($p = 0.014$) (Table 3, Models 5 and 7). This finding is consistent with the prediction that temptation bundling may be most powerful when it both increases the attractiveness of exercise and reduces guilt that can be associated with engaging in indulgent behaviors, as busier individuals are most likely to regret spending time on *wants*. This effect could also be driven by the fact that busier individuals are more in need of a reason to visit the gym.

4.3 Willingness to Pay for Temptation Bundling Devices

All 212 participants who completed a study outtake visit were told they had a chance of winning an iPod Shuffle loaded with one audio novel of their choice, which they would be able to take home and use at their leisure. We then presented them with an opportunity to pay for an exercise program: if they enrolled in the program, study staff would hold their iPod for one month in a monitored locker at the

University gym, ensuring they never listened at home and had something to look forward to during exercise. Sixty-four percent of participants indicated that this program “sound[ed] appealing” to them when asked on a “yes/no” scale ($t(211) = 19.43$; $p < 0.0001$), demonstrating that temptation bundling devices are valued and adding to the mounting evidence of demand for commitment devices (Ariely & Wertenbroch, 2002; Ashraf et al., 2006; Beshears et al., 2011; Giné et al., 2010; Halpern et al., 2012; Houser et al., 2010; John et al., 2012; Kaur et al., 2010; Volpp et al., 2008).

For the 211 participants who passed a comprehension check ensuring they understood the Becker-DeGroot-Marcshak reservation price elicitation method (Becker et al., 1964), willingness to pay for the program was next assessed (see Electronic Companion for precise elicitation procedures). Average willingness to pay for the program was \$6.91 ($t(210) = 9.34$; $p < 0.0001$), with only 39.3% ($t(210) = 18.00$; $p < 0.0001$) of participants stating a \$0 willingness to pay.¹⁰ A distribution of participants’ willingness to pay is presented in Figure 4.¹¹

We find some evidence that participants make sophisticated decisions about whether or not to pay for the program. Specifically, the effectiveness of the intermediate treatment was significantly lower for those who expressed a non-zero willingness to pay for the program ($\beta_{wtp>0_x_intermediate} = -0.59$; $p = 0.013$). In other words, the better self-imposed temptation bundling rules worked for a participant, the less likely she was to place value on the externally monitored temptation bundling program. Point estimates from our regression analyses can be used to illustrate this effect: an average person who offered to pay zero dollars for program enrollment at the end of the study experienced an initial benefit from the intermediate treatment of 0.67 gym visits per week ($p = 0.006$). However, a person who offered to pay a non-zero amount experienced a 0.30 gym visit per week *lower* benefit from the intermediate treatment. If we use a continuous measure of willingness to pay instead of an indicator for non-zero willingness to pay, our results are directionally the same but do not reach significance.

¹⁰ We did not predict differences in willingness to pay across experimental conditions, and thus do not describe between-condition analyses in detail. The likelihood of paying a non-zero amount for a temptation bundling device did not differ significantly across experimental groups (see Appendix Table A2). We did, however, observe significantly higher average willingness to pay for a temptation bundling device among participants in the control group than participants in either the full or intermediate treatment groups (see Appendix Table A2). It is unclear whether control group participants overestimate the value of the program, having never experienced it, or if treatment group participants underestimated the benefits they gained from the temptation bundling device (or appreciated its devaluation over time).

¹¹ Some of the comments and concerns participants expressed about the temptation bundling device program during early outtake sessions led us to worry that participants believed the receipt of the free iPod and audio novel might be contingent upon stating a high reservation price for the program, although our instructions explicitly stated this was not the case. Consequently, for the final 137 participants to complete an outtake visit, we added an additional comprehension check question to ensure participants knew that their chances of receiving an iPod and audio novel were unrelated to their stated willingness to pay for the temptation bundling device program. For the subpopulation who completed this extra check, average willingness to pay for the program was actually slightly higher than before: \$7.18 ($t(136) = 6.94$; $p < 0.0001$), suggesting that our results were not driven by a misunderstanding.

5. Concluding Discussion

This study provides the first evaluation of a newly-engineered type of commitment device – a temptation bundling device. It shows that in the setting explored, where exercise was bundled with tempting audio novels, this new type of commitment device is valued by a significant portion of the population studied. Further, we find that when temptation bundling is imposed on a population, it can increase gym attendance by 51% at low cost when it is initially instituted, although like most exercise interventions (Royer et al., 2012; Acland & Levy, 2011), the benefits taper off. In addition, we find that individuals are limited in their ability to self-impose temptation bundling tying rules, in line with prior findings suggesting that goal-setting has some shortcomings (Burger, Charness and Lynham, 2011). Our findings highlight that the potential for temptation-bundling to improve outcomes for those facing self-control problems is considerable, especially given that they offer a low-cost, simultaneous solution to two common willpower problems (under-engagement in *shoulds* and over-engagement in *wants*) and harness the potential motivational benefits of complementarities between *wants* and *shoulds*. They further illustrate a context where multi-tasking can be beneficial, standing in contrast to most prior research on multi-tasking (Bowman, Levine, Waite and Gendron, 2010; Fried, 2008; Strayer and Drews, 2007). However, this paper presents just one investigation of temptation bundling and its potential to change behavior, and more research is needed to better understand the effectiveness of this type of commitment device.

Our study suggests that temptations at the gym lose their allure after a period of abstinence, consistent with past research showing that temptation and cravings are reduced by distance (Hughes, Keely, & Naud, 2004; Trope, Liberman, & Wakslak, 2007). To the extent engagement with a *want* novel can induce cravings or become habit-forming, temptation bundling should theoretically be more powerful. The fact that Thanksgiving break eliminated our intervention's effectiveness is consistent with the hypothesis that engagement with the *want* (fiction) bundled with exercise drove our treatment effect, which, once eliminated by a forced period of abstinence, led the temptation bundled with exercise to lose its allure.

An important question our findings raise is how to address the decreasing effectiveness of temptation bundling over time or as a result of natural breaks in gym access that result from holidays and vacations. One solution might be to periodically take steps to renew people's appetites for the temptations bundled with exercise. For example, temptation bundling programs could be designed with rewards for people to re-engage with tempting content every several months or following holidays. Specifically, participants in our study could have been rewarded for returning to the gym after Thanksgiving to listen to the next chapter of their audio novel. Re-engaging with the indulgent novel

would potentially renew its attraction and thus the attraction of visiting the gym. Another alternative to rewarding renewed engagement would be to provide easy access to small segments of content periodically or after holidays outside of the gym (e.g., at home online). Then access to the next segment could again be withheld unless the individual visited the gym. An additional possibility would be to provide access to exercise facilities with similar temptation bundling programs during vacations. Future research exploring whether such strategies could prolong the benefits of temptation bundling programs would be extremely valuable.

Some evidence suggests that our intervention may underestimate the potential benefits of temptation bundling for promoting gym attendance. In an online survey conducted on Amazon mTurk, 54% of paid respondents (N = 89) stated that audio novels were not the most tempting or habit-forming stimuli that could be linked with exercise. A more powerful version of temptation bundling to promote exercise involving a gym could include individual television monitors attached to each machine offering members access to personalized entertainment during their workout. Members could login to their “Gymflix” account on the aerobic machine of their choice, and the television associated with their exercise equipment would grant them access (denied outside of the gym) to indulgent and suspenseful television shows (ideally ending with cliffhangers to draw viewers back, e.g., Lost and 24) or audio novels for those who prefer books. This study suggests that customers of existing entertainment streaming companies (e.g., Hulu Plus, Netflix or Blockbuster Online) might value an account that allowed them to set “gym only” permissions on certain shows, preventing them from accessing these programs anywhere except on a treadmill.

Notably, our study may also underestimate the commercial viability of selling temptation bundling devices. Specifically, our study measures an individual’s willingness to pay for a program that restricts access to her *own* iPod. Temptation bundling programs offered on the open market would likely provide such restrictions *as well as* offering consumers new electronic equipment pre-loaded with tempting content. This added benefit (providing not only restricted access to a product, but also the product itself) would likely increase consumers’ valuations of temptation bundling devices and thus could boost their commercial viability.

One potential limitation of our study is that we could not directly measure how much people exercised; and thus our intervention might be drawing participants to the gym more frequently to listen to audio novels without stimulating exercise during these visits. This explanation, however, is difficult to reconcile with either (a) the fact that the majority of participants were willing to pay for the temptation bundling device program at the end of our study, or (b) the marginal impact of our intermediate intervention on gym attendance (which provided no “illegitimate” incentive for gym visits). Thus, this explanation seems relatively unlikely.

To put our findings in context, it is helpful to compare the increase in exercise induced by temptation bundling with the increases produced by other successful behavioral economics exercise interventions. As described previously, seminal past research has shown that paying individuals to visit the gym eight times (in order to create an exercise habit) increases post-intervention exercise frequencies by between 0.26 visits per week (Acland and Levy, 2011) and 0.59 visits per week (Charness and Gneezy, 2009). Initially, temptation bundling increased gym visits by an estimated 0.48 visits per week, but the estimated weekly increase in gym visits induced by the intervention was subsequently lower: 0.31 for the 7-week, pre-Thanksgiving period, and 0.21 for the entire treatment period. Our measured treatment effects are thus of a similar magnitude to those observed in research designed to increase gym attendance through habit formation interventions.

Our research indicates that temptation bundling devices have potential for solving two problems at once – increasing engagement in desirable behaviors for which people often lack willpower while simultaneously allowing them to enjoy pleasurable activities guilt-free. Temptation bundling takes advantage of complementarities between activities, which is something that previously examined commitment devices cannot capitalize on. Finally, temptation bundling is extremely inexpensive, especially in comparison with alternative means of increasing exercise (Harland et al., 1999; Sevick et al., 2000). Our intervention's only costs were the purchase of \$15 reconditioned iPods and audio novels, which sell for as little as \$1.00 (iTunes, 2012), can be rented for \$7.49 per month (Audible.com, 2012), and can also be borrowed for free from many libraries. This is a remarkably small price to pay for an intervention with the potential to help reduce obesity, guilt, and increase a wide-range of beneficial *should* behaviors.

Tables

Table 1. Pre-treatment characteristics of the study sample reveal no significant differences between groups in pre-treatment exercise frequency, BMI or gender (standard deviations in parentheses)

	All (N = 226)	Control Group (N = 76)	Intermediate Treatment Group (N = 75)	Full Treatment Group (N = 75)
Visits to gym in first week of school	1.5 (1.6)	1.5 (1.8)	1.4 (1.5)	1.4 (1.6)
Self-reported weekly minutes of exercise	100.2 (66.6)	106.1 (69.1)	98.0 (65.5)	96.3 (65.6)
BMI (based on self-reported height and weight)	23.2 (3.9)	23.2 (4.3)	23.3 (4.0)	23.0 (3.4)
Male	34.3%	36.8%	33.3%	32.0%

Table 2. Average weekly gym attendance rates across study participants before and during our intervention period by experimental condition (standard deviations in parentheses) as well as total pre-Thanksgiving and overall gym attendance by condition.

	Control Group (N = 76)	Intermediate Treatment Group (N = 75)	Full Treatment Group (N = 75)
Baseline (First Week of School)	1.51 (1.80)	1.41 (1.50)	1.44 (1.60)
ROLLING ENROLLMENT PERIOD			
Week 1 of Intervention^a	0.75 (1.17)	0.87 (1.38)	1.16 (1.51)
Week 2 of Intervention	0.86 (1.38)	0.95 (1.40)	1.05 (1.45)
Week 3 of Intervention	0.87 (1.26)	1.13 (1.53)	1.31 (1.58)
Week 4 of Intervention	0.97 (1.36)	0.96 (1.38)	1.04 (1.34)
Week 5 of Intervention	0.82 (1.35)	0.88 (1.26)	1.01 (1.21)
Week 6 of Intervention	0.91 (1.38)	0.85 (1.34)	1.04 (1.27)
Week 7 of Intervention	0.93 (1.47)	0.87 (1.39)	1.17 (1.45)
Pre-Thanksgiving Total	6.11	6.51	7.79
THANKSGIVING BREAK			
Week 8 of Intervention	0.99 (1.56)	0.73 (1.17)	0.75 (1.22)
Week 9 of Intervention	0.67 (1.19)	0.53 (1.06)	0.47 (0.88)
Study Total	7.76	7.77	9.00

^a Note that the Week 1 averages exclude participants' gym visit made during our study's intake session, which was not a voluntary visit to the gym.

Table 3. Ordinary Least Squares (OLS) regressions predicting weekly gym attendance, with robust standard errors clustered by participant. Continuous predictors preceded by Z- and included in interaction terms were standardized before inclusion in the regression.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Primary Predictor Variables							
Full treatment	0.31* (0.14)	0.21 (0.13)	0.48** (0.17)	0.47** (0.17)	0.46** (0.16)	0.53** (0.17)	0.50** (0.17)
Intermediate treatment	0.14 (0.14)	0.09 (0.13)	0.27 [†] (0.16)	0.27 (0.17)	0.28 [†] (0.16)	0.27 [†] (0.17)	0.28 [†] (0.17)
(Weeks since intervention) x (Full treatment)	-	-	-0.07** (0.02)	-0.06** (0.02)	-0.07** (0.02)	-0.07** (0.02)	-0.07** (0.02)
(Weeks since intervention) x (Intermediate treatment)	-	-	-0.05 [†] (0.03)	-0.04 [†] (0.02)	-0.05 [†] (0.03)	-0.04 [†] (0.02)	-0.04 [†] (0.02)
Moderators							
(Z-Availability) x (Full treatment)	-	-	-	-	-0.21* (0.08)	-	-0.22** (0.08)
(Z-Availability) x (Intermediate treatment)	-	-	-	-	-0.07 (0.09)	-	-0.11 (0.09)
Z-Availability	-	-	-	-	-0.02 (0.05)	-	0.01 (0.05)
(Z-Intake workout enjoyment) x (Full treatment)	-	-	-	-	-	0.21* (0.10)	0.21* (0.10)
(Z-Intake workout enjoyment) x (Intermediate treatment)	-	-	-	-	-	0.06 (0.13)	0.08 (0.13)
Z-Intake workout enjoyment	-	-	-	-	-	0.07 (0.06)	0.07 (0.06)
Control Variables							
Visits to gym in first week of school (pre-treatment)	0.15 (0.11)	0.11 (0.10)	0.11 (0.10)	0.11 (0.10)	0.10 (0.10)	0.12 (0.10)	0.11 (0.10)
Visits to gym in first week of school (pre-treatment) ²	0.05* (0.03)	0.05* (0.02)	0.05* (0.02)	0.06* (0.02)	0.06* (0.02)	0.05* (0.02)	0.06* (0.02)
Self-reported weekly minutes of exercise x 10 ⁻³ (pre-treatment)	3.69 (2.65)	3.28 (2.45)	3.28 (2.45)	3.43 (2.52)	4.07 [†] (2.40)	2.83 (2.51)	3.66 (2.45)
Self-reported weekly minutes of exercise x 10 ⁻³ (pre-treatment) ²	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Weeks Since Intervention	-0.11*** (0.01)	-0.10*** (0.01)	-0.06*** (0.02)	-0.07*** (0.02)	-0.06*** (0.02)	-0.07*** (0.02)	-0.07*** (0.02)
Regression Statistics							
Number of Observations	1,582	2,034	2,034	1,935	2,034	1,908	1,908
Number of Clusters	226	226	226	215	226	212	212
Weeks Included	1-7 (pre-Thanksgiving)	All (Entire Period)	All (Entire Period)	All (Entire Period)	All (Entire Period)	All (Entire Period)	All (Entire Period)
R-Squared	0.27	0.24	0.28	0.28	0.29	0.30	0.31

*** p ≤ 0.001; ** p ≤ 0.01; * p ≤ 0.05; [†] p ≤ 0.1

Figures

Figure 1. The theoretical inverted shapes of the utility streams obtained from engaging in (A) *wants* and (B) *shoulds*, which are strategically combined by temptation bundling to help those who discount the future heavily both engage in *shoulds* and limit their engagement in *wants* to moments when the downstream negative consequences are minimized.

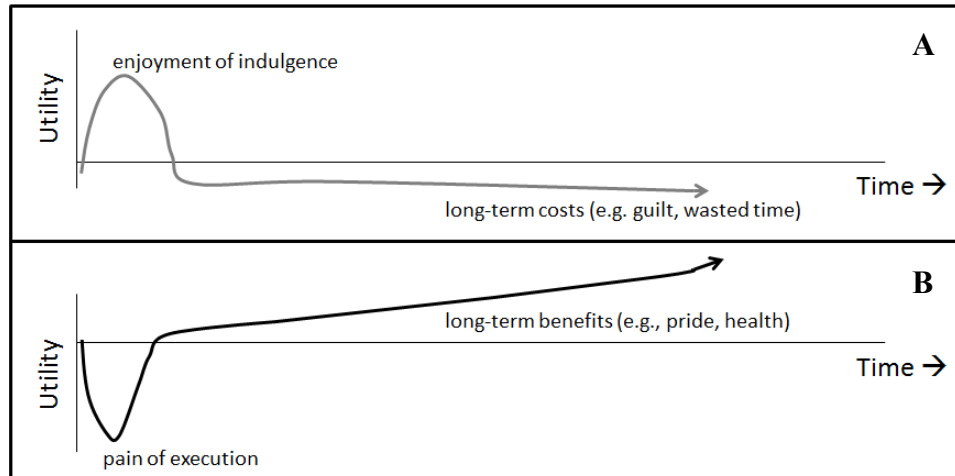


Figure 2. Flow of study participants

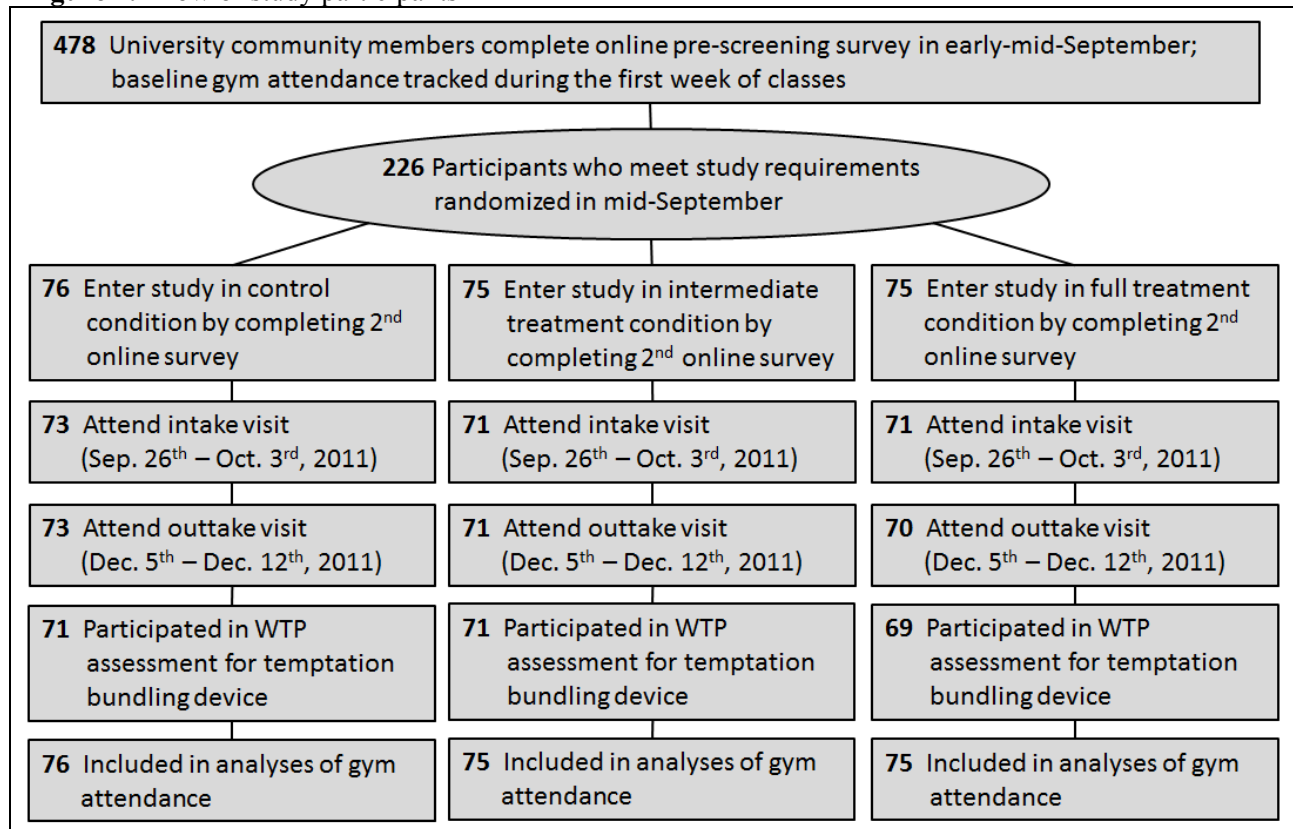
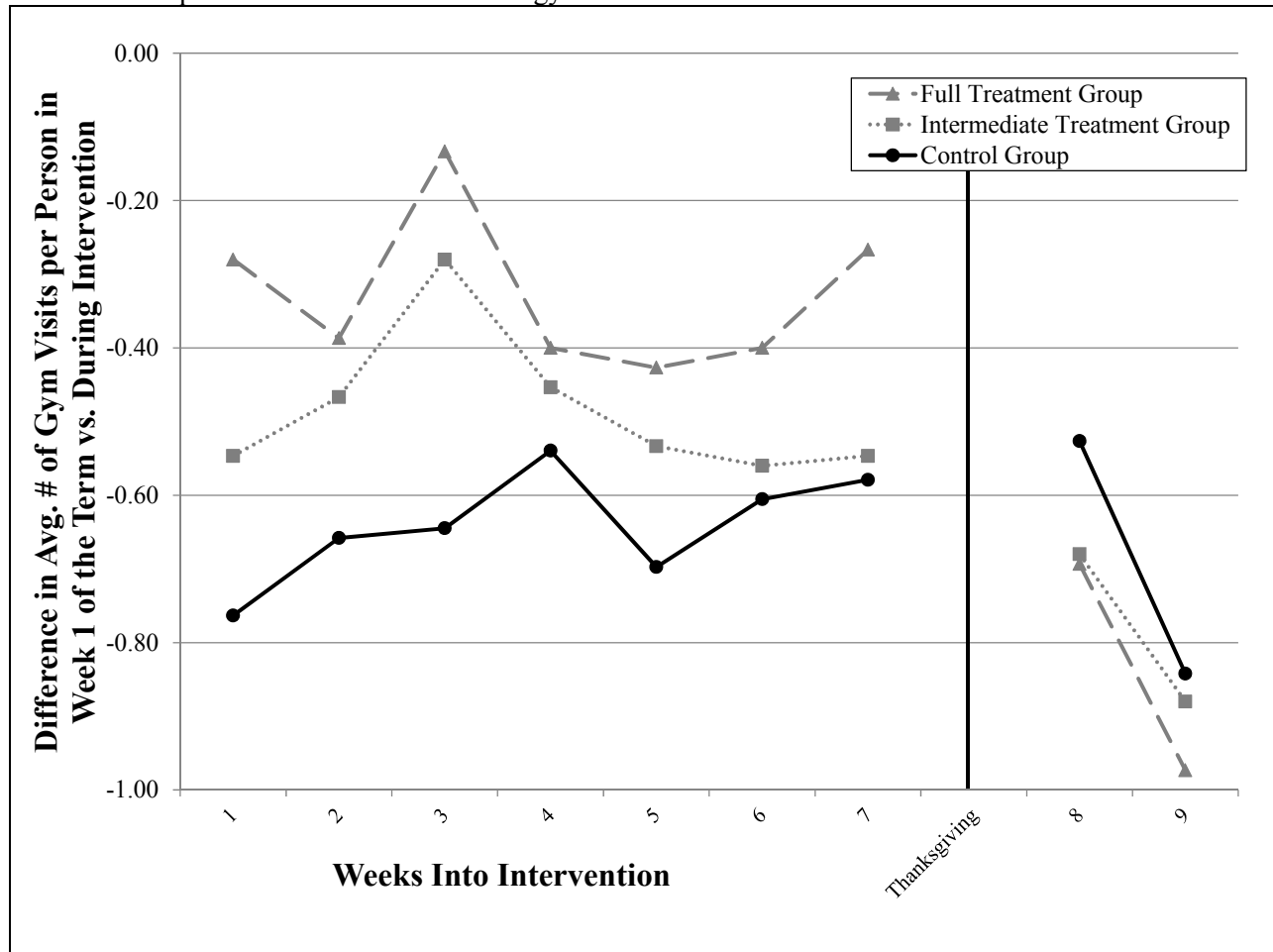
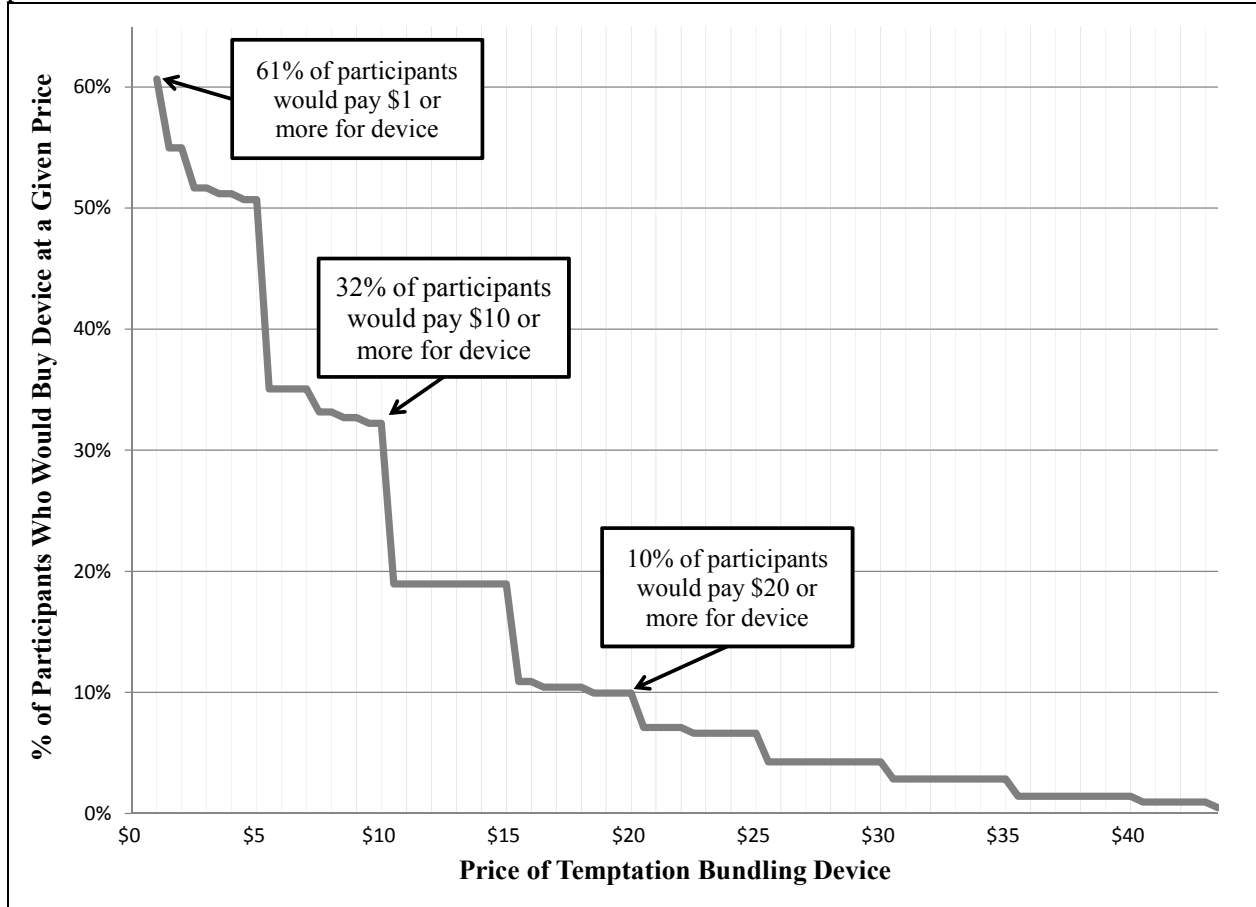


Figure 3. Gym attendance for all groups declined precipitously from baseline levels measured in the first week of the University semester, prior to the start of our intervention.¹² Participants in the treatment conditions experienced a smaller decline in gym attendance than those in the control condition.



¹² The gym provided us with a count of the total number of student visits each day throughout the semester. During our study's baseline, pre-treatment week (at the beginning of the semester), there were 12,135 visits to the gym, but during the first week of our study, there were only 8,538 visits to the gym. In short, gym attendance declined as the semester progressed and students presumably became busier.

Figure 4. Percent of study participants willing to pay for a temptation bundling device at a given price.



Appendix

Table A1. No biometric measures collected differ significantly (at alpha level of 0.10 or lower) across groups.

	Weight (pounds)	Body Mass Index	Body Fat (percentage)	Pulse Rate (per minute)	Waist Size (inches)
MEASURES AT OUTSET OF STUDY (T1)					
Control Group	144.60 (25.24) N = 72	22.94 (2.78) N = 72	28.08 (8.26) N = 66 ^a	80.88 (13.09) N = 72	31.36 (3.19) N = 71
Intermediate Treatment Group	149.15 (28.70) N = 71	23.41 (4.29) N = 71	27.89 (9.79) N = 70 ^a	80.63 ^c (14.84) N = 70 ^a	31.90 (3.91) N = 70 ^a
Full Treatment Group	143.63 (25.79) N = 71	23.14 (3.82) N = 71	28.53 (8.21) N = 71	77.30 (14.42) N = 71	31.52 (3.51) N = 71
MEASURES AT CONCLUSION OF STUDY (T2)					
Control Group	145.45 (25.53) N = 71	23.15 (2.84) N = 70 ^a	28.09 (8.33) N = 70 ^a	83.13 (14.94) N = 70 ^a	31.04 (2.92) N = 71
Intermediate Treatment Group	150.17 (29.25) N = 71	23.56 (4.38) N = 71	28.46 (9.69) N = 71	84.87 (16.25) N = 71	31.04 (3.70) N = 71
Full Treatment Group	145.13 (26.72) N = 70	23.33 (3.96) N = 70	29.09 (8.38) N = 70	83.99 (15.58) N = 70	31.34 (3.45) N = 70
Δ OVER 10 WEEK STUDY (T2 - T1)*					
Control Group	0.44 (3.78) N = 71	0.08 (0.60) N = 70	-0.08 (2.63) N = 65	2.24 (17.83) N = 70	-0.37 (1.66) N = 70
Intermediate Treatment Group	1.02 (4.49) N = 71	0.16 (0.69) N = 71	0.42 (2.35) N = 70	3.83 (18.99) N = 70	-0.75 (1.39) N = 70
Full Treatment Group	0.95 (3.94) N = 70	0.14 (0.63) N = 70	0.58 (2.35) N = 70	6.89 (18.79) N = 70	-0.27 (1.54) N = 70

^a Due to errors with biometric measurement equipment, some participants' measures were not successfully obtained.

Table A2. Raw and regression-estimated willingness to pay for a temptation bundling device by experimental condition. Average WTP is presented both including and excluding the one extreme outlier in our data set (in the control group) who stated a WTP of \$100 for a temptation bundling device.

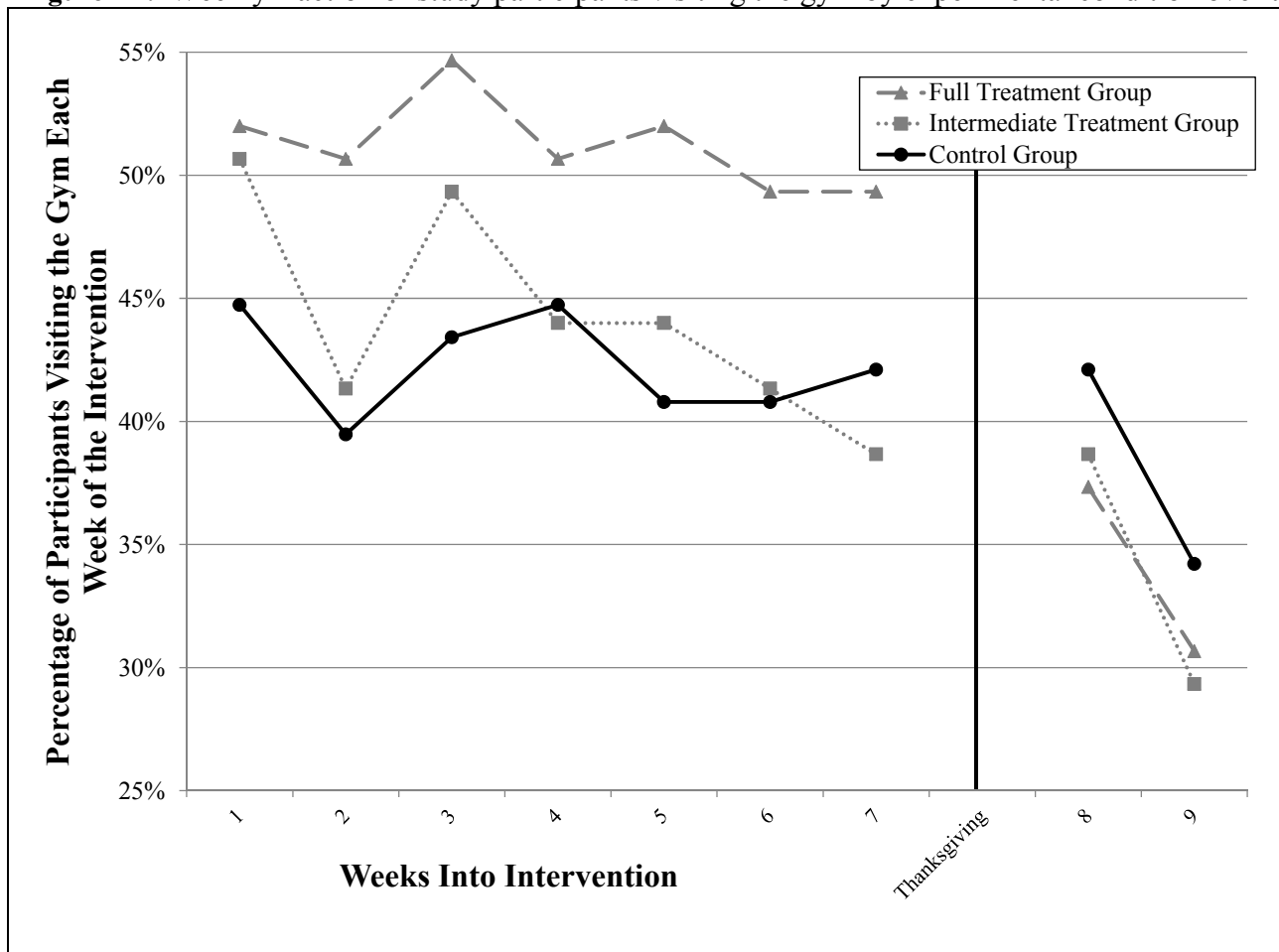
	Control Group (N = 71)	Intermediate Treatment Group (N = 69)	Full Treatment Group (N = 71)
	Average (Std. Dev.)		
WTP More than \$0	0.69 (0.47)	0.57 (0.50)	0.56 (0.50)
Average WTP	\$9.87 (14.55)	\$4.77 (7.24)	\$6.01 (8.31)
Average WTP (dropping \$100 outlier)	\$8.59 (9.76)	\$4.77 (7.24)	\$6.01 (8.31)
	Regression Estimated (Std. Err.)		
WTP More than \$0	omitted	-0.13 (0.08)	-0.12 (0.08)
Average WTP	omitted	-3.86* (1.77)	-5.11** (1.79)
Average WTP (dropping \$100 outlier)	omitted	-2.57 [†] (1.43)	-3.82** (1.44)

† p < 0.10; * p < 0.05; ** p < 0.01

Figure A1. Photographs of iPod locker (full treatment condition) in entryway of University Gym



Figure A2. Weekly fraction of study participants visiting the gym by experimental condition over the course of the intervention.



Complete List of 82 Pre-tested Tempting Audio Novels Available to Study Participants

No.	Audio Novel Title	No.	Audio Novel Title
1	Harry Potter: Book 1 - Harry Potter and the Philosopher's Stone (by J.K. Rowling)	42	The Brethren (by John Grisham)
2	Harry Potter: Book 2 - Harry Potter and the Chamber of Secrets (by J.K. Rowling)	43	The Broker (by John Grisham)
3	Harry Potter: Book 3 - Harry Potter and the Prisoner of Azkaban (by J.K. Rowling)	44	The Runaway Jury (by John Grisham)
4	Harry Potter: Book 4 - Harry Potter and the Goblet of Fire (by J.K. Rowling)	45	The Testament (by John Grisham)
5	Harry Potter: Book 5 - Harry Potter and the Order of the Phoenix (by J.K. Rowling)	46	The Chamber (by John Grisham)
6	Harry Potter: Book 6 - Harry Potter and the Half-Blood Prince (by J.K. Rowling)	47	The 6th Target (by James Patterson)
7	Harry Potter: Book 7 - Harry Potter and the Deathly Hallows (by J.K. Rowling)	48	London Bridges (by James Patterson)
8	The Hunger Games: Book 1 - The Hunger Games (by Suzanne Collins)	49	The Big Bad Wolf (by James Patterson)
9	The Hunger Games: Book 2 - Catching Fire (by Suzanne Collins)	50	Cat and Mouse (by James Patterson)
10	The Hunger Games: Book 3 - Mockingjay (by Suzanne Collins)	51	Four Blind Mice (by James Patterson)
11	The Lord of the Rings: Book 1 - The Fellowship of the Ring (by J.R.R. Tolkien)	52	I, Alex Cross (by James Patterson)
12	The Lord of the Rings: Book 2 - The Two Towers (by J.R.R. Tolkien)	53	Cross Fire (by James Patterson)
13	The Lord of the Rings: Book 3 - The Return of the King (by J.R.R. Tolkien)	54	The Bourne Trilogy: Book 1 - The Bourne Identity (by Robert Ludlum)
14	His Dark Materials: Book 1 - The Golden Compass (by Phillip Pullman)	55	The Bourne Trilogy: Book 2 - The Bourne Supremacy (by Robert Ludlum)
15	His Dark Materials: Book 2 - The Subtle Knife (by Phillip Pullman)	56	The Bourne Trilogy: Book 3 - The Bourne Ultimatum (by Robert Ludlum)
16	His Dark Materials: Book 3 - The Amber Spyglass (by Phillip Pullman)	57	The Sum of All Fears (by Tom Clancy)
17	The Lost World (by Michael Crichton)	58	Executive Orders (by Tom Clancy)
18	Dune: Book 1 - Dune (by Frank Herbert)	59	Rainbow Six (by Tom Clancy)
19	Dune: Book 2 - Dune Messiah (by Frank Herbert)	60	The Da Vinci Code Trilogy: Book 1 - Angels and Demons (Dan Brown)
20	Dune: Book 3 - Children of Dune (by Frank Herbert)	61	The Da Vinci Code Trilogy: Book 2 - The Da Vinci Code (by Dan Brown)
21	Dune: Book 4 - God Emperor of Dune (by Frank Herbert)	62	The Da Vinci Code Trilogy: Book 3 - The Lost Symbol (by Dan Brown)
22	Dune: Book 5 - Heretics of Dune (by Frank Herbert)	63	The Stephanie Plum Novels: Book 1 - One for the Money (by Janet Evanovich)
23	Dune: Book 6 - Chapterhouse: Dune (by Frank Herbert)	64	The Stephanie Plum Novels: Book 2 - Two for the Dough (by Janet Evanovich)
24	The Twilight Saga: Book 1 - Twilight (by Stephenie Meyer)	65	The Stephanie Plum Novels: Book 3 - Three to Get Deadly (by Janet Evanovich)
25	The Twilight Saga: Book 2 - New Moon (by Stephenie Meyer)	66	The Stephanie Plum Novels: Book 4 - Four to Score (by Janet Evanovich)
26	The Twilight Saga: Book 3 - Eclipse (by Stephenie Meyer)	67	The Stephanie Plum Novels: Book 5 - High Five (by Janet Evanovich)
27	The Twilight Saga: Book 4 - Breaking Dawn (by Stephenie Meyer)	68	The Stephanie Plum Novels: Book 6 - Hot Six (by Janet Evanovich)
28	The Sookie Stackhouse Novels: Book 1 - Dead Until Dark (by Charlaine Harris)	69	The Stephanie Plum Novels: Book 7 - Seven Up (by Janet Evanovich)
29	The Sookie Stackhouse Novels: Book 2 - Living Dead in Dallas (by Charlaine Harris)	70	The Stephanie Plum Novels: Book 8 - Hard Eight (by Janet Evanovich)
30	The Sookie Stackhouse Novels: Book 3 - Club Dead (by Charlaine Harris)	71	The Stephanie Plum Novels: Book 9 - To the Nines (by Janet Evanovich)
31	The Sookie Stackhouse Novels: Book 4 - Dead to the World (by Charlaine Harris)	72	The Stephanie Plum Novels: Book 10 - Ten Big Ones (by Janet Evanovich)
32	The Sookie Stackhouse Novels: Book 5 - Dead as a Doornail (by Charlaine Harris)	73	The Stephanie Plum Novels: Book 11 - Eleven on Top (by Janet Evanovich)
33	The Sookie Stackhouse Novels: Book 6 - Definitely Dead (by Charlaine Harris)	74	The Stephanie Plum Novels: Book 12 - Twelve Sharp (by Janet Evanovich)
34	The Sookie Stackhouse Novels: Book 7 - All Together Dead (by Charlaine Harris)	75	The Stephanie Plum Novels: Book 13 - Lean Mean Thirteen (by Janet Evanovich)
35	The Sookie Stackhouse Novels: Book 8 - From Dead to Worse (by Charlaine Harris)	76	The Stephanie Plum Novels: Book 14 - Fearless Fourteen (by Janet Evanovich)
36	The Sookie Stackhouse Novels: Book 9 - Dead and Gone (by Charlaine Harris)	77	The Stephanie Plum Novels: Book 15 - Finger Lickin' Fifteen (by Janet Evanovich)
37	The Sookie Stackhouse Novels: Book 10 - Dead in the Family (by Charlaine Harris)	78	The Stephanie Plum Novels: Book 16 - Sizzling Sixteen (by Janet Evanovich)
38	The Sookie Stackhouse Novels: Book 11 - Dead Reckoning (by Charlaine Harris)	79	The Stephanie Plum Novels: Book 17 - Smokin' Seventeen (by Janet Evanovich)
39	The Host (by Stephenie Meyer)	80	Bag of Bones (by Stephen King)
40	The Confession (by John Grisham)	81	The Help (by Kathryn Stockett)
41	The King of Torts (by John Grisham)	82	The Devil Wears Prada (by Lauren Weisberger)

References

- Abeler, J., & Marklein, F. (2008). *Fungibility, Labels, and Consumption*. IZA Discussion Paper, (3500).
- Acland, D., & Levy, M. (2011). Habit Formation, Naivete, and Projection Bias in Gym Attendance. *Working Paper*.
- Andersen, R.E. (2010). Exercise, an Active Lifestyle, and Obesity. *The Physician and Sportsmedicine*, 38(4), 41-50. doi: 10.3810/psm.1999.10.1.1021
- Ariely, D., & Wertenbroch, K. (2002). Procrastination, Deadlines, and Performance: Self-Control by Precommitment. *Psychol Sci*, 13(3), 219-224. doi: 10.1111/1467-9280.00441
- Ashraf, N., Karlan, D., & Yin, W. (2006). Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines. *The Quarterly Journal of Economics*, 121(2), 635-672. doi: 10.1162/qjec.2006.121.2.635
- Baumeister, R.F., Bratslavsky, E., Muraven, M., & Tice, D.M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5), 1252-1265. doi: 10.1037/0022-3514.74.5.1252
- Becker, G.M., Degroot, M.H., & Marschak, J. (1964). Measuring Utility by a Single-Response Sequential Method. *Behavioral Science*, 9(3), 1099-1743.
- Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., & Sakong, Jung. (2011). *Self control and liquidity: How to design a commitment contract*. RAND Working Paper Series.
- Bowman, L. L., Levine, L. E., Waite, B. M., & Gendron, M. (2010). Can students really multitask? An experimental study of instant messaging while reading. *Computers & Education*, 54(4), 927-931.
- Burger, N., Charness, G., & Lynham, J. (2011). Field and online experiments on self-control. *Journal of Economic Behavior & Organization*, 77(3), 393-404.
- Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor Supply of New York City Cabdrivers: One Day at a Time. *Q J Econ*, 112(2), 407-441. doi: 10.1162/003355397555244
- Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica*, 77(3), 909-931.
- Cheema, A., & Soman, D. (2008). The Effect of Partitions on Controlling Consumption. *Journal of Marketing Research*, 45(6), 665-675. doi: DOI 10.1509/jmkr.45.6.665
- Della Vigna, S., & Malmendier, U. (2006). Paying Not to Go to the Gym. *American Economic Review*, 96(3), 694-719. doi: 10.1257/aer.96.3.694
- Finkelstein, E. A., daCosta DiBonaventura, M., Burgess, S. M., & Hale, B. C. (2010). The costs of obesity in the workplace. *Journal of Occupational and Environmental Medicine*, 52(10), 971-976.
- Finkelstein, E., Fiebelkorn, I. C., & Wang, G. (2005). The costs of obesity among full-time employees. *American Journal of Health Promotion*, 20(1), 45-51.
- Flegal, K. M., Carroll, M. D., Ogden, C. L., & Curtin, L. R. (2010). Prevalence and Trends in Obesity Among US Adults, 1999-2008. *Jama-Journal of the American Medical Association*, 303(3), 235-241. doi: Doi 10.1001/Jama.2009.2014
- Flegal, K. M., Graubard, B. I., Williamson, D. F., & Gail, M. H. (2007). Cause-specific excess deaths associated with underweight, overweight, and obesity. *JAMA*, 298(17), 2028-2037. doi: 10.1001/jama.298.17.2028
- Fried, C.B. (2008). In-class laptop use and its effects on student learning. *Computers & Education*, 50(3), 906-914.

- Giné, X., Karlan, D., & Zinman, J. (2010). Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation. *American Economic Journal: Applied Economics*, 2(4), 213-235. doi: 10.1257/app.2.4.213
- Halpern, S. D., Asch, D. A., & Volpp, K. G. (2012). Commitment contracts as a way to health. *Bmj*, 344(jan30 1), e522-e522. doi: 10.1136/bmj.e522
- Harland, J., White, M., Drinkwater, C., Chinn, D., Farr, L., & Howel, D. (1999). The Newcastle exercise project: a randomised controlled trial of methods, to promote physical activity in primary care. *British Medical Journal*, 319(7213), 828-832B.
- Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology*, 38(1), 79-109.
- Houser, D., Schunk, D., Winter, J., & Xiao, E.. (2010). *Temptation and commitment in the laboratory*. Working Paper Series. Institute for Empirical Research in Economics, University of Zurich.
- Hughes, J. R., Keely, J., & Naud, S. (2004). Shape of the relapse curve and long-term abstinence among untreated smokers. *Addiction*, 99(1), 29-38.
- John, L. K., Loewenstein, G., & Volpp, K. G. (2012). Empirical observations on longer-term use of incentives for weight loss. *Prev Med*, 55 Suppl, S68-74. doi: 10.1016/j.ypmed.2012.01.022
- Kaur, S., Kremer, M., & Mullainathan, S. (2010). Self-Control and the Development of Work Arrangements. *American Economic Review*, 100(2), 624-628. doi: 10.1257/aer.100.2.624
- List, J.A., Sadoff, S. & Wagner, M. (2010). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics*. 14, 439-457.
- Milkman, K.L., & Beshears, J. (2009). Mental accounting and small windfalls: Evidence from an online grocer. *Journal of Economic Behavior & Organization*, 71(2), 384-394. doi: 10.1016/j.jebo.2009.04.007
- Milkman, K.L., Rogers, T., & Bazerman, M.H. (2008). Harnessing Our Inner Angels and Demons: What We Have Learned About Want/Should Conflicts and How That Knowledge Can Help Us Reduce Short-Sighted Decision Making. *Perspectives on Psychological Science*, 3(4), 324-338. doi: 10.1111/j.1745-6924.2008.00083.x
- Milkman, K.L., Rogers, T., & Bazerman, M.H. (2009). I'll have the ice cream soon and the vegetables later: A study of online grocery purchases and order lead time. *Marketing Letters*, 21(1), 17-35. doi: 10.1007/s11002-009-9087-0
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, 74(3), 774-789.
- Muraven, M., & Baumeister, R.F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126(2), 247-259. doi: 10.1037//0033-2909.126.2.247
- O'Donoghue, T., & Rabin, M. (1999). Doing it Now or Later. *The American Economic Review*, 89(1), 103-124.
- Read, D., & van Leeuwen, B. (1998). Predicting hunger: The effects of appetite and delay on choice. *Organizational Behavior and Human Decision Processes*, 76(2), 189-205.
- Royer, H., Stehr, M., & Sydnor, J. (2012). Incentives and Commitments for Exercise: Evidence from a Field Experiment with Workers at a Fortune 500 Company. *Working Paper*.

- Sevick, M. A., Dunn, A. L., Morrow, M. S., Marcus, B. H., Chen, G. J., & Blair, S. N. (2000). Cost-effectiveness of lifestyle and structured exercise interventions in sedentary adults - Results of project ACTIVE. *American Journal of Preventive Medicine*, 19(1), 1-8. doi: Doi 10.1016/S0749-3797(00)00154-9
- Shefrin, H.M, & Thaler, R. H. (1988). The behavioral life-cycle hypothesis. *Economic Inquiry*, 26(4), 609-643.
- Solomon, R., & Corbit, J. (1974). An Opponent-Process Theory of Motivation: Temporal Dynamics of Affect. *Psychological Review*, 81(2), 119-145.
- Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science*, 16(3), 128-131.
- Thaler, R. (1985). Mental Accounting and Consumer Choice. *Marketing Science*, 4(3), 199-214. doi: 10.1287/mksc.4.3.199
- Thaler, R. H. (1990). Anomalies: Saving, fungibility, and mental accounts. *The Journal of Economic Perspectives*, 4(1), 193-205.
- Thaler, R. H. (1999). Mental accounting matters. *Journal of Behavioral Decision Making*, 12(3), 183-206.
- Thaler, R. H., & Shefrin, H. M. (1981). An Economic-Theory of Self-Control. *Journal of Political Economy*, 89(2), 392-406.
- Trope, Y., Liberman, N., & Wakslak, C. (2007). Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *Journal of Consumer Psychology*, 17(2), 83-95.
- U.S. Physical Activity Statistics. (2007, May 21, 2007). from <http://apps.nccd.cdc.gov/PASurveillance/DemoComparev.asp>
- Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive-based approaches for weight loss: a randomized trial. *JAMA*, 300(22), 2631-2637. doi: 10.1001/jama.2008.804