Journal of Economic Behavior & Organization 77 (2011) 393–404

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

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

Field and online experiments on self-control

Nicholas Burger[a], Gary Charness[b,*], John Lynham[c]

[a] RAND Corporation, United States
[b] Department of Economics, University of California Santa Barbara, 2127 North Hall, Santa Barbara, CA 93106-9210, United States
[c] Department of Economics, University of Hawai'i at Manoa, United States

ARTICLE INFO

Article history:
Received 19 May 2010
Received in revised form
24 November 2010
Accepted 29 November 2010
Available online 8 December 2010

JEL classification:
A13
B49
C91
C93
D00

Keywords:
Experiment
Behavioral interventions
Procrastination
Willpower

ABSTRACT

Self-control problems have recently received considerable attention from economic theorists. We conducted two studies involving behavioral interventions expected to affect performance, providing some of the first experimental data in this area. In the first we investigate whether evenly spaced interim deadlines lead to higher completion rates for a lengthy task, where procrastination could be a factor. We found that these interim deadlines in fact led to lower completion rates in our set-up; we also found no evidence in the aggregate data (although there was considerable heterogeneity) of an increasing profile of the number of hours spent at the task over time, suggesting that the degree of present-bias in the participant population is not that high. In the second, we examine how willpower depletion affects behavior. Our second study isolates the effect of direct willpower depletion on the first day of a two-day period, thus exploring procrastination over a shorter time horizon. We found that depleting willpower did indeed lead to less work (and poorer quality) on the first day, but that this intervention surprisingly led to a *higher* completion rate overall. Our results help to inform ongoing efforts to understand and model self-control, willpower, and commitment mechanisms.

1. Introduction

The issues of self-control, procrastination, commitment, and willpower have important economic consequences in realms such as productivity, savings behavior, and well-being; these issues have also been the focus of a number of recent theoretical models. People experience self-control problems when their preferences are not consistent across time. One form of self-control problem concerns persistent bad habits or addictions, such as overeating or cigarette smoking. An individual knows that he or she will later regret a current self-indulgent choice, but nevertheless engages in the activity. The other side of the coin is a situation in which an individual is faced with an activity that will lead to future benefits but is unappealing at the moment. This often leads to procrastination, common in everyday life.[1] People vow to stop smoking, stop eating ice cream, or start exercising... tomorrow. An important issue is the underlying nature of the self-control problem and how one can overcome it; closely related to this issue is the notion of willpower.

* Corresponding author. Tel.: +1 805 893 2412; fax: +1 805 893 8830.
E-mail addresses: nburger@rand.org (N. Burger), charness@econ.ucsb.edu (G. Charness), lynham@hawaii.edu (J. Lynham).

[1] The Merriam-Webster dictionary defines procrastination as putting off intentionally the doing of something that should be done. Procrastination has been found to be quite pervasive among students: Ellis and Knaus (1977) find that 95% of college students procrastinate, while Solomon and Rothblum (1984) find that 46% nearly always or always procrastinate in writing a term paper.

0167-2681/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jebo.2010.11.010

In this paper, we examine aspects of commitment and willpower in two experimental studies featuring behavioral interventions. As the duration of the tasks made it rather infeasible to use simple laboratory experiments, we instead conducted field and online experiments. In Study 1, we investigate how people allocate their time over a task of significant duration (studying 75 h over a 5-week period) and whether an externally imposed commitment in the form of binding sub-goals (at least 12 h per week, on a cumulative basis) would help them to complete the overall task. The results on time allocation in Study 1 are highly cyclical, with clear troughs and peaks. While this cyclical behavior can be explained by other factors, it is consistent with self-control theories that model willpower as a depletable but renewable resource (Baumeister and Vohs, 2003; Ozdenoren et al., forthcoming).

Motivated by the literature on willpower as a cognitive resource and the results of Study 1, we designed Study 2 to consider the effects of willpower depletion on performance in a shorter duration task. In some sense, this study examines shorter term procrastination rather than longer term procrastination. In addition, Study 2 offers another perspective on the weekly cycles of effort allocation we found in Study 1. We offered participants a financial reward for completing a task within two days. On the first day, subjects were assigned either a willpower-depleting or a relatively willpower-neutral Stroop test. Psychologists have used the Stroop test (Stroop, 1935) to deplete willpower in experimental settings (see Gailliot et al., 2007 for a recent study and MacLeod, 1991 for a history). The Stroop test requires inhibiting the urge to produce an automatic response; this requires an "internally generated act of control" (Logan, 1994, p. 190) and has been linked to executive function in the frontal lobes (Perret, 1974; Kiefer et al., 1998). We expected willpower depletion using the Stroop test to negatively impact performance.

Extant empirical research on procrastination and self-control is significant (e.g., DellaVigna and Malmendier, 2004; Ashraf et al., 2006 in the economics literature), but experimental economics has contributed less to this body of work. Given that experiments offer a way to test interventions in controlled settings, they may offer the best approach to exploring these behavioral phenomena in detail. However, there is little or no experimental evidence in the economics literature on patterns of effort provision for tasks of significant duration.

One of the main motivations behind our paper is to begin to rectify this lack of evidence, by experimentally altering self-control; in one study we attempt to improve self-control, while in the other study we try to weaken self-control. Part of our research interest is to explore the effect of interim deadlines on the performance of a task for which there are no draconian consequences for non-completion and for which there are clear issues of self-control. Another interest is to test whether willpower depletion adversely affects performance. Our results represent some of the first detailed and daily data regarding performance and self-control under controlled conditions with financial incentives and tasks of significant duration and elapsed time.[2] We want to identify patterns in behavior that would inform theorists so that more relevant, descriptive models can be developed; at the same time, we also hope to aid other researchers in designing mechanisms that are effective in overcoming obstacles to performance and to explore how experimental methods might shed light on a set of robust but competing theories.[3]

There have been only a few studies that consider how one might *overcome* self-control problems. Aside from exerting willpower in the face of a disagreeable task, one approach is commitment, or binding one's own behavior with costly restrictions. Wertenbroch (1998) presents examples of binding behavior, including tactics such as putting savings into a Christmas-club account that does not pay interest or buying only small packages of goods such as cigarettes or ice cream. In a similar vein, Burger and Lynham (2010) examine weight-loss bets in England, where one could bet on achieving a weight goal by a deadline; however, the vast majority of bettors lost their bets, suggesting that this self-imposed deadline was ineffective.

Ariely and Wertenbroch (2002) present two studies in which three tasks had to be completed over a period of time, finding that externally imposed and evenly spaced deadlines during this period are more effective than self-imposed (and binding) deadlines, which in turn are more effective than having no interim deadlines. The first study involved turning in term papers for grades in a class over a 14-week period (with a 1% grade penalty for each day of delay). The second study involved proofreading papers over a 3-week period; participants were paid for each error found but fined for each day of delay. The results in Ariely and Wertenbroch (2002) suggest that exogenous deadlines may well be quite effective.

However, the results of our behavioral intervention in Study 1 differ considerably from Ariely and Wertenbroch, suggesting that some deadlines may in fact reduce flexibility enough to overcome any benefits they provide in terms of improving performance. There are some key differences in our experimental designs, such as the likelihood of completion without any deadlines and the penalty for missing a deadline. For those people who are enrolled in a class, not turning in a term paper will result in a very bad outcome; thus, everyone completes the overall task, with performance (but not completion) affected by the characteristics of the various deadlines. In their proofreading study, subjects who accept a job to be proofreaders can

[2] Two recent studies consider issues of temptation over a relatively short period of time. Houser et al. (2007) look at temptation at the check-out counter in the supermarket, while Houser et al. (2009) conduct a laboratory experiment involving temptation and commitment.

[3] There have been at least a handful of behavioral interventions designed to overcome bad habits or to form new good ones. Angrist and Lavy (2009) offer substantial cash incentives in Israel for matriculation; while this is ineffective when individual students are selected for the treatment, matriculation rates do increase when this program is school-wide. Charness and Gneezy (2009) pay students at two American universities to attend a gym during a period of time, finding that attendance rates increase substantially not only during this period, but also after the intervention ends. Angrist et al. (2009) offer merit scholarships to undergraduates at a Canadian university, with some success in improving performance, but mixed results overall.

turn in the assigned project at any point, regardless of whether the subject has finished the task. Payment only depends on the amount and quality of the work done, with a small penalty per day for being late.

Our results in both studies surprised us. In line with Ariely and Wertenbroch (2002), we had expected that the interim deadlines in Study 1 would improve completion rates by breaking the task into more manageable segments.[4] However, completion rates were actually 50% *lower* with interim deadlines, implying that our form of externally imposed commitment was ineffective.[5] While Ariely and Wertenbroch (2002) show that deadlines *can* be beneficial, our findings suggest that this is not always the case.[6] In Study 2, among those people who completed the task, those who were assigned the willpower-depleting test exerted significantly less effort on the first day, as expected. However, those subjects who were assigned the willpower-depleting test were actually *more* likely to complete the overall task within the allotted two days than the people in the willpower-neutral treatment. Thus, while willpower does appear to be depleted by the intervention on the first day, it seems to come back even stronger on the second day. This is consistent with the notion that successfully completing a willpower-depleting Stroop test serves as a signal about one's willpower.[7] We discuss this finding in more detail later. In both studies, we see considerable heterogeneity in behavior across the population.

The remainder of this paper is organized as follows: we present some related theoretical literature in Section 2, provide details of our experimental design in Section 3 and describe the results in Section 4. Section 5 concludes.

2. Related theoretical literature

There are a number of theoretical approaches that address procrastination, time-inconsistency, and willpower; these include models of present-biased quasi-hyperbolic preferences (Laibson, 1997; O'Donoghue and Rabin, 1999), temptation and commitment (Gul and Pesendorfer, 2001; Fudenberg and Levine, 2006), self-signaling of one's strength of will (Bénabou and Tirole, 2004), and willpower (Ozdenoren et al., forthcoming). We briefly summarize these models and explore their broad predictions in our design. Although our research objective was to test if (i) interim deadlines help task completion and (ii) willpower depletion hurts task completion – not to prove/disprove various theoretical explanations for time-inconsistency – our results shed empirical light on the different modeling approaches.

One notion is that people have present-biased preferences. A person puts more weight on the present than on the future, leading to dynamic inconsistency.[8] Strotz (1955), Ainslie (1992), Laibson (1997), and O'Donoghue and Rabin (1999) discuss present-biased (quasi-hyperbolic) preferences as an explanation for persistent bad habits and addictions.[9] In Study 1, these models would broadly predict that subjects with present-biased preferences should benefit from the imposition of commitment mechanisms. In other words, the evenly spaced but relatively flexible weekly deadlines should result in a higher completion rate compared to the treatment with no weekly deadlines. There are no particular predictions in Study 2. However, if impatience in the quasi-hyperbolic-discounting models is linked to willpower or cognitive resources then reducing willpower should induce procrastination.

Another notion relates to temptation and commitment. The Gul and Pesendorfer (2001) model suggests that commitments may be helpful for overcoming the temptation to avoid an unattractive task in the future. While the predictions depend on parameter values, if we view weekly restrictions as reducing the choice set of the agent, so that there is effectively less temptation, one might expect a positive impact on performance. The predictions in Study 2 are less clear, but if an agent who completes the discordant Stroop views this accomplishment as a commitment to complete the task eventually, the model would predict less temptation on the second day and so a higher completion rate overall.

Fudenberg and Levine (2006) propose a dual-self model in which a long-run self effectively ties the hands of a series of myopic short-run selves, serving as a form of commitment.[10] In the first phase, the sophisticated long-run self chooses a commitment action that affects any short-run self's utility; in the second phase, the short-run self then chooses the action (e.g., how much to spend). In a similar spirit to the Gul and Pesendorfer (2001) model, if the restrictions in Study 1 serve to reduce the self-control problem by reducing the set of options, restrictions could be beneficial. Since "increased cognitive

[4] Fischer (2001) presents arguments for breaking a task into smaller components. On p. 261, she states: "Therefore, the best way for a supervisor to. . .reduce the risk of missing the ultimate deadline may be to break it into smaller tasks with more deadlines to better compete with the other demands on the student's time."

[5] Of course, one might argue that our weekly requirements were not well chosen. However, we chose these after considerable discussion with undergraduate students and *ex post* would choose them again.

[6] In fact, in a separate study, Amir and Ariely (2008) find that discrete progress markers may "generate complacency, sway motivation away from the end goal, and decrease performance in the task" when the distance to the goal is certain, as is the case in Study 1.

[7] An alternative explanation is that participants who completed the Stroop test viewed this as a form of commitment.

[8] Overall, one of the causes for apparent reversals in preferences over time seems to be the change in the saliency of the costs and benefits of the activity in question (Akerlof, 1991).

[9] Frederick et al. (2002) provide a comprehensive review of empirical research on intertemporal choice, as well as an overview of related theoretical models. We would also like to mention two very new papers. Bisin and Hyndman (2009) investigate stopping-time problems and characterize behavior for exponential, naïve-hyperbolic and sophisticated-hyperbolic discounters. They show that an agent with standard time preferences who suffers from "temptation and self-control" would never be willing to self-impose a deadline. Suvorov and van de Ven (2008) develop a theory of self-regulation based on goal setting. They derive a condition under which proximal short-term goals are better than distal long-term goals.

[10] Thaler and Shefrin (1981) present an early agency model of self-control, involving a farsighted *planner* and a myopic *doer*.

load makes temptations harder to resist" in this model, it predicts that students who are assigned the discordant Stroop in Study 2 should procrastinate and presumably be less successful.

Two other models instead consider self-control problems in the context of strength of will or willpower.[11] Bénabou and Tirole (2004) consider a model in which self-signaling (over two periods, with imperfect knowledge of one's type) leads to self-regulation. The key is that people have imperfect knowledge of their *strength of will*, learning it only through experience and having difficulty later remembering it; one's initial choice serves as a 'signal' about one's type. The model's predictions depend on parameter values, but one feature is that "the degree of self-control an individual can achieve is shown to ... decrease with prior external constraints." A natural interpretation in our design is that a participant with weekly requirements in Study 1 will tend to be less likely to complete the studying task. In Study 2, if willpower depletion enhances a student's recall of their own strength of will, then students subjected to this test will tend to be *more* successful on the second day. Of course, alternative explanations also exist. For example, having completed the discordant Stroop could be a form of "investment" or a psychological "sunk cost" that subjects do not want to waste by not completing the task.

Ozdenoren et al. (forthcoming) provide an explanatory model based on willpower as a depletable (but renewable) resource. A willpower-constrained individual will regard seemingly unrelated activities as linked because he or she uses the same cognitive resources to exercise self-control in different activities. If willpower is depleted during one part of the week and replenished during another, we could see weekly studying cycles in Study 1. Weekly requirements will tend to induce better performance, since requirements substitute for using willpower. In Study 2, students who have their willpower depleted should tend to postpone answering questions until the second day. This model does not predict a higher success rate for those people who have their willpower depleted.

3. The experiments

3.1. Study 1

This experiment was conducted at the University of California at Santa Barbara. We obtained permission to have anonymous access to the grade records of students in a large introductory undergraduate class and then recruited as many people as possible from this class. We then advertised the session to first-year students in the general experimental subject pool.[12] All students were told that they could attend an introductory meeting about an experiment that would involve a non-trivial amount of money to be earned over time. Interested students were randomly assigned to one of two introductory meetings.[13] Participation was voluntary and everyone who showed up was guaranteed $5 even if they were not interested in participating in the study. At these meetings, we explained the nature and rules of the experiment. This process lead to a total of 74 eventual participants (out of 87 students who showed up to the meetings); 42 were from the class and 32 were from the campus-wide experimental subject pool. As we show later, there was no appreciable difference in behavior across these two sets of participants.

We chose the task of studying because it is a common activity for students but one that is susceptible to procrastination. Studying has obvious long-term benefits, but it is costly in the short-run insofar as other activities have more immediate appeal.[14] Nevertheless, there are already incentives in place for the studier; thus, we did not pay the usual average per-hour rates for experiments but chose to pay $95 for 75 h of monitored studying. Ideally one could record 'natural' studying completed by students, but this was not feasible. Instead, we recorded the time students spent studying in a popular, easy-to-access location: the central campus library.

We showed participants the studying location, a room in the library that was frequently (but intermittently) monitored. This study area was available for between 14 and 16.5 h each day. Subject to the availability constraint, students were free to log in and out by handing over an ID card to the monitor who would then log the student in or out on a computer. In addition, students were each given a large numbered placard, unique to each individual. This was visible to the monitor at all times. The studying area was monitored hourly at a varying time each hour to ensure students were present at the studying location when signed in.

We conducted two treatments. In one treatment, there were no requirements other than logging in the 75 h of study over the five-week period. In the other treatment, each student was required to log in a cumulative total of at least 12 h per week (12 h by the end of the first week, 24 h by the end of the second week, etc.); the idea was that this would prevent participants from falling too far behind, while still offering some scope for procrastination (one could log 48 h by the end of the fourth week, but would still need to log in 27 h in the last week). The aim was to provide evenly spaced deadlines (which have been shown by Ariely and Wertenbroch, 2002 to be very effective) but still allow students some flexibility in how they allocated their studying time over the five weeks. If the weekly deadlines were not met, one did not receive any payment. Note that

[11] These models relate closely to the experimental work by Baumeister et al. (1994) and Baumeister and Vohs (2003), which finds that people who have just been required to exercise self-restraint tend to exhibit a lower degree of self-control.

[12] We thank ORSEE (Greiner, 2004) for the free recruiting software, which permitted selective invitations.

[13] All the students in a particular informational meeting were assigned to the same treatment group. This was done to reduce social interaction threats (Cook and Campbell, 1979), i.e., the possibility that students in the control group become aware of the treatment group and vice versa.

[14] Students may therefore wish to do more studying than they actually manage; this is similar to self-control problems such as dieting or smoking.

this 'forfeiture' penalty is quite different from the daily penalties in Ariely and Wertenbroch (2002). Subjects who missed a deadline in their design could still participate in the experiment but faced a small grade or financial penalty for each day past the deadline.

Each student was assigned a web page where he or she could check on the number of hours logged and could then contact us in the case of any discrepancy. In addition, students who satisfied weekly studying requirements 'banked' their contingent earnings; their web pages had a check-like graphic showing the credit already amassed (of course, this credit was only to be paid if the student completed the overall 75-h requirement). Students who failed to meet a weekly requirement were notified at the end of the applicable week that they were no longer eligible to earn the $95. At the end of the five-week period, those students who had completed the requirement(s) received their earnings and filled out a short questionnaire.

We would like to immediately address two possible concerns. First, students had access to both computers and wireless Internet, so we cannot be certain how much of their time in the library was devoted to studying. However, the anecdotal evidence from the monitors is that, although students occasionally just checked e-mail or Facebook, etc., apparent studying was by far the most common activity observed. Moreover, our results indicate some improvement in grades among those students who completed the studying task, so there is also an inference that significant studying occurred. Second, one might also be concerned about contamination, since people from both treatments studied in the same area. Again, we have only anecdotal evidence against this: (1) Exit interviews of the people who completed the study task indicate that students who spent over 75 h in each other's company were not aware of the other treatment group, and (2) monitors did not observe students conversing with one another.

3.2. Study 2

Where our first study collected detailed data on long-term behavior in a quasi-realistic setting, our second study focuses more closely on behavior over a shorter time period. Study 2 also adds the element of willpower depletion, as some of the patterns observed in Study 1 were suggestive of willpower effects. Study 2 was conducted online with multiple-choice economics questions.[15] We recruited participants from a micro principles class and two intermediate micro classes at UCSB. We supplemented the 135 participants from these classes with 23 economics or business economics majors who were not in these classes but who were in the campus-wide subject pool. In all cases, we described the task (in general terms) and the payment scheme. The task consisted of answering, over a two-day period, 20 multiple-choice questions taken from microeconomics exams. The participant could answer all 20 of these on the first day or spread them over the two days.[16] We paid each person $7.50 for completing the 20 multiple-choice questions, with each correct answer earning an additional $0.75 for the student.

Each participant was randomly assigned to either a Tuesday–Wednesday group or a Friday–Saturday group; these different two-day periods presumably reflect either different opportunity costs or different stocks and/or flows of willpower. We also required people to do 250 rounds of a Stroop test on the first day before proceeding to answer the multiple-choice questions.[17] Discordant Stroop tests are used in psychology experiments to deplete willpower and consist of showing words that are the names of colors, although the actual words are printed in a color of ink different from the color name they represent. For example, the word "blue" might be printed in red ink. One is asked to respond by typing the color seen and ignoring the word itself. Resisting the urge to read the word and instead focus on the color of the ink requires considerable use of the brain's executive functions. We randomly assigned each participant to process either discordant Stroop exercises or concordant ones (where the word color matches the ink color and the task is trivial). Thus, one treatment group had its willpower depleted before beginning work on the multiple-choice questions and the other group performed a simple task that required minimal cognitive control and should not have affected willpower levels.

3.3. Questionnaire

The predictions of some of the leading theoretical models within the context of our experiments depend on parameter values. Since these cannot be divined from our experimental results, we conducted an anonymous survey of 69 undergraduate students in an economics class in an attempt to gain insight into some of the actual parameter values. The full questionnaire is presented in Appendix D. While there was no way to make these questions incentive compatible, we did pay each student $2 for completing the questionnaire, as anecdotal evidence suggests that such payment induces more serious thinking about the questions and responses.

The questions focus on aspects of the tasks in both of our studies. One question concerns how people would allocate 75 h of library study time over each week of a 5-week period. In one version, people were asked how they would *prefer* to

[15] We also conducted a study involving tasks of varying length to be completed online over seven consecutive days. However, the overall completion rate was only 15% and there was evidence of substantial confusion on the part of the participants. We observe smaller day-of-the-week effects and a slight tendency to delay the longer tasks until later. Details of this study are reported in Burger et al. (2008), and earlier version of this paper.

[16] It was made clear to each participant that being inactive for 15 min after logging on would result in one being logged off for that day; thus, people could not return to the task on the same day after taking a long break.

[17] To control for possible differences in the time spent on the two Stroop tasks, we required an elapsed time of 15 min to complete the concordant Stroop; this was roughly the time it took to complete the discordant Stroop.

allocate their time during this period, while another version asked how they thought they would *end up* allocating their time during this period. The first version can be used as a benchmark against which procrastination can be measured: if people would ideally have an increasing profile over time, the presence of such a profile would not indicate that they are in fact procrastinating. The comparison provides some evidence of whether students are aware of any tendency to delay relative to their optimal path, *i.e.* are they "sophisticated" or "naïve" about their tendency to procrastinate? Other questions asked students how they would allocate 15 h of library study over the days of the week, how likely it was they would be able to complete a project that required periodic work over a 5-week period (this provides at least a proxy for uncertainty in the students' lives over this amount of time), how much they would have to be compensated for various amounts of studying, and whether they would rather do 90 min of homework all at once or do more homework spread out over two days.

4. Experimental results

4.1. Study 1

One of our research objectives was to test whether providing interim targets or requirements would help people to achieve the overall task. In the weekly requirements treatment, 15 of 38 students (39.5%) were successful. When there were no weekly requirements, 22 of 36 students (61.1%) completed the mandated 75 h. Clearly, there is no support for the view that externally imposed restrictions helped students to achieve the goal. In fact, the test of the difference of proportions (see Glasnapp and Poggio, 1985) gives $Z = 1.86$, so that there is a marginally significant difference in success rates across treatments in the opposite direction to what we expected ($p = 0.062$, two-tailed test).[18]

We also find that the completion rate for the 45 female participants was more than 50% higher than for the 29 male participants.[19] However, this difference is only marginally significant; the test of the difference of proportions gives $Z = 1.67$ ($p = 0.095$ with a two-tailed test). This result is similar in flavor to the finding in Angrist et al. (2009) that providing incentives for good grades is much more effective for female students.

One might speculate as to what the completion rate for the students with no weekly requirements would have been if they had faced the requirements. While it is impossible to construct the appropriate counterfactual, we can conduct a simple thought experiment. We take the data for the subjects without weekly requirements and calculate their total studying hours by week. We then calculate which students completed enough studying each week to satisfy the requirements in the weekly requirements treatment, even though they were not subject to them. This provides our 'counterfactual' for the students with no weekly requirements. We find that the completion rates are quite similar: the completion rate for the no-requirements treatment would now be 14 of 36 (38.9%), compared to 15 of 38 (39.5%) for the weekly requirement treatment. In fact, the nearly identical completion rates suggest the difference in success rates is driven by the presence of the restrictions and the corresponding reduction in flexibility, implying that procrastination *per se* was unaffected by imposing weekly restrictions.[20]

One interesting aspect of our data is the cyclical weekly patterns in the number of study hours logged. Fig. 1 shows these patterns for those students who completed the project (referred to as "winners"); the pattern is actually slightly stronger for the group without weekly requirements. In the aggregate, the cumulative number of study hours logged for winners were very close to a 'target' line of 15 h per week.[21] The average number of study hours (75.35) for the winners was very close to the minimum of 75, ranging from 75.02 to 76.74. This suggests that students did not find this studying task to be innately pleasurable.

Is there a difference in study hours across weeks for those who completed the 75 h of study? Table 1 presents the average number of hours for the winners by week and by group.

There is no clear trend over time for either treatment. Regressions of hours against weeks yield insignificant coefficients for the time trends (0.09 and 0.30, respectively, with corresponding t-statistics of 0.11 and 0.76). This does not appear to be evidence of procrastination. However, if we look at the study patterns for each individual, there is evidence that some people front-load studying hours while others delay.[22] Appendix A shows the study hours by individual, both for all participants and for only those people who completed the 75 h. We classify people who finished as front-loaders (back-loaders) if they

[18] Recall that some participants were from an introductory class and others were first-year students from the general subject pool. Thirteen of 20 students from the class succeeded in the no-weekly treatment, compared to 9 of 16 students from the subject pool ($Z = 0.54$), while eight of 22 students from the class succeeded in the weekly requirements treatment, compared to 7 of 16 students from the subject pool ($Z = 0.46$). Neither difference (nor the overall comparison of 21 of 42 versus 16 of 32) is close to statistical significance.

[19] Twenty-six of the 45 females (57.8%) completed the requirement(s), compared to 11 of the 29 males (37.9%).

[20] On the other hand, a portion of the difference in completion rates across treatments is driven by the difference in the proportions of participants who never log any study hours after signing up for the experiment. This applies to nine of the 38 people (23.7%) in the weekly treatment, compared to two of the 36 people (5.6%) in the no-weekly treatment; this difference in rates is significant ($Z = 2.19$, $p = 0.028$, two-tailed test). This suggests that the realization that there were multiple hurdles may have deterred people from even starting the task.

[21] This is reflected in the fact that both the weekly and no weekly requirement lines intersect with the 15-h target line on Sundays – the last day of each week in the experiment.

[22] In addition, one might classify those people who did not finish (or did not even start) as more serious procrastinators (we thank Jeroen van de Ven for this comment). On this interpretation, we in fact do observe a considerable degree of procrastination. However, it is not clear that these people are really delaying (in the sense of the definition in footnote 1) the task rather than abandoning it. In any case, we are primarily considering the treatment effect of interim deadlines on completion, rather than committing to an interpretation.

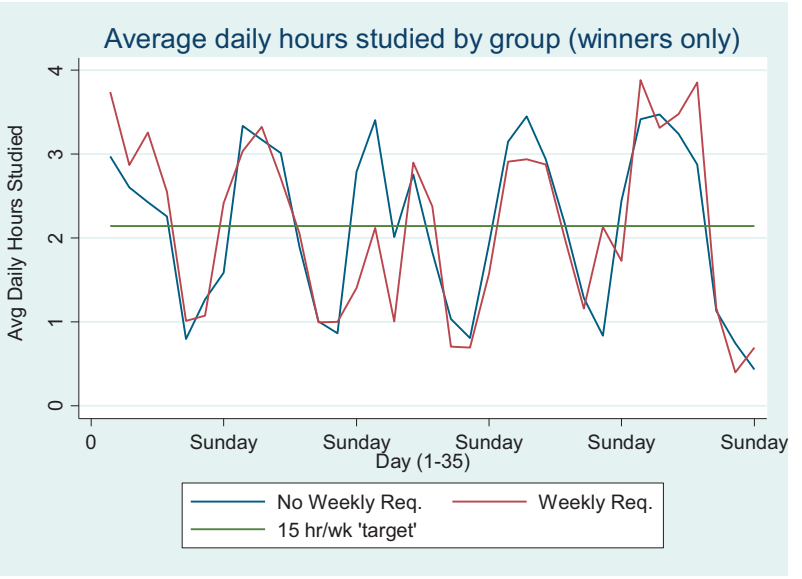

Fig. 1. Average daily hours studied by group (winners only).

Table 1
Average weekly study hours (winners), by group.

Week	Weekly requirements	No weekly requirements
1	16.92 (0.41)	13.91 (0.50)
2	14.53 (0.43)	16.08 (0.39)
3	11.38 (0.50)	13.78 (0.44)
4	15.75 (0.62)	16.27 (0.61)
5	16.78 (0.69)	15.31 (0.71)

Note: Standard errors are in parentheses.

Table 2
Regressions for weekly studying hours (winners only).

Variables	Group		
	(1) Weekly requirements	(2) No weekly requirements	(3) Pooled data
Week 2	−2.39 [1.57]	2.17 [1.41]	0.32 [1.09]
Week 3	−5.54*** [1.62]	−0.13 [1.36]	−2.32* [1.11]
Week 4	−1.18 [2.24]	2.36 [2.82]	0.92 [1.89]
Week 5	−0.14 [2.82]	1.40 [2.93]	0.78 [2.04]
Constant	16.92*** [1.15]	13.91*** [1.36]	15.13*** [0.95]
# observations	75	110	185
R^2	0.12	0.02	0.03

Notes: Week 1 is the omitted variable in these regressions. Robust standard errors clustered by subject are in brackets. * and *** indicate significance at the 10% and 1% level, respectively (two-tailed tests). The pooled regression includes a control for treatment group (not reported).

logged at least half of their hours in the first (last) two weeks of the five-week period. Of the 37 people who completed the task, four people front-loaded, while 11 people back-loaded.[23,24]

We present OLS regressions showing study patterns for both treatments in Table 2. The regressions confirm that there is no clear increasing or decreasing profile over the course of the experiment. Only the coefficient for the dummy for Week 3 has any statistical significance, and this would appear to reflect the effect of a closure of the library during evening peak study time due to a power outage (see the dip around day 16 in Fig. 1).

Table 3 shows the average hours of studying logged by winners on each day of the week. The number of study hours shows a tendency to decrease over the course of the week (from Monday to Thursday), with a dramatic drop on Friday and

[23] As some of the front-loaders may be people who study a great deal and who simply continued extensive studying at home once 75 h were logged, this is an upper bound on the number of front-loaders.

[24] We see little difference across gender in this regard – one (three) of the 11 (26) male (female) winners chose to front-load, and three (eight) of the 11 (26) male winners chose to back-load; neither of these differences is close to statistical significance ($Z = 0.22$ and 0.21 for the respective comparisons).

Table 3
Average study hours (winners), by day and group.

Day	Weekly requirements	No weekly requirements	Pooled data	Test of weekly vs. no weekly (Z)
Monday	3.14 (0.23)	3.25 (0.20)	3.21 (0.15)	0.168
Tuesday	2.69 (0.24)	2.94 (0.20)	2.84 (0.15)	0.718
Wednesday	3.04 (0.23)	2.87 (0.19)	2.94 (0.15)	−0.409
Thursday	2.57 (0.30)	2.21 (0.22)	2.35 (0.18)	−0.939
Friday	1.01 (0.17)	1.05 (0.15)	1.03 (0.11)	−0.101
Saturday	1.06 (0.21)	0.90 (0.16)	0.97 (0.13)	−0.658
Sunday	1.56 (0.24)	1.84 (0.24)	1.73 (0.17)	0.325

Note: Standard errors are in parentheses. The final column reports a Wilcoxon–Mann–Whitney test of the difference between day-of-week results for the two treatments.

Table 4
Mean scores on tests, by group.

Group	N	Quiz	Midterm	Final
Non-participants	403	3.13 (0.05)	8.78 (0.13)	7.56 (0.15)
Participants, non-winners	21	3.21 (0.16)	9.57 (0.51)	7.71 (0.4)
Participants, winners	21	3.67 (0.15)	10.71 (0.49)	9.24 (0.5)

Note: Standard errors are in parentheses

Saturday, and some recovery beginning Sunday afternoon. The patterns are essentially similar across the two treatments. Students appear to start the week fairly fresh and run out of steam as it progresses. The weekend appears to be the time when students 're-charge' (perhaps their willpower); alternatively, students might have a higher opportunity cost of studying on the weekend.

We see a strong and significant cyclical pattern, common to both treatments.[25] In regressions (not shown, but available by request), there is virtually no difference between the estimated coefficients for the two groups (when an indicator variable for *group* is added, its coefficient is 0.00; when we include week × group interaction indicators, none of these has a coefficient that is close to statistical significance – the lowest p-value is 0.47).

Although the regression results find no significant differences from Monday to Wednesday, analysis of each individual's logged study hours suggests that people tend to log the most hours on Monday, with a steadily declining rate until it increases on Sunday. We can examine how many people logged the most and second-most study hours by day of the week (see Appendix C). For the full participant population, 22 of the 63 people (34.9%) who logged any study hours logged the highest number of study hours on Monday.[26] The pattern is even stronger if we only include the winners, with 15 of 37 people (40.5%) logging the highest number of study hours on Monday.[27] On the other hand, the non-winners have a higher proportion of study hours later in the week, with this being particularly true for the group with weekly restrictions; it seems plausible that these people tried to catch up late in the week but did not succeed.

A final question of importance is whether completing the study task was helpful in terms of performance. As mentioned earlier, 42 of our participants were in an introductory undergraduate class. We were given permission to access the grade records for the course, matching student ID numbers for the participants. There were quizzes, a midterm, and a final exam in the course. Our study commenced in the fourth week of the quarter, with two quizzes preceding our study. There is more variability in the quiz grades, with a number of people missing them, particularly after the midterm. We therefore trust the midterm (taken after the first week of the five-week experiment) and final-exam (taken a week after the experiment ended) scores more, but we nevertheless include an average for the first two quizzes. Table 4 shows the mean scores by group.

We see that the differences between the non-participants and the non-winners are generally small, although slightly larger for the midterm.[28] In fact, Wilcoxon–Mann–Whitney (two-tailed) ranksum tests confirm that none of these differences are significant (for the midterm comparison, we find that $Z = 1.37$, $p = 0.171$). On the final exam, there was no difference between non-participants and non-winners ($Z = 0.10$). There is also no significant difference between non-winners and winners on the midterm scores ($Z = 1.58$, $p = 0.115$), although the p-value is close to the 10% significance level. The difference between final scores, however, is highly significant ($Z = 2.38$, $p = 0.017$). Thus, the difference in test scores across participants

[25] Since the winners only represent 50% of the participants (37 of 74), it is natural to wonder whether there is more back-loading among those students who did not manage to complete the task. Naturally, the data for this group are far less complete, as most people who did not complete the studying task stopped logging hours early on. Appendix B shows a pattern somewhat similar to Fig. 1 for the first two weeks of the study (70% of the people who did not complete the task logged no hours after the second week) for people who did not attain the studying target.

[26] This compares to 15 on Tuesday, 13 on Wednesday, six on Thursday, one on Friday, one on Saturday, and five on Sunday; the pattern is similar if we consider highest and second-highest days of the week.

[27] This compares to six on Tuesday, eight on Wednesday, five on Thursday, and five on Sunday; again, the pattern is similar if we consider highest and second-highest days of the week.

[28] Given that only 42 of 445 students in these classes chose to participate, one could certainly argue that there was a higher degree of procrastination among those students who did not participate. While we are primarily interested in treatment effects (and use random assignment to these treatments), we caution that the levels of the behavior observed may potentially not be representative of the overall student population.

Table 5
Determinants of the probability of success and number of questions answered.

Dependent variables	Independent variable	
	(1) Success rate	(2) Questions, 1st day
Tuesday	.539** (.212)	1.59 (3.34)
Discordant Stroop	.406* (.212)	−6.92** (3.42)
Male	.515** (.213)	−8.04** (3.45)
Constant	−.384* (.222)	22.11*** (4.17)
Observations	158	100
Pseudo R^2	0.068	0.018

Notes: Standard errors in parentheses. Specification (1) is a Probit regression and (2) is a two-sided Tobit regression, including only those people who answered all 20 questions. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively (two-tailed tests). We omit interaction terms, as these are not significant and do not qualitatively change the main results.

increased over the course of the quarter, suggesting that people who logged in 75 h of library study improved their grades relative to those who signed up and did not complete the task.[29]

4.2. Study 2

Of the 158 people who signed up online, 100 (63%) completed the task successfully. In a departure from our earlier results, 60 of 85 males (71%) completed this shorter task, compared to 40 of 73 females (55%); this difference is statistically significant ($Z = 2.05$, $p = 0.040$, two-tailed test). Success rates were significantly higher for the Tuesday–Wednesday group, as 57 of 79 people (72%) in this group finished and 43 of the 79 people (54%) in the Friday–Saturday group finished ($Z = 2.31$, $p = 0.021$, two-tailed test).[30]

With respect to the people who completed the task, the average number of questions answered on the first (second) day was 11.53 (8.47); similarly, for the nine people who answered more than zero, but less than 20, questions in total, the average number of questions answered on the first (second) day was 5.56 (3.33).[31] We see some tendency for males to procrastinate more than females, as the number of questions answered on the first day for people who finished was 9.97 for males and 13.88 for females ($Z = -2.25$, $p = 0.025$, two-tailed test).

We observe differences in behavior according to the type of Stroop test that was assigned. Twenty-two of the 45 people (49%) who were assigned the concordant Stroop and succeeded at the task answered all 20 questions on the first day, compared to 17 of the 55 people (31%) who completed the task and were assigned the discordant Stroop ($Z = 1.83$, $p = 0.034$, one-tailed test). This implies some degree of willpower depletion.[32]

A rather surprising result is that people who were assigned the discordant Stroop were actually slightly *more* likely to eventually finish the task. Fifty-five of the 79 people (70%) assigned the discordant Stroop finished, while 45 of the 79 people (57%) assigned the concordant Stroop finished; the test of proportions gives $Z = 1.65$, $p = 0.099$, two-tailed test. Specification (2) in Table 5 confirms that people who completed the task after performing the discordant Stroop answered significantly fewer questions on the first day, and that males who completed the task delayed more than did females.

Another measure that may reflect willpower depletion, but in any case certainly reflects the quality of work, is the percentage of questions answered correctly. There is no significant difference in the percentage correct across gender (62% for males vs. 60% for females) or type of Stroop test (60% in both cases). In addition, the percentage of correct answers on the first day does not predict success rates (the t-statistic on the coefficient for number of questions is 0.21). However we do find that the percentage of correct answers on the first day is significantly higher than on the second day.[33] This is robust to whether we consider the whole sample of 158 people, the 100 people who finished, or the 47 people who answered questions on both days.[34] The percentage

[29] One issue is whether the study hours in the monitored location were simply a substitute for study hours elsewhere. The data suggest that perhaps this is not completely the case. In addition, the results from our pre- and post-experiment questionnaires reveal that only 24% of the eventual winners studied more than 15 h per week before the experiment started and 64% of winners reported reducing their weekly study hours once the experiment ended. This provides some evidence that the experiment increased total hours studied over the five-week period.

[30] Thirty-nine people answered all 20 questions on the first day, 15 people answered all 20 questions on the second day, and the other 46 people who completed the task answered questions on both days. Ninety-two people answered some questions on the first day, while 64 people answered questions on the second day. Forty-nine people never answered any questions, while the remaining nine people answered between four and 15 questions. Sixty-one of the 64 people (95%) who answered some questions on the second day completed the task successfully.

[31] Since more questions are answered on the first day, at first glance there is no evidence of procrastination in the aggregate. However, one must keep in mind that each participant was required to answer the Stroop tests on the first day, so that coming back on a second day involved an additional transaction cost.

[32] We are not wedded to the term "willpower", as this could be seen as, for example, depletion of cognitive resources. We use this term simply to reflect the standard terminology in the psychology literature.

[33] The order of questions was randomized, so there should be no difference in difficulty levels across days. One possible explanation for the lower success rate on the second day is that students may have "sprinted to the finish" once the finish was within sight.

[34] The respective comparisons are 0.66 vs. 0.49 ($Z = 4.35$, $p = 0.000$), 0.66 vs. 0.50 ($Z = 4.10$, $p = 0.001$), and 0.67 vs. 0.47 ($Z = 3.43$, $p = 0.001$). All of these tests are two-tailed.

Table 6
Optimal and expected weekly study hours, questionnaire.

Week	Optimal	Expected
1	13.85 (1.37)	13.56 (1.20)
2	15.45 (1.07)	12.88 (1.00)
3	16.30 (1.08)	14.15 (0.91)
4	13.67 (1.07)	15.85 (1.04)
5	15.24 (2.17)	18.85 (1.92)

Note: Standard errors are in parentheses.

of correct answers is also higher for the Tuesday–Wednesday group (0.63 vs. 0.56, $Z = 1.78$, $p = 0.038$, one-tailed test).

Two features of our data are the pronounced weekly studying cycles in Study 1 and the day-of-the-week effects in Study 2. Although Study 2 (and studies in the psychology literature) provide evidence that willpower is a resource that can be depleted, it is not clear whether the weekly cycles in Study 1 reflect the depletion of willpower, changes in the opportunity cost of studying, or some other influence. Nevertheless, data from Study 1 offer some insight into this question. The opportunity-cost story would be most consistent with a difference in behavior during the workweek and during the weekend, rather than a daily difference (at least during the workweek). On this basis, we should expect similar behavior for the period from Monday through Thursday. The individual data suggest that Monday is the favorite day for logging study hours. In addition, we see a significant decline in study hours on Thursday.[35]

4.3. Questionnaire results and a calibration

The main question in the survey concerns how people would allocate 75 h of library study time over each week of a 5-week period. In one version, people were asked how they would *prefer* to allocate their time during this period, while another version asked how they thought they would *end up* allocating their time during this period. The results are shown in Table 6:

The aggregate optimal studying profile is essentially flat, with the hours for each week never significantly different from 15. The aggregate expected studying profile does show evidence that people expect to delay somewhat in the task, or procrastinate. A second question asked how people would allocate 15 h of library study over the days of the week. Here the pattern resembles that seen in Table 3.[36] Thus, people do appear to be aware of their actual intra-week preferences, as only the expected hours for Sunday study differ much from the actual numbers in Table 3.

Regarding how likely it was that students would not be able to complete a project requiring periodic work over a 5-week period (a proxy for the degree of uncertainty in the students' lives), there is overall only a modest degree of uncertainty, with a median response of 10%. Two additional questions examined the convexity of effort costs. In one question, students were asked for the minimum amount of money they would require to study in the library for 1, 2, or 4 h in a day; the median responses were \$8, \$18, and \$40, respectively, indicating a slight degree of convexity. A second question asked students for the minimum amount of money they would require to study in the library for 10, 20, or 30 h during a one-week period; the median responses were \$55, \$120, and \$300. Both sets of responses indicate at least a moderate degree of convexity in effort cost.

In Study 1, if students procrastinate, study hours should be logged near the end of each week in the weekly requirement treatment; without weekly requirements, there should be a surge in hours logged late in the five-week period. However, since uncertainty could potentially overwhelm the tendency to procrastinate, the predictions for the studying profile in Study 1 will depend on assumptions about parameter values. We use the results from the questionnaire to perform a simulation regarding the quasi-hyperbolic model of procrastination with different values of β (either 1.0, 0.9, or 0.7), with and without uncertainty over the cost of future studying. Appendix F shows how we calibrate the cost function for study hours in a week; we describe how we estimate the weekly profile for various parameter combinations and present the simulations in Appendix G.

We see that the calibrated cost function predicts a weekly profile that is essentially flat for an exponential discounter ($\beta = 1$) without uncertainty. This pattern (which we observe in the aggregate) is also predicted for an agent with mild time-inconsistency ($\beta = 0.9$, as suggested by Shapiro, 2005; Meier and Sprenger, 2010) and uncertainty. However, when we instead use the value of $\beta = 0.7$, as suggested by McClure et al. (2007) and Laibson et al. (forthcoming), the model predicts that even with uncertainty, studying hours increase sharply over time. Thus, while the aggregate pattern that we observe

[35] While some might argue that the weekend starts on Thursday evening, time-of-day data (shown in Appendix E) show more of a decrease (relative to Wednesday) for Thursday morning and afternoon, rather than Thursday evening; the percentage decrease in logged study hours for the morning and afternoon combined is 36.4%, while this percentage decrease is only 14.2% for the evening. While this suggests that willpower is being depleted through the workweek, it is hardly conclusive.

[36] There were 2.82, 2.59, 2.87, 1.81, 0.91, 1.08, and 2.90 h projected respectively for Monday–Sunday.

is not fundamentally inconsistent with the quasi-hyperbolic models, it appears to rule out population values for β that are substantially less than 1.

5. Conclusion

Our two studies consider self-control, procrastination and willpower effects on tasks of long and short duration. In Study 1, we thought that what we considered to be well-chosen deadlines would help people to overcome their tendency to procrastinate and be more likely to complete the studying task than people not facing restrictions; however, we find the opposite result. In Study 2, while we do find that the discordant Stroop tends to induce people to delay more of the task until the second day, the overall success rate is actually higher for the people who faced the willpower-depleting Stroop on the first day. In both studies, we see evidence of day-of-the-week effects, whereby there is less effort put forward on weekends and more effort on weekdays in the earlier part of the week. The results of both studies suggest a degree of procrastination in the population, although there is considerable heterogeneity in behavior. However, the simulation based on our questionnaire data indicates that the degree of procrastination is not severe overall.

The fact that the behavioral interventions in both studies were ineffective points out the lack of existing data on how incentives affect behavior on tasks involving significant duration. Our Study 1 results suggest that commitment devices (at least externally imposed ones) may in fact be counter-productive if they involve imposing additional hurdles to overcome. That our studying task and the tasks considered by Ariely and Wertenbroch (2002) produce decidedly different results suggests that the tendency to procrastinate varies significantly with incentives and context. Further, in relation to the effectiveness of commitment devices, we strongly suspect that details matter. In some cases, there may be a preference for flexibility while in other cases there may be a preference for commitment.

Study 2 also highlighted that procrastination is not a simple phenomenon – and not one easily reproduced in a controlled experimental setting. For example, while the discordant Stroop exercise did deplete willpower on the first day, a natural prediction would seem to be that this depletion would reduce overall success rates. Yet, overall success rates were higher for the willpower-depleted group. Is exerting willpower in the discordant Stroop test self-signaling, so that this leads to higher (eventual) completion rates, or does this result represent some sort of commitment due to having suffered through this exercise, inducing the determination to complete the task? Is there a reduction in willpower due to an excess of cognitive load in an alternative task, as suggested by Ozdenoren et al. (forthcoming) and Fudenberg and Levine (2006)?

It is quite clear that our results are only a beginning. We consider the area of time structuring, procrastination, and incentives to be just coming into its first full flowering. We hope that researchers in this area can build upon our results in conducting both more empirical research and more theoretical work. Our studies could provide motivation for the design of other studies aimed at mapping out the boundaries of effective deadlines, and could form the starting point for designing a series of studies to test the various behavioral theories mentioned, as well as others yet to be formulated.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jebo.2010.11.010.

References

Ainslie, G., 1992. Picoeconomics: The Interaction of Successive Motivational States within the Individual. Cambridge University Press, Cambridge.
Akerlof, G., 1991. Procrastination and obedience. American Economic Review 81, 1–19.
Amir, O., Ariely, D., 2008. Resting on laurels: the effects of discrete progress markers as subgoals on task performance and preferences. Journal of Experimental Psychology: Learning, Memory, and Cognition 34, 1158–1171.
Angrist, J., Lang, D., Oreopoulos, P., 2009. Incentives and services for college achievement: evidence from a randomized trial. American Economic Journal: Applied Economics 1, 136–163.
Angrist, J., Lavy, V., 2009. The effects of high stakes high school achievement awards: evidence from a randomized trial. American Economic Review 99, 1384–1414.
Ariely, D., Wertenbroch, K., 2002. Procrastination, deadlines, and performance: self-control by precommitment. Psychological Science 13, 219–224.
Ashraf, N., Karlan, D., Yin, W., 2006. Tying Odysseus to the mast: evidence from a commitment savings product in the Philippines. Quarterly Journal of Economics 121, 635–672.
Baumeister, R., Heatherton, T., Tice, D., 1994. Losing Control: How and Why People Fail at Self-regulation. Academic Press, San Diego.
Baumeister, R., Vohs, K., 2003. Willpower, choice and self-control. In: Loewenstein, G., Read, D., Baumeister, R. (Eds.), Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice. Russell Sage Foundation, New York, pp. 201–216.
Bénabou, R., Tirole, J., 2004. Willpower and personal rules. Journal of Political Economy 112, 848–886.
Bisin, A., Hyndman, K., 2009. Procrastination, self-imposed deadlines and other commitment devices. Mimeo.
Burger, N., Lynham, J., 2010. Betting on weight loss. . . and losing: personal gambles as commitment mechanisms. Applied Economics Letters 17, 1161–1166.
Burger, N., Charness, G., Lynham, J., 2008. Three field experiments on procrastination and willpower. Mimeo.
Charness, G., Gneezy, U., 2009. Incentives to exercise. Econometrica 77, 392–406.
Cook, T., Campbell, D., 1979. Quasi-experimentation: Design & Analysis Issues for Field Settings. Houghton Mifflin, Boston.
DellaVigna, S., Malmendier, U., 2004. Contract design and self-control. Quarterly Journal of Economics 119, 353–402.
Ellis, A., Knaus, W., 1977. Overcoming Procrastination. Institute for Rational Living, New York.
Fischer, C., 2001. Read this paper later: procrastination with time-consistent preferences. Journal of Economic Behavior and Organization 46, 249–269.
Frederick, S., Loewenstein, G., O'Donoghue, T., 2002. Time discounting and time preference: a critical review. Journal of Economic Literature 40, 351–401.
Fudenberg, D., Levine, D., 2006. A dual-self model of impulse control. American Economic Review 96, 1449–1476.
Gailliot, M., Baumeister, R., DeWall, C., Maner, J., Plant, E., Tice, D., Brewer, L., Schmeichel, B., 2007. Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. Journal of Personality and Social Psychology 92, 325–336.

Glasnapp, D., Poggio, J., 1985. Essentials of Statistical Analysis for the Behavioral Sciences. Merrill, Columbus.
Greiner, B., 2004. An online recruitment system for economic experiments. In: Kremer, K., Macho, V. (Eds.), Forschung und Wissenschaftliches Rechnen 2003. GWDG Bericht 63, Ges. für Wiss. Datenverarbeitung, Göttingen, pp. 79–93.
Gul, F., Pesendorfer, W., 2001. Temptation and self-control. Econometrica 69, 1403–1435.
Houser, D., Reiley, D., Urbancic, M., 2007. Checking out temptation. Mimeo.
Houser, D., Schunk, D., Winter, J., Xiao, E., 2009. Temptation, commitment, and self-control in the laboratory. Mimeo.
Kiefer, M., Marzinzik, F., Weisbrod, M., Scherg, M., Spitzer, M., 1998. The time course of brain activations during response inhibition: evidence from event-related potentials in a go/no go task. NeuroReport 9, 765–770.
Laibson, D., 1997. Golden eggs and hyperbolic discounting. Quarterly Journal of Economics 112, 443–477.
Laibson, D., Repetto, A., Tobacman, J., forthcoming. Estimating discount functions with consumption choices over the lifecycle. American Economic Review.
Logan, G.D., 1994. On the ability to inhibit thought and action: a user's guide to the stop signal paradigm. In: Dagenbach, D., Carr, T.H. (Eds.), Inhibitory Processes in Attention, Memory, and Language. Academic Press, San Diego, pp. 189–239.
MacLeod, C., 1991. Half a century of research on the Stroop effect: an integrative review. Psychological Bulletin 109, 163–203.
McClure, S., Ericson, K., Laibson, D., Loewenstein, G., Cohen, J., 2007. Time discounting for primary rewards. Journal of Neuroscience 27, 5796–5804.
Meier, S., Sprenger, C., 2010. Present-biased preferences and credit card borrowing. American Economic Journal: Applied Economics 2, 193–210.
O'Donoghue, T., Rabin, M., 1999. Doing it now or later. American Economic Review 89, 103–124.
Ozdenoren, E., Salant, S., Silverman, D., forthcoming. Willpower and the optimal control of visceral urges. Journal of the European Economic Association.
Perret, E., 1974. The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. Neuropsychologia 12, 323–330.
Shapiro, J., 2005. Is there a daily discount rate? Evidence from the food stamp nutrition cycle. Journal of Public Economics 89, 303–325.
Solomon, L., Rothblum, E., 1984. Academic procrastination: frequency and cognitive-behavioral correlates. Journal of Counseling Psychology 31, 503–509.
Stroop, J., 1935. Studies of interference in serial verbal reactions. Journal of Experimental Psychology 18, 643–662.
Strotz, R., 1955. Myopia and inconsistency in dynamic utility maximization. Review of Economic Studies 23, 165–180.
Suvorov, A., van de Ven, J., 2008. Goal setting as a self-regulation mechanism. Mimeo.
Thaler, R., Shefrin, H., 1981. An economic theory of self-control. Journal of Political Economy 89, 392–406.
Wertenbroch, K., 1998. Consumption self-control by rationing purchase quantities of virtue and vice. Marketing Science 17, 317–337.