Temptation and Commitment in the Laboratory

Daniel Houser, Daniel Schunk, Joachim Winter and Erte Xiao

November 2017

Contact Details:

Daniel Houser
George Mason University
ICES, 4400 University Drive, MSN 1B2
Fairfax, VA, 22030
USA

dhouser@gmu.edu

Daniel Schunk
Chair of Public Economics
Johannes Gutenberg University Mainz
Jakob-Welder-Weg 4
55128 Mainz
Germany

daniel.schunk@uni-mainz.de

Joachim Winter
Chair of Empirical Economic Research
Ludwig-Maximilians-Universität München
Ludwigstraße 33
80539 München
Germany

winter@lmu.de

Erte Xiao
Department of Economics
Monash University
Clayton, VIC, 3800
Australia

erte.xiao@monash.edu

Temptation and Commitment in the Laboratory

Daniel Houser
George Mason University
dhouser@gmu.edu

Daniel Schunk
Johannes Gutenberg University Mainz
daniel.schunk@uni-mainz.de

Joachim Winter
University of Munich
winter@lmu.de

Erte Xiao
Monash University
erte.xiao@monash.edu

This paper is forthcoming in Games and Economic Behavior.

Abstract: We report data from a novel laboratory experiment on economic decisions under persistent temptations. This type of temptation is ubiquitous, as it refers to any temptation that is present until one either gives in or makes a costly commitment decision to have it removed. Subjects in our experiment are repeatedly offered an option with instantaneous benefit that also entails a substantial reduction to overall earnings. We show that this option is tempting in the sense that a substantial fraction of our subjects incur pecuniary costs to eliminate the choice, and thus commit not to choose this alternative. We find that commitment and giving in to temptation generally occur at the first opportunity, though a non-negligible fraction of subjects delay either making the commitment decision or giving in to temptation. This delay is consistent with the costs of self-control increasing with its use.

Keywords: self-control; willpower; temptation; commitment; laboratory experiment

JEL classification: D11; C91

1. Introduction

Because temptation can interfere with attaining long-term goals, the ability to resist temptation is an important skill. In this paper, we investigate behavior under "persistent" temptations. That is, temptations that appear repeatedly until one either succumbs to the temptation or uses a (possibly costly) commitment device to remove it. For example, saving money requires consistently controlling the impulse to consume immediately. Also, to complete a time consuming task can require continuous effort to resist the temptation to procrastinate. Since self-control is difficult, people frequently turn to external devices to assist them in resisting. One such device is to intentionally exclude the tempting good from a choice set, as has been studied in a number of contributions (e.g. Fudenberg and Levine 2006; Gul and Pesendorfer, 2001; Noor 2007). For instance, a person trying to avoid eating meat may commit to dining only at vegetarian restaurants, or people may choose to set up a deadline to help overcome procrastination (Ariely and Wertenbroch, 2002).

The questions of when people choose to use a commitment device when facing a persistent temptation and how that choice is affected by the cost of commitment are of significant economic importance. In particular, economic theory suggests that the timing of commitment, or the time at which one succumbs to temptation, can depend on how the cost of self-control changes with the use of self-control (e.g., Fudenberg and Levine, 2012). Here we present a novel laboratory experiment that informs how people make decisions under persistent temptation, and thus provides evidence on the specification of models of behavior under temptation.[1]

How people make decisions under temptation has long been a topic of interest in both psychology and economics. Strotz (1955-1956), develops theoretical explanations for time-inconsistent decision-making.[2] More recently, it has been shown that temptation and commitment can be modeled in a time-consistent manner.[3] These and related theoretical

---

[1] An early draft of this paper (Houser et al., 2010) reported data from a similar design that included a "surprise" temptation, as compared to the full-information design analyzed below. We are grateful to an anonymous referee for encouraging this alternative design and analysis.

[2] Subsequent theoretical approaches and reviews of this vast literature include Thaler and Shefrin (1981), Laibson (1997), O'Donoghue and Rabin, (1999, 2000), Bénabou and Tirole (2004), Heidhues and Köszegi (2009), and Caillaud and Jullien (2000).

[3] Important contributions to the theoretical literature on temptation and commitment include Gul and Pesendorfer, (2001, 2004, 2005), Fudenberg and Levine (2006, 2010, 2012), Noor (2007, 2011), Miao (2008), Ozdenoren et al. (2008), Dekel et al. (2009), and Ali (2011).

developments are grounded in a broad empirical literature demonstrating that humans are affected by temptation and sometimes succumb to it (see, e.g., Mischel et al. 1989).[4,5] Further, data from natural and field experiments provide evidence for the importance of temptation in real economic choices.[6] In psychology, substantial evidence suggests cognitive load (which is thought to reduce the ability to exercise "willpower") leaves subjects more likely to choose tempting goods.[7]

The psychic cost of exercising self-control can encourage people to use commitment devices to resist temptation. Early research considered a variety of commitment devices that restricted individuals' choice sets, for instance, by visiting a restaurant with fewer tempting foods (Wertenbroch, 1998), by using specific ordering strategies that enforce watching "high-brow" movies (Read et al., 1999), by removing the ability to make choices in the future (Casari, 2009), or by committing to saving plans (Ashraf et al., 2006; Benartzi and Thaler, 2004; Beshears, 2015). In addition to commitment, other strategies to avoid succumbing to temptation can involve increasing the cost giving in (Schelling, 1992; Gine et al., 2010; Kaur et al. 2010; http://www.stickk.com/[8]), distracting oneself from the temptation, or forcing oneself to use certain personal rules of conduct (see Baron, 2000; Elster, 2000, Ariely and Wertenbroch, 2002).

Given the vast empirical evidence on people's difficulties in resisting temptation and their effort to use commitment devices, it is important for economists to develop an improved understanding of whether these devices are adopted with delay, and whether and how the timing of their use is affected by the benefit of resisting the temptation and the cost of commitment. That is, it is important to know more about the timing and elasticity of decisions in environments with persistent temptations. To shed light on this, we designed a new laboratory experiment that allows the cost of commitment and the cost of giving in to a temptation to be manipulated

---

[4] Additional empirical results in this literature are due to Hoch and Loewenstein (1991), Ainslie (1992), Loewenstein and Prelec (1992), Baumeister et al. (1994), Metcalfe and Mischel (1999), Ameriks et al. (2007), Charness and Gneezy (2009), Casari (2009), Augenblick et al. (2015), Bonein and Denant-Boémont (2015), and Corgnet et al. (2015); see also Bryan et al. (2010) for a survey.

[5] It is worth emphasizing that the early "marshmallow task" delay of gratification experiments (see the survey by Mitchell et al., 1989) as well as the later derivatives of these experiments (Bucciol et al., 2011) generally include a "persistent temptation" of the sort studied in this paper. The same is true of early studies of procrastination and deadlines (Lowenstein and Prelec, 1992). The advantage to our study is that it offers a carefully controlled environment that allows a more direct test of theory by narrowing the scope for alternative explanations for the patterns in our data.

[6] See Della Vigna and Malmendier (2006), Houser et al. (2008), Burger et al. (2009) and Bucciol et al. (2009).

[7] See Shiv and Fedorikin (1999), Baumeister et al. (1994), Baumeister and Vohs (2003), Hinson et al. (2003), Dewitte et al. (2005), and Vohs and Faber (2007).

[8] This is a company launched by Ian Ayres and Dean Karlan. A participant signs a contract requiring him/her to donate a self-specified amount of money to charity if he/she fails to achieve a self-specified goal.

exogenously and independently. This control enables us to discover systematic patterns in commitment under temptation that would be difficult to observe in field data.

In an environment with persistent temptation it can be optimal, under perfect foresight, to delay one's decision to commit depending on whether the costs of exercising self-control increase with its use, as shown for instance by Fudenberg and Levine (2012). They first note that in a model without cognitive resources, "it is always cheaper to commit now to avoid a future temptation than to do so later when the temptation is more imminent" (p. 29). In their Example 9 they then show that once a model with cognitive resource is augmented with a commitment cost, then it might pay to use the cognitive resources to exert self-control until its marginal benefit is sufficiently high and then to pay the commitment cost, taking the temptation off the table. In order to provide evidence on whether self-control costs might change in this way, participants in our laboratory environment are exposed to the same tempting stimulus multiple times. On each exposure, they can either: (i) give in to the temptation; (ii) make a (perhaps costly) commitment decision that removes any future exposures to the temptation; or (iii) resist the temptation without choosing to commit, thus being be exposed to it again in the future. Participants are provided complete information at the beginning of the experiment about all aspects of the environment, including the specific way that temptation and opportunities to give-in or commit will evolve over time. Consequently, our experiment sheds light on whether delay in commitment occurs in a perfect foresight environment.[9]

We find that the majority of our participants either commit or give in immediately, or resist temptation until the end of the experiment. This result is consistent with Fudenberg and Levine (2012) under the condition that there are unchanging (positive) costs to exercising self-control.[10] This behavior is also consistent with predictions from many other models of temptation, including Noor (2007, 2011), Fudenberg and Levine (2006, 2010) , Gul and Pesendorfer (2001), and in general models that rely on quasi-hyperbolic discounting (Strotz, 1955; Phelps and Pollack, 1968; Laibson, 1997; O'Donoghue and Rabin, 1999).[11]

---

[9] Fudenberg and Levine (2012), at the conclusion of their Example 9, suggest the importance of data from these types of environments.

[10] In the sense of Fudenberg and Levine, we mean by unchanging costs of self-control, that the exertion of self-control in period t increases the self-control costs at $t+1$, and this should not be confused with a temporal convex costs of self-control. In the remainder of the paper, we also call this property "increasing (marginal) costs of self-control".

[11] See Fudenberg and Levine (2012) for detailed discussion of connections among various leading economic theories of temptation.

At the same time, we observe a non-negligible fraction of participants to delay their decisions either to commit or give in to the temptation. In particular, after initially resisting without commitment, 37% of the subjects either give in or commit before the end of the experiment. Finally, we find that the demand for commitment is highly price elastic, a result that could help to explain the paucity of markets for commitment devices.

The remainder of this paper is organized as follows. Section 2 illustrates a key finding from Fudenberg and Levine (2012) using a simple example. This also serves to motivate our experiment design, which is described in Section 3. In Section 4, we use the framework of Fudenberg and Levine (2012) to derive hypotheses for the behavior of the subjects in our experiment. Sections 5 and 6 contain results and discussion, respectively. The last section of the paper provides a summary and suggestions for future research.

2. Theory Background and a Motivating Example

Our analysis is built on the idea that self-control is an "exhaustible resource", in the sense that using self-control may leave immediately subsequent use of self-control more costly. This insight is far from new (see, e.g., Hoch and Loewenstein, 1991; Shiv and Fedorikhin, 1999; Baumeister et al., 1994; Baumeister and Vohs, 2003; Hinson et al., 2003; Dewitte et al., 2005; Vohs and Faber, 2007; Ozdenoren et al. 2012). Our development below is inspired by Fudenberg and Levine (2012, henceforth, FL) which shows that increasing marginal costs of self-control in dual-self models can imply rational delay of commitment or "giving in" decisions in temptation environments with perfect foresight.

Although we develop our hypotheses within a dual-self framework, delays in committing or giving in to temptation could also be modeled by extending any existing static temptation models in the literature to allow for dynamically increasing marginal costs (e.g., Noor, 2007, 2011)[12]. The intuition is that a person's decisions under temptation depend not only on the current cost of exercising self-control but also expected costs of exercising self-control in the future. We next offer an example that strengthens this intuition and motivates our experiment design.

---

[12] We are grateful to an anonymous referee for detailing a model of our experimental decision environment that builds on Noor (2007, 2011) and proves that delays in commitment can occur in that framework due to increasing marginal costs of self-control.

Consider a dual-self model in the spirit of FL. There is a single patient long-run self along with a sequence of myopic short-run selves. The long and short-run selves share the same preferences over stage outcomes, but differ in how they view the future. For the purposes of our example, we consider the case where the long-run selves have a positive discount rate over future periods, while short-run selves are perfectly myopic in that they care only about stage-game outcomes.[13]

Each period consists of two parts. In the first part of each period, the long-run self can exert self-control to affect the preferences (thus choices) of the short-run self. Exerting self-control is costly for both parties. In the period's second part, after preferences have been chosen, the short-run player makes the period's final decision. In this framework: (i) the long-run self cannot pre-commit for the entire dynamic game, but instead begins each new period needing to make a (costly) self-control decision; and (ii) the long-run self has the same stage-game preferences as the short-run selves, and so wishes to serve the interests of future short-run selves.

More specifically, consider a three-period consumption/savings problem where each period a person can either save or consume some, part or all of their resource. Each unit saved doubles the next period, while a unit consumed provides a unit value. In this environment the ability to consume is the persistent temptation, as the short-run self will always prefer to consume everything immediately, while the long-run self will see the value of investing. In order to make the example precise, suppose the long-run self values future payoffs using a discount-factor of 0.8 per period, meaning from the perspective of time-period one the utility of consuming two in the second period is 1.6, and the utility of consuming 4 in the third period is 2.56.

Suppose the long-run self can change the short-run self's preferences at a cost equal to a scalar multiple a of the short-run self's forgone period utility. Suppose further that this multiple increases with each use. In particular, suppose the first time self-control is used the multiple $a_1 = 0.05$, while the second use the multiple increases to $a_2 = .20$. Under these parameter values, for the long-run self, we find: the utility of exercising no self-control and consuming in the first period is 1.0; the value of exercising self-control in the first period and then consuming 2 in the second period is 1.6-0.05=1.55; and the value of exercising self-control once in the first period and once in the second period and then consuming 4 in the third period is 2.56-0.05*1-

[13] Fudenberg and Levine (2006) develop this special case in detail, while their 2012 paper analyzes the general case where the short-run self may care about the future positively, but always less than the long-run self.

0.2*1.6=2.19. Consequently, an agent with perfect-foresight will plan to use self-control twice when in period 1, and when in period 2 will again prefer to exercise self-control.

Suppose further that, in addition, there is a commitment device available that can take the temptation to consume off the table at the beginning of a period and for all subsequent periods, before self-control needs to be exercised. Suppose the nominal cost of this commitment, whenever used, is 0.30. Consequently, from the perspective of the first period, the utility earned by using commitment immediately is 2.56-0.3=2.26, while the utility gained by exercising self-control in the first period, and then commitment in the second, is 2.56-0.05-0.8*0.3 = 2.27. Consequently, an agent with perfect-foresight will plan in period one to exercise self-control and delay commitment to period 2. Likewise, in period 2, the utility of exercising no self-control and consuming immediately is 2; the value of exercising self-control and then consuming 4 in the next period is 4*0.8-0.2*2=2.8; and the value of committing and consuming 4 in the next period is 4*0.8-0.3=2.9. Thus, the agent will commit in period 2, after exercising self-control in period 1, and doing so is time-consistent.

In this simple environment, it is easy to see that increasing costs of self-control are necessary to generate delayed commitment (under the natural assumptions that discounting and costs of commitment are fixed). It is important to emphasize however that, so long as costs of self-control are increasing with repeated use, delayed commitment is robust to many different parameterizations.[14] For example, both zero initial costs of self-control and a discount factor equal to one can be consistent with delayed commitment in a perfect foresight model. To see this, note that if $a_1 = 0$ and the discount factor is 1 in the example above (with all other parameter values remaining the same), then the agent is indifferent between committing immediately or in the second period, but would strictly prefer delayed commitment for any discount factor less than one. Consequently, by examining individual decisions in an environment with persistent temptations, and in particular discovering whether there is delay in decisions to commit or succumb to temptation, one can provide evidence regarding whether increasing costs of self-control are relevant for models of economic decision-making under temptation.

---

[14] Appendix 1 generalizes the example and shows that both impatience and increasing costs of self-control are necessary conditions, within this framework, for an agent to strictly prefer to delay to commit. For instance, using the notation of Appendix 1, the utility of exercising self-control in period 1 followed by commitment in period 2 is written as $\delta^2 B - a_1 - \delta c = 2.27$. We thank an anonymous referee for suggesting this generalization.

Unfortunately, we are unable to design our experiment around an environment as simple as the one described above. The reason is that during a relatively short laboratory experiment nobody would have trouble reinvesting until they earned maximum profits. Our design takes a different tack, one that has not previously appeared in the literature, that turns out to be quite effective at producing a tempting good. We now turn to describing this design.

3. Experiment Design

We design a laboratory experiment that includes a tempting good and the (costly) option to commit not to choose this tempting good. Our design allows us to manipulate exogenously both the commitment cost as well as the benefits from resisting temptation. Finally, in order to test directly specific predictions from FL, we observe subjects' decisions over a sequence of repeated exposures to a tempting good. We use a perfect foresight environment. Subjects are given full information about the number and timing of temptations they can experience as well as the consequences of succumbing (or not) to temptation. We create temptation by using an avoidable repeated counting task designed to be boring.

Counting task

In each repetition of the counting task, subjects are asked to count the number of ones in a series of nine digits, either zeros or ones, that are displayed on the computer screen for 15 seconds (see Figure A.1 for a screenshot). Counting tasks are displayed at random time intervals. Between counting tasks, subjects face an otherwise empty screen with a digital clock that displays the elapsed time of the experiment (see Figure A.2 for a screenshot). Subjects reported that they found this task distasteful.

The time between counting screens is either 1 minute, 2 minutes, or 3 minutes. To ensure that subjects focus their attention on the computer screen (and hence not get involved in other activities), they are informed that the lengths of these waiting periods are drawn randomly such that each of these time spells occurs, on average, equally often during the counting task (i.e., the waiting time can be considered as being drawn from a uniform distribution over the time spells).

At the beginning of each session, participants are informed that the experiment will last for 120 minutes and that they have to remain in the laboratory for the entire duration of the experiment. Then, the counting task and the random nature of the waiting periods between counting tasks are explained and participants are informed that their payoff is higher (up to an

additional $15), the more counting tasks they perform in the two hours on the condition that the accuracy rate is 70% or greater. To make more likely that subjects attended to the computer screen for the duration of the experiment, we told subjects that they would earn only $3 in addition to their show-up fee if they provide less than 70% of correct responses.[15] (see the instructions in Appendix 2 for details)

The tempting good and the temptation screen

All subjects were required to deposit all backpacks, phones and any other items with the experimenter prior to the experiment beginning. Consequently, they were unable to access the internet or become involved in any other electronic activity for the total duration of the experiment.

In our experiment, the tempting good is the option to stop performing the counting task and to instead obtain internet access. We refer to this temptation as "surfing the internet" in the sequel. The reason we chose this is that we expected all subjects to experience positive utility from internet access[16]: one can check e-mail, chat, visit social networks, read the news, or even study.

Surfing the internet is offered to subjects during the two-hour counting task at up to 12 specific points in time (depending on their decisions). The points in time at which the temptation screens would appear were announced to subjects in the instructions, at the beginning of the experiment. In this context, it is important to stress the following two details of our design. First, when subjects surf the internet, they lose the possibility to return to the counting task during the remainder of the experiment. Second, when subjects are offered the opportunity to surf, they are simultaneously offered the option to commit to counting for the remainder of the experiment. In particular, the use of this commitment device eliminates one's ability to surf the internet during the experiment.

We operationalize the surfing option by showing to the subjects an internet browser window, referred to as the "temptation screen" (for a screenshot see Figure A.3). This screen displays the payoff the subject has earned so far and offers her three choices represented by two buttons and a default option:

---

[15] None of our participants nearly violated this restriction.
[16] The utility of surfing the internet might vary between subjects. Such preference heterogeneity does not affect our analysis as we randomly assign (i) opportunity costs of surfing; and (ii) commitment costs.

- *Continue*: Continue the counting task, having the chance to earn an additional payoff depending on their performance in the counting task, and keep the option to surf the internet.
- *Commit*: Continue the counting task, having the chance to earn an additional payoff depending on their performance in the counting task, but without being given the opportunity to surf the internet any more. Choosing this commitment option involves a cost that is taken from the subject's total payoff.
- *Surf*: Surf the internet right away. This is the default if no other choice is made by the time the temptation screen disappears (after 120 or 60 seconds).[17]

It is worth emphasizing that the instructions (see Appendix 2) made subjects aware of the timing of the temptation screens as well as the three alternatives available on each temptation screen.

Phases of the experiment

A key advantage of our laboratory environment is the control it affords. For example, subjects might resist surfing the internet and perform the counting task for the entire two hours without using the commitment device even when it is free, suggesting that "surfing" is not sufficiently tempting to overcome subjects self-control. Alternatively, subjects could choose to commit to counting by removing the surfing option. Especially when commitment is costly, this suggests that surfing is indeed a tempting good. Yet another possibility is that subjects could choose to surf the internet at one of the pre-announced points of time, evidently giving up future rewards for an immediate benefit. As we will see below, the timing and presence of these possible behaviors informs economic theory.

Note that at the beginning of the experiment subjects may not correctly anticipate neither the boredom associated with the counting task nor the disutility it creates. Our design minimizes this as a potential explanation for decisions by dividing the experiment into three phases which every subject passes through in the same order. In the first phase, phase 0, subjects perform the counting task—as described above—for 30 minutes, and, importantly, no surf or commitment

---

[17] Our interest is in showing that people will pay to avoid exposure to temptation. Because we wanted this to be an active choice, we chose surfing as the default option. Note that the default was almost never exercised. When the temptation screen appears for the first time, it is shown for 120 seconds, ensuring that subjects had enough time to understand the text and the payoffs. The remaining temptation screens are displayed for 60 seconds.

options are provided; in this phase subjects receive repeated exposure to, and thus gain experience with, the counting task.[18]

The subsequent two phases, 1 and 2, offer surf and commitment options using the "temptation screen" at specific points in time, with various opportunity costs of surfing and commitment costs depending on the randomly assigned experimental conditions (see Experimental Conditions section below for details). In all experimental conditions payoffs are structured such that the total amount earned by a subject who performs the counting task successfully for the three phases of the experiment is $15 (in addition to the show-up bonus). The timing of the experiment is summarized in Table 1.

- *Phase 0* lasts for 30 minutes. In this phase, subjects are exposed to 15 counting tasks with empty screens shown in between the counting tasks, as described to them at the beginning of the experiment.

- *Phase 1* lasts for 45 minutes (i.e., it starts at minute 31 and ends after minute 75). Subjects are exposed to up to 12 counting tasks with empty screens appearing in between the counting tasks as described above. Additionally, in this phase there are six instances at which subjects are shown the temptation screen (described above). On this screen, subjects can choose whether they would like to continue performing the counting task, surf the internet, or use a commitment device that removes the opportunity to surf the internet for the remainder of the experiment. Subjects are made aware, at the beginning of the experiment, both that these screens will appear as well as the time at which each screen will appear.

- *Phase 2* lasts for another 45 minutes (i.e., it starts at minute 76 and ends after minute 120). As in phase 1, up to 12 counting tasks, separated by empty screens, are displayed. Additionally, subjects are shown the temptation screen at six instances. The temptation screen is identical with the temptation screen shown in phase 1, only the payoffs differ (see Tables 1 and 2). Again, subjects are made aware, at the beginning of the experiment, both that these screens will appear as well as the time at which each screen will appear.

** Include table 1 here **

---

[18] While this helps participants understand the boredom associated with counting, it is possible that even with this experience they may not be able to anticipate their willingness to trade money for counting as the experiment proceeds.

** Include table 2 here **

When a subject enters a new phase, the payoff structure changes according to the description in Tables 1 and 2, enabling tests of the behavioral hypotheses developed in section 3. This change in the payoff structure, of course, only concerns those subjects who have not yet decided to surf the internet or use the commitment option. If a subject had already chosen to surf the internet in the previous phase, she would be required to continue to surf the internet for the entire remaining time of the experiment, without possibility to do the counting task or use the commitment option. Similarly, if a subject had already chosen the commitment option, she would continue with the counting task for the duration of the experiment without seeing further temptation screens.

Experiment conditions

At the beginning of the experiment, each subject is randomly assigned to one of 12 experimental conditions, denoted by G1 through G12. The condition determines the values of three design parameters—commitment cost (P), value of completing the counting task successfully in Phase 1, W_1, and value of completing the counting task successfully in Phase 2, W_2—which in turn determine the payoff a subject can achieve conditional on her performance in the counting task and additional decisions made during the experiment. Subjects are given complete information about their decision environment, but are not aware that there are multiple experiment conditions. The values of the design parameters are listed in Table 2.

These conditions serve two purposes: First, they allow us to test whether observed behavior is robust to differences in the parameterization. Second, they test whether the surfing option is a temptation and whether subjects understand the structure of the task: Specifically, in conditions G2, G5, G8, and G11, subjects receive the maximum additional payoff of $15 already if they successfully complete the counting tasks through phase 1. That is, in those conditions, all subjects for whom surfing has more utility than counting should stop counting at beginning of Phase 2. Indeed, all subjects in these treatments did this.

Earnings and sample

All subjects received a show-up fee of $5. Thus, the maximum amount that a subject could earn (by performing the counting task for the entire two hours with more than 70% correct answers

and filling in the questionnaire) was $20. The entire amount was paid in cash at the end of the experiment.

The experiment was conducted in the experimental laboratory of the Interdisciplinary Center for Economic Science (ICES) at George Mason University. A total of 108 subjects participated in 17 sessions; the corresponding descriptive statistics describing subjects' characteristics are reported in Table 3. The number of subjects within sessions was kept small to minimize distractions due to other subjects in the laboratory.

<center>** Include table 3 here **</center>

4. Hypotheses

To develop our hypotheses, note first that each phase of our game includes a finite number of periods, where each period is either a "counting" task or a "temptation" task, but not both. Players know the exact number and distribution of the two tasks. Recall also that all subjects are exposed to a 30-minute phase 0 consisting of repeated counting tasks before moving to Phase 1 where they see the first temptation screen. In addition, all subjects have experience with surfing the internet. Hence, subjects know what utility they derive from counting, or from surfing. Players also know that if they finish all of the counting in a given phase (with some easy-to-obtain level of accuracy), they earn $W_H > 0$ at the end of the experiment, but if they choose instead to enjoy the internet, they earn $W_L < W_H$. (The payoffs W_L may change between Phase 1 and Phase 2, but for ease of exposition subscripts denoting this dependence are dropped). Players evaluate payoffs according to the strictly monotonically increasing and concave utility function $U(\cdot)$.

Participating in the counting task is the only alternative during a "counting" period. Denote the utility of counting in period t by $U(\text{count}_t)$. Without loss of generality, assume $U(\text{count}_t) = 0$ for all t.[19]

Players have several choices during a "temptation" period. The temptation screen informs them that they may: (i) stop counting and surf the internet for the duration of the experiment; (ii) pay a one-time non-negative amount to make a commitment to count for the duration of the

---

[19] This assumes there is no risk of failure when performing the counting task. As an empirical matter, nobody failed to achieve the 70% accuracy rate required for success.

experiment; or (iii) continue counting without committing. Players' decisions during a temptation period are modeled as follows.

We suppose the short-run self evaluates the temptation period perfectly myopically. It considers only the period utility of surfing the internet, in comparison to the period utility of deciding to count in the future (which they will not experience). Denote the utility of surfing in period t as $V_t(\text{surf})$. Because the utility of counting is zero, the short-run player will prefer to surf whenever $V_t(\text{surf}) > 0$. If surfing carries negative utility (is less enjoyable than counting), the short-run player is indifferent between continuing and committing when commitment cost is zero, and otherwise strictly prefers to continue without cost.

The long-run self uses all short-run selves' period-utilities to evaluate decisions in this game. In any period t, the long-run value to finishing the counting task is equal to $U(W_H)$, the amount earned by the period t short-run player, plus $\tau(t)U(\text{count})$, where $\tau(t)$ is the number of counting periods a subject believes remain in the experiment at period t. However, because of the normalization $U(\text{count})=0$, the expectation term can be dropped and the long-run value of finishing the counting task reduces to $U(W_H)$. The long-run value of surfing is $U(W_L)+V_t(\text{surf})$, where $V_t(\text{surf})$ is the expected value of surfing from period t to the end of the experiment.

Thus, the long-run self compares $U(W_H)$ to $U(W_L)+V_t(\text{surf})$. If the value of surfing exceeds $U(W_H)$, then the long-run and short-run selves have aligned interests and the subject chooses to surf. However, if $U(W_H) > U(W_L)+V_t(\text{surf})$ then the long-run self might exercise self-control, at some utility (psychic) cost C $\geq$ 0, in order to ensure that the short-run player does not surf. In particular, FL assumes that by exercising self-control, the long-run self can change the short-run self's choice, in our case to ensure that surfing is viewed as less attractive than counting, so that $V_t(\text{surf}) < 0$.

Let $\tau'(t)$ denote the known remaining number of temptation screens in period t (including the current screen). Because self-control must be exercised at each temptation screen, and assuming additive separability and no time-discounting, the expected utility cost of exercising self-control each time it is necessary to do so throughout the remainder of the experiment is $\tau'(t)C$. Further, as we show in Hypotheses 1 and 2 below, any decision to commit or to surf

occurs immediately in this linear environment. In view of this, it is straightforward to show that the long-run self is willing to use exercise self-control if

$$U(W_H) - \tau'(t)C \geq U(W_L) + V_t(\text{surf}).$$

During a temptation period, the long-run self might also prefer to commit. The long-run self is willing to choose commitment preferences for the short-run self whenever one of the two following conditions hold:

$$U(W_H - P) - C \geq U(W_H) - \tau'(t)C \geq U(W_L) + V_t(\text{surf}) \qquad (1)$$

or

$$U(W_H - P) - C \geq U(W_L) + V_t(\text{surf}) \geq U(W_H) - \tau'(t)C \qquad (2)$$

where $P \geq 0$ is the pecuniary price of the commitment device. People who satisfy (1) will count without surfing if the cost of commitment is sufficiently high, while people with preferences satisfying (2) will surf if the cost of commitment is sufficiently high.

Finally, note that because participants have full information about the way the environment will evolve, $\tau'(t)$ is strictly monotonically decreasing in t. Similarly, assume that $V_t(\text{surf})$ is strictly monotonically decreasing in t, implying that surfing for a shorter amount of time is less valuable than surfing longer.

We are now in a position to state our main hypotheses.

Hypothesis 1: *At least some people exhibit a demand for costly commitment.*

While we cannot observe individuals' preferences, support for Hypothesis 1 suggests that surfing is a temptation for at least some people. It is easy to see why. Suppose surfing is not tempting (is less enjoyable than counting), so that $V_t(\text{surf}) \leq 0$. For such a person, the short-run self will prefer to count, so that the long-run self does not incur self-control costs. Thus, because $U(W_H) > U(W_H - P)$, no such person would commit. It follows that any person who makes a costly commitment necessarily finds surfing a temptation. Note this is true regardless of whether the costs of exercising self-control increase the more one uses self-control.

It is worth noting that if exercising self-control does not change subsequent costs to exercise self-control, then the decision to commit occurs at the first opportunity. If costs of exercising self-control increase with the use of self-control, then it can be rational to delay commitment.

To see this, note the value of committing after observing t temptation screens is $U(W_H - P) - tC$, which is maximized when t is minimized. Consequently, in an environment with unchanging costs of self-control, commitment always pays the most when exercised at the first opportunity.

It is also easy to see that this result need not hold in the presence of increasing costs of exercising self-control. Building on the intuition developed through the example in Section 2 (and Appendix 1), if committing has positive cost but a person can exercise self-control without cost at least once, and subsequent use of self-control becomes more costly than the commitment device, then a person would (at least weakly) prefer to delay commitment to the period where the cost of exercising self-control increases. Moreover, if the person discounts the future at any positive rate then the preference for delayed commitment can become strict.

Hypothesis 2: *Any decision to surf occurs at the first opportunity within a phase (either screen 1 or screen 7).*

The reason is that the decision to surf always pays the most when exercised immediately. This follows by reasoning exactly analogous to that above, and our assumption that $V_t(\text{surf})$ is strictly decreasing in t. Further, while counting may be preferred to surfing in phase 1, after a payoff change in phase 2 surfing may be preferred to counting. This occurs, for example, when the experiment's entire earnings are realized at the conclusion of phase 1, or more generally if the person has a threshold earnings requirement and the realized amount meets or exceeds that threshold. In these cases, one would predict surfing to be chosen at the first opportunity in phase 2 (screen 7). Finally, note that these predictions are invariant to whether there are increasing costs to using self-control. Because surfing always pays most when exercised earlier, any decision to surf should be made at the first opportunity (even when the costs of self-control are vanishingly small, and regardless whether the discount rate is positive).

Combining Hypotheses 1 and 2, we see that if the costs of exercising self-control are positive and constant, then any commitment should occur immediately (screen 1) and surfing decisions should occur on the first opportunity of Phase 1 or Phase 2 (screens 1 or 7). On the other hand, if exercising self-control leaves its subsequent use more costly, then we would expect surfing decisions on screen 1 or 7, while commitment could occur on any screen. We summarize this as Hypothesis 2a.

Hypothesis 2a: *Assuming the decision error-rate is the same for surfing and commitment, then delayed commitment (any commitment after screen 1) should be at least as frequent as delayed surfing (any surfing outside of screen 1 or screen 7).*

Hypothesis 3: *The frequency of commitment (weakly) monotonically decreases as its price increases.*

$$U(W_H - P) - C \geq U(W_H) - \tau'(t)C \geq U(W_L) + V_t(\text{surf})$$

This simple comparative static follows directly from (1) or (2). If people are heterogeneous with respect to their utility function, and in light of our assumption that utility is strictly increasing in its arguments, the fraction of the population for whom (1) or (2) is satisfied decreases monotonically as P increases, regardless of whether self-control costs exhibit nonlinearities.

Hypothesis 4: *The number of people who choose to surf will not decrease as commitment costs increase.*

It is easy to see that commitment costs can only affect surfing behavior when the following holds. Let commitment costs be ordered according to $P < P'$. Then, suppose

$$U(W_H - P) - C \geq U(W_L) + V_t(\text{surf}) > U(W_H - P') - C > U(W_H) - \tau'(t)C.$$

Subjects with such preferences avoid surfing by committing, because the cost to them of expending willpower is very high. However, when commitment becomes sufficiently expensive, it is no longer an attractive option, and the subject surfs.

5. Results

In order to structure the discussion of our results, we begin by describing the frequency of commitment and surfing decisions in each phase of the experiment. It turns out there are seven patterns of interest (see Table 4). We first show that the treatments generate differences in the distribution of these seven behaviors. After establishing this, we return to test the hypotheses detailed above.

Table 4 shows that upon seeing the temptation screen for the first time, 20.4% used the commitment device (in Table 4, *Behavior 1*), while 13.9% of the subjects started to surf (*Behavior 2*). Together, 34.3% of the subjects made one of these two decisions at the first

temptation screen. The next two lines of the table refer to subjects who used the commitment device (*Behavior 3*, 5.6% of all subjects) or started to surf (*Behavior 4*, 0.9%) later during phase 1 at one of the temptation screens 2 through 6. The fifth and sixth lines of the table list the subjects who used the commitment device in Phase 2 (*Behavior 5*, 3.7%), and the subjects (*Behavior 6,* 26.9%) who started to surf in Phase 2. This relatively large frequency of surfing at the beginning of Phase 2 is almost entirely due to subjects who had no remaining value for counting. The remaining 31 subjects (*Behavior 7,* 28.7%) never used a commitment device and resisted the temptation of surfing at all screens. For these subjects, utility from counting was not so low relative to that of surfing the internet to justify the lower payoff at the end of the experiment.

<center>** Include table 4 here **</center>

Table 4 also contains the test statistics and *p*-values for χ^2-tests for the effects of the randomly assigned treatments on the distribution of stylized behaviors. There is statistically significant difference for the contrast between zero vs. positive commitment cost ($p=0.012$). There is no significant effect of the size of the additional payoff for counting to the end of phase 1 ($p=0.564$). Considering only phase 2, however, we see that for those still counting, there is a significant effect ($p<0.000$) of the additional payoff for counting to the end of phase 2. The reason is that 25 of the 29 subjects who chose to surf in phase 2 had no remaining value for counting (three subjects with no remaining value counted to the end).

From the findings in Table 4, it is clear that our treatment manipulations have significant effects on the behaviors of subjects. In particular, an increase in commitment costs decreases the frequency of commitment (11.5% use commitment when it costs $1, as compared to 28.6% when it is free), while an increase in the value of using the commitment device (i.e., an increase in R_1) increases the frequency of commitment (22.2% use commitment at the first temptation screen if the remaining value was $7, and 18.5% if it was $5).

We next analyze the behavior in the first phase in more detail. Figure 1 shows, by temptation screen, the *fraction* of subjects who decided to commit or surf, respectively, during Phase 1 of the experiment. Figure 2 shows, by temptation screen, the *hazard rates* of subjects who committed and began surfing, respectively. Both figures are stratified by commitment cost (P) and the value of counting (R_1). These figures convey the five key findings that emerge from our data.

<center>18</center>

First, the two top panels of Figures 1 and 2 show that a substantial fraction of subjects pay the commitment costs and hence are willing to pay to remove the surfing option from their choice set. This supports Hypothesis 1.

Second, a majority of commitment decisions are taken at the first opportunity (in temptation screen 1); the increase in the share of commitment on screens 2 to 6 is small as compared to the large fraction of commitment decisions at screen 1. This finding suggests that costs of self-control are linear for most participants (see our remarks following the statement of Hypothesis 1). Third, surfing decisions are almost always taken at the first opportunity, thus supporting Hypothesis 2.

However, Figures 1 and 2 further indicate that in Phase 1 a number of subjects (a total of seven) do make commitment ($n=6$) and surfing ($n=1$) decisions after screen 1. Moreover, if we include those in Phase 2 whose decisions were payoff relevant (that is, those who could still earn positive amounts in Phase 2), we find three delayed commitment decisions (two at screen 11, one at screen 12) and one delayed surfing decision (at screen 8). Overall then, we observed a total of 11 payoff-relevant delayed decisions, 9 of which were to commit. With the caveat that the sample is small, this difference (9 vs. 2) is statistically significant (two-tail proportions (z) test against the null of 0.5, $p<0.01$). The finding that delayed commitment is substantially more likely than delayed surfing confirms Hypothesis 2a, and is additional evidence supporting nonlinear costs of self-control.

Fourth, supporting Hypothesis 3, we observe that the frequency of commitment decreases with an increase in commitment cost. Finally, we observe that the fraction of people who choose to surf increases with an increase in commitment cost, supporting Hypothesis 4.

** Include figure 1 here **

** Include figure 2 here **

In Table 5, we present results from six separate Probit regressions that corroborate all our findings above.[20] The dependent variables are binary and capture the decisions to surf and commit at three different points in the experiment: (i) at the first temptation screen; (ii) at one of the other five temptation screens in phase 1; and (iii) by the end of the experiment (i.e., after Phase 2). All six regressions include as independent variables a set of dummies for the randomly assigned treatments, as well as controls for the socio-demographic characteristics of the subjects:

---

[20] Each participant made independent individual decisions, and there was no communication among people within a session. Consequently, each regression reports results from $n=108$ independent observations at the individual level.

female, age above 20 years, and major (natural sciences, economics or business, and social sciences, with reference category "other").

Table 5 reports marginal effects. The personal characteristics are expressed as dummy variables, so that their coefficients correspond to the effect of a change of the dummy from zero to one, again expressed in percentage points. Finally, the commitment cost and value variables are expressed in dollars, so their coefficients measure the percentage-point effect of a one-unit change.

The first row in Table 5 shows that the commitment cost dummy is statistically significant in all regressions with commitment as the dependent variable. Commitment is less likely when it involves a positive cost. The impact is sizeable; changing commitment costs from zero to one dollar reduces the probability of ever *committing* during the experiment by about 29 percentage points (column 6). Commitment costs are also statistically significant predictors of choosing to surf at the first temptation screen (column 1). In the other two regressions (columns 2 and 3), commitment costs have no significant effect on the likelihood of choosing to surf. The reason is that, in line with hypotheses 1 and 2, virtually all decisions are made immediately. To summarize, the regression results with respect to commitment costs are in line with Hypotheses 3 and 4, derived in section 3.

** Include table 5 here **

We see from Table 5 that, in Phase 1, there is an insignificant effect of the remaining value to counting R_1 on the decision to commit. On the other hand, the remaining value of resisting temptation R_2 is a powerful predictor of whether subjects complete Phase 2 counting. The reason, as noted above, is that 29 participants faced a zero value of counting in phase 2, and nearly all of them chose to surf in this case.

6. Discussion

The majority of decisions we observe are consistent with an unchanging positive cost of self-control. This is consistent with specifications in Noor (2007, 2011), Fudenberg and Levine (2012) as well as a variety of other specifications in the literature. In addition to the qualitative patterns described by Hypotheses 1-4, all of which found support in our data, frameworks including an unchanging cost of self-control make the stark prediction that commitment decisions should occur at the first screen in Phase 1 (screen 1), while surfing decisions should be

made either at the first Phase 1 screen or the first screen of Phase 2 (screen 7). Indeed, from Table 4 one observes that of the 44 total surfing or commitment decisions made in Phase 1, 37 (84%) were made at the first screen. In Phase 2, of the 29 people who had no remaining value for participating in the game, 22 chose to surf on the first screen. Further, of the seven people in the second phase who chose to surf or commit and who had positive earnings available from counting, three made their decisions on the first screen (all to surf).

It is worthwhile to emphasize that our participants knew the precise distribution of future temptations and were never surprised by the occurrence of a temptation screen. Further, the initial 30 minute "training" phase helped to ensure subjects were aware of the distasteful nature of the counting task prior to making decisions in Phase 1. Consequently, the decision to commit implies surfing is a tempting good, particularly when there is a positive commitment cost. Our experimental design thus provides a useful paradigm for future research on temptation.

A possible alternative explanation for commitment is that the "temptation" screen could perhaps be perceived as an annoyance people will pay to remove. This would suggest that commitment decisions should monotonically increase with the number of exposures to the temptation screen. It turns out that while we do observe some delay in commitment decisions, commitment does not increase over time, and thus this possibility does not seem to explain the patterns in our data.

While we have emphasized that delay in commitment can be a rational response to nonlinearities in cognitive depletion, there are other potential explanations for our data. For example, a subject may be unsure about whether she will be able to maintain concentration, or whether boredom will become worse over the course of the experiment. In order to gain more certainty regarding the correct decision, it may be better to wait.[21] A related idea is that commitment itself may require self-control. In particular, a subject may believe that committing is optimal and yet may be tempted to maintain the surfing option. Resolving this internal conflict may take time, making it look like the subject delayed the decision. Our design narrows the scope for these explanations by giving participants full information at the beginning of the experiment regarding the way temptation and opportunities to commit will evolve over time, as well as giving them 30 minutes of exposure to the environment prior to any exposure to temptation. While it could in principle be that even with this experience and information, they

---

[21] The model by Dekel et al (2009), for example, could perhaps be used to model this possibility.

may not be able to anticipate their willingness to trade money for counting as the experiment proceeds, we were unable to discover an experiment design that would improve in this regard.

Finally, Tables 4 and 5 make clear that the demand for commitment is relatively low even when its price is zero, and also highly price elastic. While we are cautious in generalizing from this single experiment, it is worth noting that these findings may help to explain why the market for commitment devices has been slow to emerge, and leaves as an open question the importance of self-control in driving economic behavior in natural environments[22].

7. Conclusion

Our goal was to study behavior under temptation, and a primary contribution of this paper is the design of an experiment that, we showed, induces a persistent temptation while maintaining high experimenter control. In our experiment, subjects were paid to pursue a distasteful counting task, but tempted to forgo their wage for counting and instead surf the internet. Subjects were also offered the option to commit to counting by choosing to remove the option to surf. We found that surfing as well the commitment device are both normal goods, in the sense that succumbing to temptation is more likely when the cost of doing so is lower, and further that the price elasticity of the demand for commitment is high. In particular, when commitment comes at zero cost, 28.6% of subjects choose to eliminate the surfing option at the first opportunity. When the cost is $1, this fraction decreases to 11.5%.

We used the patterns in our data to test predictions derived from a dual-self model in the spirit of Fudenberg and Levine (2012). We noted that these predictions are consistent with and informative for a large number of other frameworks in the literature, such as Gul and Pesendorfer (2001) and Noor (2007, 2011). Under regularity conditions including that the cost of self-control is unchanging with the use of self-control, these models make the stark prediction that commitment or giving in to temptation should occur at the first opportunity. On the other hand, if using self-control leaves immediately subsequent use of self-control more costly, then we pointed out that delay in commitment can be rationalized in all the theoretical frameworks mentioned above. We showed further that in our specific environment only delay in commitment, but not delay in surfing decisions, can be rationalized by increasing costs to self-control. Consequently, our observation of more delayed commitment than delayed surfing

---

[22] We thank anonymous referees for suggesting this and several of the alternative explanations for commitment and delay in commitment noted in this section.

decisions is evidence favoring specifications that allow for increasing costs to exercising self-control.

As indicated in the discussion surrounding Hypothesis 1, delay to commit in our environment can be rationalized if self-control rapidly changes from very easy to very difficult (see also the example in Section 2, or Appendix 1). This observation appears consistent with rapid cognitive depletion combined with very slow replenishment. It would be valuable to know more about this process, and particularly how it might differ among individuals. For example, Table 5 reveals that females are significantly less likely to surf than males in Phase 1, a finding in line with the other research showing that females display greater self-control than males in other economic environments (Houser et al, 2012; Wang et al. 2017).

In view of our results, it would seem to be useful to model cognitive resource depletion using a mixture of flexible functional forms that nest linear within non-linear specifications, thus accommodating heterogeneity at the individual level. Such approaches are widely used in modeling beliefs and expectations (see, e.g., Houser et al, 2004). We believe these same methods might be profitably applied to the analysis of willpower and self-control.

Our paper provides a promising experimental paradigm to study temptation. We detailed a choice environment in which a substantial fraction of participants (undergrads at a large state university) were willing to pay to remove a "tempting" item from their menu of alternatives. Discovering this design was challenging, and we are aware of no other published experiment in which this behavior has been demonstrated. We intend to continue to use this paradigm to investigate why people delay commitment. Some have conjectured that non-stationarity is key, whether in costs of self-control or, for example, the preference to maintain opportunities for future indulgence. We expect future research using our experiment design to shed light on these and other mechanisms that may underlie the patterns we observe in behavior under temptation.

References

Ainslie, G. (1992): *Picoeconomics: The Interaction of Successive Motivational States within the Person*. Cambridge and New York: Cambridge University Press.

Ali, S. N. (2011): Learning self-control. *Quarterly Journal of Economics*, 126, 857–893.

Ameriks, J., A. Caplin, J. Leahy, and T. Tyler (2007): Measuring self-control problems. *American Economic Review*, 97(3), 966–972.

Ariely, D. and K. Wertenbroch (2002): Procrastination, deadlines, and performance: Using precommitment to regulate one's behavior. *Psychological Science*, 13, 219–224.

Ashraf, N., D. Karlan and W. Yin (2006): Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quarterly Journal of Economics*, 121, 673–697.

Augenblick, N., M. Niederle, and C. Sprenger (2015): Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, 130, 1067–1115.

Baron, J. (2000): *Thinking and Deciding*, 3rd edition. Cambridge and New York: Cambridge University Press.

Baumeister, R. F., T. F. Heatherton, and D. M. Tice (1994): *Losing Control: How and Why People Fail at Self-Regulation*. San Diego, CA: Academic Press.

Baumeister, R. and K. Vohs (2003): Willpower, choice, and self-control. In G.F. Loewenstein, D. Read, and R. Baumeister (eds), *Time and Decision*. Russell Sage Foundation, New York.

Bénabou, R. and J. Tirole (2004): Willpower and personal rules. *Journal of Political Economy*, 112, 848–886.

Benartzi, S., and R. H. Thaler (2004): Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112, S164–S187.

Beshears, J., J. J. Choi, V. Harris, D. Laibson, B. C. Madrian, and J. Sakong (2015): Self control and commitment: Can decreasing the liquidity of a savings account increase deposits? NBER Working Paper No. 21474.

Bonein, A., and L. Denant-Boémont (2015): Self-Control, commitment and peer pressure: A laboratory experiment. *Experimental Economics*, 18, 4, 543-568.

Bucciol, A., Houser, D., and M. Piovesan (2011): Temptation and productivity: A field experiment with children. *Journal of Economic Behavior and Organization*. 77, 126-136.

Burger, N., Charness, G. and J. Lynham (2011): Field and online experiments on self-control. *Journal of Economic Behavior and Organization*, 77, 393-404.

Bryan, G., Karlan, D., and S. Nelson (2010): Commitment devices. *Annual Review of Economics*, 2, 671–698.

Caillaud, B. and B. Jullien (2000): Modelling time-inconsistent preferences. *European Economic Review*, 44, 1116–1124.

Casari, M. (2009): Pre-commitment and flexibility in a time-decision experiment. *Journal of Risk and Uncertainty*, 38, 117–141.

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

Corgnet, B., Hernán-González, R., Schniter, E. (2015). Why real leisure really matters: Incentive effects on real effort in the laboratory. *Experimental Economics*, 18, 2, 284-301.

Dekel, E., B. L. Lipman, and A. Rustichini (2009): Temptation-Driven Preferences. *Review of Economic Studies*, 76(3), 937–971.

DellaVigna, S. and U. Malmendier (2006): Paying not to go to the gym. *American Economic Review*, 96, 694–719.

Dewitte, S., M. Pandelaere, B. Briers and L. Warlop (2005): Cognitive load has negative after effects on consumer decision-making. Working paper, Katholieke Universiteit Leuven.

Elster, J. (2000): *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge University Press, New York.

Fudenberg, D. and D. K. Levine (2006): A dual-self model of impulse control. *American Economic Review*, 96, 1449–1476.

Fudenberg, D. and D.K. Levine (2011): Risk, Delay, and Convex Self-Control Costs. *American Economic Journal: Microeconomics*, 3, 34-68.

Fudenberg, D. and D. K. Levine (2012): Timing and self-control. *Econometrica*, 80, 1–42.

Gine, X., D. Karlan, and J. Zinman (2010): Put your money where your butt is: A commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4), 213–235.

Gul, F. and W. Pesendorfer (2001): Temptation and self-control. *Econometrica*, 69, 1403–1435.

Gul, F. and W. Pesendorfer (2004): Self-control and the theory of consumption. *Econometrica*, 72, 119–158.

Gul, F. and W. Pesendorfer (2005): The revealed preference theory of changing tastes. *Review of Economic Studies*, 72, 429–448.

Heidhues, P. and B. Köszegi (2009): Futile attempts at self-control. *Journal of the European Economic Association*, 7(2–3), 423–434.

Hinson, J., T. Jameson, and P. Whitney (2003): Impulsive decision-making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 298–306.

Hoch, S. J. and G. F. Loewenstein (1991): Time-inconsistent preferences and consumer self-control. *Journal of Consumer Research*, 17, 492–507.

Houser, D., M. Keane and K. McCabe (2004). Behavior in a dynamic decision problem: An analysis of experimental evidence using a Bayesian type classification algorithm. *Econometrica*. 72, 781-822.

Houser, D., D. H. Reiley, and M. B. Urbancic (2008): Checking out temptation: A natural experiment with purchases at the grocery register. Unpublished manuscript, University of California, Berkeley, University of Arizona, and George Mason University.

Houser, D., D. Schunk, J. Winter and E. Xiao (2010): Temptation and Commitment in the Laboratory. University of Zurich, Institute for Empirical Research in Economics. Working Paper No. 488.

Houser, D., S. Vetter, and J. Winter (2012): Fairness and Cheating, European Economic Review 56 (8), 645-1655.

Kaur, S., M. Kremer, and S. Mullainathan (2010): Self-control and the development of work arrangements. *American Economic Review*, 100(2), 624–628.

Loewenstein, G. and D. Prelec (1992): Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics*, 107, 573–597.

Laibson, D. (1997): Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112, 443–478.

Metcalfe, J. and W. Mischel (1999): A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, 106, 3–19.

Miao, J. (2008): Option exercise with temptation. *Economic Theory*, 34, 473–501.

Mischel, W., Y. Shoda, and M. Rodriguez (1989): Delay of gratification in children. *Science*, 244, 933–938.

Noor, J. (2007): Commitment and self-control. *Journal of Economic Theory*, 135, 1–34.

Noor, J. (2011): Temptation and revealed preferences. *Econometrica*, 601-644.

O'Donoghue, T. and M. Rabin (1999): Doing it now or later. *American Economic Review*, 89, 103–124.

O'Donoghue, T. and M. Rabin (2000): The economics of immediate gratification. *Journal of Behavioral Decision-Making*, 13, 233–250.

Ozdenoren, E., S. Salant, and D. Silverman (2012): Willpower and the optimal control of visceral urges. *Journal of the European Economic Association*, 10, 342–368.

Phelps, E.S. and R. A. Pollack (1968). On second-best national saving and game-equilibrium growth. *Review of Economic Studies*. 35, 185-199.

Read, D., G. F. Loewenstein, and S. Kalyanaraman. (1999): Mixing virtue and vice: The combined effects of hyperbolic discounting and diversification. *Journal of Behavioral Decision-Making*, 12, 257–273.

Schelling, T. C. (1992): Addictive drugs: The cigarette experience. *Science*, 255, 430–434.

Shiv, B. and A. Fedorikhin (1999): Heart and mind in conflict: The interplay of affect and cognition in consumer decision-making. *Journal of Consumer Research*, 26, 278–292.

Strotz, R. H. (1955-1956): Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23, 165–180.

Thaler, R. H. and H. M. Shefrin (1981): An economic theory of self-control. *Journal of Political Economy*, 89, 392–406.

Vohs, K. D. and R. J. Faber (2007): Spent resources: Self-regulatory resource availability affects impulse buying. *Journal of Consumer Research*, 33, 537–547.

Wang, J., Y. Rao and D. Houser (2017): An experimental analysis of acquired impulse control among adult humans intolerant to alcohol, *Proceedings of the National Academy of Sciences of the U.S.A.*, 114 (6), 1299-1304.

Wertenbroch, K. (1998): Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Science*, 17, 317–337.

Table 1: Phases of the experiment and design parameters characterizing each phase.

Phase	Duration	Number of counting tasks	Number of temptation screens	Commitment cost [in $]	Final payoff if surfing [in $]	Additional payoff for continuing to count to end of experiment [in $]
0	30 min	15	0			
1	45 min	12	6	P	W_{L1}	$R_1 = 15 - W_{L1}$
2	45 min	12	6	P	W_{L2}	$R_2 = 15 - W_{L2}$

Note: P, W_{L1} and W_{L2} are design parameters whose values are randomly assigned to the 12 experimental conditions. As described in the model in Section 4, P is the cost of commitment, and the values of W_{L1} and W_{L2} are the "low" payoff values associated with choosing to surf in Phase 1 or Phase 2, respectively. In all experiment conditions, the payoff for counting to the end of the experiment is W_H=$15.

Table 2: Experimental design: values of design parameters by experimental condition

Experimental condition G	Commitment cost (P)	If completing phase 0 but surfing in phase 1		If completing phase 0 and 1 but surfing in phase 2		N
		Final payoff (W_{L1})	Foregone additional payoff (R_1)	Final payoff (W_{L2})	Foregone additional payoff (R_2)	
1	$ 0	$ 10	$ 5	$ 10	$ 5	8
2	$ 0	$ 10	$ 5	$ 15	$ 0	8
3	$ 0	$ 8	$ 7	$ 12	$ 3	8
4	$ 1	$ 10	$ 5	$ 10	$ 5	10
5	$ 1	$ 10	$ 5	$ 15	$ 0	9
6	$ 1	$ 8	$ 7	$ 12	$ 3	9
7	$ 0	$ 8	$ 7	$ 10	$ 5	12
8	$ 0	$ 8	$ 7	$ 15	$ 0	9
9	$ 0	$ 10	$ 5	$ 12	$ 3	11
10	$ 1	$ 8	$ 7	$ 10	$ 5	7
11	$ 1	$ 8	$ 7	$ 15	$ 0	9
12	$ 1	$ 10	$ 5	$ 12	$ 3	8

Note: The total number of subjects is 108. The values of W_{L1} and W_{L2} are the "low" payoff values associated with choosing to surf in Phase 1 and Phase 2, respectively, as described in the model in Section 4. The "high" payoff W_H is always $15.

Table 3: Descriptive statistics on subjects' characteristics

Variable	Value	N	%
Age	18–20	40	37.0
	21–29	68	63.0
Sex	Male	62	57.4
	Female	46	42.6
Major subject	Economics or business	18	16.7
	Natural sciences or mathematics	40	37.0
	Social sciences	21	19.4
	Other	29	28.9

Note: The total number of subjects is 108.

**Table 4:** Frequency of behaviors, overall and by values of treatment variables

				Commitment cost (P)				Additional payoff in phase 1 (R_1) for counting to end of experiment				Additional payoff in phase 2 (R_2) for counting to end of experiment					
				$0		$1		$5		$7		$0		$3		$5	
Behavior		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1	Committed at first temptation screen	22	20.4	16	28.6	6	11.5	10	18.5	12	22.2						
2	Surfed at first temptation screen	15	13.9	5	8.9	10	19.2	10	18.5	5	9.3						
3	Committed at later screen in Phase 1	6	5.6	5	8.9	1	1.9	2	3.7	4	7.4						
4	Surfed at later screen in Phase 1	1	0.9	1	1.8	0	0	0	0	1	1.8						
5	Committed in Phase 2	4	3.7	4	7.1	0	0	3	5.6	1	1.8	1	3.4	0	0	3	20.0
6	Surfed in Phase 2	29	26.9	14	25.0	15	28.9	15	27.8	14	25.9	25	86.2	4	20.0	0	0
7	Never committed and never surfed (so counted) at all screens	31	28.7	11	19.6	20	38.5	14	25.9	17	31.5	3	10.3	16	80.0	12	80.0
Column total		108		56		52		54		54		29		20		15	
Test for independence of columns				$\chi^2(6) = 16.4, p = 0.012$				$\chi^2(6) = 4.8, p = 0.564$				$\chi^2(4) = 42.4, p < 0.000$					

Table 5: Probit regressions of the decision to surf or commit

|	Surfing 1st screen (1)	Surfing screens 2-6 (2)	Surfing at end (3)	Committed 1st screen (4)	Committed screen 2-6 (5)	Committed at end (6)
Commitment cost ($P=1$)	0.129*	0.0888	0.0544	-0.151*	-0.205**	-0.291***
	(0.0659)	(0.0734)	(0.114)	(0.0771)	(0.0831)	(0.0832)
Additional payoff for counting to end (R_1)	-0.0290	-0.0242	-0.0727	0.0193	0.0272	-0.00403
	(0.0305)	(0.0330)	(0.0577)	(0.0365)	(0.0409)	(0.0415)
Additional payoff for counting to end (R_2)			-0.148***			0.0792***
			(0.0285)			(0.0221)
Female	-0.133**	-0.112*	-0.0178	-0.0374	0.0303	0.0163
	(0.0570)	(0.0635)	(0.111)	(0.0793)	(0.0912)	(0.0931)
Age: older than 20	0.0251	-0.00814	-0.0481	0.0679	0.0678	0.0298
	(0.0593)	(0.0688)	(0.116)	(0.0800)	(0.0933)	(0.0973)
Natural sciences	-0.0769	-0.0846	-0.248*	-0.00137	0.127	0.127
	(0.0653)	(0.0753)	(0.136)	(0.100)	(0.116)	(0.125)
Economics or business	-0.0619	-0.0714	-0.164	0.172	0.235	0.139
	(0.0605)	(0.0690)	(0.143)	(0.148)	(0.165)	(0.172)
Social sciences	-0.0615	-0.0172	0.123	0.0534	0.0654	0.0565
	(0.0618)	(0.0878)	(0.158)	(0.126)	(0.143)	(0.150)
Log Likelihood	-38.13	-41.77	-53.09	-50.58	-55.61	-52.40
Pseudo R^2	0.12	0.08	0.27	0.07	0.10	0.20
Observations	108	108	108	108	108	108

Notes: Coefficients are expressed as marginal effects; for continuous covariates, they are evaluated at their means, for binary covariates, they correspond to a discrete change from 0 to 1. The symbol * denotes statistical significance at the 10% level; ** at the 5% level; and *** at the 1% level.

Figure 1: Shares of commitment and surfing in Phase 1, by commitment cost and additional payoff for counting to end.
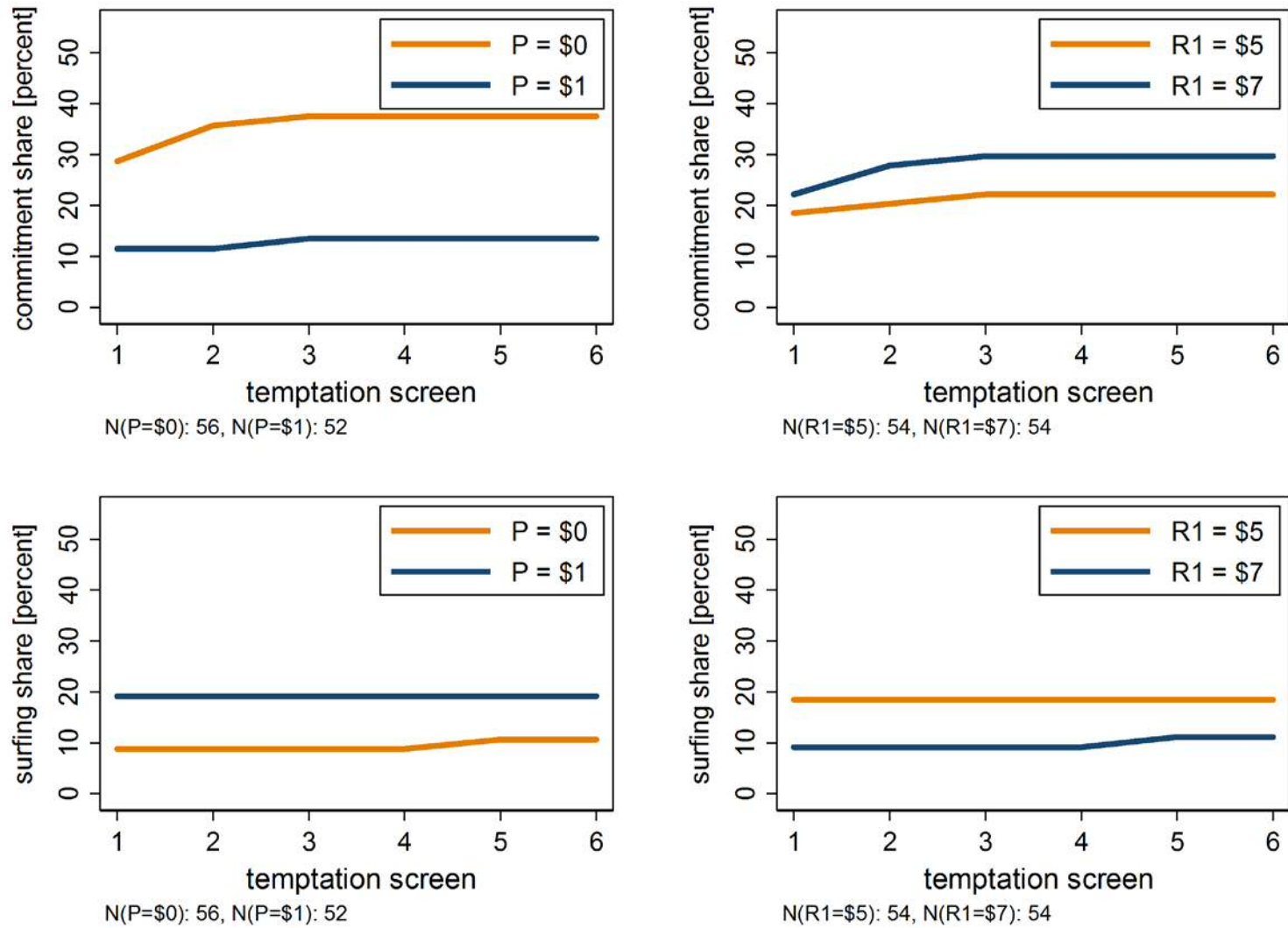
Legend: P = Commitment costs [in $]; R1 = *Additional* payoff gained from counting to the end of the experiment [in $].

Figure 2: Hazard rates of commitment and surfing in Phase 1, by commitment cost and additional payoff for counting to end.
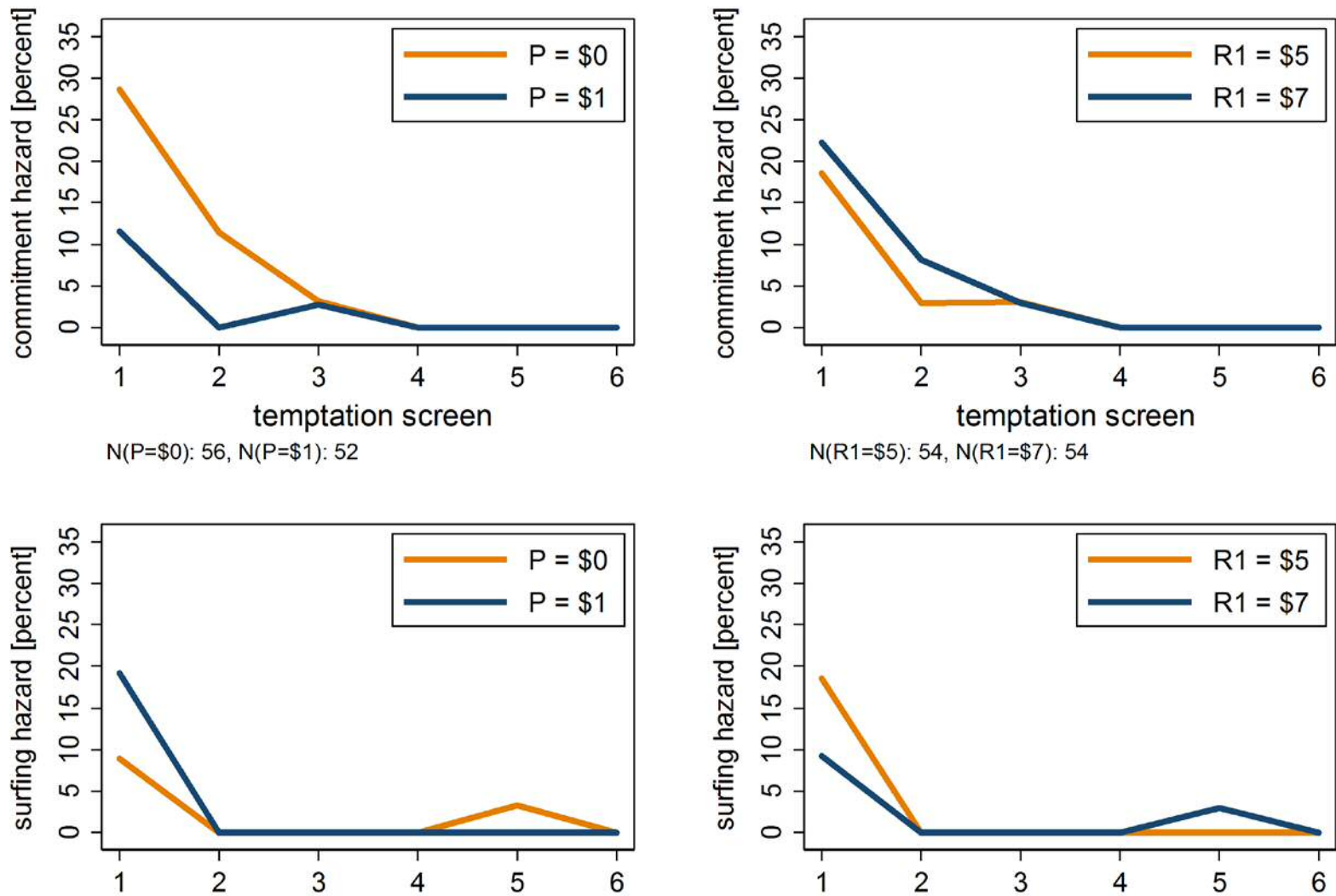
Legend: P = Commitment costs [in $]; R1 = *Additional* payoff gained from counting to the end of the experiment [in $].

Appendix 1: Proof that both increasing cost of self-control and impatience are necessary conditions for delay in commitment

The example presented in the body of the text can be generalized as follows. There are 3 periods, the normative discount factor is $\delta \leq 1$. The short-run self always prefers immediate consumption. The return to consuming in the third period is $B > 0$. The cost of self-control in period 1 is $a_1 \geq 0$, and assuming it is used in period 1 then in period 2 the cost of self-control is $a_2 > a_1$. The cost of commitment in any period is $c > 0$. The utility benefit of consumption in period 1 is $b_1 > 0$, and in period 2 it is $b_2 > b_1$. Thus, commitment in period 1 gives utility $\delta^2 B - c$, and planning in period 1 to commit in period 2 gives utility $\delta^2 B - a_1 - \delta c$.

It follows that, in period 2, the agent commits if $\delta B - c > \delta B - a_2$, and $\delta B - c > b_2$. Thus, the following four conditions are necessary and sufficient for the agent in period 1 to strictly prefer to plan to commit in period 2, and then strictly prefer to follow this plan and commit in period 2.

$$\delta^2 B - a_1 - \delta c > b_1 \qquad \text{(A1)}$$
$$\delta^2 B - a_1 - \delta c > \delta^2 B - c \qquad \text{(A2)}$$
$$a_2 > c \qquad \text{(A3)}$$
$$\delta B - c > b_2. \qquad \text{(A4)}$$

From (A2), we have that $(1-\delta)c > a_1$ which in turn necessitates $\delta < 1$. Note also that, from combining (A2) and (A3), it is also necessary that $a_2 > \dfrac{a_1}{1-\delta}$. The interpretation is that in order for delayed commitment to be strictly preferred it is necessary that the cost of self-control should increase sufficiently quickly, and the agent must be normatively impatient.

Appendix 2: Instructions provided at the beginning of the experiment

Thank you for coming. You have already earned a show-up bonus of $5 for arriving on time. These instructions explain how you can earn more money during the experiment.

Today's experiment involves a counting task. From time to time you will see a screen which we call the "*counting screen*". This screen displays nine digits, either zeros or ones. Your task is to count the number of ones, and report that number in a box provided. You will have 15 seconds to provide an answer. Not providing an answer, or providing an incorrect answer, is counted as a mistake. If you make mistakes on more than 30% of the counting screens that are presented to you during the counting task, you will earn $3.

If your share of mistakes remains below 30%, then your earnings depend on how much time you spend on the counting task: **[P1]** if you spend less than 75 minutes on the counting tasks; **[P2]** if you spend more than 75 minutes but less than 120 minutes on the counting tasks; or **$15** if you spend 120 minutes on the counting tasks. These payments are in addition to your show-up bonus.

In addition to the *counting screen*, another kind of screen is displayed up to 12 times during the experiment. This "*choice screen"* will appear for the first time after about 33 minutes, and it may appear again around minutes 41, 51, 58, 64, 69, 78, 83, 89, 97, 109, and 114 of the experiment, depending on which options you choose. Whenever the *choice screen* appears, it will present you with the following three options:

1. You can end the counting tasks. For the remaining time of the experiment, you can surf the internet on the computer here in the laboratory room. If you choose

this option, the *choice screen* will <u>not</u> appear again during the experiment. You select this option by not pressing any of the buttons on the *choice screen*.

2. You can decide to continue the counting tasks. If you choose this option, you may see the *choice screen* again, as indicated by the timeline described above. You select this option by pressing the corresponding button on the *choice screen*.

3. You can decide to continue the counting tasks with the following change: in contrast to option 2, the *choice screen* will <u>never</u> appear again for the rest of the experiment. Choosing this option reduces your earnings by **[C]**. You select this option by pressing the corresponding button on the *choice screen*. Choosing this option means that the counting task will continue until the end of the experiment.

This experiment will end at different times for different participants. Please do not leave the room, talk or otherwise distract other participants in any way until you are told that all participants have completed the experiment and you have left the laboratory.
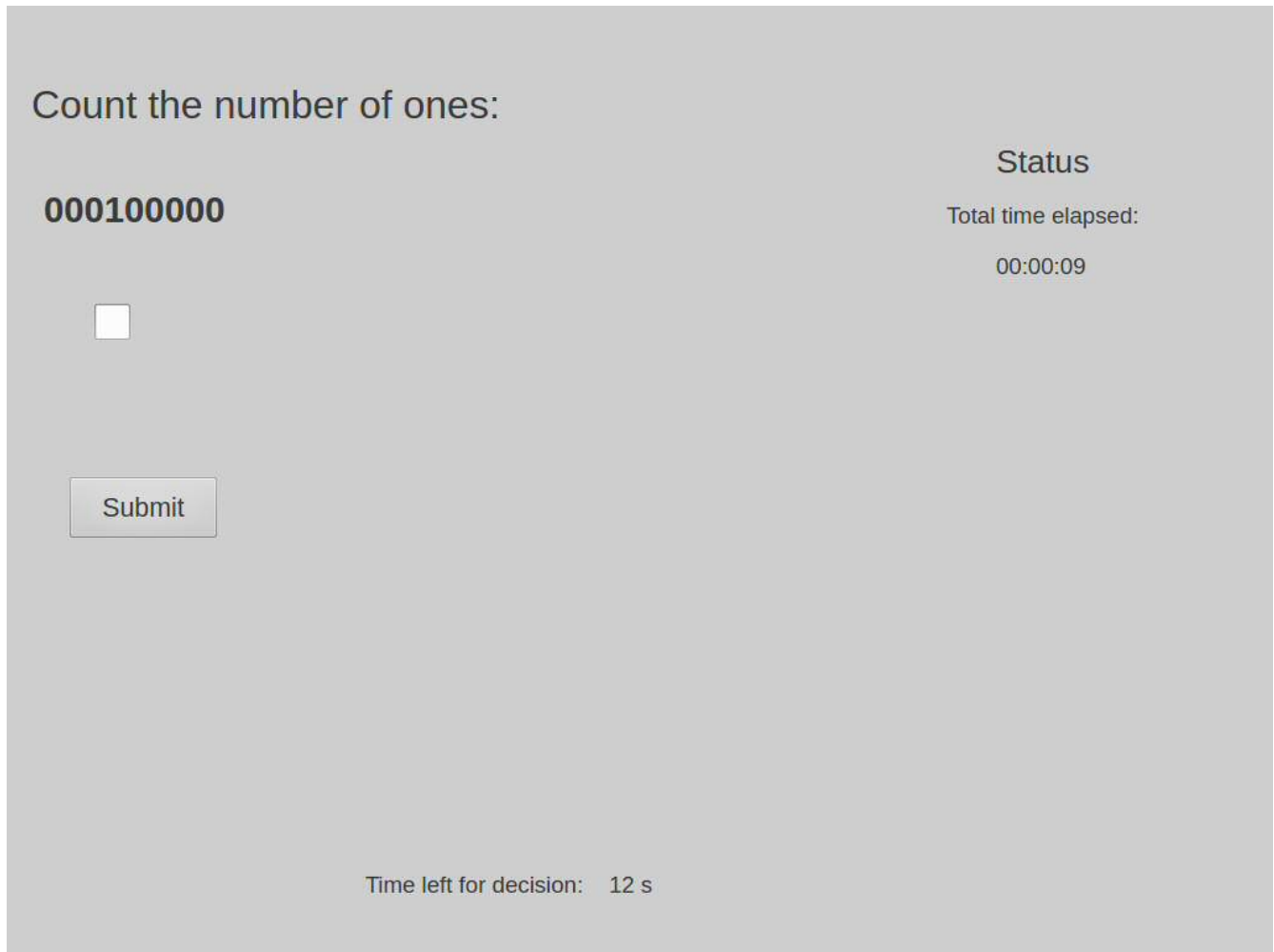
Figure A.1: Counting screen

Count the number of ones:

000100000

Status

Total time elapsed:

00:00:09

☐

Submit

Time left for decision: 12 s

Figure A.2: Empty screen

Status

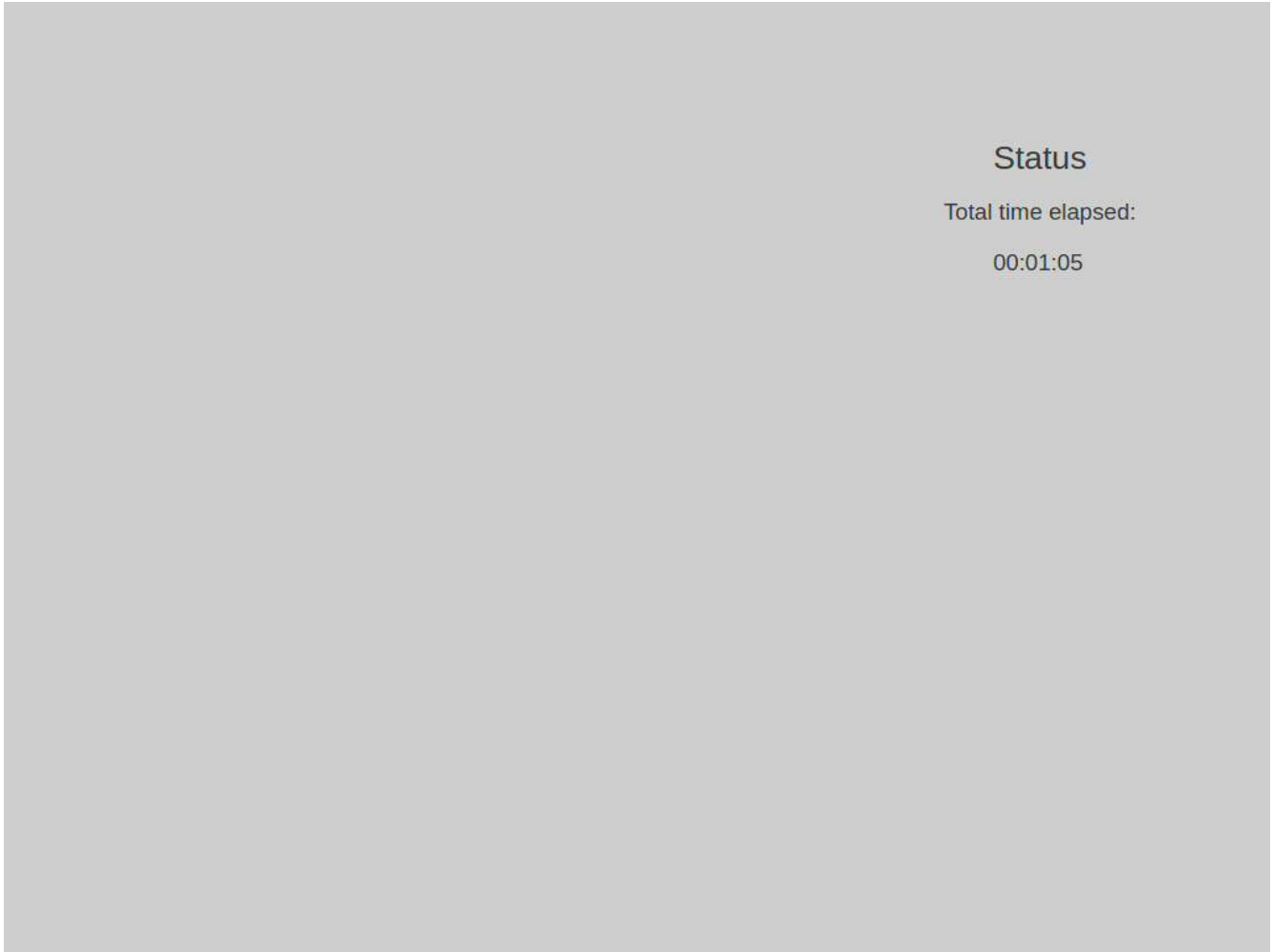Total time elapsed:

00:01:05

Figure A.3: Temptation screen

Thank you for participating in today's experiment. If you have answered at least 70% of the counting tasks correctly, your earnings are $10.

Status

You will now be given access to the internet, so that you can pass the time until the experiment ends for all participants. If you like, you can also continue with the counting experiment. If you continue with the counting experiment you can earn up to an additional $5. You will be given access to the internet unless you press one of the two buttons below.

Total time elapsed:

00:41:40

| Continue Counting |

Click here if you want to continue counting without any more opportunities to access the internet. There is a $1 charge for clicking this button. You will continue counting until the experiment ends. You can earn up to $5 in addition to your current earnings.

| Continue Counting and Remove Internet-Option |

Time left for decision: 96 s