Séverine Toussaert

Eliciting temptation and self-control through menu choices: a lab experiment

Article (Accepted version) (Refereed)

Original citation:
Toussaert, Séverine (2018) *Eliciting temptation and self-control through menu choices: a lab experiment*. Econometrica, 86 (3). pp. 859-889. ISSN 0012-9682
DOI: 10.3982/ECTA14172

Eliciting temptation and self-control through menu choices: a lab experiment[*]

Séverine Toussaert[†]

February 22, 2018

Abstract

Unlike present-biased individuals, agents who suffer self-control costs as in Gul and Pesendorfer (2001) may choose to restrict their choice set even when they expect to resist temptation. To identify these self-control types, I design an experiment in which the temptation was to read a story during a tedious task. The identification strategy relies on a two-step procedure. First, I measure commitment demand by eliciting subjects' preferences over menus that did or did not allow access to the story. I then implement preferences using a random mechanism, allowing to observe subjects who faced the choice yet preferred commitment. A quarter to a third of subjects can be classified as self-control types according to their menu preferences. When confronted with the choice, virtually all of them behaved as they anticipated and resisted temptation. These findings suggest that policies restricting the availability of tempting options could have larger welfare benefits than predicted by standard models of present bias.

JEL classification: C91, D03, D83, D99

Keywords: temptation; self-control; menu choice; curiosity; experiment

[*] I would like to thank my advisors David Cesarini and Guillaume Fréchette for their continuous support in the development of this project, Anwar Ruff for teaching me how to program experiments, and Margaret Samahita for her assistance during the experimental sessions. I would also like to thank three anonymous referees and the co-editor, as well as Marina Agranov, Ned Augenblick, Mark Dean, Kfir Eliaz, Judd Kessler, Yves Le Yaouanq, Barton Lipman, Daniel Martin, Pietro Ortoleva, Emanuel Vespa, and Sevgi Yüksel for their detailed comments on the manuscript. Finally, I wish to thank B. Douglas Bernheim, Andrew Caplin, Micael Castanheira, Andrew Ellis, Erik Eyster, Xavier Gabaix, Yoram Halevy, David Laibson, Efe Ok, Matthew Rabin, Giorgia Romagnoli, Ariel Rubinstein, Tobias Salz, Rani Spiegler, Roberto Weber, and other seminar participants at U. of Amsterdam, the CESifo Area Conference on Behavioral Economics 2017, City University, ESA Dallas, Harvard, LMU, LSE, Melbourne, Michigan, Nova, NYU, Oxford, Stanford SITE Workshop on Psychology and Economics 2016, Sussex, Sydney, UCL, UCSB, UNSW, UTS, Vienna, Warwick, and Zürich for their useful comments. All remaining errors are mine.

[†] London School of Economics, Department of Psychological and Behavioural Science. Contact: S.Toussaert@lse.ac.uk.

1. Introduction

Models of dynamically inconsistent time preferences (Strotz (1956), Laibson (1997), O'Donoghue and Rabin (1999)) are by far the most popular framework in the literature on self-control problems. A central implication of those models is that present-biased agents may demand commitment devices to constrain the choices of their future selves. As an alternative approach, models of menu-dependent preferences à la Gul and Pesendorfer (2001) (henceforth GP 2001) generate commitment demand by modeling agents whose preferences not only depend on actual consumption, but also on the most tempting alternative in the choice set.[1] One key distinction between these two classes of models pertains to the motives that drive a decision maker to restrict his choice set. Whereas a present-biased agent will choose to eliminate a temptation from his choice set only if he expects to succumb to it, an agent with menu-dependent preferences may value commitment even if he expects to resist temptation, because commitment eliminates the cost of exerting self-control. The present paper takes a first step to quantify the importance of these "self-control types" who may prefer to remove a temptation from their choice set, despite expecting not to succumb to it.

Assessing the prevalence of self-control types is important from a policy perspective: if unchosen alternatives affect utility, the welfare benefits of policies that restrict access to temptations could be much larger than what the usual calculations would suggest.[2] To see why, consider the welfare implications of introducing smoking bans in public spaces. Both of the above classes of models predict that a ban would benefit current smokers who are trying to quit; what the second class of models further suggests is that a ban could also increase the welfare of *former* smokers by alleviating the self-control costs of remaining smoke-free.[3] To evaluate the welfare benefits of a smoking ban, one could in principle elicit each individual's willingness to pay to implement such policy, and then aggregate values across all individuals. However, in practice, various limitations of such ex-ante valuations - including hypothetical bias, individual budget constraints, and a lack of sophistication of respondents - often constrain policy appraisers to instead perform calculations

[1] Since GP 2001, several axiomatic models of menu choice have extended and/or relaxed the original framework, with some variations on the set of primitives and axioms. A few examples include Dekel et al. (2009), Noor and Takeoka (2010), Noor (2011), or Kopylov (2012); see Lipman and Pesendorfer (2013) for a general review. In the class of models of menu-dependent preferences, one can also include the dual-self framework of Fudenberg and Levine (2006, 2012), which presents close connections with GP 2001 and further extensions.

[2] Besides lower self-control costs, other benefits of smaller choice sets include less choice overload and minimal regret; see *The Paradox of Choice* by B. Schwartz for a general discussion of why more can end up being less.

[3] In addition, a ban would likely decrease the expected costs of resuming a former smoking habit; this scenario would be particularly likely for recent quitters, who face a higher probability of relapse.

based on observable behaviors (e.g., number of failed quit attempts × health and financial costs).[4] One major downside of this ex-post approach is that if agents suffer non-consequentialist costs from resisting temptation (e.g., if relapses are prevented by exerting self-control), the welfare benefits of smoking bans will be substantially underestimated. Furthermore, ignoring self-control costs may not only bias our estimate of the *effect size* of a given policy, but also our assessment of the *type* of policy tools likely to be most effective. If self-control is high enough such that tempted agents rarely succumb to temptation, then price policies such as proportional taxes or subsidies will be ineffective, for their aim is to alter consumption behavior. On the other hand, policies that impose a cap on consumption of the tempting good may improve welfare even for those whose consumption would be below the cap in the absence of restrictions.[5]

While the above discussion illustrates the importance of measuring self-control costs, it also hightlights the empirical challenge pertaining to the identification of the population incurring those costs: to identify self-control types, one not only needs to observe whether they would prefer to restrict their choices, but also what they would do in a counterfactual world in which no form of commitment is available. However, with naturally occurring data, we rarely observe individuals having a preference for a restricted choice set A and yet receiving a larger choice set B. To tackle this empirical challenge, I design and implement an experimental method that tests for the prevalence of self-control types and implement it in a laboratory setting.

In the experiment, the potential temptation was to forego additional earnings to read a sensational story during a tedious attention task for which subjects received payment. I adopt a two-step procedure to identify subjects who suffer from self-control costs. First, using an incentive-compatible mechanism, I elicit subjects' preference ordering over a set of menus that either did or did not allow access to the story during the task, and classify subjects into types according to their menu preferences. A *self-control type* is a subject who would strictly prefer to (*i*) remove the temptation from his choice set instead of facing the choice, and (*ii*) face the choice instead of receiving the tempting option for sure, because he expects to resist it. Second, I implement subjects' preferences

[4] Ex-ante valuations of intangibles such as health or environmental benefits typically rely on stated preference methods (also called contingent valuation) to elicit willingness to pay (WTP) for that benefit; these methods are unincentivized and suffer from a number of biases (Diamond and Hausman (1994)). Furthermore, WTP measures will fail to capture the true benefits of a policy if (*i*) those benefits exceed what the respondent can afford, and (*ii*) respondents wrongly perceive the true returns to the policy (e.g., because they underestimate their self-control problems).

[5] For a more extensive discussion of the implications for policy design, see Gul and Pesendorfer (2007), Krusell et al. (2009, 2010), and Online Appendix Section E.3. For instance, Krusell et al. (2009, 2010) show in a dynamic general equilibrium model that proportional subsidies on investment are an effective policy instrument only if self-control is low and agents usually succumb to the temptation to overconsume, as is the case of present-biased agents.

using a random implementation rule. This mechanism allows me to observe the behavior of subjects who faced the choice yet preferred commitment, and to contrast *perceived* self-control with *actual* self-control. Finally, I use two types of auxiliary data to further refine the interpretation of menu preference orderings and subsequent choices from the flexible menu. First, I measure subjects' beliefs about their anticipated choice in the absence of commitment to test whether those classified as self-control types indeed expected to resist temptation. Second, I contrast the task performance of subjects who faced the choice with those who did not, in order to study whether those confronted with the choice incurred self-control costs in the form of a productivity loss.

Depending on how conservative one wants to be, I find that 23% to 36% of subjects can be classified as self-control types according to their menu preferences. This preference pattern is by far the most common one among those who preferred to eliminate their access to the story. By contrast, only 2.5% of subjects exhibit commitment preferences consistent with standard models of dynamic inconsistency. In line with theories of costly self-control, virtually all subjects classified as self-control types predicted they would resist the temptation to read the story in the absence of commitment. Finally, *perceived* self-control, as measured by subjects' menu preferences and anticipated choices, almost entirely coincides with *actual* self-control: when confronted with the choice, only one subject with self-control preferences decided to read the story; by contrast, over 20% of the other subjects did so. At the same time, task performance in the full sample was lower in the absence of commitment, which provides suggestive evidence that resisting temptation opportunities might have entailed a self-control cost.

The idea that exerting self-control entails a cost is of course not new; in fact, it speaks to a long literature in psychology, which proposes that self-control is a limited resource that can be exhausted after repeated efforts to resist temptation (Baumeister et al. (1994), Baumeister and Vohs (2003)). The paper is also connected to a vast literature in economics that explores the link between self-control problems and commitment demand, both in laboratory experiments (Houser et al. (2018), Augenblick et al. (2015)) and in field settings (Ashraf et al. (2006), Kaur et al. (2015), John (2015), Sadoff et al. (2015)). Finally, the paper contributes to a burgeoning literature studying commitment and flexibility through menu choice. Dean and McNeill (2015) explore the relationship between preference uncertainty and preference for larger choice sets by linking preferences over menus of work contracts to subsequent choices of contracts; they find no evidence of a preference for commitment in their setting. In the context of a weight-loss challenge, Toussaert (2016) studies participants' preferences over lunch reimbursement options differing in their food coverage, and finds a strong demand for eliminating unhealthy foods from the coverage; however, the actual food selections were not observed.

The paper is organized as follows. Section 2 introduces the theoretical framework used to construct the dataset. Section 3 outlines the experimental design and Section 4 presents the results. Section 5 concludes with a summary and discussion of the main findings. Additional results are reported in the Appendix at the end of this paper as well as in a detailed Online Appendix (OA).

2. Temptation and self-control through menu choices

The analysis of this paper is grounded in the theory of menu choice originally introduced by Gul and Pesendorfer (2001) to study costly self-control. This section describes how temptation and self-control are elicited in this framework, explains key distinctions and connections with other models of temptation and discusses the restrictions imposed by the theory on choice behavior.

2.1 Costly self-control in GP 2001

GP 2001 consider a two-period expected utility model, $t \in \{1, 2\}$. Their primitive is a preference relation $\succeq_1$ defined on a set $\mathcal{M}$ of menus (of lotteries). In Period 1, a decision maker (DM) chooses among menus according to $\succeq_1$, with the interpretation that in Period 2, he will make a choice from the selected menu according to $\succeq_2$. In addition to the usual assumptions,[6] GP 2001 impose a new behavioral axiom on $\succeq_1$ called *Set Betweenness*, which states that for any two menus A and B,

$$A \succeq_1 B \text{ implies } A \succeq_1 A \cup B \succeq_1 B$$

This axiom allows to capture behaviorally the notions of temptation and self-control. To see how, consider a simple choice situation with two options a (for apple) and b (for brownie) and assume the ex-ante preferences of the DM are such that $\{a\} \succ_1 \{b\}$. A standard DM (STD) evaluates a menu by its best element(s) and is unaffected by the presence of dominated options, implying $\{a\} \sim_1 \{a, b\} \succ_1 \{b\}$. On the other hand, a DM who is tempted by the brownie would prefer to commit to a menu that excludes b than to face the choice between a and b in Period 2. In other

[6] $\succeq_1$ is required to be a weak order, which satisfies the standard expected utility axioms of continuity and independence adapted to a menu choice setting. These technical axioms are not tested in this paper and are treated as maintained assumptions; incentive-compatibility of the elicitation procedure for menu preferences requires some version of these assumptions. The requirement that $\succeq_1$ be a weak order is also assumed away in the experiment, because subjects are required to provide a full ranking (allowing for ties) of the alternatives, thus automatically satisfying completeness and transitivity. See Section 3.2 for more details about the elicitation procedure.

words, b is a *temptation* for a if $\{a\} \succ_1 \{a, b\}$. In this model, there are two reasons why a tempted DM may favor commitment to a. First, the DM may expect to give in to b if offered a choice from $\{a, b\}$, thus assigning the same value to $\{b\}$ and $\{a, b\}$. Alternatively, the DM may anticipate that he will resist b when facing $\{a, b\}$ by exerting self-control, which makes $\{a, b\}$ more valuable than $\{b\}$. In formal terms, say (i) b is an *overwhelming temptation* if $\{a\} \succ_1 \{a, b\} \sim_1 \{b\}$, and (ii) b is a *resistible temptation* if $\{a\} \succ_1 \{a, b\} \succ_1 \{b\}$. In the experiment, a DM with the menu preferences $\{a\} \succ_1 \{a, b\} \succ_1 \{b\}$ will be called *self-control type.* GP 2001 show that under their axioms, $\succeq_1$ admits the following self-control representation:

$$V_{GP}(A) := \max\nolimits_{x \in A} [u(x) + v(x)] - \max\nolimits_{y \in A} v(y)$$

The *commitment* utility u measures utility in the absence of temptation, that is, when committed to a singleton choice. The *temptation* utility v measures the temptation value of an alternative and $\max_{y \in A} v(y) - v(x)$ is the self-control cost of choosing x over the most tempting alternative in A.[7] In Period 2, the DM chooses as if he maximized the compromise utility $u + v$.

2.2 Connections and differences with other theories

Models of menu-dependent preferences à la GP 2001 present several distinguishing features, which guide the identification of self-control types. First, commitment in this framework can be rationalized through two channels: either by the DM's belief that he will give in to temptation or because commitment eliminates the cost of exerting self-control. In contrast, standard models of dynamic inconsistency can only rationalize the case of overwhelming temptation, $\{a\} \succ_1 \{a, b\} \sim_1 \{b\}$.[8] The reason is that the preferences of a present-biased agent only depend on final consumption and not on the specific set from which consumption is taken; as a result, commitment can only be valuable if the agent expects to deviate from the ex-ante optimal consumption path. As such, models of present bias can be understood as a limit case of the GP model when the self-control cost becomes arbitrarily large, so that the agent never exercises self-control.[9]

[7] To see why u can be interpreted as a commitment utility, let $A = \{a\}$ and notice that $V_{GP}(A) = u(a)$. To see why v measures temptation, notice that if $u(a) > u(b)$ and $v(b) > v(a)$, then $V_{GP}(\{a\}) > V_{GP}(\{a, b\})$; that is, the agent is tempted by b.

[8] By standard, I mean models that assume a fixed present bias parameter and degenerate beliefs about the size of this bias, the most common assumptions in this literature.

[9] GP 2001 show that the limit case in which the agent never exercises self-control can be obtained in their framework by relaxing continuity; in this case, the DM's preferences have a Strotz representation $V_S(A) := \max_{x \in A} u(x)$ subject to $v(x) \geq v(y)$ for all $y \in A$. In words, the DM chooses in Period 2 as if he lexicographically

Second, although observing the preference ordering $\{a\} \succ_1 \{a,b\} \succ_1 \{b\}$ is generally enough to distinguish costly self-control from dynamic inconsistency in a deterministic world, this is no longer true if Period 2 choice is allowed to be stochastic. To see this point, suppose the DM is uncertain about his future temptation: with probability p, he expects to succumb to temptation and select b, while with probability $(1-p)$, he believes he will face no temptation and choose a. For such a DM, the preference ordering $\{a\} \succ_1 \{a,b\} \succ_1 \{b\}$ does not reflect costly self-control; rather, it is explained by a probability $p \in (0,1)$ of indulgent behavior.[10] Therefore, to be able to distinguish between these two interpretations (costly self-control vs. random indulgence), enriching the dataset to include expectations about Period 2 choice from $\{a,b\}$ is necessary: only a DM who suffers from random indulgence will expect to give in with positive probability.

Third, theories of costly self-control à la GP 2001 typically model a sophisticated agent who correctly anticipates the choice he will make in Period 2 from the selected menu and chooses a menu in Period 1 accordingly.[11] Formally, say that a DM is *sophisticated* if $A \cup \{x\} \succ_1 A$ implies $x \succ_2 y$ for all $y \in A$. In other words, if a DM values the addition of an alternative x to menu A, it must be because he correctly anticipates that he will choose x over any element of A in Period 2. It can be shown that sophistication is a necessary condition for $\succeq_2$ to comply with the interpretation of $\succeq_1$ provided in 2.1, that is, for $\succeq_2$ to be represented by the utility $u + v$ (Kopylov (2012), Thm 2.2). As a consequence, the GP model cannot capture the behavior of a (partially) naive agent for whom $\{a\} \succeq_1 \{a,b\} \succ_1 \{b\}$ and yet $b \succ_2 a$. In the experiment, it will be useful to distinguish *perceived* self-control (identified by $\{a\} \succ_1 \{a,b\} \succ_1 \{b\}$) and *actual* self-control (identified by $\{a\} \succ_1 \{a,b\} \succ_1 \{b\}$ *and* $a \succ_2 b$). This will be done by first eliciting subjects' menu preferences and then contrasting these preferences with the actual choices made from the flexible menu.

Finally, Set Betweenness imposes several restrictions on choice behavior, which preclude two interesting phenomena. First, a DM who satisfies this axiom can never express a strict preference for flexibility (i.e., $\{a,b\} \succ_1 \{a\},\{b\}$). As a result, the GP model cannot accommodate the fact that an agent who feels uncertain about his future tastes may want to keep his options open, an idea originally motivated by Kreps (1979). Second, Set Betweenness gives a special structure to the

maximized the temptation utility and then the commitment utility. Under specific functional-form assumptions, Krusell et al. (2010) show the GP model nests the multiple-selves model of Laibson (1997), which corresponds to the case in which their temptation-strength parameter γ - governing the cost of self-control - tends to infinity.

[10] This point has been formally addressed by Dekel and Lipman (2012), who show that any menu preference $\succeq_1$ that admits a (possibly random) GP representation also has a random Strotz representation (see previous footnote), where the temptation utility v is uncertain.

[11] One exception is Kopylov (2012) who considers a weakening of sophistication in order to model self-deception. Also see Ahn et al. (2017a,b) for behavioral definitions of naiveté in models of dynamic inconsistency and costly self-control.

form of temptation by excluding the possibility that $\{a\} \succ_1 \{b\} \succ_1 \{a, b\}$. Such a preference profile could be motivated by the agent's anticipated feeling of guilt if he chooses the tempting option b from $\{a, b\}$, whereas he could have acted virtuously by selecting a. This interpretation has been formalized by Kopylov (2012) who proposes a relaxation of the Set Betweenness axiom allowing to capture guilt. These preferences ($FLEX$, $GUILT$) will be incorporated in the taxonomy of types presented in the results section, the prevalence of which will be assessed against the one of self-control types.

3. Experimental design

The experiment was divided into two periods, followed by an exit survey. Period 1 comprised 5 sections (A-E) described below, pertaining to the elicitation of a temptation (Sections A & B), of menu preferences (Sections C & D), and of beliefs about choice in Period 2 (Section E). Details about the exit survey are provided at the end of this section, as well as a summary of the structure of the experiment (Fig.1); see OA-F for the instructions.

3.1 Description of the tempting good

The first part of the experiment was devoted to the elicitation of temptation. Generating temptation in the lab poses several challenges. First, one needs to find a good that is tempting to a majority of subjects, that is, a good that subjects think they should not consume and yet find enticing.[12] Second, the goods commonly considered in the literature, such as surfing the internet (Bonein and Denant-Boèmont (2015), Houser et al. (2018)) or watching an entertaining TV show (Bucciol et al. (2015)), can be easily consumed outside the lab, which reduces their immediate appeal. In this experiment, I exploit subjects' curiosity and, in particular, the human tendency to like gossiping and hearing gossip about others, which is present in virtually all human societies (Dunbar (2004)).

The potential temptation was to forfeit money to read a personal story from one subject in the room, while performing a tedious task. In Section A, subjects were asked to describe an incredible or strange life event they had personally experienced. As an aid, they were given three hypothetical examples. Subjects were given 10 minutes to write their story by hand on a form and place it back in a blank envelope. An assistant then collected the stories, went through them in a separate room, and came back with the story she found most entertaining (see OA-G.1 for the selected stories).

[12] For instance, note that for chocolate to qualify as a tempting good, the subject must (i) find chocolate appealing, *and* (ii) perceive that consuming chocolate is bad.

To stimulate subjects' curiosity, the experimenter opened the envelope with the selected story in front of them and expressed surprise while taking a look at the winning story. Finally, the assistant recorded the story in the system while the experimenter read the next part of the instructions. Subjects were told at the end of Section A that an envelope containing a secret code would be distributed in Period 2, allowing them to potentially display the story on their screen.

In Section B, subjects were introduced to the main task of Period 2. They were told that they would have to focus for a period of up to 60 minutes on a four-digit number updated on their screen every second.[13] At random times, they received a prompt to enter the last number they saw, and the number was reinitialized after every prompt (see screenshots in OA-F.3). All subjects received 5 prompts and could earn $2 per correct answer. After describing the task, subjects were told that two options could be potentially available in Period 2 depending on their choices in later sections:

Option 0: Do the task without reading the story and receive payment for all 5 prompts.

Option 1: Read the story during the task and receive payment for 4 randomly selected prompts.

The two options were referred to as "No Learning" (for 0) and "Learning" (for 1). Regardless of the option, subjects worked on the task for the same duration and received feedback about their performance and earnings only at the end of the experiment. To minimize communication opportunities after the experiment, subjects were told that they would be requested to leave the lab one at a time; furthermore, no student could a priori know who read the story in their session. As a result, it was difficult for subjects to satisfy their curiosity for this specific piece of information outside of the context of the experiment. At the end of Section B, subjects practiced with the task for two minutes and received feedback about their performance during that practice period.

3.2 Elicitation of menu preferences

To identify temptation and perceived self-control, Sections C & D elicited subjects' preferences over a set of three "menus," one of which was assigned to them at the start of Period 2:

Menu {0}: Eliminates the chance to read the story and pays for all 5 prompts; practically, the box where the secret code could be entered to access the story was removed from the subject's screen.

[13] During the first session, Period 2 was announced to last exactly 60 minutes; however, given the tediousness of the task and the overall length of the session, the duration of the task was reduced to 45 minutes. The other 5 sessions had the same task duration of 45 minutes with prompts occurring at the same time; the only difference was that subjects were told that the task could last "up to" 60 minutes. Since no major differences in behavior were observed relative to Session 1, all sessions are pooled in the data analysis. The econometric analysis systematically controls for session fixed effects.

Menu $\{\mathbf{1}\}$: Guarantees access to the story and pays for only 4 prompts; the story could be read at any time during the task but was automatically displayed at the end if not displayed before.

Menu $\{\mathbf{0},\mathbf{1}\}$: Offers the chance to decide during the task whether and when to read the story by entering the secret code.

To avoid strong word connotations, the three menus were called "Pre-Select No Learning," "Pre-Select Learning," and "Decide in Period 2". The elicitation of subjects' weak ordering $\succeq_1$ over the set $\mathcal{M} = \{\{0\},\{1\},\{0,1\}\}$ was performed in two steps (Sections C & D). In Section C, subjects were asked to assign a rank number 1, 2, or 3 to the three menus presented in a list.[14] To allow for the expression of indifferences, subjects could assign the same rank number to two or all three menus. Before providing their ranking, subjects were told they would be assigned a menu at the start of Period 2 based on the following procedure:

1. With probability 1/2, a subject received $\{0,1\}$ regardless of his ranking.

2. With probability 1/2, a subject's ranking was implemented stochastically such that the odds of receiving a given menu were increasing in its ranking, as displayed in the following table:

Ranking of (X,Y,Z)	% chance of being drawn $(\%_X,\%_Y,\%_Z)$
(1,2,3)	(50,30,20)
(1,1,2)	(40,40,20)
(1,2,2)	(50,25,25)
(1,1,1)	(33.3,33.3,33.3)

The above elicitation procedure has two important properties. First, it makes it incentive compatible for a DM with a strict rank ordering $\succ_1$ (satisfying independence) to report his true preferences. Second, because preferences are only implemented probabilistically, one can observe the behavior of subjects who faced the choice and yet preferred commitment. As a result, one can contrast perceived self-control, as revealed by subjects' rank ordering, with actual self-control when facing the flexible menu.[15]

[14] To minimize order effects, subjects were randomly assigned to one of two list orders, $l_1 = (\{0,1\},\{1\},\{0\})$ or $l_2 = (\{1\},\{0\},\{0,1\})$, meaning the flexible menu was presented either at the top or at the bottom, and $\{0\}$ never appeared at the top. Because options listed first are in general more likely to be assigned rank 1 than those listed last, this design feature should have if anything reduced the likelihood of observing temptation (understood as a strict preference for $\{0\}$). However, there were no significant differences in ranking across the two lists; see OA-A.1.

[15] Random implementation rules have been used in a variety of settings in order to elicit full rank orderings, incentivize potential choice revisions, and/or create a wedge between expressed preferences and actual choices; see for instance Casari and Dragone (2015), Augenblick et al. (2015), or Karlan and Zinman (2009)

However, the procedure so far does not strictly incentivize subjects to report indifferences since an expected utility maximizer who is indifferent between two menus would also take any probability distribution over these menus.[16] To disentangle indifferences from strict preferences, one needs a cardinal measure of preferences. Such a measure was collected in Section D by asking subjects for their willingness to pay (WTP) to replace their second choice with their top choice and their last choice with their second choice. If a subject was indifferent between two menus, one of them was selected to be the replaceable option. Subjects were randomly assigned within a session to express their WTP either in terms of money or in terms of time via a Multiple Price List mechanism:

$ *WTP* : Subjects made 8 decisions between [*their second (last) choice*] and [*their top (second) choice* - $$X$] where $X = \{0.01, 0.02, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50\}$. The money was taken from a subject's show-up fee of $10.

Time *WTP* : Subjects made 8 decisions between [*their second (last) choice*] and [*their top (second) choice + N minutes on the attention task*] where $N = \{1, 2, 3, 4, 5, 6, 8, 10\}$. Subjects spent additional minutes on the task at the end, for no additional payment.

To enforce monotonicity, subjects were not allowed to make multiple switches between the two options. If a subject's ranking was implemented and his second (last) choice was drawn, then one of the 8 decisions was chosen for implementation, thus ensuring incentive compatibility.

The purpose of contrasting willingness to pay for time versus money was to assess the extent to which the expression of a strict preference (in particular, for commitment) might be sensitive to the unit of payment. Indeed, so far, very few studies have found that individuals are willing to pay even the smallest amount of money for commitment.[17] For instance, Augenblick et al. (2015) find that while 59% of their subjects favor commitment when it is free, the demand is close to zero at a price as low as $0.25. Although these findings could raise the concern that a demand for commitment at a price of zero does not reveal a true preference for commitment, another interpretation is that individuals think differently about money and time (Ellingsen and Johannesson (2009)) and would be more inclined to pay in terms of their time. Testing for differences in WTP across domains offers a way to assess the robustness of the elicitation procedure.

[16] I thank Sevgi Yüksel for pointing out to me this difficulty at the design stage.

[17] Two exceptions are Milkman et al. (2014) and Schilbach (2017).

3.3 Elicitation of beliefs

Finally, Section E gathered data on subjects' beliefs about their likelihood of reading the story in Period 2 if offered $\{0, 1\}$. The measurement of these beliefs served two objectives. First, although beliefs about ex-post choice are generally not a primitive of models of menu preferences, they play a central role in the interpretation of those models. A GP agent with the preference ordering $\{0\} \succ_1 \{0, 1\} \succ_1 \{1\}$ expects to resist the temptation to read the story if offered $\{0, 1\}$ (i.e., $0 \succ_2 1$), while a DM who suffers from random indulgence expects to succumb some of the time. Similarly, a "Krepsian" DM with a preference for flexibility $\{0, 1\} \succ_1 \{0\}, \{1\}$ should express uncertainty about his willingness to read the story if offered $\{0, 1\}$. Gathering belief data allows to gain further insights into the interpretation of subjects' preference orderings.

A second reason to collect belief data is to obtain a measure of the gap between predicted and actual behavior. So far, few papers in the self-control literature have attempted to measure sophistication, understood as the ability to predict one's own behavior in the future. Yet the prediction that agents with self-control problems should demand commitment crucially relies on the assumption of sophistication. It is therefore important to understand the degree to which subjects mispredict their future behavior and how this might affect their menu preferences.

The elicitation of individuals' predictions about their future behavior, however, poses a methodological challenge. Indeed, any payment scheme designed to incentivize subjects to truthfully report their beliefs will also incentivize changes in the behavior to be predicted. This point has been acknowledged by Acland and Levy (2015) and further investigated by Augenblick and Rabin (2017).[18] As an alternative route, a few papers measure sophistication through the use of an unincentivized survey instrument such as the one proposed in Ameriks et al. (2007). In this paper, I propose a third, incentivized, method to elicit an individual's beliefs about his future choices: instrumenting beliefs about oneself with beliefs about a similar other. In the present context, the relevant dimension of similarity was the menu preference ordering: subjects were asked to guess the future choice (0 or 1) of a participant who submitted the same ranking as them in Section C; provided such a participant existed and he could make a choice from $\{0, 1\}$ in Period 2, a subject received \$2 for a correct guess.

[18] Acland and Levy (2015) measure predictions about future gym attendance by eliciting WTP for a coupon that pays contingent on attending the gym. With this mechanism, a sophisticated individual with self-control problems may have an incentive to overstate his WTP for the coupon as a commitment device to attend the gym more often than initially expected, thus providing a biased estimate of expected gym attendance. Augenblick and Rabin (2017) use accuracy payments of various sizes to elicit beliefs about future task completion; they find no evidence that stake size affects reported beliefs for the range of payments considered in their study.

A priori, there are two reasons to believe that incentivized beliefs about somebody with the same rank ordering could be a strong predictor of beliefs about oneself. First, if subjects interpret menu rankings in a way consistent with theories of menu choice, one should observe a higher proportion of "1" guesses for rankings where $\{0,1\} \succ_1 \{0\}$ and/or $\{0,1\} \sim_1 \{1\}$ relative to rankings where $\{0,1\} \succ_1 \{1\}$ and/or $\{0,1\} \sim_1 \{0\}$; therefore, the belief of a subject who conditions his guess on a ranking identical to his own should be highly correlated with what he expects his future choice to be. Second, there is large evidence in economics and psychology that individuals tend to form beliefs about the behavior of others by extrapolating from their own type (Ross et al. (1977), Rubinstein and Salant (2016)). As a result, subjects are likely to form their guess regarding the other participant assuming similarity on other - possibly unobservable - dimensions than the preference ordering.[19]

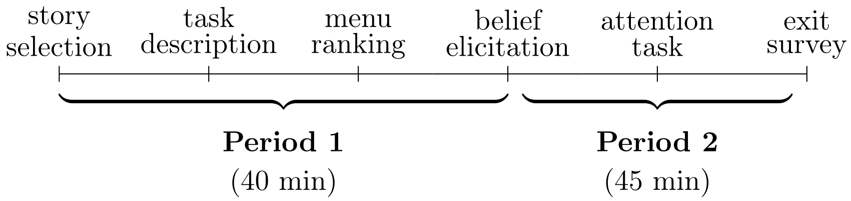
To test the strength of the above instrument, subjects were also asked an unincentivized question about their own likelihood of reading the story in Period 2 if given the chance. Answers were expressed on a 5-item scale (*very unlikely*, *quite unlikely*, *unsure*, *quite likely*, *very likely*); thus, the structure of this question differed from the binary choice frame adopted for the incentivized guess. This choice was made to minimize the chances of observing a mechanical correlation between answers simply due to subjects' exposure to identically-framed questions. To further gauge subjects' interest in the story, the end of Section E also asked them to rate their interest on a 5-item scale (*completely indifferent*, *somewhat indifferent*, *somewhat interested*, *very interested*, *dying to learn*) along two dimensions: (*i*) interest in learning the best story among the other subjects, and (*ii*) interest in knowing whether the selected story was theirs. In addition, subjects were asked to give a subjective assessment of the likelihood that their story was selected (see OA-B.1 and -F.2 for more details).

3.4 Exit survey

At the end of the session, subjects replied to a short survey designed to better understand (*i*) their ranking of the menus, and (*ii*) their interest in the story. In addition, the survey gathered some basic demographic and academic information (gender, major, GPA), and subjects were evaluated on three psychometric scales designed to measure conscientiousness and trait curiosity. More information about the exit-survey variables can be found in OA-B.3 and -F.4.

[19] In psychology, several theories emphasize the importance of self-similarity in the formation of perceptions; see, for instance, the "vicarious self-perception" theory of Goldstein and Cialdini (2007).

Figure 1: Timeline of the Experiment



4. Results

In this section, I present results from 6 experimental sessions conducted at the Center for Experimental Social Science (CESS) of New York University. A total of 120 subjects participated in the experiment and average earnings were \$18.70 per subject (including a \$10 show-up payment). The experiment lasted a little less than two hours.

The first part of this section studies perceived self-control by analyzing the distribution of menu preferences elicited in Period 1 through the initial rank-ordering procedure and subsequent WTP decisions, and by relating these preferences to beliefs about Period 2 choice. The second part turns to actual self-control by comparing subjects' menu preferences and beliefs with their actual choices in Period 2, and by studying task performance under commitment versus flexibility. Bringing all pieces of data together, the end of the section discusses support for models of costly self-control relative to other theories of temptation. Detailed power calculations for the key results presented in this section are provided in OA-H.

4.1 Perceived self-control: menu preferences

4.1.1 Initial rank orderings

Using data from the rankings submitted in Section C, I classified subjects into menu types, the distribution of which is presented in Table 1. In principle, subjects could have ranked the three menus $\{0\}, \{1\}$ and $\{0, 1\}$ in 13 different ways.[20] In actuality, 90% of subjects can be grouped in one of 7 menu types. As a benchmark, the observed frequency of each menu type is contrasted with the limit frequency that would be observed if subjects had picked a rank ordering at random.

The first two types ranked $\{0, 1\}$ strictly in between the other two menus and are labelled SSB_{-i}, for *Strict Set Betweenness* with singleton $i \in \{0, 1\}$ ranked first. In line with the intuition

[20] In addition to the full indifference ordering (1,1,1), there are 6 permutations of the ranks (1,2,3), 3 permutations of (1,1,2), and 3 permutations of (1,2,2).

Table 1: Main preference orderings

Preference ordering	menu type	% subjects	(N)	random benchmark	p-value
$\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$	SSB_{-0}	**35.8%**	**(43)**	7.7%	< 0.001
$\{1\} \succ_1 \{0,1\} \succ_1 \{0\}$	SSB_{-1}	4.2%	(5)	7.7%	0.171
$\{0,1\} \succ_1 \{0\} \succ_1 \{1\}$	$FLEX_{-0}$	**20.8%**	**(25)**	7.7%	< 0.001
$\{0,1\} \succ_1 \{1\} \succ_1 \{0\}$	$FLEX_{-1}$	7.5%	(9)	7.7%	1.000
$\{0,1\} \succ_1 \{0\} \sim_1 \{1\}$	$FLEX_{-0\vee 1}$	5.8%	(7)	7.7%	0.605
$\{0\} \sim_1 \{0,1\} \succ_1 \{1\}$	STD_{-0}	9.2%	(11)	7.7%	0.494
$\{0\} \succ_1 \{1\} \succ_1 \{0,1\}$	$GUILT$	6.7%	(8)	7.7%	0.863
other ordering		10.0%	(12)	46.1%	< 0.001
Total		100%	(120)	100%	

Notes: The reported p-values correspond to the result of a two-sided binomial test that the observed frequency is equal to the benchmark frequency of selecting one of the 13 rank orderings at random. Option 1 (0) refers to reading (not reading) the story.

that reading the story is the source of temptation in this experiment, 90% of subjects who satisfy Strict Set Betweenness are of type SSB_{-0}. The ordering of self-control types is also the most represented category, with a proportion more than 4 times larger than what would be observed under the random benchmark (35.8% vs. 7.7%, $p < 0.001$). The second category of types denoted $FLEX_{-i}$ corresponds to subjects who expressed a strict preference for $\{0,1\}$ with $i \in \{0, 1, 0 \vee 1\}$ as their second-best choice. Only the proportion of $FLEX_{-0}$ is significantly higher than what would be expected under the benchmark (20.8% vs. 7.7%, $p < 0.001$). The last two categories corresponding to the standard DM with no temptation to read the story, STD_{-0}, and the flexibility-averse type $GUILT$ represent a small fraction of the sample. Interestingly, the rank ordering capturing temptation with no self-control $\{0\} \succ_1 \{0,1\} \sim_1 \{1\}$ (included in the "other ordering" category) is underrepresented in this sample (2.5%, $p = 0.026$ against benchmark). In other words, models of sophisticated present bias with no uncertainty - which can rationalize $\{0\} \succ_1 \{0,1\} \sim_1 \{1\}$ but not $\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$ - have low explanatory power in this environment.

4.1.2 Refinement of menu rankings through WTP decisions

The above classification may overestimate the proportion of subjects with a strict preference ordering as it relies only on the initial ranking procedure, which does not strictly incentivize subjects to truthfully report an indifference. To obtain a lower-bound estimate on the proportion of self-control types, I now examine WTP decisions for replacing the second (last) choice in the ranking with the top (second) choice.

In total, 67 (53) subjects were assigned to the \$ (time) WTP condition. No significant differences were observed across the two conditions: subjects had a positive WTP in 70% (75%) of the menu comparisons in the money (time) condition ($F(1,119) = 0.52$, $p = 0.472$); the average number of rows (out of 8) at which subjects preferred to pay was 4.01 for money and 3.69 for time ($F(1,119) = 0.56$, $p = 0.456$). Differences across conditions also appear to be marginal when breaking down the distribution of WTP by comparison of ranks (top vs. second choice and second vs. last choice); see OA-A.1 for more details. For the rest of the analysis, I therefore convert the time WTP into a \$ $WTP \in [0, 0.50]$ to evaluate decisions on a single scale. For each of the 7 major preference orderings, Table 2 shows the average WTP to replace one menu with a (weakly) better-ranked menu, as well as the percentage of subjects who had a strictly positive WTP.

Overall, there is a high degree of consistency between subjects' initial ordering ($\succ_1$ or $\sim_1$) and subsequent WTP (> 0 or $= 0$), which are coherent with each other in more than 70% of the cases. First, 62% (87%) of subjects who ranked their top (second) choice strictly above their second (last) choice also had a strictly positive WTP. For all types except $FLEX_{-0}$, a majority of subjects were willing to pay for an option they strictly ranked higher. In particular, 58% of the SSB_{-0} subjects of Table 1 were willing to pay to receive $\{0\}$ instead of $\{0, 1\}$; furthermore, their WTP for commitment is increasing in their level of curiosity for the story (see Section 4.3). Second, as would be expected from subjects who are indifferent, those who gave the same rank to their top (bottom) two options had a significantly lower WTP than subjects with a strict preference for their top (second-best) option ($t_{118} = -2.22$, $p = 0.028$ for top; $t_{118} = -1.74$, $p = 0.084$ for bottom).[21] Table 3 presents an alternative classification, which accounts for subjects' WTP decisions by replacing $\succ_1$ with $\sim_1$

[21] However, 10 of the 12 subjects who gave the same rank to their bottom two options reported a positive WTP for one of the options. This high percentage is mostly due to subjects with menu type $\{0, 1\} \succ_1 \{0\} \sim_1 \{1\}$ who might have expressed their indecisiveness (rather than an indifference) by assigning the same rank to $\{0\}$ and $\{1\}$; I thank Giorgia Romagnoli for this interpretation. Some of their comments seem to go in this direction (see OA-G.2):
- "I was undecided so I ranked to make my decision later." (Session 3, ID 31)
- "I had put Decide in period 2 first so that I could have some choice and effect on which menu I would receive. I ranked the other two options both as 2 because I was unsure at the time of which menu I wanted." (Session 3, ID 40)

Table 2: Distribution of WTP by rank ordering

	top choice versus second choice		**second choice versus last choice**	
Preference ordering	average WTP (all)	% with $WTP > 0$ (freq.)	average WTP (all)	% with $WTP > 0$ (freq.)
$\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$	**\$0.14**	**58.1% *(25/43)***	\$0.31	88.4% *(38/43)*
$\{1\} \succ_1 \{0,1\} \succ_1 \{0\}$	\$0.30	80.0% *(4/5)*	\$0.38	80.0% *(4/5)*
$\{0,1\} \succ_1 \{0\} \succ_1 \{1\}$	\$0.07	40.0% *(10/25)*	\$0.28	96.0% *(24/25)*
$\{0,1\} \succ_1 \{1\} \succ_1 \{0\}$	\$0.23	88.9% *(8/9)*	\$0.11	88.9% *(8/9)*
$\{0,1\} \succ_1 \{0\} \sim_1 \{1\}$	\$0.10	57.1% *(4/7)*	\$0.25	85.7% *(6/7)*
$\{0\} \sim_1 \{0,1\} \succ_1 \{1\}$	\$0.06	27.3% *(3/11)*	\$0.37	81.8% *(9/11)*
$\{0\} \succ_1 \{1\} \succ_1 \{0,1\}$	\$0.25	100.0% *(8/8)*	\$0.20	62.5% *(5/8)*
Strict ranking	\$0.15	**62.4% *(63/101)***	\$0.28	**87.0% *(94/108)***
Indifference	\$0.05	31.6% *(6/19)*	\$0.17	83.3% *(10/12)*

Notes: Average WTP is subjects' mean WTP pooling money and time conditions; time WTP converted into dollars according to the following formula: $\tilde{WTP}$= 0.01 (=0.50) if WTP_t=1 (=10) and $\tilde{WTP}$= 0.01 + 0.5($\frac{t-1}{10-1}$) if $WTP_t \in \{2,3,4,5,6,8\}$. "Strict ranking" refers to subjects who assigned rank 1 and 2 (2 and 3) to their top (bottom) two choices, while "Indifference" refers to those who gave rank 1 (2) to their top (bottom) two choices. For $FLEX_{-0\vee 1}$, the last option was taken to be $\{1\}$; for STD, the top option was taken to be $\{0\}$. Option 1 (0) refers to reading (not reading) the story.

whenever $WTP = 0$ and $\sim_1$ with $\succ_1$ whenever $WTP > 0$.

The fraction of subjects with SSB_{-0} preferences drops to 23.3% (relative to 35.8% in Table 1), but remains about three times higher than what would be observed if subjects had ranked menus at random. The standard DM with no temptation to read the story, STD_{-0}, is now the most represented category (30% of the sample), while the proportion of subjects with a preference for flexibility is divided by two. In particular, the category $FLEX_{-0\vee 1}$ almost disappears from the sample and is replaced in the table by subjects classified as indifferent (IND). However, besides STD_{-0} and SSB_{-0}, no other menu type is present in a proportion significantly higher than what would be observed if orderings were picked at random. Finally, as with the initial classification, the rank ordering capturing temptation with no self-control $\{0\} \succ_1 \{0,1\} \sim_1 \{1\}$ remains underrepresented (2.5%, $p = 0.026$ against benchmark).[22]

[22] OA-A.2 presents the distribution of types for two other classifications. The first one excludes the 16 subjects who assigned the same rank to two menus and yet were willing to pay for one over the other i.e., ($\sim_1$, $WTP > 0$), since this behavior can be regarded as anomalous if subjects' preferences are complete and respect monotonicity in money. The second classification excludes the 60 subjects who presented some inconsistency between their initial

Table 3: Alternative classification accounting for WTP choices

Preference ordering	menu type	% subjects	(N)	random benchmark	p-value
$\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$	SSB_{-0}	**23.3%**	**(28)**	7.7%	< 0.001
$\{1\} \succ_1 \{0,1\} \succ_1 \{0\}$	SSB_{-1}	4.2%	(5)	7.7%	0.171
$\{0,1\} \succ_1 \{0\} \succ_1 \{1\}$	$FLEX_{-0}$	10.8%	(13)	7.7%	0.226
$\{0,1\} \succ_1 \{1\} \succ_1 \{0\}$	$FLEX_{-1}$	5.8%	(7)	7.7%	0.605
$\{0\} \sim_1 \{0,1\} \succ_1 \{1\}$	STD_{-0}	**30.0%**	**(36)**	7.7%	< 0.001
$\{0\} \succ_1 \{1\} \succ_1 \{0,1\}$	$GUILT$	8.3%	(10)	7.7%	0.732
$\{0\} \sim_1 \{1\} \sim_1 \{0,1\}$	IND	9.2%	(11)	7.7%	0.494
other ordering		8.3%	(10)	46.1%	< 0.001
Total		100%	(120)		

Notes: The reported p-values correspond to the result of a two-sided binomial test that the observed frequency is equal to the benchmark frequency of selecting one of the 13 rank orderings at random. Option 1 (0) refers to reading (not reading) the story.

The next findings will be presented for the full sample and for both types of classifications, $\succeq_1^{rank}$ and $\succeq_1^{WTP}$ (i.e., based on the initial ranking and based on WTP). It is indeed important to note that although it was not strictly incentive compatible for subjects to truthfully report an indifference with the initial rank-ordering procedure, it was nevertheless a weakly dominant strategy; furthermore, it remains to understand how one should interpret a zero WTP, for instance if specific dimensions of the elicitation procedure such as the unit of payment or the range of payments in the MPL affect WTP behavior.[23]

4.1.3 Link between menu preferences and beliefs about Period 2 behavior

Another way to refine the interpretation of the elicited preference orderings is to study subjects' beliefs about their likelihood of reading the story if offered $\{0,1\}$ in Period 2. Remember that beliefs about Period 2 behavior were measured in two ways by asking subjects to (i) guess the Period 2

rank ordering and their WTP behavior, that is, subjects for whom either ($\sim_1$, $WTP > 0$) or ($\succ_1$, $WTP = 0$) at least once; since the incentive structure a priori allowed for ($\succ_1$, $WTP = 0$), this is a much stricter requirement. Nonetheless, the previous findings are robust to these alternative classifications with, respectively, 24.0% (25/104) and 41.7% (25/60) of SSB_{-0} subjects (forming 20.8% of the whole sample; $p < 0.001$ against benchmark).

[23] Although one might question the informational content of a demand for commitment at a price of 0, Augenblick et al. (2015) find that subjects who prefer commitment over flexibility when both are free are more likely to exhibit present bias in effort.

choice, 0 or 1, of someone with the same rank ordering as them (incentivized), and (*ii*) report their own subjective likelihood of reading the story on a 5-item scale (*very unlikely*, *quite unlikely*, *unsure*, *quite likely*, *very likely* - unincentivized).

As shown in the Appendix (Figure 3 & Table 8), subjects' answers to (*i*) and (*ii*) are highly correlated. Among those who said they were very unlikely (likely) to read the story, only 4% (over 90%) guessed that a similar other would read the story. Excluding those who reported being unsure, close to 90% (91/102) of subjects made guesses consistent with their own subjective likelihood of reading the story (likely or unlikely). To increase comparability between the two measures, below I dichotomize the subjective measure, taking 1 (0) if the subject reported being either quite or very likely (unlikely) to read the story if given the chance; for subjects who reported being unsure, answers to the incentivized question are used as a tie breaker.

For both types of classification ($\succeq_1^{rank}$, $\succeq_1^{WTP}$) and both belief measures, Table 4 shows the proportion of subjects who anticipated the choice of Option 1 (i.e., reading the story) as a function of their menu type. As a benchmark, the third column reports the distribution of Period 2 choices inferred from $\succeq_1$ under the assumptions of *Sophistication* (S) *and No Preference Reversals* (NPR). To define these notions in a general (possibly stochastic) environment, denote by λ_x the DM's propensity to choose x from $\{0,1\}$ in Period 2, that is, $\lambda_x := \mathbb{P}\{x \in c(\{0,1\},\succeq_2)\}$ where $c(A,\succeq_2) := \{x \in A \,|\, x \succeq_2 y,\ \forall y \in A\}$. Then *Sophistication* means $\{x,y\} \succ_1 \{y\}$ implies $\lambda_x > 0$, with the additional restriction that $\lambda_x = 1$ in a deterministic world such as GP 2001.[24] In other words, a DM who strictly values the addition of an option to a menu must choose this option at least some of the time. In addition, say the DM exhibits *No Preference Reversals* between Periods 1 & 2 if $\{x\} \succ_1 \{y\}$ implies $\lambda_x > \lambda_y$, which is equivalent to $\{x\} \succ_1 \{y\}$ implies $x \succ_2 y$ in a deterministic setting.[25]

Regardless of the classification and belief measure used, subjects' beliefs are highly consistent with the restrictions imposed by *Sophistication* and *No Preference Reversals*. First, while all the SSB_{-1} subjects expected to read the story if given the chance, virtually none of the SSB_{-0} subjects

[24] This condition is also referred to as Consequentialism in the model of Ahn and Sarver (2013), which connects the DM's desire for flexibility to his preference uncertainty.

[25] It is worth noting that NPR is generally *not* a restriction of axiomatic models of preference for flexibility such as Dekel et al. (2001, 2007). To see this, suppose that the DM expects to be in one of two states during the task: with probability p, he expects to choose according to utility v such that $v(1) > v(0)$; with probability $1-p$, he expects to choose according to u such that $u(0) > u(1)$. For this DM, $\{1\} \succ_1 \{0\}$ provided that $pv(1) + (1-p)u(1) > pv(0) + (1-p)u(0)$, that is, $\frac{v(1)-v(0)}{u(0)-u(1)} > \frac{1-p}{p}$. Therefore, as long as $v(1) - v(0) > u(0) - u(1)$, one can have $\{1\} \succ_1 \{0\}$ and $p < \frac{1}{2}$ (i.e., $\lambda_0 > \lambda_1$). As such, NPR may be viewed as a rather strong requirement. In the same vein, $GUILT$ preferences as in Kopylov (2012) need not satisfy NPR (see OA-E.2).

Table 4: Relationship between initial preference ordering and beliefs

Preference ordering	menu type	dist. of Period 2 choices	Incentivized $\bar{\lambda}_1$		Unincentivized $\bar{\lambda}_1$	
$\succeq_1$ on $\mathcal{M}$		under S and NPR	$\succeq_1^{rank}$	$\succeq_1^{WTP}$	$\succeq_1^{rank}$	$\succeq_1^{WTP}$
$\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$	SSB_{-0}	$\lambda_0 > \lambda_1 \geq 0$	0.023 *(1/43)*	0 *(0/28)*	0.023 *(1/43)*	0 *(0/28)*
$\{1\} \succ_1 \{0,1\} \succ_1 \{0\}$	SSB_{-1}	$\lambda_1 > \lambda_0 \geq 0$	1 *(5/5)*	1 *(5/5)*	1 *(5/5)*	1 *(5/5)*
$\{0,1\} \succ_1 \{0\} \succ_1 \{1\}$	$FLEX_{-0}$	$\lambda_0 > \lambda_1 > 0$	0.12 *(3/25)*	0.385 *(5/13)*	0.12 *(3/25)*	0.308 *(4/13)*
$\{0,1\} \succ_1 \{1\} \succ_1 \{0\}$	$FLEX_{-1}$	$\lambda_1 > \lambda_0 > 0$	0.667 *(6/9)*	0.571 *(4/7)*	0.778 *(7/9)*	0.714 *(5/7)*
$\{0,1\} \succ_1 \{0\} \sim_1 \{1\}$	$FLEX_{-0\vee 1}$	$\lambda_0, \lambda_1 > 0$	0.714 *(5/7)*	–	0.714 *(5/7)*	–
$\{0\} \sim_1 \{0,1\} \succ_1 \{1\}$	STD_{-0}	$\lambda_1 = 0$	0 *(0/11)*	0.083 *(3/36)*	0 *(0/11)*	0.056 *(2/36)*
$\{0\} \succ_1 \{1\} \succ_1 \{0,1\}$	$GUILT$	$\lambda_0 > \lambda_1 \geq 0$	0.125 *(1/8)*	0.30 *(3/10)*	0.25 *(2/8)*	0.20 *(2/10)*
$\{0\} \sim_1 \{1\} \sim_1 \{0,1\}$	IND	$\lambda_0, \lambda_1 \geq 0$	–	0.364 *(4/11)*	–	0.455 *(5/11)*

Notes: Incentivized $\bar{\lambda}_1$ is the fraction of subjects who guessed that someone with the same rank ordering would read the story if offered {0,1} in Period 2. Unincentivized $\bar{\lambda}_1$ is the fraction of subjects who reported being quite or very likely to read the story if offered {0,1} in Period 2; for subjects reporting being "unsure," answers to the *Incentivized* question are used as a tie breaker. The distribution of Period 2 choices inferred from $\succeq_1$ relies on the assumptions of *Sophistication* (S) and *No Preference Reversals* (NPR).

expected to do so. This latter finding provides some support for the interpretation of the ordering $\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$ as reflecting costly self-control rather than random indulgence (see Sections 2.2 and 4.3). Second, for all $FLEX$ types, the fraction of subjects who expected not to read the story is strictly positive and below 1; furthermore, those who preferred $\{1\}$ to $\{0\}$ ($\{0\}$ to $\{1\}$) were more likely to anticipate reading (not reading) the story. Finally, nearly all subjects with standard preferences STD_{-0} expected not to read the story, which was also the case for most subjects with $GUILT$ preferences. Looking at all preference orderings, the adjusted R^2 of a regression of the incentivized guess on indicators $\mathbb{1}_{(\{0\} \succ_1 \{1\})}$, $\mathbb{1}_{(\{1\} \succ_1 \{0\})}$, $\mathbb{1}_{(\{0,1\} \succ_1 \{1\})}$ and $\mathbb{1}_{(\{0,1\} \succ_1 \{0\})}$ is 0.62 using $\succeq_1^{rank}$ and 0.37 using $\succeq_1^{WTP}$; the corresponding numbers are 0.59 and 0.47 for the unincentivized guess (see OA-E.1). In other words, menu preferences encode a lot of information about beliefs.

4.2 Actual self-control: Period 2 behavior

I now turn to the analysis of Period 2 behavior. First, I compare perceived self-control with actual self-control by examining the relationship between the menu preferences and beliefs elicited in Period 1 and subjects' actual propensity to read the story in Period 2. I then present results from an exploratory analysis linking task performance to menu assignment in order to suggest one possible interpretation of self-control costs in this experiment.

4.2.1 Link between menu preferences and propensity to read the story in Period 2

Out of the 120 subjects, 87 were asked to make a choice from the flexible menu $\{0, 1\}$; of the remaining subjects, 29 received menu $\{0\}$, which removed the opportunity to read the story, while the last 4 subjects were assigned menu $\{1\}$, thus accessing the story for sure. The analysis of this subsection focuses on the 87 subjects who were offered to make a choice from $\{0, 1\}$.

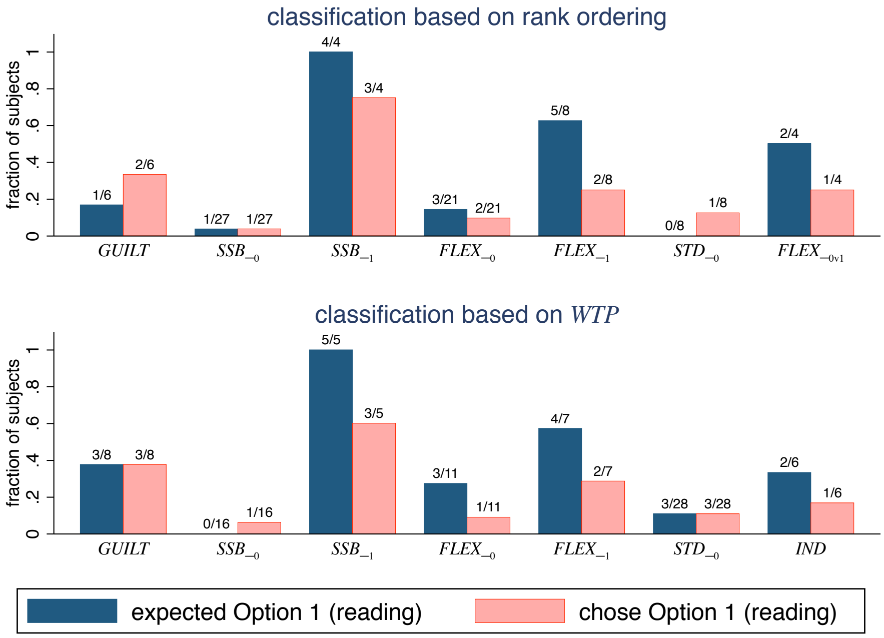
Overall, 18.4% (16/87) of the subjects assigned $\{0, 1\}$ chose to read the story at some point during the attention task, with some heterogeneity in the timing of access (see OA-C.1). For both of the classifications presented earlier, Figure 2 shows the proportion of subjects who chose to read the story during the task as a function of their menu preferences; as a benchmark, actual behavior is contrasted with subjects' expectations.

As is immediately apparent from the figure, there is a lot of heterogeneity across types in their propensity to access the story. The restrictions of *Sophistication* and *No Preference Reversals* (see Table 4 column 3) capture some of this heterogeneity, although the predictive power of menu preferences is significantly weaker for ex-post choice than for beliefs.[26] Among those who ranked $\{1\}$ strictly above $\{0\}$, slightly less than half chose to read the story, thus departing from NPR. Their propensity to read the story is, however, 3 to 4 times higher than those who strictly preferred $\{0\}$ to $\{1\}$.[27] Furthermore, of the 7 menu types of Figure 2, $FLEX_{-1}$ is the only type that violates NPR. At the individual level, $\{x\} \succ_1 \{y\}$ implies $x \succ_2 y$ for about 80% of subjects (for both $\succeq_1^{rank}$ and $\succeq_1^{WTP}$). Most discrepancies between menu preferences and ex-post choice come from the $FLEX$ and $GUILT$ types (10/15 for $\succeq_1^{rank}$ and 9/17 for $\succeq_1^{WTP}$) and, as noted earlier, existing models that rationalize those types allow in principle for violations of NPR. Looking at

[26] The adjusted R^2 of a linear regression of an indicator for whether the subject read the story on indicators $\mathbb{1}_{(\{0\}\succ_1\{1\})}$, $\mathbb{1}_{(\{1\}\succ_1\{0\})}$, $\mathbb{1}_{(\{0,1\}\succ_1\{1\})}$, and $\mathbb{1}_{(\{0,1\}\succ_1\{0\})}$ is 0.19 using $\succeq_1^{rank}$, and 0.12 using $\succeq_1^{WTP}$ (see OA-E.1).

[27] Using $\succeq_1^{rank}$, 40.0% (6/15) of subjects with preference $\{1\} \succ_1 \{0\}$ chose to read the story compared to 9.4% (6/64) of those with preference $\{0\} \succ_1 \{1\}$ ($t_{77} = -3.12$, $p = 0.003$); using $\succeq_1^{WTP}$, the corresponding numbers are 42.9% (6/14) and 13.8% (9/65) ($t_{77} = -2.58$, $p = 0.012$).

Figure 2: Beliefs versus ex-post choice by menu type



Notes: "expected Option 1 (reading)" refers to the proportion of subjects who guessed that someone with the same rank ordering as them would choose to read the story if offered {0,1}; patterns are very similar for the unincentivized belief measure (see OA-E.1). Means were computed for each menu type using the classifications presented in Table 1 (for top panel) and Table 3 (for bottom panel).

the relationship between beliefs and ex-post choice gives a similar picture. About three quarters of subjects behaved in a way consistent with their beliefs (regardless of the measure), and the majority of inconsistencies come from the $FLEX$ and $GUILT$ types ($13/20$ for $\succeq_1^{rank}$ and $11/20$ for $\succeq_1^{WTP}$). Although Figure 2 seems to indicate that subjects overestimated on average their propensity to read the story, mispredictions go both ways: among those who read the story eventually, nearly half expected not to do so. I discuss observed discrepancies between menu preferences and beliefs on the one hand, and ex-post choice on the other hand, in the conclusion section.

Most importantly, the fraction of subjects with self-control preferences who read the story is very close to zero: of the 27 (16) subjects classified as SSB_{-0} according to $\succeq_1^{rank}$ ($\succeq_1^{WTP}$), only one chose to access the story; this finding contrasts with the 25% (21%) proportion of other types who did so ($t_{85} = 2.42$, $p = 0.018$ for $\succeq_1^{rank}$; $t_{85} = 1.39$, $p = 0.169$ for $\succeq_1^{WTP}$).[28] The pattern

[28] Furthermore, among those classified as SSB_{-0} according to the WTP classification, the subject who read the story turned out to be the one with the lowest WTP for replacing $\{1\}$ with $\{0, 1\}$ (and also, $\{0, 1\}$ with $\{0\}$): while 90% of the other subjects selected at least 4 rows in the MPL when comparing $\{0, 1\}$ to $\{1\}$, this subject only selected one row. Therefore, this subject could have been classified as $\{0\} \sim_1 \{0, 1\} \sim_1 \{1\}$ according to $\succeq_1^{WTP}$. I thank Roberto Weber for his suggestion to study WTP for replacing $\{1\}$ with $\{0, 1\}$ as a robustness check (see OA-D.1).

of behavior of the SSB_{-0} subjects is also very consistent with their ex-ante beliefs about their propensity to access the story. In other words, perceived self-control almost entirely translated into actual self-control, as would be expected under *Sophistication.* In light of this evidence, I discuss support for theories of costly self-control relative to other temptation models in Section 4.3.

4.2.2 Is there a cost of self-control?

While virtually none of the SSB_{-0} subjects ended up reading the story, models of costly self-control à la GP suggest that resisting temptation may involve utility costs, despite remaining silent about the nature of those costs. Below I present the results of an exploratory analysis, which suggests one possible way of interpreting and measuring self-control costs in the context of this experiment, namely by testing whether subjects' productivity was impacted by the menu they were assigned.

In psychology, self-control is often defined as "the capacity to regulate attention, emotion, and behavior in the presence of temptation" (Duckworth and Gross (2014)). In this experiment, subjects were paid for correctly answering a series of 5 prompts, which appeared on their screen at random times over a period of 45 minutes. Success in the task required subjects to constantly direct their attention resources to the number on their screen and to suppress their thoughts about the story. In the context of this experiment, I therefore interpret self-control as the costly self-regulation of attention. With this interpretation in mind, one indirect way to test for the presence of self-control costs is to measure whether the availability of temptation opportunities affected productivity. Indeed, if (*i*) attention is limited and costly to regulate, and (*ii*) a tempting alternative competes for the attention of the decision maker, then productivity should be higher when all temptation opportunities are removed. In other words, subjects who were assigned the flexible menu $\{0, 1\}$ should have a lower productivity than those who were assigned the commitment menu $\{0\}$.

To test this hypothesis, I consider two measures of productivity: (a) whether a subject correctly answered all 5 prompts and (b) the number of prompts correctly answered. Overall, 70% of subjects provided 4 or 5 correct answers, and 37% answered all prompts correctly (see OA-C.2). Looking at raw averages, subjects assigned $\{0\}$ were about 20 ppts more likely to obtain a perfect score than those who were assigned $\{0, 1\}$ (51.7% vs. 32.2%, one-sided $p = 0.030$, $t_{114} = -1.90$); furthermore, they gave 0.4 more correct answers on average (4.2 vs. 3.8, one-sided $p = 0.057$, $t_{114} = -1.59$). Although these raw comparisons are in line with the main hypothesis, menu assignment was only random conditional on a subject's initial ordering and WTP choices, which determined the probability of facing each of the three menus. If subjects who strictly prefer $\{0\}$ to $\{0, 1\}$ tend to be more productive than others (e.g., because they care more about their earnings), a naive comparison of productivities based on menu assignment will overestimate the detrimental impact on productivity of facing $\{0,1\}$. To address this issue, Table 5 presents results from linear regressions that control for a subject's probability $\mathbb{P}_m$ of facing menu $m \in \{\{0\},\{1\},\{0,1\}\}$. As columns (2) & (5) show, those who strictly preferred $\{0\}$ (and thus faced a higher probability $\mathbb{P}_{\{0\}}$ of receiving that menu) were indeed more productive on average than the other subjects. Although only marginally significant, the effect of being assigned $\{0,1\}$ remains negative after controlling for menu preferences and of a similar magnitude as the effect measured without controlling for preferences.[29]

Table 5: Effect of flexible menu on productivity

	Obtained perfect score			Number of correct answers		
	(1)	(2)	(3)	(4)	(5)	(6)
assigned {0,1}	-0.225**		-0.194*	-0.429*		-0.392*
	(0.105)		(0.107)	(0.228)		(0.235)
$\mathbb{P}_{\{0\}}$		1.419**	1.260**		2.140*	1.818
		(0.551)	(0.553)		(1.212)	(1.218)
$\mathbb{P}_{\{0,1\}}$		0.975	1.049*		1.539	1.689
		(0.629)	(0.624)		(1.383)	(1.375)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	116	116	116	116	116	116
Mean dependent variable	0.37	0.37	0.37	3.93	3.93	3.93

Notes: Columns (1)-(3) are linear probability models where the dependent variable *Obtained perfect score* is equal to 1 if the subject correctly answered all 5 prompts; probit models give similar results. The variable $\mathbb{P}_m$ is the subject's probability of receiving menu $m \in \{\{0\},\{0,1\},\{1\}\}$ given his rank ordering and WTP; * $p < 0.1$ and ** $p < 0.05$.

While the previous analysis suggests that the mere presence of opportunities to read the story might have impaired subjects' productivity, the specific mechanism driving those productivity differentials remains unclear. If productivity losses are driven by self-control costs, then one should expect a productivity gap only among those who truly experienced a choice conflict between maximizing their earnings (initial plan) and reading the story (immediate desire).[30] In OA-C.3.3, I therefore test whether differences in productivity depend on whether reading the story conflicted with a subject's original plan. To this end, I consider 4 measures of conflict based on whether read-

[29] Power calculations indicate that the study was not well powered to detect small productivity differences; therefore, this finding should be interpreted with caution; see OA-H.2.3. To complement this econometric analysis, OA-C.3.2 reports estimates of productivity differences based on matching methods, taking subjects with the same rank ordering as counterfactuals. Results are both qualitatively and quantitatively similar.

[30] I thank an anonymous referee for suggesting this idea.

ing the story conflicted with subjects' initial beliefs (if they did not anticipate reading the story) or with their initial preferences (if they strictly preferred $\{0\}$ to $\{1\}$). For 3 of the 4 measures, I find that conflicted subjects were significantly less likely to obtain a perfect score when they faced $\{0, 1\}$; on the other hand, productivity losses are smaller and insignificant among subjects who faced no conflict.[31] Although the evidence is more suggestive, conflicted subjects were also more likely to report that the story occupied their mind during the task when they faced $\{0, 1\}$ rather than $\{0\}$; again, no such finding emerged for subjects who faced no conflict (see OA-C.3.4). Since conflicted subjects were less likely to envision reading the story and to read it eventually, observed productivity differentials cannot be simply due to the contemplation costs of deciding when to access the story. Instead, they appear to be consistent with a cost of self-control, coming from subjects' efforts to suppress their thoughts about the story in order to stay focused on the task. This interpretation also resonates with a large literature in psychology, which proposes that prior acts of self-restraint may impair subsequent self-control, similar to a muscle that gets tired from exertion (Baumeister et al. (1994), Baumeister and Vohs (2003)).[32]

4.3 Costly self-control or random indulgence?

4.3.1 Comparing temptation models

The unique combination of data on menu preferences, beliefs about Period 2 behavior, and actual Period 2 behavior provides a way to assess the explanatory power of theories of costly self-control relative to other temptation models. Table 6 contrasts the data with the predictions made by 4 classes of temptation models under the assumption of sophisticated behavior. To make comparisons, I look at the subset of 54 subjects who ex ante preferred not to read the story but expressed being tempted by it, that is, those for whom $\{0\} \succ_1 \{1\}$ and $\{0\} \succ_1 \{0, 1\}$ according to $\succeq_1^{rank}$; among them, 35 made a choice from $\{0, 1\}$ in Period 2 (see OA-E.2 for a similar table based on $\succeq_1^{WTP}$). The first class of models corresponds to standard models of dynamic inconsistency with no uncertainty (Strotz (1956), Laibson (1997), O'Donoghue and Rabin (1999)). As discussed in Section 2.2, present-biased agents who are sophisticated will choose to restrict their choice set if and only if they expect

[31] However, conflict does not appear to explain productivity differences for the second productivity measure, namely number of correct answers. One conjecture is that the two productivity measures capture something different about a subject's motivation to complete the task: since most prompts were easy to answer, subjects with low scores likely had a low motivation to perform the task ex ante; on the other hand, obtaining a perfect score may better capture determination and persistence during the task.

[32] See Hagger and Chatzisarantis (2016) and Dang (2016) for recent debates about the existence and the size of the ego-depletion effect.

to succumb to temptation. The next two classes are deterministic models of costly self-control à la Gul and Pesendorfer (2001) and models of random indulgence in which temptation is uncertain (Chatterjee and Krishna (2009), Eliaz and Spiegler (2006), Duflo et al. (2011)). As explained in Section 2.2, both classes of models can rationalize the ordering $\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$, but models of random indulgence also predict a strictly positive probability of giving in. Finally, the model of Kopylov (2012), which nests GP 2001 as a special case, can rationalize a form of temptation induced by guilt or fear of making the wrong choice.[33]

Table 6: Explanatory power of existing temptation models

Temptation model	menu preferences	expected propensity to read the story λ_1	actual propensity to read the story ρ_1
Dynamic Inconsistency (Strotz preferences)	$\{0\} \succ_1 \{0,1\} \sim_1 \{1\}$	$\lambda_1 = 1$	$\rho_1 = 1$
Costly Self-Control (GP 2001)	$\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$	$\lambda_1 = 0$	$\rho_1 = 0$
Random Indulgence (Models w/ temptation uncertainty)	$\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$	$\lambda_1 \in (0,1)$	$\rho_1 \in (0,1)$
Temptation with Guilt (Kopylov 2012)	$\{0\} \succ_1 \{1\} \succ_1 \{0,1\}$	$\lambda_1 \in \{0,1\}$	$\rho_1 \in \{0,1\}$
Observed	$\{0\} \succ_1 \{0,1\} \succ_1 \{1\}$ for 79.6% (43/54)	$\lambda_1 = 0.023$ (1/43)	$\rho_1 = 0.037$ (1/27)
	other temptation ranking for 20.4% (11/54)	$\lambda_1 = 0.091$ (1/11)	$\rho_1 = 0.25$ (2/8)

Notes: Predictions and findings for the set of 54 subjects for whom $\{0\} \succ_1 \{1\}$ and $\{0\} \succ_1 \{0,1\}$ according to $\succeq_1^{rank}$. Observed frequency λ_1 corresponds to the proportion of tempted subjects who predicted that someone with the same ranking would read the story, and ρ_1 is the fraction of tempted subjects who indeed read the story.

As can be seen from the table, the only two classes of theories that are broadly consistent with the data are those of costly self-control and random indulgence. However, for the latter to rationalize observed behavior, the (perceived) probability of indulgence would have to be very close to zero, thus making temptation uncertainty a less compelling rationalization than costly self-control. The next findings provide further evidence in favor of theories of costly self-control.

[33] In Kopylov (2012), choice is deterministic and a DM with the ordering $\{0\} \succ_1 \{1\} \succ_1 \{0,1\}$ may choose either option from $\{0,1\}$ (i.e., $\rho_1 \in \{0,1\}$). See OA-E.2 for a discussion of the different temptation models.

4.3.2 Can temptation uncertainty explain commitment demand?

Although temptation uncertainty in the aggregate appears to be minor, a perhaps more important question is whether any residual uncertainty can explain the preference for commitment of the SSB_{-0} subjects. To address this question, I next study the determinants of WTP for $\{0\}$ of the 43 subjects classified as SSB_{-0} based on their initial ranking of the three menus, $\succeq_1^{rank}$. Subjects who suffer from random indulgence will only pay for $\{0\}$ if they expect to succumb with positive probability ($\lambda_1 > 0$). Furthermore, their WTP will be increasing in λ_1 i.e.,

$$u(0) - WTP = \lambda_1 u(1) + (1 - \lambda_1)u(0)$$

$$\Leftrightarrow WTP_{-RI} = \lambda_1[u(0) - u(1)]$$

On the other hand, the WTP for commitment of self-control types should be increasing in how tempting they find the story (irrespective of their beliefs) i.e.,

$$V_{GP}(\{0\}) - WTP = V_{GP}(\{0, 1\})$$

$$\Leftrightarrow \; u(0) - WTP = u(0) - [v(1) - v(0)]$$

$$\Leftrightarrow \; WTP_{-GP} = v(1) - v(0)$$

Below I therefore test whether subjects' WTP for replacing $\{0, 1\}$ with $\{0\}$ depends on (i) their perceived chances of reading the story (proxy for λ_1) and/or (ii) how enticing they find the story (proxy for $v(1)$). To measure (i), I exploit variation in subjects' answers to the unincentivized belief question, which allowed them to express their subjective likelihood of reading the story on a 5-item scale (coded below as: 1 = "very unlikely", 2 = "quite unlikely", 3 = "unsure", 4 = "quite likely", 5 = "very likely"). Appendix Figure 4 shows that while 65% (28/43) of the SSB_{-0} subjects reported being very unlikely to read the story, the rest expressed more uncertainty, with 28% (12/43) selecting "quite unlikely" and 5% (2/43) selecting "unsure".[34] To measure (ii), I use subjects' responses to two questions pertaining to their interest in learning whether the selected story was theirs (Q1) and what the best story was among the other subjects in the room (Q2). Answers were also expressed on a 5-item scale (coded below as: 1 = "completely indifferent", 2 = "somewhat indifferent", 3 =

[34] In addition, one subject selected "very likely"; incidentally, this subject is not the same as the one who guessed that a similar other would read the story or the one who actually chose to read the story.

"somewhat interested", 4 = "very interested", 5 = "dying to learn"). Over half (23/43) of the SSB_{-0} subjects reported being at least somewhat interested in the story in Q1 and/or Q2 (70% in the entire sample; see OA-B.1).

Appendix Figure 5 shows the distribution of WTP for replacing $\{0,1\}$ with $\{0\}$ of the SSB_{-0} subjects as a function of their beliefs (Panel A) and interest in the story (Panel B). Looking at beliefs, the WTP distributions appear very similar if one compares subjects who expressed uncertainty about their chances of reading the story with those who did not; this is confirmed by a Kolmogorov-Smirnov test ($D = 0.162$, $p = 0.928$). Looking at other statistics gives a similar picture: the mean WTP is \$0.13 for those who expressed uncertainty and \$0.15 for those who did not ($t_{41} = 0.35$, $p = 0.732$), and the Spearman correlation between WTP and beliefs (score from 1 to 5) is $\rho = -0.077$ ($p = 0.622$, 95% CI = [-0.369, 0.228]). On the other hand, curiosity for the story does predict WTP for commitment. Consistent with the idea that reading the story is a temptation that should be avoided, those who expressed interest in the story had a higher WTP for eliminating it from their choice set than those who did not (\$0.20 vs. \$0.07, $t_{41} = -2.28$, $p = 0.028$). Furthermore, WTP behavior differs significantly at the boundaries of the Multiple Price List: while only 5% (1/20) of those who expressed no real interest in the story exhibited maximal WTP, this was the case of nearly a third (7/23) of those who expressed interest ($t_{41} = -2.21$, $p = 0.033$).

Although a positive link appears to exist between interest in the story and WTP for commitment, the more enticing the story is, the higher the likelihood might be of succumbing to temptation. Indeed, a strong positive correlation exists between subjects' interest in the story and their beliefs that they will read it.[35] As a result, a more convincing test is whether the positive relationship between interest in the story and WTP still exists after controlling for beliefs. I therefore analyze the robustness of this relationship in a regression framework. Because WTP is censored to the right for a non-trivial proportion of SSB_{-0} subjects (see Appendix Figure 5 and previous discussion), Table 7 presents results from Tobit regressions, using \$ WTP as the outcome variable.[36] Regressions include either a dichotomic measure of beliefs and interest in the story or a multi-level measure

[35] Among the SSB_{-0} subjects ($N = 43$), the Spearman correlation coefficient between a subject's level of interest in the story and his perceived chances of reading it (variables *Interest level* and *Chances of reading* in Table 7) is $\rho = 0.5$, $p < 0.001$ (95% CI = [0.231, 0.694]). Comparing the dichotomic measures of interest and beliefs (*Interested* and *May read* in Table 7 & Appendix Figure 5), $p = 0.023$ (Fisher's exact test). See OA-D.1 for more details.

[36] Although, in principle, WTP for replacing $\{0,1\}$ with $\{0\}$ should be weakly positive if subjects ranked $\{0\}$ strictly above $\{0,1\}$, I run two-limit tobit regressions to allow for negative WTP. Results are very similar for one-limit Tobit models assuming WTP is only censored to the right. As a reminder, \$ WTP is only a conversion for subjects in the time WTP condition; see Footnote of Table 2 for the construction of this variable. As a complement, OA-D.1 presents results from simple linear regressions in which WTP is measured as the number of rows in the Multiple Price List (out of 8) at which a subject preferred to pay to replace $\{0,1\}$ with $\{0\}$; findings are very similar with this measure. Other robustness checks addressing concerns regarding the small sample size are also presented in OA-D.1.

(score from 1 to 5); all regressions also control for the WTP condition (time vs. money), as well as session fixed effects. The first set of regressions (1-4) shows the effect of interest and beliefs separately, while the second set of regressions combines the two dimensions.

Table 7: Determinants of normalized WTP for replacing $\{0, 1\}$ with $\{0\}$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Interested	0.371**				0.459***		0.461***	
	(0.136)				(0.141)		(0.144)	
Interest level		0.133*				0.188**		0.197**
		(0.073)				(0.079)		(0.082)
May read			-0.102				-0.269*	-0.276*
			(0.144)				(0.136)	(0.156)
Chances of reading				-0.107	-0.208*	-0.202*		
				(0.10)	(0.103)	(0.116)		
Time WTP	0.177	0.141	0.101	0.09	0.173	0.143	0.196	0.165
	(0.129)	(0.133)	(0.135)	(0.135)	(0.120)	(0.127)	(0.122)	(0.128)
Session FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	43	43	43	43	43	43	43	43
Mean dependent variable	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14

Notes: Two-limit Tobit regressions of WTP for replacing $\{0, 1\}$ with $\{0\}$ (converted in \$ for time WTP). The indicator *Time WTP* is equal to 1 for subjects in the time WTP condition, *Interested* is equal to 1 for subjects who reported being at least somewhat interested in learning whether their own story was selected (Q1) and/or what was the most incredible story among others (Q2); *Interest level* $\in \{1, 1.5, ..., 5\}$ is a subject's mean answer to Q1 and Q2 (coded as: 1 = "completely indifferent", 2 = "somewhat indifferent", 3 = "somewhat interested", 4 = "very interested", 5 = "dying to learn"); the indicator *May read* is equal to 1 for subjects who did not answer that they were very unlikely to read the story. Finally, *Chances of reading* refers to the belief category number for the unincentivized guess (coded as: 1 = "very unlikely", 2 = "quite unlikely", 3 = "unsure", 4 = "quite likely", 5 = "very likely" to read the story); * $p < 0.1$, ** $p < 0.05$ and *** $p < 0.01$.

Confirming the above impression, interest in the story (both on the extensive and intensive margins) predicts a higher WTP for commitment (columns 1 & 2). On the other hand, beliefs about the likelihood of succumbing have no predictive power when considered alone (columns 3 & 4). Interestingly, the effect of interest in the story is even stronger and more precisely estimated after controlling for beliefs. Furthermore, the effect of beliefs is, if anything, negative: subjects who are more confident that they will refrain from reading the story have a higher WTP for commitment. Although this finding may appear surprising at first sight, it is in line with several studies showing that those who choose commitment tend to be the ones who would resist temptation anyway (Sadoff et al. (2015), Royer et al. (2015)).[37] One possible interpretation is that avoiding temptation through commitment is “a meta-regulation strategy,” which itself requires self-efficacy and/or self-control (Ent et al. (2015)). Alternatively, since some level of sophistication is a necessary condition for commitment, it could be that individuals with less severe self-control problems are more likely to be aware of their problems and/or acknowledge them. Given the lack of evidence on this issue, more research is needed to understand how an individual’s desire for commitment relates to his belief that he can resist temptations in the future.

Due to the small sample size, the findings presented in this section should be interpreted with caution. With this caveat in mind, the evidence seems to suggest that the commitment decisions of the SSB_{-0} subjects were guided by temptation concerns, with a higher WTP for commitment among those who found the story most enticing. However, WTP for commitment does not seem to be explained by subjects’ fear of succumbing to temptation, as the random-indulgence hypothesis would predict; if anything, those who exhibited the strongest commitment demand appear to be the ones who anticipated the most self-control.[38]

5. Discussion

In this paper, I propose a new experimental method designed to identify and document costly self-control. The method is grounded in the theory of menu choice originally developed by Gul and Pesendorfer (2001) to study the behavioral implications of temptation and self-control. While present-biased agents will choose to restrict their choice set only if they expect to succumb to temptation, self-control types may demand commitment despite expecting to resist, in order to avoid the non-consequentialist cost of exerting self-control.

To identify self-control types, I conduct a laboratory experiment in which the temptation was to

[37] In a field experiment on food choices conducted in Chicago, Sadoff et al. (2015) find that those who commit to their advance choices are less likely to have exhibited time inconsistency on prior choices (by switching to more unhealthy items for immediate consumption); the authors recently replicated this finding in a new field experiment in Los Angeles. In a field experiment with employees of a large company, Royer et al. (2015) observe that the individuals who commit to increasing their gym attendance are more likely to be the ones who already exercised quite frequently. In contrast, a few other studies find a positive relationship between commitment demand and some measure of dynamic inconsistency (Ashraf et al. (2006), Kaur et al. (2015), Augenblick et al. (2015)), but the relationship tends to be fairly weak. In addition, Alan and Ertac (2015) find no relationship.

[38] In OA-D.2, I also perform a calibration exercise relating subjects’ WTP for $\{0\}$ to their expected loss in earnings from facing $\{0, 1\}$, $E(L)$. Under the random-indulgence hypothesis, WTP for $\{0\}$ should be positively correlated with $E(L) = 2\lambda_1\pi$, where λ_1 is a subject’s belief that he will read the story when facing {0,1}, and π is his perceived likelihood of correctly answering the \$2 prompt excluded from the payment. Unsurprisingly given the above findings, there is virtually no relationship between WTP for $\{0\}$ and the estimated $E(L)$, although the two quantities are on average very close in levels. However, given the small sample size and the noisy estimate of $E(L)$, confidence intervals are very wide and large positive correlations cannot be discarded; see OA-D.2 for more details.

forfeit money to read a sensational story during a tedious task. The identification strategy relies on a two-step procedure. First, I elicit subjects' preferences over a set of menus, which either did or did not allow access to the story during the task. Second, I implement preferences only probabilistically so as to observe subjects who faced the choice yet preferred commitment. With this rich dataset containing menu preferences, beliefs about ex-post choice, and actual choices from the flexible menu, I assess the explanatory power of theories of costly self-control against other temptation models.

In this specific setting, a quarter to a third of subjects can be classified as self-control types according to their menu preferences, a proportion that is 3 to 4 times higher than what would be observed if subjects had picked a rank ordering at random. Consistent with costly self-control, those subjects expected to resist the temptation to read the story in the absence of commitment. Furthermore, perceived self-control, as measured by subjects' menu preferences and anticipated choices, almost entirely coincides with actual self-control when facing the choice: whereas over 20% of the other subjects read the story when offered the choice, only one subject with self-control preferences succumbed to temptation.

While not entirely conclusive given the small sample size, I find two additional pieces of evidence suggesting that self-control costs, more than temptation uncertainty, influenced behavior in this experiment. First, controlling for menu preferences, I find that subjects who faced the choice to read the story during the task were less productive than those who did not; furthermore, observed productivity differentials appear to mostly come from subjects for whom reading the story would have conflicted with their initial preferences and/or beliefs. Second, while the commitment decisions of self-control types seem to have been guided by temptation concerns, they do not appear to have been driven by a fear of succumbing to temptation, as models of random indulgence would predict; in fact, if anything, subjects with the highest WTP for commitment appear to be those who most expected to have self-control ex post.

Although the above evidence suggests the importance of self-control costs in the specific decision environment of this study, an open question is whether the findings of this paper would extend to different settings. Below I discuss several distinctive features of the decision task and the experimental setup, which may individually contribute to explaining the level of temptation and self-control observed in this experiment. In light of this discussion, I suggest several ways in which this paper could be extended and complemented by future studies in order to gather new insights on the nature of self-control.

One specificity of the task is that succumbing to temptation entailed an explicit monetary cost:

subjects lost the chance to earn an additional $2 if they read the story. This direct cost was imposed in order to increase the chances that subjects anticipate and experience a decision conflict between maximizing their earnings and satisfying their curiosity. The formulation of a clear trade-off is likely to have encouraged commitment demand, thus making it possible to study the motives behind a preference for restricting choice sets. One legitimate question is whether the subjects classified as self-control types would have still paid for commitment had there been no explicit cost for reading the story. As discussed in Section 4.3, the WTP for commitment of the self-control types cannot be simply explained by their expected monetary loss from reading the story; therefore, the value of commitment must reside beyond the perceived cost of giving in. One possibility is that subjects anticipated a potential productivity loss due to reduced attention on the task, as suggested in Section 4.2.2. A complementary interpretation is that they paid to avoid the psychological discomfort of being confronted with the temptation. Both interpretations point toward a psychic cost of self-regulation, and future research could aim to properly measure and decompose this cost.

A second specificity of the task was the random sequencing of prompts and the unknown time length. Subjects were told they would have to answer 5 prompts occurring at random times during the task. Subjects' uncertainty about the waiting time between any two prompts meant that success in the task required their continuous attention. Furthermore, while subjects were prepared to work on the task for up to 60 minutes, the actual task duration was only 45 minutes; this effectively ensured that subjects had no information about the timing of the final prompt, which concluded the task. Although this attention task appears to have been depleting (see 4.2.2), the task might have been too short and/or too absorbing for subjects to widely succumb to temptation. Given the uncertainty, it is also possible that subjects mispredicted the difficulty of the task, for instance, if they expected the task to be longer than the actual duration. Together with the short practice time (2 minutes), this uncertainty could explain why subjects seem to have overestimated on average their propensity to read the story (see 4.2.1). To gain insights on the dynamics of self-control, future work could look at how the joint distribution of menu preferences, beliefs, and ex-post choice changes as the task becomes longer and subjects gain experience with it.[39]

Besides the decision task, several features of the experimental setup could explain the relatively high proportion of self-control types in this study, a finding that contrasts with several studies in the

[39] For instance, one could think of offering subjects the choice to perform the task a second time and observe whether their preferences and behavior change. The implied dynamics would, however, introduce many complexities, as subjects would not only learn about the nature of the task, but also about their own self-control. Furthermore, self-control capacities could be greatly reduced after the first trial, making temporal distance between the first and second trial a key parameter of the experiment.

literature on dynamic inconsistency and/or partial naiveté (Read and Van Leeuwen (1998), Acland and Levy (2015), Augenblick and Rabin (2017), John (2015), Bai et al. (2017)). First, there was a short temporal distance between the menu preferences and beliefs elicited in Period 1, and the choices made from the flexible menu in Period 2: all decisions in this experiment occurred within two hours on a single day. While this particular design choice was made to minimize attrition, time compression may give less leeway for dynamic preference reversals to occur on a large scale. Similarly, subjects may be less likely to mispredict their future behavior if the future is close. Second, this study was conducted in the tightly controlled environment of the lab, in which uncertainty cannot be too large. Most notably, subjects worked on a task for which the time commitment was clearly bounded and no outside distraction was available besides reading the story. By contrast, most of the above studies are field experiments with a rich temporal dimension and in which many outside considerations, some unanticipated, could have easily diverted subjects from fulfilling their experiment-related goals. In more unpredictable decision environments, observing more time-inconsistent choices and more mispredictions seems less surprising. More generally, models of stochastic present bias with partial naiveté may be particularly relevant in environments with high uncertainty and many delay opportunities. On the other hand, models of costly self-control à la GP may be more suitable to analyze behavior in relatively stable and/or familiar environments with minimal delay between decisions (see Fudenberg and Levine (2006) for a similar point).[40]

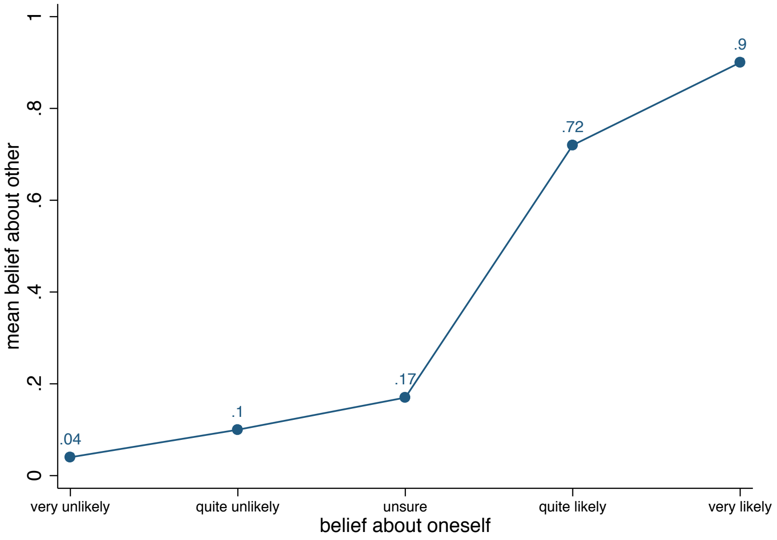
Given the specificity of the environment, it is legitimate to wonder how the proportion of self-control types would vary if the methodology were implemented in a field context with more common temptations and large stakes. More generally, one may question the stability of menu preferences (and their interpretation) across decision environments and within individuals: should we see the notion of "menu type" as referring to a stable individual trait or instead a highly context-dependent construct? While the present study remains silent on these issues, results from a companion paper may bring some preliminary answers. In Toussaert (2016), I use menu choice to study the commitment demand of participants in a weight-loss challenge. The menus were lunch reimbursement options that differed in the range of foods included in the coverage. I elicit participants' preference ordering over the various options and test whether their preference for a restricted coverage can

[40] Two other considerations could explain the lack of temptation-driven preference reversals and/or naive choices in this experiment. First, subjects may have tried to be consistent with their prior preferences and/or beliefs; however, since ex-post choice was inconsistent with beliefs (menu preferences) for about 25% (20%) of subjects, a preference for consistency cannot be the primary explanation. Second, subjects who felt observed during the task may have tried to behave in a more rational way. However, because the experimental design was double-blind, neither the experimenter nor the other subjects could a priori tell whether somebody succumbed to temptation.

predict commitment behavior in some related domain (exercise and participation in the challenge). The distribution of menu preferences is quite similar across the two papers; furthermore, I observe some cross-domain consistency in participants' preference for commitment. Of course, this companion paper only provides limited answers to the broader question of the stability of menu preferences, but suggests an interesting avenue for future research.

Appendix

Figure 3: Relationship between belief about other and belief about oneself



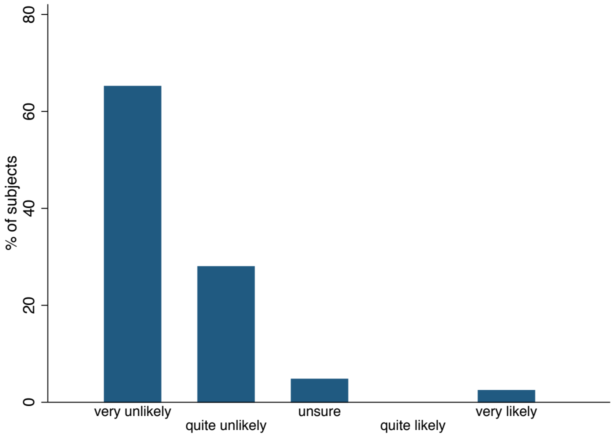
Notes: Proportion of subjects who guessed that a similar other would read the story as a function of their answer to the unincentivized question ($N = 120$).

Table 8: Relationship between belief about other and belief about oneself

	said unlikely to read	said likely to read	Total
	[Self]		
	69	6	75
expected other not to read	**92.0%**	8.0%	100%
	93.2%	21.4%	73.5%
[Similar other]			
	5	22	27
expected other to read	18.5%	**81.5%**	100%
	6.8%	**78.6%**	26.5%
	74	28	102
Total	72.6%	27.4%	100%
	100.0%	100.0%	100.0%

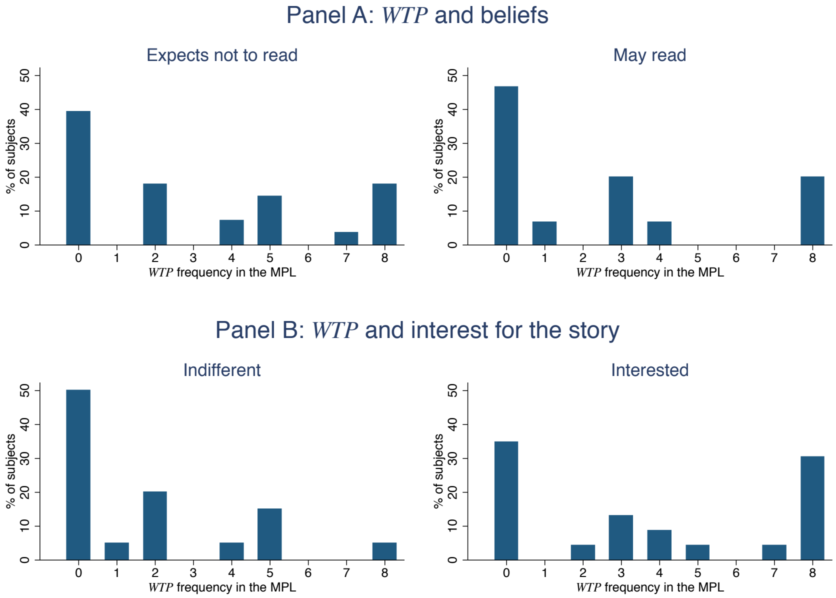
Notes: The categories “expected other not to read’ and “expected other to read” refer to the incentivized guess of a subject regarding the choice from $\{0, 1\}$ made in Period 2 by someone with the same rank ordering as them. The category “said unlikely to read” (“said likely to read”) includes subjects who reported being quite or very unlikely (likely) to read the story if offered {0,1} in Period 2; subjects who reported being “unsure” (18/120) are excluded. Fisher’s exact test gives $p < 0.001$.

Figure 4: Distribution of answers to the unincentivized belief measure among the SSB_{-0} subjects



Notes: Distribution of answers of the SSB_{-0} subjects ($N = 43$) to the unincentivized belief question "How likely are you to choose to learn the selected story in Period 2 if given the chance?" (answers: *very unlikely*, *quite unlikely*, *unsure*, *quite likely*, *very likely*).

Figure 5: Distribution of WTP for $\{0\}$ as a function of beliefs and interest for the story



Notes: Distributions of the number of rows (out of 8) in the Multiple Price List at which the SSB_{-0} subjects ($N = 43$) preferred to pay to replace $\{0, 1\}$ with $\{0\}$. In Panel A, a subject belongs to the category "Expects not to read" ("May read") if he reported (did not report) being very unlikely to read the story. In Panel B, a subject is classified as "Interested" if he reported being at least somewhat interested in learning the most incredible story among others and/or learning whether his own story was selected, and classified as "Indifferent" otherwise.

References

Acland, D. J. and M. Levy (2015): "Naiveté, Projection Bias, and Habit Formation in Gym Attendance," *Management Science*, 61, 146–160.

Ahn, D. S., R. Iijima, Y. Le Yaouanq, and T. Sarver (2017a): "Behavioral Characterizations of Naiveté for Time-Inconsistent Preferences," Working paper.

Ahn, D. S., R. Iijima, and T. Sarver (2017b): "Naiveté about Temptation and Self-Control: Foundations for Naive Quasi-Hyperbolic Discounting," Working paper.

Ahn, D. S. and T. Sarver (2013): "Preference for Flexibility and Random Choice," *Econometrica*, 81, 341–361.

Alan, S. and S. Ertac (2015): "Patience, Self-Control, and the Demand for Commitment: Evidence from a Large-Scale Field Experiment," *Journal of Economic Behavior and Organization*, 115, 111–122.

Ameriks, J., A. Caplin, J. Leahy, and T. Tyler (2007): "Measuring Self-Control Problems," *American Economic Review*, 97, 966–972.

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, 635–672.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 130, 1067–1115.

Augenblick, N. and M. Rabin (2017): "An Experiment on Time Preference and Misprediction in Unpleasant Tasks," *Review of Economic Studies*, forthcoming.

Bai, L., B. Handel, T. Miguel, and G. Rao (2017): "Self-Control and Chronic Illness: Evidence from Commitment Contracts for Doctor Visits," Working paper.

Baumeister, R. F., T. F. Heatherton, and D. M. Tice (1994): *Losing control: how and why people fail at self-regulation*, San Diego: Academic Press.

Baumeister, R. F. and K. D. Vohs (2003): "Willpower, Choice, and Self-Control," in *Time and Decision: Economic and psychological perspectives on intertemporal choice*, ed. by G. Loewenstein, D. Read, and R. F. Baumeister, New York: Russell Sage Foundation.

Bonein, A. and L. Denant-Boèmont (2015): “Self-Control, Commitment and Peer Pressure: A Laboratory Experiment,” *Experimental Economics*, 18, 543–568.

Bucciol, A., D. Houser, and M. Piovesan (2015): “Temptation at Work,” *PLoS ONE*, 8, 1–5.

Casari, M. and D. Dragone (2015): “Choice reversal without temptation: A dynamic experiment on time preferences,” *Journal of Risk and Uncertainty*, 50, 119–140.

Chatterjee, K. and R. V. Krishna (2009): “A "Dual Self" Representation for Stochastic Temptation,” *American Economic Journal: Microeconomics*, 1, 148–67.

Dang, J. (2016): “Commentary: A Multilab Preregistered Replication of the Ego-Depletion Effect,” *Frontiers in Psychology*, 7.

Dean, M. and J. McNeill (2015): “Preference for Flexibility and Random Choice: an Experimental Analysis,” Working paper.

Dekel, E. and B. L. Lipman (2012): “Costly Self-Control and Random Self-Indulgence,” *Econometrica*, 80, 1271–1302.

Dekel, E., B. L. Lipman, and A. Rustichini (2001): “Representing Preferences with a Unique Subjective State Space,” *Econometrica*, 69, 891–934.

——— (2009): “Temptation-Driven Preferences,” *Review of Economic Studies*, 76, 937–971.

Dekel, E., B. L. Lipman, A. Rustichini, and T. Sarver (2007): “Representing Preferences with a Unique Subjective State Space: A Corrigendum,” *Econometrica*, 75, 591–600.

Diamond, P. A. and J. A. Hausman (1994): “Contingent Valuation: Is Some Number Better than No Number?” *Journal of Economic Perspectives*, 8, 45–64.

Duckworth, A. and J. Gross (2014): “Self-control and grit: Related but separable determinants of success,” *Current Directions in Psychological Science*, 23, 319–325.

Duflo, E., M. Kremer, and J. Robinson (2011): “Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya,” *American Economic Review*, 101, 2350–90.

Dunbar, R. I. M. (2004): “Gossip in Evolutionary Perspective,” *Review of General Psychology*, 8, 100–110.

Eliaz, K. and R. Spiegler (2006): "Contracting with Diversely Naive Agents," *Review of Economic Studies*, 73, 689–714.

Ellingsen, T. and M. Johannesson (2009): "Time Is Not Money," *Journal of Economic Behavior and Organization*, 72, 96–102.

Ent, M. R., R. F. Baumeister, and D. M. Tice (2015): "Trait self-control and the avoidance of temptation," *Personality and Individual Differences*, 74, 12–15.

Fudenberg, D. and D. K. Levine (2006): "A Dual-Self Model of Impulse Control," *American Economic Review*, 96, 1449–1476.

——— (2012): "Timing and Self-Control," *Econometrica*, 80, 1–42.

Goldstein, N. J. and R. B. Cialdini (2007): "The spyglass self: A model of vicarious self-perception," *Journal of Personality and Social Psychology*, 92, 402–417.

Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403–1435.

——— (2007): "Harmful Addiction," *Review of Economic Studies*, 74, 147–172.

Hagger, M. and N. Chatzisarantis (2016): "A multilab preregistered replication of the ego-depletion effect," *Perspectives on Psychological Science*, 11, 546–573.

Houser, D., D. Schunk, J. K. Winter, and E. Xiao (2018): "Temptation and Commitment in the Laboratory," *Games and Economic Behavior*, 107, 329–344.

John, A. (2015): "When Commitment Fails - Evidence from a Regular Saver Product in the Philippines," EOPP Discussion Papers 55, London School of Economics.

Karlan, D. and J. Zinman (2009): "Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment," *Econometrica*, 77, 1993–2008.

Kaur, S., M. Kremer, and S. Mullainathan (2015): "Self-Control at Work," *Journal of Political Economy*, 123, 1227–1277.

Kopylov, I. (2012): "Perfectionism and Choice," *Econometrica*, 80, 1819–1843.

Kreps, D. (1979): "A Representation Theorem for Preference for Flexibility," *Econometrica*, 47, 565–576.

Krusell, P., B. Kuruşçu, and A. A. Smith (2009): "How Much Can Taxation Alleviate Temptation and Self-Control Problems?" Working paper.

——— (2010): "Temptation and Taxation," *Econometrica*, 78, 2063–2084.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112, 443–478.

Lipman, B. and W. Pesendorfer (2013): "Temptation," in *Advances in Economics and Econometrics: Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, Cambridge University Press, vol. 1.

Milkman, K., J. Minson, and K. Volpp (2014): "Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling," *Management Science*, 60, 283–299.

Noor, J. (2011): "Temptation and Revealed Preference," *Econometrica*, 79, 601–644.

Noor, J. and N. Takeoka (2010): "Uphill self-control," *Theoretical Economics*, 5, 127–158.

O'Donoghue, T. and M. Rabin (1999): "Doing It Now or Later," *American Economic Review*, 89, 103–124.

Read, D. and B. Van Leeuwen (1998): "Predicting hunger: The effects of appetite and delay on choice," *Organizational behavior and human decision processes*, 76, 189–205.

Ross, L., D. Greene, and P. House (1977): "The False Consensus Phenomenon: An Attributional Bias in Self-Perception and Social Perception Processes," *Journal of Experimental Social Psychology*, 13, 279–301.

Royer, H., M. Stehr, and J. Sydnor (2015): "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company," *American Economic Journal: Applied Economics*, 7, 51–84.

Rubinstein, A. and Y. Salant (2016): ""Isn't everyone like me?" On the presence of self-similarity in strategic interactions," *Judgment and Decision Making*, 11, 168–173.

Sadoff, S., A. Samek, and C. Sprenger (2015): "Dynamic Inconsistency in Food Choice: Experimental Evidence from a Food Desert," Working paper.

Schilbach, F. (2017): "Alcohol and Self-Control: A Field Experiment in India," Working paper.

Strotz, R. H. (1956): “Myopia and Inconsistency in Dynamic Utility Maximization,” *Review of Economic Studies*, 23, 165–180.

Toussaert, S. (2016): “Connecting commitment to self-control problems: Evidence from a weight loss challenge,” Working paper.