UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2019

Pessimism and Overcommitment

Claes Ek, University of Gothenburg
Margaret Samahita, University College Dublin

WP19/21

September 2019

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Pessimism and Overcommitment[*]

Claes Ek[†] Margaret Samahita[‡]

September 13, 2019

Abstract

Economic agents commonly use commitment devices to limit impulsive behavior in the interest of long-term goals. We provide evidence for excess demand for commitment in a laboratory experiment. Subjects are faced with a tedious productivity task and a tempting option to surf the internet. Subjects state their willingness-to-pay for a commitment device that removes the option to surf. The commitment device is then allocated with some probability, thus allowing us to observe the behavior of subjects who demand commitment but have to face temptation. We find that a significant share of the subjects overestimate their demand for commitment when compared to their material loss from facing the temptation. This is true even when we take into account the potential desire to avoid psychological costs from being tempted. Assuming risk aversion does not change our conclusion, though it suggests that pessimism in expected performance, rather than psychological cost, is the main driver of overcommitment. Our results suggest there is a need to reconsider the active promotion of commitment devices in situations where there is limited disutility from the tempting option.

JEL classification: C91, D03, D91.
Keywords: Commitment device, pessimism, self-control.

1 Introduction

Economic agents commonly use commitment devices to limit future choices. For example, people pay for long-term gym memberships hoping that the sunk cost would motivate

[*] We thank Giuseppe Attanasi and participants at the SABE ECR Workshop (Dublin) for helpful comments and suggestions. We also thank Miloš Fišar and the team at Masaryk University Experimental Economics Laboratory (MUEEL) for their assistance during the experimental sessions. Financial support from the Torsten Söderberg Foundation [grant number E25/18] is gratefully acknowledged.

[†] Department of Economics, University of Gothenburg. E-mail: claes.ek@economics.gu.se

[‡] School of Economics, University College Dublin. E-mail: m.samahita@gmail.com

them to exercise (DellaVigna and Malmendier, 2006), forgo quantity discounts by buying smaller packs of junk food to limit consumption (Wertenbroch, 1998), and bind themselves to inflexible savings accounts to stop themselves from spending too much (Ashraf et al., 2006). Like the classical story of Odysseus tying himself to the mast, these are attempts at resisting impulsive behavior in the interest of long-term goals. While there has been a lot of effort studying and encouraging the adoption of commitment devices, it is an empirical question whether or not commitment is over- or underdemanded in different situations. For example, demand for the inflexible savings account could be said to be excessive if it becomes a hassle for the agent to manage their money and if the agent would have been able to resist spending with a more flexible account anyway.[1] Thus, in general, interventions designed to promote commitment may have opposing welfare implications, depending on the relative share of over- and underdemanding agents that they induce in a population. The aim of our paper is to examine whether or not excess demand for commitment is a widespread phenomenon.

We conduct a laboratory experiment where we study the demand for a commitment device that removes a tempting option to surf the internet during a tedious productivity task. Subjects state their willingness-to-pay (WTP) for the commitment device, which is then allocated with some probability, thus allowing us to observe the behavior of subjects who demand commitment but have to face temptation. We also elicit subjective beliefs regarding expected self- and peer productivity, as well as the *ex ante* expected and *ex post* actual experienced difficulty of resisting temptation. These measures allow us to decompose subjects' valuation of the commitment device into expected material loss from being exposed to temptation and any non-material 'psychological costs', such as the mental burden of maintaining self-control while tempted.

We find that 23% of subjects overestimate their demand for commitment when compared to their actual material loss from facing the temptation. By contrast, and perhaps surprisingly, fewer than 5% underestimate their demand for commitment by the same measure. Even when we take into account both material loss and psychological costs, WTP is still overestimated by a significant share of subjects at 17%. While these figures are based on the assumption that subjects are risk-neutral, assuming very strong risk aversion still yields 16% of subjects who overestimate WTP relative to material loss. This is driven by subjects' pessimism regarding their productivity under temptation, as WTP appears to accurately capture subjects' *expected* material loss. When psychological cost is considered as well, the lower bound on the number of overestimators is not much lower at 13%.

[1] The same goes for long-term gym memberships for any agent who is only interested in short-term membership but does not require a sunk cost to feel motivated, and for the junk food consumer who may well have been able to limit her portion size and thus unnecessarily forgoes the monetary savings from buying in bulk.

The literature on self-control identifies two main reasons why people may demand commitment (see Bryan et al. (2010) for a review). The first of these is because of present-biased preferences, whereby decision-makers have a strong bias towards current outcomes and heavily discount future benefits and costs relative to now with a multiplicative factor $\beta < 1$, the present-bias parameter (Laibson, 1997). The decision-maker can be either sophisticates, who know their true $\beta < 0$, naifs, who think their $\beta = 1$, or partially naïve who fall somewhere in between (O'Donoghue and Rabin, 1999). Sophisticates and partially naïve agents are aware of their self-control problems and will thus have an *ex ante* positive demand for a commitment device that prevents choice reversal in the future.

A second reason for demanding commitment is captured in the model of menu-dependent utility (Gul and Pesendorfer, 2001): agents face a psychological cost for being exposed to the most tempting item in a choice set, regardless of whether they expect to actually choose the tempting item or not. Agents may thus choose to restrict their choice set in order to avoid the tempting item. Such behavior was recently demonstrated in the lab by Toussaert (2018), where around a quarter to a third of subjects preferred a restricted menu without the tempting option to the full choice set, and the full menu to the tempting option by itself. Strikingly, even the subjects that successfully resisted temptation when confronted with the full menu exhibited such preferences, suggesting they anticipated psychological costs of resisting temptation.

In either theory, it is possible to imagine situations where the degree of commitment chosen by an agent is sub-optimal. So far, the literature has focused exclusively on explaining stylized facts that are consistent with *under*demand of commitment devices (John, forth., Heidhues and Kőszegi, 2009), such as too little pension savings (Thaler and Benartzi, 2004). Within the present-bias framework, Heidhues and Kőszegi (2009) model the behavior of a partially naïve agent who is overconfident (in the sense of overestimating β) and purchases too little commitment, thus incurring a cost of purchasing commitment while not being prevented from taking the tempting but harmful option anyway. For a real-world example of this, see DellaVigna and Malmendier (2006): agents are overconfident about their gym attendance and buy 'too little' gym membership. The optimal 'amount' of gym membership, to ensure commitment to regular exercise, would have been one that perfectly enforces gym attendance. Myrseth and Wollbrant (2013) model the decision problem of a similarly naïve agent who is overconfident in the sense that he underestimates the strength of the impending temptation and thus mistakenly chooses to confront temptation instead of buying a commitment device.[2]

However, it does not follow that underdemand obtains in all situations and for all

[2] This model uses a belief parameter between 0 and 1 to capture the extent to which the agent is aware of the strength of the temptation, in contrast to O'Donoghue and Rabin (1999) where the belief parameter captures the awareness of the present-bias problem.

people. Agents may plausibly sometimes *over*estimate their demand for commitment, and hence commitment becomes sub-optimal not because it is insufficient, but because it is unnecessary.[3] Indeed, some of the theoretical models discussed above focusing on insufficient commitment also allow an analysis of overcommitment. In Heidhues and Kőszegi (2009), for instance, the agent with $\hat{\beta} < \beta < 1$ who is overly pessimistic about self-control purchases *too much* commitment (see Figure 1 in Heidhues and Kőszegi (2009)). Furthermore, the theoretical model of Myrseth and Wollbrant (2013) admits situations where the agent demands commitment *at all* even when it is unnecessary. If an agent overestimates the strength of the temptation or underestimates his own willpower,[4] he will sub-optimally choose commitment despite the fact that his expected utility from facing temptation (even with some probability of succumbing) is higher than the utility from using the commitment device.[5] These extensions, however, are only implicit in the cited papers; as far as we are aware, ours is the first study of potential overdemand for commitment.

We describe the experimental setting in Section 2 and derive our hypotheses in Section 3. The results are presented in Section 4 and Section 5 concludes.

2 Experimental Design

The experiment was conducted at Masaryk University Experimental Economics Laboratory (MUEEL) in Brno, Czech Republic during the period 27-30 May 2019 and programmed using z-Tree (Fischbacher, 2007). Participants were recruited from the laboratory subject pool consisting of students at Masaryk University. In total we ran 12 sessions with 289 subjects. Each session lasted around 2 hours and average earnings were 707 CZK per subject, including 100 CZK participation fee.

The experiment consists of two stages. In Stage 1, subjects complete an attention task without temptation (Task 1). At the start of Stage 2, subjects learn that they will complete the same task again but with temptation (Task 2). They are told about the possibility

[3] In a sense, the partially naïve agent who purchases too little commitment to fully prevent him from choosing the tempting option can also be said to overestimate his demand for commitment, since he buys too much relative to the resulting outcome: he could have done as the fully naïve does and purchase no commitment at all, since they both end up giving in to temptation. In this paper we focus instead on those who buy more commitment than is sufficient to achieve the goal of resisting the temptation.

[4] We use the terms "willpower" and "self-control" interchangeably in our setting.

[5] By buying commitment when it is unnecessary, he thus behaves as though he estimates $\hat{\beta} < 1$ when in actual fact $\beta = 1$ (or even $\beta > 1$). In a similar spirit, Bénabou and Tirole (2004) develop a theory where an agent who doubts his own willpower instead adopts personal rules. In their setting, overregulation using these personal rules can make agents behave "*as though* they overweighed the future rather than the present" (p. 850, emphasis in original).

to purchase a commitment device that removes the temptation. We elicit their WTP for this commitment device and their beliefs about productivity in the second attention task. Subsequently subjects complete the second attention task, followed by an exit survey. Subjects are informed before Task 1 that one of the two tasks will be randomly chosen for payment. This randomization is performed after Task 2. The use of within-subject design is motivated by wanting all subjects to have experienced the attention task prior to stating a WTP for the commitment device.[6]

2.1 The attention task, temptation, and commitment device

We use an attention task similar to that used in Toussaert (2018): for a period of 30 minutes, subjects are asked to pay attention to their computer screen where a four-digit number updates every third second. At five random times, they are prompted to enter the last number they saw, after which the number is reinitialized. Subjects can earn tokens for each correct answer at a rate of 120 tokens per question.[7] The potential earnings from this task are set to be relatively high to induce subjects to be interested in completing the task.

In Stage 1 subjects complete the attention task as described above (Task 1). At the start of Stage 2, subjects are informed that they will do the attention task again (Task 2), but this time there will be an additional button on the screen which allows internet access.[8] Clicking the internet access button means that the subject surfs the internet for the remainder of the period instead of continuing with the attention task. The subject will forfeit the chance to earn any more money from the attention task, but will retain any money earned from correct answers up until the point of clicking the internet access button. Subjects are thus aware that to get the highest possible monetary payoff they would have to exercise willpower to overcome the temptation. This temptation, as also used in Houser et al. (2018) and Bonein and Denant-Boèmont (2015), has immediate appeal given the tedious attention task and should be perceived to be bad since subjects choosing this option forfeit the possibility to earn more money from the attention task.[9]

We then offer subjects a commitment device: the possibility of paying to remove the internet access button which would guarantee participation in Task 2 for the whole 30-minute period. WTP for the commitment device is elicited using the Becker-DeGroot-Marschak (BDM) mechanism as depicted in Figure 1. Subjects state a price between 0 and 100 tokens representing the maximum price they are willing to pay to remove the option

[6] Instructions are included in Section D in the Appendix.

[7] 1 token corresponds to 1 CZK. 1 CZK corresponded to 0.039 EUR at the time of the experiment. The Czech minimum wage is 13,000 CZK, or 500 EUR, per month.

[8] A screenshot is included in Section D in the Appendix.

[9] As revealed in subjects' feedback elicited at the end of the experiment, they appear to consider the internet as a temptation to be avoided.

to surf. The computer will then simulate a coin toss. If Heads comes up, the internet button continues to be present regardless of the subject's WTP. Only if Tails comes up will the WTP be taken into account. The computer will draw a random number between 0 and 100 and if this number is less than or equal to the stated price then the internet option will be removed. Hence the probability of getting the commitment device to remove the option to surf is equal to $WTP/200$ and increases linearly with subject WTP, up to a value of 0.5.[10]

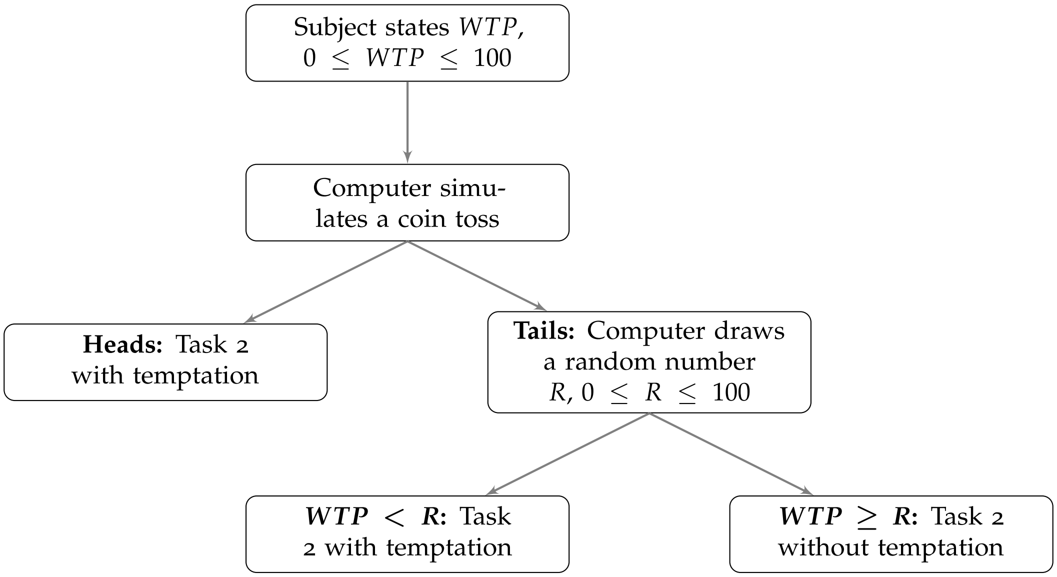

Subject states WTP, $0 \leq WTP \leq 100$
Computer simulates a coin toss
Heads: Task 2 with temptation
Tails: Computer draws a random number R, $0 \leq R \leq 100$
$\boldsymbol{WTP < R}$: Task 2 with temptation
$\boldsymbol{WTP \geq R}$: Task 2 without temptation

Figure 1: Elicitation of WTP for commitment device

WTP is elicited twice. The initial measure WTP_0 is not used in our main analysis. The second elicitation (WTP_1) occurs after subjects have been asked to reflect on their own productivity (as will be described in the next section), which may promote more accurate preferences. WTP_1 will later be compared against various measures of subjects' productivity in the attention task.

2.2 Measures of subjects' productivity

The subject's actual productivity is measured as y_1 and y_2, the number of correct answers in Tasks 1 and 2 respectively. We also elicit unincentivized subject beliefs about their productivity. Directly after the conclusion of Task 1, we ask subjects how many questions

[10] The use of a random implementation rule is motivated by the following objectives: to ensure incentive-compatibility, to observe the performance under temptation of subjects who have a positive demand for commitment, and to maximize the number of subjects who in fact face temptation. See other uses in Karlan and Zinman (2009), Augenblick et al. (2015) and Toussaert (2018).

they expect to answer correctly if they were to redo Task 1, where there was **n**o **t**emptation (y_s^{nt}); this is followed by the elicitation of WTP_0.[11] Next, we ask how many questions they expect to answer correctly if the **t**emptation is present in Task 2 (y_s^t). We decompose this value and ask subjects to state the percentage likelihood that they would succumb to temptation (p_s), and how many questions they expect would be answered correctly if they **s**uccumb (y_s^s) or **n**ot **s**uccumb to the temptation (y_s^{ns}).[12] This allows us to construct a second measure of expected productivity $\hat{y}_s^t = p_s(y_s^s) + (1 - p_s)(y_s^{ns})$. While we expect that $y_s^t = \hat{y}_s^t$ on average, the two measures allow us to check for the presence and source of misestimation of performance. We do not incentivize these measures of self-productivity to prevent subjects hedging with their stated beliefs against adverse performance in Task 2.

After the above variables have been measured, we repeat the WTP elicitation, framing it as an option to revise the amount stated previously. We explain that this second measure (WTP_1, henceforth WTP) will be the one that determines whether the internet button is present in Task 2.

Next, in order to obtain a payment-contingent measure of productivity, we elicit and incentivize subjects' belief of the performance of a similar peer.[13] Each subject is matched with another participant whose WTP is closest to the subject's own WTP. Subjects with a positive WTP first estimate how many questions this peer will answer correctly without temptation. Then, all subjects estimate how many questions the peer will answer correctly if the peer faces temptation and succumbs (y_p^s) or does not succumb to temptation (y_p^{ns}). The subject is paid 20 tokens if their answer matches the actual outcome of the chosen peer. Finally, the subject is matched to a group of five participants whose WTP values are closest to the subject's own WTP and who face temptation in Task 2. The subject is asked to estimate how many of these five participants would succumb to temptation and press the internet button (n_p). The third measure of expected productivity is then $\hat{y}_p^t = (n_p/5)(y_p^s) + (1 - n_p/5)(y_p^{ns})$.

[11] The subscript s refers to **s**elf-productivity, as contrast to **p**eer productivity whose elicitation is described later.

[12] There are some potential concerns here. Although we ask subjects to state a single expected value, it may be that they instead consider a distribution of cases, across which they are risk-averse; or report some other statistic, such as a modal outcome; or round values up or down. Such 'behind-the-scenes' considerations may imply that their behavior will appear less rational than it actually is. In this paper, we make the assumption that subjects do state a single expected value when asked to do so, and that this expected value can be treated as certain conditional on the outcome of the coin toss and the BDM mechanism (and, in Section 4.3 and for y_s^s, y_s^{ns}, additionally whether or not the subject succumbs to temptation).

[13] See, for example, Gächter and Renner (2010) who find that incentivizing beliefs significantly improves accuracy.

2.3 Other variables

To better understand subjects' estimation of the psychological cost of temptation, we elicit the perceived strength of the temptation (θ) before Task 2 with the question "How tempted do you think you would be by internet access?". Subjects respond on a scale from 1 to 4 (not at all tempted, not that tempted, quite tempted, very tempted). Subjects' actual experience of the temptation is elicited in the post-experiment questionnaire, where those who did face temptation are asked to respond whether they think the difficulty of ignoring the internet button was easier, more difficult, or neither easier nor more difficult than expected. We derive the variable v, representing actual temptation, and set it equal to 1 for those who respond that ignoring internet access was easier than expected, that is, who overestimated the psychological cost of temptation. We set v equal to -1 for whom the temptation was more difficult than expected to ignore, and 0 otherwise.

Subjects then proceed to the second stage where they do the attention task with or without the internet option, as determined by the outcome of the coin toss and BDM mechanism.

In the post-experiment questionnaire we also elicit subjects' perception of their willpower using the brief self-control measure (Tangney et al., 2004). This questionnaire consists of 13 statements, to each of which the subject indicates their agreement on a five-point scale. These values are aggregated to give ω, the perceived general level of willpower. We collect data on demographic variables such as age, gender, degree program (1 for Bachelor, 2 Master, 3 PhD), major, and GPA. Finally, we include an optional question asking subjects to comment on their choice of WTP in order to better understand their motivation.

3 Hypotheses

Our aim is to investigate whether a substantial share of subjects overstate their WTP to remove temptation. WTP can be decomposed as the sum of expected material losses due to productivity reduction, either from succumbing to temptation or purely from being tempted (e.g. if devoting cognitive resources to self-control reduces productivity in the task); and non-material psychological costs from facing temptation.

This can be shown within a simple expected-utility model. Recall that Task 2 is selected for payment with probability 1/2, and that in the BDM mechanism, the probability of commitment is $WTP/200$. Furthermore, payments for correct answers in Task 2 are made only if that task is selected for payment, whereas any positive WTP may be paid by subjects regardless of the task selected.[14] We assume that Bernoulli utility is additively separable, so $u = u(x + PC)$, with wealth x and psychological costs PC. Then, for some fixed (possibly accurate) expectations on earnings with and without temptation (y^t, y^{nt}), subjects are taken to maximize expected utility as

$$\begin{aligned} U(WTP) = & \frac{1}{2}\left[\frac{WTP}{200}u(100 + 120y_1 - WTP) + \left(1 - \frac{WTP}{200}\right)u(100 + 120y_1 - PC)\right] \\ & + \frac{1}{2}\left[\frac{WTP}{200}u(100 + 120y^{nt} - WTP) + \left(1 - \frac{WTP}{200}\right)u(100 + 120y^t - PC)\right] \end{aligned}$$

and the solution under risk neutrality is given by[15]

$$WTP = 30(y^{nt} - y^t) + \frac{PC}{2} \tag{1}$$

The first term (expected material loss) is the subject's expected productivity without temptation less their expected productivity with temptation.[16] The model is silent on whether any positive difference $y^{nt} - y^t$ arises from succumbing to temptation or from performing worse due to having to exercise more self-control when the internet button is present, even if it is never clicked. For example, in an exploratory analysis of the effort task of Toussaert (2018), subjects who were exposed to temptation despite preferring to commit to a temptation-free choice set were found less productive, suggesting that self-control is indeed costly. In any case, these values are elicited in the experiment, as noted in Section 2.2. The second term (expected psychological cost) is likely to consist of (i) the effort of resisting temptation as well as (ii) any self-image loss should the subject succumb to temptation. The former is only elicited inexactly, using θ, and the latter component is entirely unknown.

Similarly, concerning actual values (i.e. if expectations are accurate), only actual material loss can be measured in the experiment while the true psychological costs cannot. Our approach is to ask the subjects whether they think resisting temptation was easier than expected, yielding v. If $v \geq 0$ this would indicate that psychological costs were no larger than expected (though this measure admittedly misses the self-image loss from succumbing).

[14] Since this is not made explicit in the instructions, there is a possibility that subjects assume that WTP is only paid if Task 2 is chosen for payment. We show that the analysis based on this assumption is unchanged in Section A.1 in the Appendix.

[15] The case of risk aversion is considered in Section 4.3.

[16] The multiplication of material loss $60(y^{nt} - y^t)$ and psychological cost PC by half in equation (1) reflects the probabilistic nature of the BDM mechanism: the likelihood of having to actually pay for the button increases with stated WTP.

The analysis has the following steps.[17]

1. We begin by comparing WTP with **actual** performance in the attention tasks. As suggested above, subjects might overstate WTP relative to actual material loss because they (i) inaccurately estimate future material payoffs or if (ii) they (correctly or not) anticipate non-material psychological costs of facing temptation. In either case, the WTP to remove the internet button will be greater than material losses, which we are able to observe for the sub-sample of subjects who face temptation in Task 2.

   Thus, among the sub-sample of subjects who face temptation in Task 2, we classify subjects according to whether they exhibit $WTP > 30(y_1 - y_2)$ or $WTP < 30(y_1 - y_2)$ or neither, and test the following null hypothesis. All tests of proportions are based on the standard normal approximation of binomial parameters.

   Hypothesis 1a. *Among the subjects who face temptation in Task 2, no more than 10% have* $WTP > 30(y_1 - y_2)$.[18]

   If this hypothesis is rejected, we conclude that a 'substantial' share of subjects overestimate WTP compared to material losses.

   Hypothesis 1 compares performance in Task 2 under temptation with the same subjects' performance in Task 1, where there was no temptation. Thus, performance in Task 1 is effectively used as the counterfactual. Although we cannot check the validity of this approach directly, we attempt to provide supporting evidence by testing the following hypothesis. All tests involving means are based on the standard t test for the equality of two population means or, whenever at least one group has fewer than 30 observations, the non-parametric Wilcoxon rank-sum test.

   Hypothesis 1b. *Among the sub-sample of subjects who do* not *face temptation in Task 2,* $\bar{y}_1 = \bar{y}_2$.

   Henceforth, all bars denote average values across the relevant set of subjects. For example, $\bar{y}_1$ denotes the average number of correct answers in Task 1.

2. Next, we compare WTP with **expected** performance in the attention tasks. As already stated, we measure expected material losses from temptation in three different ways (y_s^t, $\hat{y}_s^t$, and $\hat{y}_p^t$). First, we perform the following preliminary test to see whether it is worthwhile looking separately at the first two measures.

[17] Our hypotheses were pre-specified in an analysis plan that was pre-registered online: see `https://osf.io/h7x9v/`. All differences between the plan and the analysis in this paper will be clearly stated.

[18] 10% is assumed to be the share that can be attributed to subject confusion. Our results for Hypotheses 1-3 are robust at the 5% level of significance assuming up to 15% of subjects overestimating due to confusion.

Hypothesis 2a. $\bar{y}_s^t = \bar{\hat{y}}_s^t$ *in the sub-sample of subjects who face temptation in Task 2.*

We then test:

Hypothesis 2b. *Among the subjects who face temptation in Task 2, no more than 10% have* $WTP > 30(y_s^{nt} - y_s^t)$.

If Hypothesis 2a is rejected, we additionally test:

Hypothesis 2c. *Among the subjects who face temptation in Task 2, no more than 10% have* $WTP > 30(y_s^{nt} - \hat{y}_s^t)$.

And in any case, we also repeat the procedure for our incentivized 'peer' measure:

Hypothesis 2d. *Among the subjects who face temptation in Task 2, no more than 10% have* $WTP > 30(y_s^{nt} - \hat{y}_p^t)$.

Thus, we perform either two or three tests of the same hypothesis, using different outcome variables. To adjust for this, p values for Hypotheses 2b-2d will be corrected for multiple hypothesis testing. We expect p values for different expectation measures to be correlated and are interested in being able to pinpoint the overall significance of several small effects rather than a single large one. Because of this, we will use the approach proposed by Brown (1975), which is a version of Fisher's method appropriate to our setting. This method requires the analyst to specify correlations between different outcomes; we will calculate and use the correlation matrix for the various dummy variables associated with overestimators.

3. As noted, a WTP higher than expected material losses may reflect expected psychological costs as well as 'true' overestimation (of both material losses and psychological costs).[19] Because the actual PC is unknown, we cannot test directly whether stated WTP is larger than the entire realized right-hand side of equation (1). However, an indirect test is possible. Starting from equation (1) and denoting expected quantities by subscript e and actual values (i.e. accurate expectations) by subscript a, true overestimation in the expected-utility model with risk neutrality would be characterized by

$$\begin{aligned} WTP\,(\cdot_e) > WTP\,(\cdot_a) &\iff 30(y_e^{nt} - y_e^t) + \frac{PC_e}{2} > 30(y_a^{nt} - y_a^t) + \frac{PC_a}{2} \\ &\iff 60(y_e^{nt} - y_e^t - (y_a^{nt} - y_a^t)) > -(PC_e - PC_a) \qquad (2) \end{aligned}$$

[19] Subjects who underestimate WTP relative to expected material loss thus have implied $PC_e < 0$. Using y_s^t, these subjects account for 13% of those facing temptation, and 31% of them are overestimators as defined in Hypothesis 3.

Thus the overestimation of material losses needs to exceed any *under*estimation of psychological costs. The above condition can be satisfied in two ways. For subjects with $v = 1$, indicating that $PC_e > PC_a$, the RHS of the inequality is negative, implying that a sufficient condition for overestimation is that the LHS is greater than or equal to 0. Second, for subjects who find that ignoring the temptation was exactly as easy or difficult as expected ($v = 0$), the RHS of the above inequality is zero, implying that a sufficient condition for overestimation is that the LHS is strictly greater than 0.[20] Testing this on the sub-sample of subjects who are exposed to temptation, we set $y_a^t = y_2$; however, we also need to choose an appropriate counterfactual y_a^{nt}. We suggest to use $y_e^{nt} = y_a^{nt} = y_s^{nt}$, in which case the LHS reduces to $y_2 - y_e^t$.[21] Additionally, given that subjects cannot state negative WTP, $WTP\left(\cdot_e\right) > 0$ is a necessary condition for overestimation. Note that Hypothesis 3b is tested only if Hypothesis 2a is rejected.

Hypothesis 3a. *Among the subjects who face temptation in Task 2, no more than 10% have* $y_2 - y_s^t \geq 0$ *and* $v \geq 0$*, with at least one strict inequality, and* $WTP > 0$.

Hypothesis 3b. *Among the subjects who face temptation in Task 2, no more than 10% have* $y_2 - \hat{y}_s^t \geq 0$ *and* $v \geq 0$*, with at least one strict inequality, and* $WTP > 0$.

Hypothesis 3c. *Among the subjects who face temptation in Task 2, no more than 10% have* $y_2 - \hat{y}_p^t \geq 0$ *and* $v \geq 0$*, with at least one strict inequality, and* $WTP > 0$.

As with Hypotheses 2b-2d, p values for Hypotheses 3a-3c will be corrected for multiple hypothesis testing by the Brown (1975) method. Rejection suggests that a substantial share of subjects have overestimated WTP; however, this holds only if it is valid to use y_s^{nt} as counterfactual. While we cannot test this assumption directly, we attempt to support it by testing:

Hypothesis 3d. *Among the sub-sample of subjects who do* not *face temptation in Task 2,* $\bar{y}_2 = \bar{y}_s^{nt}$.

Only if this is not rejected while Hypotheses 3a-3c are rejected (adjusting for multiple tests), do we conclude that a substantial share of subjects have overestimated WTP.

4. We hypothesize that subjects' pessimism, in the form of their perception of how strong the temptation will be and of their own willpower, will drive the overesti-

[20] Note that this implies a lower bound on the number of overestimators.

[21] Our results are robust to using y_1 in place of y_a^{nt}.

mation of WTP.[22] We test these ideas within a regression framework by estimating variants of[23]

$$overestimator_s = \beta_1 + \beta_2\theta_s + \beta_3\omega_s + \beta_4 y_{1s} + \beta_5 y_s^{nt} + \boldsymbol{\beta}'\mathbf{X}_s + \epsilon_s \tag{3}$$

where *overestimator* is, as derived from (2) above, a dummy which equals 1 if the subject states $WTP > 0$ and satisfies ($y_2 - y_s^t \geq 0$ and $v = 1$) or ($y_2 - y_s^t > 0$ and $v = 0$), or 0 otherwise. $\boldsymbol{X}$ is a set of subject-specific demographic variables. Subjects who are low performers (low y_1) or generally expect low performance even without temptation (low y_s^{nt}) are hypothesized to have higher WTP for the commitment aid, and hence also more likely to be overestimators. All variants of regression (3) will use only data from subjects who face temptation in Task 2 and (separately) test the null hypotheses:

Hypothesis 4a. $\beta_2 = 0$ *(effect of perceived temptation).*

Hypothesis 4b. $\beta_3 = 0$ *(effect of perceived willpower).*

4 Results

4.1 Main results

Summary statistics from the experiment are presented in Table 1.

The majority of subjects have little willingness to pay for the commitment device, consistent with previous studies such as Augenblick et al. (2015). 211 subjects (73%) state $WTP = 0$. The average WTP is 6.94 tokens, or 25.73 for those with positive WTP. The distribution of positive WTP is given in Figure 2.[24] In total, 12 subjects are successful in getting the temptation removed.[25] Overall, subjects are successful in resisting temptation even without the commitment device, as only 4 subjects decide to access the internet. It appears that subjects are sufficiently incentivized to complete the tedious (but easy) task, 87% of subjects get all 10 correct answers in Task 1 and Task 2 combined.

[22] Our original analysis plan used the excess in material loss and psychological cost as outcome variables, however we consider the current specification to be better able to answer the question we are studying. Results for the original specifications are included in Section B in the Appendix.

[23] Where relevant, we have added subject subscripts s to conform with standard regression notation.

[24] Subjects' final WTP values, which we have used for the analysis, are not significantly different from those elicited before the questions about expected self-performance (6.94 vs 7.78, $p = 0.1885$). Our results are robust to using WTP_0 instead of WTP_1.

[25] Due to a coding error, it was possible to get the commitment device despite bidding 0 if the computer also drew the number 0. In our sample, a single subject received commitment in this way.

Table 1: Summary statistics.

Variable	N	Mean	SD	Min	Max
WTP, final stated maximum price (in tokens) for removing internet button	289	6.94	18.11	0	100
Actual self-productivity					
y_1, number of correct answers in Task 1	289	4.92	0.38	0	5
y_2, number of correct answers in Task 2 if temptation is NOT present	12	4.75	0.45	4	5
y_2, number of correct answers in Task 2 if temptation is present	277	4.88	0.54	0	5
Beliefs about self-productivity in Task 2					
y_s^{nt}, predicted self-productivity if temptation is NOT present	289	4.72	0.62	0	5
y_s^t, predicted self-productivity if temptation is present	289	4.60	0.72	0	5
p_s, predicted likelihood of succumbing to temptation	289	0.09	0.15	0	1
y_s^s, predicted self-productivity if subject succumbs	289	3.27	1.25	0	5
y_s^{ns}, predicted self-productivity if subject does NOT succumb	289	4.69	0.59	2	5
$\hat{y}_s^t$, expected self-productivity if temptation is present	289	4.56	0.68	0.33	5
Beliefs about peer productivity in Task 2					
y_p^{nt}, predicted peer productivity if temptation is NOT present	78	4.58	0.57	3	5
y_p^s, predicted peer productivity if peer succumbs	289	3.28	1.18	0	5
y_p^{ns}, predicted peer productivity if peer does NOT succumb	289	4.73	0.48	3	5
n_p, predicted number of peers (out of 5) who succumb to temptation	289	0.87	1.08	0	5
$\hat{y}_p^t$, expected peer productivity if temptation is present	289	4.49	0.65	1.8	5
θ, how tempted subject expects to be	289	1.76	0.76	1	4
v, ignoring temptation was easier than expected	277	0.47	0.58	-1	1
ω, score on brief self-control scale (Tangney et al., 2004)	289	38.61	8.64	19	64
Age	289	23.03	2.71	18	36
Male	289	0.52	0.50	0	1
Degree	289	1.53	0.56	1	3
Econ major	289	0.37	0.48	0	1
GPA	289	1.99	0.52	1	4

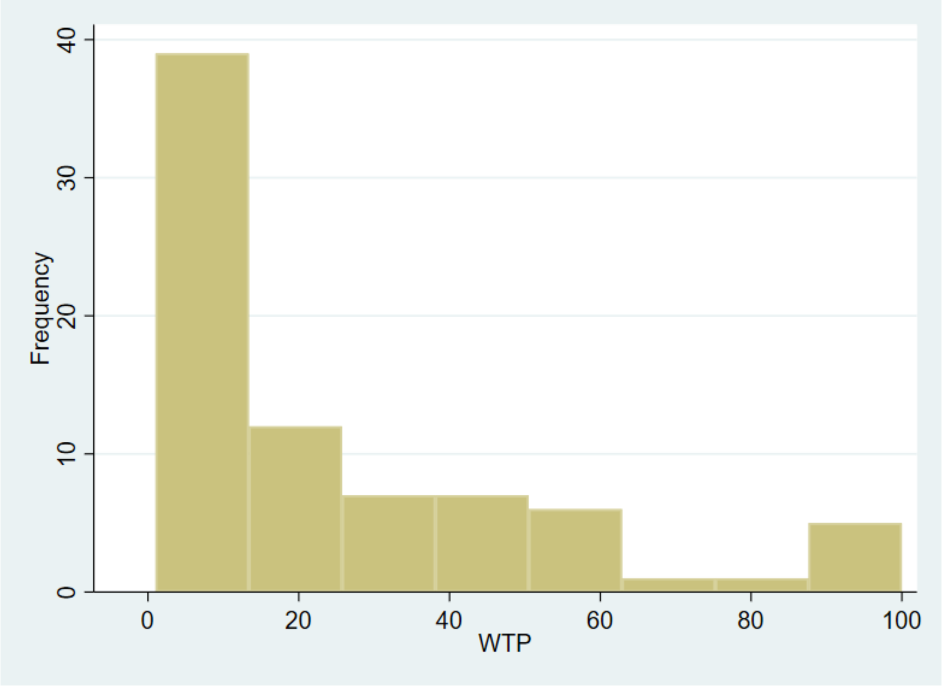


Figure 2: WTP for commitment device, for 27% of subjects stating $WTP > 0$

We start by comparing subjects' WTP for removing temptation with their *actual* productivity loss when exposed to temptation, for the 266 subjects who do not get the commitment device.[26] As described in Section 3, we classify subjects as overestimator, accurate estimator or underestimator according to whether their WTP is above, equal to, or below what would maximize utility when only material loss is considered (this is distinct from the *overestimator* dummy variable as used in Hypothesis 4, which also takes into account psychological cost). An accurate estimator would state $WTP = 30(y_1 - y_2)$. Table 2 summarizes these classifications. Around 71.8% of subjects accurately estimate their WTP for the commitment device, the vast majority unwilling to pay and not incurring material loss from being tempted.

As shown in Table 2, around 23.3% of subjects are overestimators, greater than the 10% attributed to confusion ($p < 0.0001$), thus rejecting Hypothesis 1a. As seen in the column $y_1 - y_2$, these overestimators on average even perform better in Task 2, thus making their positive WTP for commitment device (on average 18.94) seemingly irrational. Interestingly, and in contrast to the usual narrative of underdemand for commitment devices, there are more WTP overestimators than underestimators (two-sample test of proportion, $p < 0.0001$). Fewer than 5% of subjects underestimate their need for the

[26] We exclude 11 subjects with WTP equal to 0 and $y_1 < y_2$. According to equation (1) these subjects' WTP should have been negative but the experiment only allows for values greater than or equal to zero.

Table 2: Classification of subjects who face temptation.

Classification	N	Average WTP	$y_1 - y_2$	Frequency
Overestimator	62	18.94	-0.05	23.31%
$WTP > 0$	62	18.94	-0.05	
Accurate estimator	191	0.31	0.01	71.80%
$WTP > 0$	2	30	1	
$WTP = 0$	189	0	0	
Underestimator	13	2.54	1.62	4.89%
$WTP > 0$	3	11	1	
$WTP = 0$	10	0	1.8	
Total	266	4.76	0.08	100%

commitment device: they are unwilling to pay (or have low WTP) and yet, when facing temptation, perform worse by nearly 2 questions.

Given that we have used Task 1 performance as the counterfactual and assume that this is how the subject would have performed without temptation in Task 2, we check if this is indeed the case for those who do manage to buy their way out of temptation. Among these 12 subjects, we cannot reject Hypothesis 1b that $\bar{y}_1 = \bar{y}_2$ ($p = 0.5901$). Hence, we state our first result:

Result 1. *A substantial share of subjects overestimate WTP compared to actual material losses.*

Do subjects overestimate WTP because they are unable to accurately predict their performance in the attention task? To check this, we next compare WTP with subjects' *expected* performance in Task 2 as elicited in the experiment. We first check whether subjects are consistent in their estimate of y_s^t, their predicted performance in Task 2 when exposed to temptation, when elicited directly and when this value is decomposed into the number of correct answers in case of succumbing or resisting temptation and the corresponding probabilities (thus yielding $\hat{y}_s^t$). We do not reject Hypothesis 2a at the 5% level: among subjects who face temptation, $\bar{y}_s^t$ is not significantly different from $\bar{\hat{y}}_s^t$ (4.61 vs 4.56, $p = 0.0696$).[27] We will therefore use y_s^t and the incentivized peer measure $\hat{y}_p^t$ as our measures of expected performance in the face of temptation.[28][29]

[27] The same hypothesis is also not rejected when using the whole sample (4.60 vs 4.56, $p = 0.1637$).

[28] $\bar{y}_s^t$ is not equal to $\bar{\hat{y}}_p^t$ for subjects who face temptation (4.61 vs 4.52, $p = 0.0263$) or for all subjects (4.60 vs 4.49, $p = 0.0107$).

[29] Compared to actual performance, subjects generally underestimate the number of correct answers they would get in Task 2 regardless of the way in which beliefs are elicited. Actual performance $\bar{y}_2 = 4.88$ is higher than $\bar{y}_s^t = 4.61$, $\bar{\hat{y}}_p^t = 4.52$ and $\bar{\hat{y}}_s^t = 4.56$, with $p < 0.0001$ in all cases.

Among subjects who face temptation, 20.9% have WTP greater than 30 times the expected productivity difference ($y_s^{nt} - y_s^t$) using the direct measure of y_s^t ($p < 0.0001$).[30] Using the incentivized peer measure, we also find that 16.9% of subjects overstate their WTP ($p = 0.0001$).[31] Brown's correction for multiple hypothesis testing yields a combined $p < 0.0001$. We therefore conclude the following:[32]

Result 2. *A substantial share of subjects overestimate WTP compared to expected material losses.*

A possible explanation for the discrepancy between WTP and the expected material losses is that subjects expect to experience psychological cost when faced with temptation which they seek to avoid by stating a higher WTP. We next check if WTP is still overestimated even when accounting for subject's expectation of the psychological cost of temptation. As derived from equation (2) above, a sufficient condition, given $WTP > 0$, is that the subject's overestimation of material losses exceeds their underestimation of psychological cost ($y_2 - y_e^t \geq 0$ and $v \geq 0$, with at least one strict inequality). As before, we use the two measures of expected performance when tempted: y_s^t and $\hat{y}_p^t$, as stated in Hypotheses 3a and 3c. The proportion of subjects who overestimate WTP using these measures are 17.3% and 19.1% respectively, with combined $p < 0.0001$.[33]

Before deriving conclusions for Hypotheses 3a and 3c, we check whether y_s^{nt} is a valid counterfactual using Task 2 performance for the 12 subjects who do not face temptation. We do not reject Hypothesis 3d that $\bar{y}_2 = 4.75$ is not significantly different from $\bar{y}_s^{nt} = 4.58$ ($p = 0.7303$). We therefore conclude that:

Result 3. *A substantial share of subjects overestimate WTP compared to expected material losses and psychological temptation costs.*

We next explore what may be the drivers of subjects' overestimation of their WTP. We hypothesize that overestimation is driven by subjects' pessimism in terms of perceiving the temptation as strong (increase in θ) or their willpower as weak (decrease in ω), and potentially other controls as outlined in model (3) above. The results are presented in Tables 3 and 4, using y_s^t and $\hat{y}_p^t$ respectively.

[30] In the first session, a coding error resulted in subjects not learning of their Task 1 performance until the end of the experiment. However, their estimate of y_s^{nt} is not significantly different from subjects in the remaining sessions (4.68 vs 4.73, $p = 0.7765$).

[31] Our results are robust to using $\hat{y}_p^t \pm 0.5$ to account for possible rounding in n_p.

[32] It is also possible to test Hypotheses 2b and 2d for all subjects, including those who do not face temptation. Using y_s^t and $\hat{y}_p^t$ respectively, the proportion of subjects who overestimate their WTP are 23.3% and 19.6% respectively, with combined $p < 0.0001$.

[33] In some cases, the stated WTP with expected material losses implies a negative PC_e. However, this is assumed to be due to a random error as there are more positive implied PC_e values such that the average is positive.

Table 3: Marginal effects from logistic regressions of WTP overestimation, using y_s^t.

	(1)	(2)	(3)	(4)	(5)
θ	0.0489*	0.0464	0.0494*	0.0403	0.0439
	(1.70)	(1.62)	(1.68)	(1.41)	(1.50)
ω	0.000919	0.00118	0.000899	0.000277	-0.0000137
	(0.35)	(0.45)	(0.34)	(0.10)	(-0.00)
y_s^{nt}		-0.0451		-0.0432	
		(-1.44)		(-1.42)	
y_1			0.00491		0.0112
			(0.09)		(0.22)
Age				0.0157*	0.0160*
				(1.68)	(1.70)
Male				-0.0687	-0.0709
				(-1.59)	(-1.64)
Degree				0.0237	0.0225
				(0.52)	(0.49)
Econ				0.0270	0.0277
				(0.59)	(0.60)
GPA				-0.00212	0.0000367
				(-0.05)	(0.00)
Session FE	X	X	X	X	X
N	277	277	277	277	277

Dependent variable *overestimator*: dummy variable which equals 1 if subject states $WTP > 0$ and satisfies both $y_2 - y_s^t \geq 0$ and $v \geq 0$ with one strict inequality, and 0 otherwise. θ how much subject expects to be tempted on a scale from 1 to 4. ω score on the brief self-control scale which ranges from 13 to 65 (Tangney et al., 2004). y_s^{nt} predicted number of correct answers out of 5 in Task 2 if temptation is present. y_1 number of correct answers out of 5 in Task 1. *Age* subject's age in years. *Male* dummy variable which equals 1 for male subjects and 0 otherwise. *Degree* subject's degree program, 1 Bachelor, 2 Master, 3 PhD. *Econ* dummy variable which equals 1 for economics majors and 0 otherwise. *GPA* subject's GPA on a scale from 1 to 4. t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Marginal effects from logistic regressions of WTP overestimation, using $\hat{y}_p^t$.

	(1)	(2)	(3)	(4)	(5)
θ	0.0561*	0.0544*	0.0583*	0.0496*	0.0546*
	(1.87)	(1.81)	(1.91)	(1.67)	(1.81)
ω	-0.00125	-0.00110	-0.00132	-0.00273	-0.00300
	(-0.45)	(-0.40)	(-0.48)	(-0.93)	(-1.02)
y_s^{nt}		-0.0267		-0.0246	
		(-0.76)		(-0.71)	
y_1			0.0233		0.0317
			(0.37)		(0.54)
Age				0.00742	0.00723
				(0.74)	(0.72)
Male				-0.0829*	-0.0858*
				(-1.81)	(-1.87)
Degree				0.0438	0.0457
				(0.91)	(0.95)
Econ				-0.00231	-0.00161
				(-0.05)	(-0.03)
GPA				-0.0222	-0.0193
				(-0.49)	(-0.42)
Session FE	X	X	X	X	X
N	277	277	277	277	277

Dependent variable *overestimator*: dummy variable which equals 1 if subject states $WTP > 0$ and satisfies both $y_2 - \hat{y}_p^t \geq 0$ and $v \geq 0$ with one strict inequality, and 0 otherwise. θ how much subject expects to be tempted on a scale from 1 to 4. ω score on the brief self-control scale which ranges from 13 to 65 (Tangney et al., 2004). y_s^{nt} predicted number of correct answers out of 5 in Task 2 if temptation is present. y_1 number of correct answers out of 5 in Task 1. *Age* subject's age in years. *Male* dummy variable which equals 1 for male subjects and 0 otherwise. *Degree* subject's degree program, 1 Bachelor, 2 Master, 3 PhD. *Econ* dummy variable which equals 1 for economics majors and 0 otherwise. *GPA* subject's GPA on a scale from 1 to 4. t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The likelihood that the subject overestimates their demand for the commitment device appears to be weakly driven by their perception of the temptation's strength, θ. The more subjects expect to be tempted by internet access (higher θ), the more likely they are to state a higher-than-justified WTP for the commitment device. In Tables 3 and 4, an extra point on θ increases the likelihood of being a WTP overestimator by around 5%. In a sense, this confirms our definition of an overestimator as someone whose perception of the temptation strength is stronger than it actually is. In contrast, subjects' perception of their own willpower does not appear to affect their excess demand for commitment, as observed from the insignificant coefficients of ω. Neither confidence in future performance without temptation (y_s^{nt}) nor past performance (y_1) appears to play a role in WTP misestimation.

What can explain the insignificant role played by perceived willpower in predicting excess commitment demand? We have used the self-control scale (Tangney et al., 2004) to elicit subjects' perception of their willpower, but it may have been more accurate in capturing subjects' *actual* than their *perceived* willpower – which is not predicted to directly influence subjects' pessimism about their self-control. In this setting, asking subjects about their expectation of the strength of the temptation (θ) may be a more salient way of capturing pessimism.

Result 4. *WTP overestimation is weakly driven by pessimism in perceived temptation strength, but not perceived willpower.*

4.2 Does subject WTP reflect random indulgence or costly self-control?

Thus far, our analysis has rested on the distinction between material and psychological costs of temptation.[34] This is not the only way to conceptualize separate drivers of subject WTP. For example, we might make the distinction between (i) costs from the risk of succumbing, i.e. from 'random indulgence' (Chatterjee and Krishna, 2009; Dekel and Lipman, 2012), and (ii) costs from pure exposure to temptation. Both of these likely have material as well as non-material components; for example, the non-material component of costs from pure exposure is what we have called the psychological cost, PC.

The two types of subjects could be identified according to the following:

- **Random Indulgence (RI)**-type: demand commitment since they expect to succumb with positive probability. WTP should increase with the expected likelihood of succumbing, as captured by p_s for self-prediction and n_p for peer prediction.

[34] The analysis from this section onward is not pre-registered. Nevertheless, we believe the topics covered are important to study.

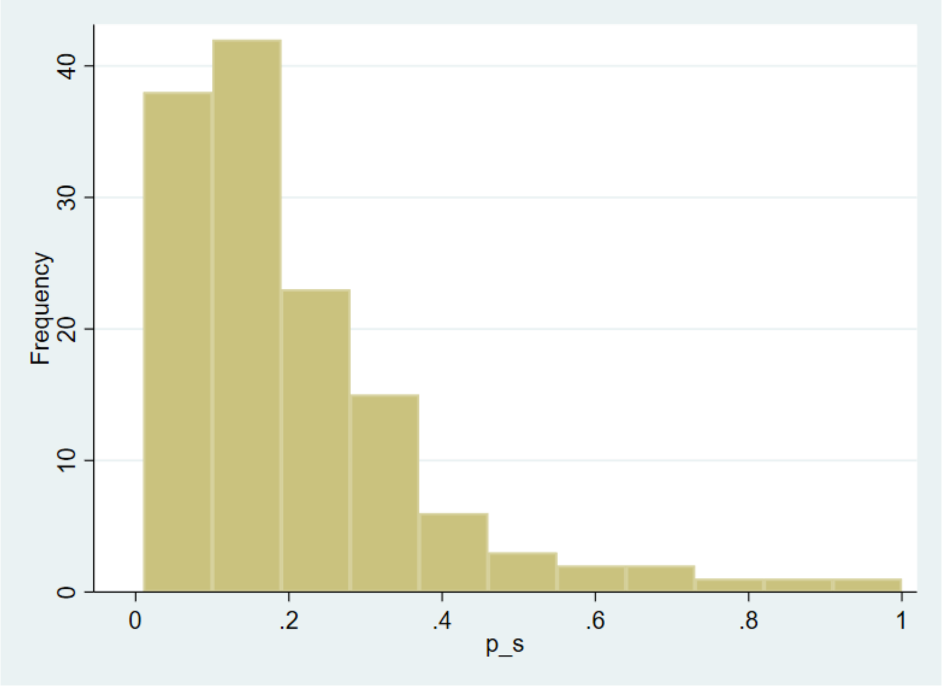


Figure 3: Predicted likelihood of succumbing, for 46% of subjects stating $p_s > 0$

- **Self-Control (SC)**-type: demand commitment to avoid the cost of self-control, both psychological and material. WTP should increase with perceived strength of temptation, as captured by θ.

As a preliminary, we first check how expected likelihood of succumbing compares with actual behavior under temptation. 46% of subjects state a positive p_s, as shown in Figure 3. On the other hand, as mentioned above, very few (only 4) subjects actually succumb to internet surfing. Indeed, the actual probability of succumbing is lower than p_s (1.44% vs 8.22%, $p < 0.0001$) in the group that face temptation but even more so for the subjects who face temptation despite having positive WTP for commitment device (1.49% vs 15%, $p < 0.0001$). How about the cost of exerting self-control when resisting temptation? In terms of psychological cost, of the 277 subjects who face temptation, only 12 (4%) experience it to be harder than expected to ignore. Neither does exerting self-control translate into lower productivity: comparing performance in Task 2 between those who face temptation and those who do not, we do not find any difference ($y_2 = 4.88$ vs 4.75, $p = 0.1347$). It appears that subjects are overly pessimistic about both expecting to succumb more often and anticipating the temptation to be more tempting than is actually the case.

We then study whether the fear of succumbing or costly self-control, though unwarranted, is the main driver of WTP. We find evidence for the presence of RI-types in our sample. The correlation between WTP and p_s is 39% ($p < 0.0001$), while the correlation between WTP and n_p is 40% ($p < 0.0001$). However, we also find positive and significant correlation between WTP and θ (19%, $p = 0.0014$). These correlations are also confirmed in regressions of WTP as shown in Table 5. The coefficients of θ, p_s and n_p are each separately significant in columns (1-3), controlling for demographic variables.

To control for the fact that a higher θ might cause subjects to estimate a higher likelihood of succumbing, we regress WTP on θ controlling for beliefs about likelihood to succumb in Table 5.[35] The results are consistent with subjects being of RI-type: subjects state a high WTP not because they expect to suffer the cost of exerting self-control while resisting temptation, but because they anticipate succumbing to temptation. The coefficients for likelihood of succumbing are still highly significant whether we use self- or peer prediction, while the significance of θ disappears. An extra percentage point in subject's expected likelihood of succumbing increases WTP by around 1 experimental token (CZK). The effect from using the peer measure is slightly lower: expecting that 1 extra person out of 5 (an extra 20 percentage points) in the likelihood of a similar peer succumbing leads to only 16-17 token increase in WTP.

We cannot exclude that some subjects may also state a high WTP because they fear costly self-control, as the experiment was not designed to separate these two mechanisms. However, in sum, the model of random indulgence seems to rationalize the behavior of our subjects. Subjects' WTP for commitment device is driven by their expected likelihood of succumbing to temptation, despite a low probability of actually succumbing.

4.3 Assuming risk aversion

Given that subjects' demand for commitment appears to stem from the fear of randomly indulging their preference for surfing the internet, the temptation may have been perceived as a risk which subjects would like to avoid. Because of this, what looks like an overstated WTP for commitment compared to the optimal WTP of a risk-neutral agent may become rationalizable or even understated when compared to the optimal choice under risk aversion. This section therefore checks whether WTP remains overstated if subjects are assumed to be (strongly) risk-averse.

[35] This may admittedly be a bad control since θ may be a function of p_s or vice versa. The direction of the selection bias is unknown since it is unclear whether WTP should be higher or lower than average for those who do not expect to succumb and yet anticipate the internet access to be very tempting.

Table 5: Tobit regressions of WTP.

	(1)	(2)	(3)	(4)	(5)
θ	14.34***			3.699	6.899
	(3.10)			(0.78)	(1.62)
p_s		106.2***		98.64***	
		(5.45)		(4.57)	
n_p			17.25***		15.92***
			(5.95)		(5.39)
Age	2.635*	3.024**	1.969	2.999**	2.061
	(1.87)	(2.31)	(1.55)	(2.29)	(1.62)
Male	-14.15**	-12.10*	-12.61*	-11.94*	-12.05*
	(-1.98)	(-1.84)	(-1.95)	(-1.81)	(-1.87)
Degree	2.892	-3.083	2.014	-2.838	1.501
	(0.40)	(-0.45)	(0.30)	(-0.42)	(0.23)
Econ	-1.876	1.246	1.723	1.150	1.645
	(-0.24)	(0.17)	(0.25)	(0.16)	(0.24)
GPA	5.205	0.840	2.683	0.873	2.199
	(0.76)	(0.13)	(0.44)	(0.14)	(0.36)
Constant	-110.2***	-88.31***	-83.53**	-93.44***	-93.77***
	(-2.98)	(-2.67)	(-2.58)	(-2.76)	(-2.83)
Session FE	X	X	X	X	X
N	289	289	289	289	289

Dependent variable *WTP*: willingness-to-pay to remove internet access, from 0 to 100. θ how much subject expects to be tempted on a scale from 1 to 4. p_s predicted likelihood of succumbing to temptation, from 0 to 1. n_p predicted number of peers (out of 5) who succumb to temptation. *Age* subject's age in years. *Male* dummy variable which equals 1 for male subjects and 0 otherwise. *Degree* subject's degree program, 1 Bachelor, 2 Master, 3 PhD. *Econ* dummy variable which equals 1 for economics majors and 0 otherwise. *GPA* subject's GPA on a scale from 1 to 4. t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Recall that, in Section 3, the subject maximized utility as given by

$$\begin{aligned}U(WTP) = &\frac{1}{2}\left[\frac{WTP}{200}u(100+120y_1-WTP)+\left(1-\frac{WTP}{200}\right)u(100+120y_1-PC)\right]\\&+\frac{1}{2}\left[\frac{WTP}{200}u(100+120y^{nt}-WTP)+\left(1-\frac{WTP}{200}\right)u(100+120y^{t}-PC)\right]\end{aligned}$$

The above utility function only captures risk aversion in the BDM 'lottery' and misses the second 'lottery' faced by the subject: the possibility of earning a lower amount if she succumbs to temptation. We therefore change the last term in the utility function which gives

$$\begin{aligned}U(WTP) = &\frac{1}{2}\left[\frac{WTP}{200}u(100+120y_1-WTP)+\left(1-\frac{WTP}{200}\right)u(100+120y_1-PC)\right]\\&+\frac{1}{2}\left[\frac{WTP}{200}u(100+120y^{nt}-WTP)\right]\\&+\frac{1}{2}\left[\left(1-\frac{WTP}{200}\right)(p_s u(100+120y^{s}-PC)+(1-p_s)u(100+120y^{ns}-PC))\right]\end{aligned}$$

assuming that PC is the same regardless of whether the subject succumbs or not – there is, for example, no self-image loss or guilt from succumbing. The solution under risk neutrality, denoted WTP_{RN}, is unchanged:

$$WTP_{RN} = 30(y^{nt}-y^{t})+\frac{PC}{2}$$

where $y^t = \hat{y}^t_s$.[36]

Assume now that the subject is risk-averse and has CRRA utility function defined as:

$$u(x) = \begin{cases}\frac{x^{1-\eta}-1}{1-\eta} & \eta \neq 1\\ \ln(x) & \eta = 1\end{cases}$$

[36] Since we use subjects' expectation of their performance when succumbing or not succumbing, we necessarily use $\hat{y}^t_s$, the decomposed measure of self-performance, in our analysis.

No closed-form solution for WTP then exists, but we may derive the first-order condition

$$\frac{dU}{dWTP} = \frac{1}{200\,(1-\eta)}\left\{\frac{100+120y_1+(\eta-2)\,WTP}{(100+120y_1-WTP)^{\eta}} + \frac{100+120y^{nt}+(\eta-2)\,WTP}{(100+120y^{nt}-WTP)^{\eta}}\right.$$
$$-\left(100+120y_1-PC\right)^{1-\eta} - p_s\left(100+120y^{s}-PC\right)^{1-\eta}$$
$$\left. -\left(1-p_s\right)\left(100+120y^{ns}-PC\right)^{1-\eta}\right\} = 0 \qquad (4)$$

To show the robustness of our results under risk aversion, our strategy is the following. We seek to calculate the optimal WTP for the risk-averse agent, denoted WTP_{RA}, and show that there are still a significant number of subjects who overestimate WTP. We obtain values for WTP_{RA} using numerical simulations of (4) with the relevant y_1, y^{nt}, y^{s}, y^{ns} and p_s values inserted for each individual subject. η, the coefficient of relative risk aversion, has been estimated in different studies to be around 1.[37] To be conservative, we present results for several values of η up to $\eta = 4$, though as will be shown our results do not change drastically.

We start by asking whether risk-averse subjects overestimate their WTP when only considering actual material loss (corresponding to Hypothesis 1a in the risk-neutral case). In equation (4), y^{nt} is thus interpreted as the *actual* number of correct answers when the subject is not tempted; as per the risk-neutral case above, we use y_1 as the counterfactual. p_s is obtained using the percentage of subjects who succumb out of all subjects exposed to temptation, this equals 1.44%.[38] For subjects who do not succumb, $y^{ns} = y_2$, while y^{s}, the counterfactual had they succumbed, is obtained using the average productivity of subjects who do succumb, which is $y^{s} = 2$. In the same way, for subjects who succumb, $y^{s} = y_2$ while the counterfactual $y^{ns} = 4.93$, the average productivity for those who do not succumb. Comparing the resulting WTP_{RA} with the WTP stated by each subject, the proportion of overestimators under different values of η are given in the first row of Table 6. Just under 20% of subjects are still considered to be overestimators, stating WTP greater than what should be optimal when considering the actual material loss. A much higher number of subjects are now underdemanders of commitment (77% under η =0.5, 1 or 1.5). Nevertheless, our first result is robust to assuming CRRA with $\eta \leq 4$.

We next turn to subjects' WTP considering expected material loss (corresponding to

[37] For example, in one of the most widely cited lab experiments on risk aversion, Holt and Laury (2002) find that almost all subjects have $\eta \leq 1.37$. In a field experiment in Denmark, Harrison et al. (2007) find the mean η to be 0.67. The estimate is 0.74 in Andersen et al. (2008), who also estimate a population standard deviation for η of 0.056. Chiappori and Paiella (2011) estimate a median of 1.7 with a quarter of the population exhibiting a coefficient larger than 3.

[38] The proportion of succumbers among those who state $WTP > 0$ is not substantially different at 1.49%.

Table 6: Proportion of overestimators under risk aversion.

Relative to	$\eta = 0.5$	$\eta = 1$	$\eta = 1.5$	$\eta = 2$	$\eta = 3$	$\eta = 4$
(1) Actual material loss	18.41%	18.41%	18.41%	17.33%	16.97%	15.52%
(2) Expected material loss	15.16%	14.80%	14.08%**	13.72%**	12.27%^{ns}	11.91%^{ns}
(3) Actual material loss and psychological cost	16.25%	16.25%	16.25%	14.44%	13.72%**	13.36%**

All coefficients significant at the 1% level unless otherwise indicated, ** $p < 0.05$, ns $p > 0.1$.

Hypothesis 2c). We proceed as above, except that we now use each subject's predictions of their own performance y^{nt}, y^{s}, y^{ns} and p_s. As shown in the second row of Table 6, we find a lower number of risk-averse subjects overestimate their WTP, compared to the case with actual material loss above. The proportion of overestimators is around 15% for low values of η, and decreases to 12% and no longer significant when $\eta \geq 3$. Putting aside psychological cost, at these higher levels of risk aversion many subjects appear to underestimate their performance relative to how well they actually do in the face of temptation and their WTP is a relatively accurate reflection of this pessimism (see also Footnote 29).

Finally, we check whether WTP is still overestimated by risk-averse subjects when allowing for psychological costs of temptation. Since we do not know the actual PC faced by each subject, our strategy is analogous to the test of Hypotheses 3a-3c under risk neutrality. First, we note that optimal WTP is strictly increasing in PC under any degree of risk aversion; the proof is given in Section A.2 in the Appendix. Given this fact, we may proceed as follows.

For all subjects with $WTP > 0$ and $v \geq 0$, and for all PC values consistent with $0 < WTP < 100$, we plug in appropriate outcome variables in (4) to calculate what the WTP should have been for a risk-averse subject based on *actual* material losses. We then repeat the exercise for the subject's *expected* material loss; denote these two (sets of) WTP values WTP_a and WTP_e for actual and expected WTP, respectively. Now, suppose for some particular expected psychological cost PC_e, $WTP_e \geq WTP_a$ while $v \geq 0$ (implying $PC_e \geq PC_a$), with at least one of these two inequalities strict. Since WTP is increasing in PC, we then have $WTP_e(PC_e) \geq WTP_a(PC_e) \geq WTP_a(PC_a)$, again with at least one strict inequality. Thus, WTP has been strictly overestimated in relation to both actual material losses and actual psychological costs. To be conservative, we classify as overestimators those subjects who have $v \geq 0$ and $WTP_e \geq WTP_a$, with at least one strict inequality, for *all* values of PC_e consistent with $0 < WTP_e < 100$.[39]

[39] In principle, since both stated WTP and all parameters related to expected material losses are known,

As shown in row (3) of Table 6, we find that such subjects make up between about 13-16% of all subjects who face temptation ($p < 0.05$ for all values of η). Hence, our result of overestimation relative to both actual material and psychological costs is also robust to assuming CRRA with $\eta \leq 4$.

Overall, even assuming a very strong degree of risk aversion, our conclusion that a significant share of subjects overstate their demand for the commitment device is unchanged. The subjects in question appear to state a higher WTP than motivated either by actual material losses, or when including actual psychological costs.

5 Concluding Remarks

Excess demand for commitment is one aspect of self-control misestimation that has so far not been studied in the literature. We take a first step towards investigating this possibility in the lab and make the following contributions. First, we show under both risk neutrality and risk aversion that a significant share of subjects overdemand commitment devices. This is true when we compare WTP with material loss, but also when we take into account expected psychological cost. Thus we provide the first evidence for the need to focus not only on encouraging the take-up of commitment devices, but also potentially putting an upper limit on this in situations when facing the temptation may not actually be that harmful both materially and psychologically. Second, we show the excess WTP for commitment in a lab setting where, if anything, subjects should be unwilling to pay: internet access is easily available outside the lab and should have less immediate appeal. Additionally, shortly before the productivity task with temptation, subjects did the exact same task without temptation. Hence they should have a good idea about the difficulty of the task – and even so they still overestimate their WTP.

We find suggestive evidence that subjects' WTP for the commitment device is driven by their fear of succumbing to the temptation, with WTP increasing with subjects' predicted likelihood of succumbing. Taking into account risk aversion in this domain does not rationalize subjects' overestimation of WTP. While WTP correctly values subjects' expectations of their material loss, these expectations are overly pessimistic thus resulting in WTP which still exceed what would be optimal given the actual material and psychological costs faced.

A recent paper by Carrera et al. (2019) suggests that, with some uncertainty about the future, commitment demand can also be driven by noise and demand effects. In a field

we might use them in (4) to solve for a single implied value of PC_e. The reason why we do not check whether $WTP_e \geq WTP_a$ only at this implied PC_e is because, as in Footnote 33, it is sometimes negative, which we interpret as there being some random error in subjects' WTP responses.

experiment with members of a fitness facility, the authors find that a substantial number of participants demand commitment for *both* more and fewer gym visits. Together with our results, these findings suggest that commitment contracts may not always be the optimal policy tool.

The negative consequence of choosing to face the temptation less often than optimal is not limited to overpaying for the commitment device. If the agent derives signals about what kind of person he is from his actions, then achieving the goal by resisting temptation gives higher self-image or self-confidence that is not obtained when the agent achieves the same goal using the commitment device. Hence, this is yet another reason why, from a welfare point of view, one should consider limiting the extent to which commitment is encouraged or imposed on economic agents.

One limitation of the present research is that it has been calibrated to be conducted in the lab. Using different parameters may yield different results: for example, paying subjects a smaller amount may cause them to prefer to succumb to the temptation of internet access. We see our findings as a first step towards showing the existence of excess demand for temptation. The empirical applications and their policy implications will need to be investigated in future research.

References

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3):583–618.

Ashraf, N., Karlan, D., and Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, 121(2):635–672.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, 130(3):1067–1115.

Bénabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy*, 112(4):848–886.

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, 19(5):547–556.

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

Brown, M. B. (1975). 400: A method for combining non-independent, one-sided tests of significance. *Biometrics*, 31(4):987–992.

Bryan, G., Karlan, D., and Nelson, S. (2010). Commitment devices. *Annual Review of Economics*, 2(1):671–698.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (2019). How are preferences for commitment revealed? NBER Working Paper 26161.

Chatterjee, K. and Krishna, R. V. (2009). A "dual self" representation for stochastic temptation. *American Economic Journal: Microeconomics*, 1(2):148–67.

Chiappori, P.-A. and Paiella, M. (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, 9(6):1021–1052.

Dekel, E. and Lipman, B. L. (2012). Costly self-control and random self-indulgence. *Econometrica*, 80(3):1271–1302.

DellaVigna, S. and Malmendier, U. (2006). Paying not to go to the gym. *The American Economic Review*, 96(3):694–719.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Gächter, S. and Renner, E. (2010). The effects of (incentivized) belief elicitation in public goods experiments. *Experimental Economics*, 13(3):364–377.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6):1403–1435.

Harrison, G. W., Lau, M. I., and Rutström, E. E. (2007). Estimating risk attitudes in Denmark: A field experiment. *Scandinavian Journal of Economics*, 109(2):341–368.

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

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92(5):1644–1655.

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

John, A. (forth.). When commitment fails – Evidence from a field experiment. *Management Science*.

Karlan, D. and Zinman, J. (2009). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77(6):1993–2008.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Myrseth, K. O. R. and Wollbrant, C. E. (2013). A theory of self-control and naïveté: The blights of willpower and blessings of temptation. *Journal of Economic Psychology*, 34:8–19.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *The American Economic Review*, 89(1):103–124.

Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2):271–324.

Thaler, R. H. and Benartzi, S. (2004). Save more tomorrow$^{\text{TM}}$: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1):S164–S187.

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, 86(3):859–889.

Wertenbroch, K. (1998). Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Science*, 17(4):317–337.

Appendices

A Proofs

A.1 That results are robust to assuming WTP is paid conditional on Task 2 being chosen for payment

In this subsection we show that our analysis is robust to assuming that WTP is paid only conditional on Task 2 being chosen for payment. In this case subjects maximize expected utility, with equal probabilities of either Task 1 or Task 2 being paid, as

$$
\begin{aligned}
U(WTP) = \frac{1}{2}&\left[\frac{WTP}{200}u(100+120y_1-0)+\left(1-\frac{WTP}{200}\right)u(100+120y_1-PC)\right]\\
&+\frac{1}{2}\left[\frac{WTP}{200}u(100+120y^{nt}-WTP)+\left(1-\frac{WTP}{200}\right)u(100+120y^{t}-PC)\right]
\end{aligned}
$$

and the solution under risk neutrality is given by

$$WTP = 60(y^{nt}-y^{t})+PC$$

Clearly, the analysis for Hypothesis 3, and thus Hypothesis 4, is equivalent. For Hypothesis 1a, the proportion of overestimator using the new WTP threshold given above is 22.9% ($p < 0.0001$). For Hypothesis 2, the proportion of overestimators using y_s^t is 19.3%, while using $\hat{y}_p^t$ it is 13.8% with combined $p < 0.0001$.

A.2 That WTP under risk aversion is increasing in psychological cost

Assuming CRRA with $\eta > 1$, the utility function can be written as:

$$
\begin{aligned}
U(WTP) = \frac{1}{2}&\left[\frac{WTP}{200}\frac{(100+120y_1-WTP)^{1-\eta}-1}{1-\eta}+\left(1-\frac{WTP}{200}\right)\frac{(100+120y_1-PC)^{1-\eta}-1}{1-\eta}\right]\\
&+\frac{1}{2}\left[\frac{WTP}{200}\frac{(100+120y^{nt}-WTP)^{1-\eta}-1}{1-\eta}\right.\\
&\left.+\left(1-\frac{WTP}{200}\right)\left(p_s\frac{(100+120y^{s}-PC)^{1-\eta}-1}{1-\eta}+(1-p_s)\frac{(100+120y^{ns}-PC)^{1-\eta}-1}{1-\eta}\right)\right]
\end{aligned}
$$

Multiplying by 1/2 and differentiating, this gives the first-order condition (4) stated in the main text. The second derivative is

$$\frac{d^2U}{dWTP^2} = \frac{1}{200(1-\eta)}\left[\frac{1}{(100+120y_1-WTP)^\eta}\left(-2+\eta\left(1+\frac{100+120y_1+(\eta-2)WTP}{100+120y_1-WTP}\right)\right)\right.$$

$$\left.+\frac{1}{(100+120y^{nt}-WTP)^\eta}\left(-2+\eta\left(1+\frac{100+120y^{nt}+(\eta-2)WTP}{100+120y^{nt}-WTP}\right)\right)\right]$$

$$< 0$$

because each (outer) parenthesis is strictly larger than 0 iff $\eta > 1$. The partial derivative of the first-order condition with respect to PC is

$$\frac{\partial^2U}{\partial WTP \partial PC} = \frac{1}{200}\left[\frac{1}{(100+120y_1-PC)^\eta}+\frac{p_s}{(100+120y^s-PC)^\eta}+\frac{1-p_s}{(100+120y^{ns}-PC)^\eta}\right]$$

$$> 0$$

Using the implicit function theorem,

$$\frac{dWTP}{dPC} = -\frac{\frac{\partial^2U}{\partial WTP\partial PC}}{\frac{d^2U}{dWTP^2}} > 0$$

Hence, WTP is strictly increasing in PC. The proof for $0 < \eta < 1$ and $\eta = 1$ is similar and is left to the reader.

B Sources of excess WTP valuation

We show below the drivers of the individual components of WTP overestimation: misestimation of performance in Table 7 and misestimation of temptation experience in Table 8.

Table 7: OLS regressions of misestimation of performance $y_2 - y_s^t$.

	(1)	(2)	(3)	(4)	(5)
θ	0.247***	0.239***	0.271***	0.232***	0.260***
	(3.48)	(3.37)	(3.77)	(3.26)	(3.59)
ω	0.00354	0.00433	0.00299	0.00249	0.00131
	(0.57)	(0.70)	(0.48)	(0.38)	(0.20)
y_s^{nt}		-0.157*		-0.149*	
		(-1.84)		(-1.75)	
y_1			0.251*		0.213
			(1.81)		(1.52)
Age				0.00561	0.00474
				(0.23)	(0.19)
Male				-0.0976	-0.104
				(-0.91)	(-0.97)
Degree				-0.119	-0.110
				(-1.07)	(-0.99)
Econ				0.110	0.109
				(0.97)	(0.96)
GPA				-0.180*	-0.166
				(-1.74)	(-1.59)
Constant	0.0249	0.738	-1.237	1.244	-0.523
	(0.07)	(1.39)	(-1.57)	(1.56)	(-0.53)
Session FE	X	X	X	X	X
N	277	277	277	277	277
R^2	0.085	0.097	0.097	0.119	0.117

Dependent variable $y_2 - y_s^t$: misestimation of performance, defined as actual performance in Task 2 less expected performance in Task 2, both when temptation is present. θ how much subject expects to be tempted on a scale from 1 to 4. ω score on the brief self-control scale which ranges from 13 to 65 (Tangney et al., 2004). y_s^{nt} predicted number of correct answers out of 5 in Task 2 if temptation is present. y_1 number of correct answers out of 5 in Task 1. *Age* subject's age in years. *Male* dummy variable which equals 1 for male subjects and 0 otherwise. *Degree* subject's degree program, 1 Bachelor, 2 Master, 3 PhD. *Econ* dummy variable which equals 1 for economics majors and 0 otherwise. *GPA* subject's GPA on a scale from 1 to 4. t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Marginal effects from logistic regressions of misestimation of temptation experience v.

	(1)	(2)	(3)	(4)	(5)
θ	0.165***	0.165***	0.164***	0.159***	0.157***
	(3.65)	(3.66)	(3.60)	(3.53)	(3.45)
ω	-0.00201	-0.00206	-0.00199	-0.00176	-0.00164
	(-0.53)	(-0.54)	(-0.53)	(-0.44)	(-0.41)
y_s^{nt}		0.00961		0.0148	
		(0.19)		(0.28)	
y_1			-0.00692		-0.0226
			(-0.07)		(-0.22)
Age				0.0191	0.0192
				(1.24)	(1.24)
Male				-0.0959	-0.0958
				(-1.45)	(-1.45)
Degree				-0.0832	-0.0837
				(-1.19)	(-1.20)
Econ				0.0790	0.0789
				(1.13)	(1.13)
GPA				0.0161	0.0153
				(0.25)	(0.24)
Session FE	X	X	X	X	X
N	277	277	277	277	277

Dependent variable v: misestimation of temptation experience, defined as dummy variable which equals 1 if subject experiences temptation to be easier than expected and 0 otherwise. θ how much subject expects to be tempted on a scale from 1 to 4. ω score on the brief self-control scale which ranges from 13 to 65 (Tangney et al., 2004). y_s^{nt} predicted number of correct answers out of 5 in Task 2 if temptation is present. y_1 number of correct answers out of 5 in Task 1. *Age* subject's age in years. *Male* dummy variable which equals 1 for male subjects and 0 otherwise. *Degree* subject's degree program, 1 Bachelor, 2 Master, 3 PhD. *Econ* dummy variable which equals 1 for economics majors and 0 otherwise. *GPA* subject's GPA on a scale from 1 to 4. t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C Power calculations

Our power calculations are based on the concept of minimum detectable effect (MDE) (see, for example, Bloom (1995)). Using 90% power and 5% level of statistical significance, the MDE is the smallest effect that, if true, has a 90% chance of producing an estimate that is significant at the 5% level. The MDE is calculated to be

$$\begin{aligned} MDE &= (t_{\alpha/2} + t_{\beta})\sigma \\ \hat{\theta} - \theta &= 1.645\sigma + 1.282\sigma \\ \hat{\theta} - \theta &= 2.927\sigma \end{aligned}$$

where σ is the standard error of the estimator $\hat{\theta}$. We use the score test for binomial proportion which is based on the null standard error, yielding

$$\begin{aligned} \hat{\theta} - \theta &= 2.927\sqrt{\frac{\theta(1-\theta)}{n}} \\ \hat{\theta} &= 0.1 + 2.927\sqrt{\frac{0.1(0.9)}{289}} \\ &= 0.1517 \end{aligned}$$

with the null $\theta = 0.1$ (attributable to subject confusion) and a sample size of 289 ($\hat{\theta} = 0.1528$ for $n = 277$, for tests using the sample who face temptation). This means that if the true proportion of WTP overestimators is at least 15.17%, our experimental design will detect this with 90% probability at a 5% level of statistical significance.

For the purpose of our hypothesis tests, the critical value above which the hypothesis would be rejected would thus be $\theta + t_{\alpha/2}\sigma = 12.90\%$ for tests with $n = 289$ and 12.97% for tests with $n = 277$.

D Instructions

Begins on next page.

General instructions

You are about to participate in an experiment on decision-making. Before we start, please make sure your phones are on silent and put away all personal belongings.

The experiment will take place through your computer terminals. Please do not talk or try to communicate with other participants during the session. If you have any question, please raise your hand and the experimenter will approach you to answer it.

This experiment consists of 2 stages plus a short questionnaire at the end. The whole session will last up to 2 hours. After the session, you will receive your experimental payment. This payment consists of a **participation fee of 100 CZK** plus your **experiment earnings.** Your experiment earnings will depend on your own decisions, on the decision of another participant, and on chance. It is therefore important to think about each of your decisions carefully.

During the experiment, your payoff will be denominated in experimental tokens that will be converted to CZK at the end at the following rate:

1 CZK = 1 token

You are about to begin with Stage 1. Out of Stage 1 and Stage 2, only one will be used for payment. Which stage is chosen will be determined by a random draw at the end of the experiment.

We will now read together the instructions for Stage 1.

Stage 1

During Stage 1, your main task will be to focus attentively on a four-digit number that will appear on your computer screen for a period of up to 30 minutes. This number will increment every 3 seconds. At random times during the 30 minute period, you will be prompted to **enter the last number you saw on your screen.** The number will be reinitialized after every prompt and you will receive a total of 5 prompts during the period. You will earn 120 tokens per correct answer, should this stage be chosen for payment.

Besides performing the attention task, no other activity will be allowed (including checking your phone, surfing the internet, studying…). If you are caught doing something else, you will not be paid for your participation in the experiment.

Are there any questions at this point?

You will now practice the attention task for 1 minute on the computer before moving on to the real task.

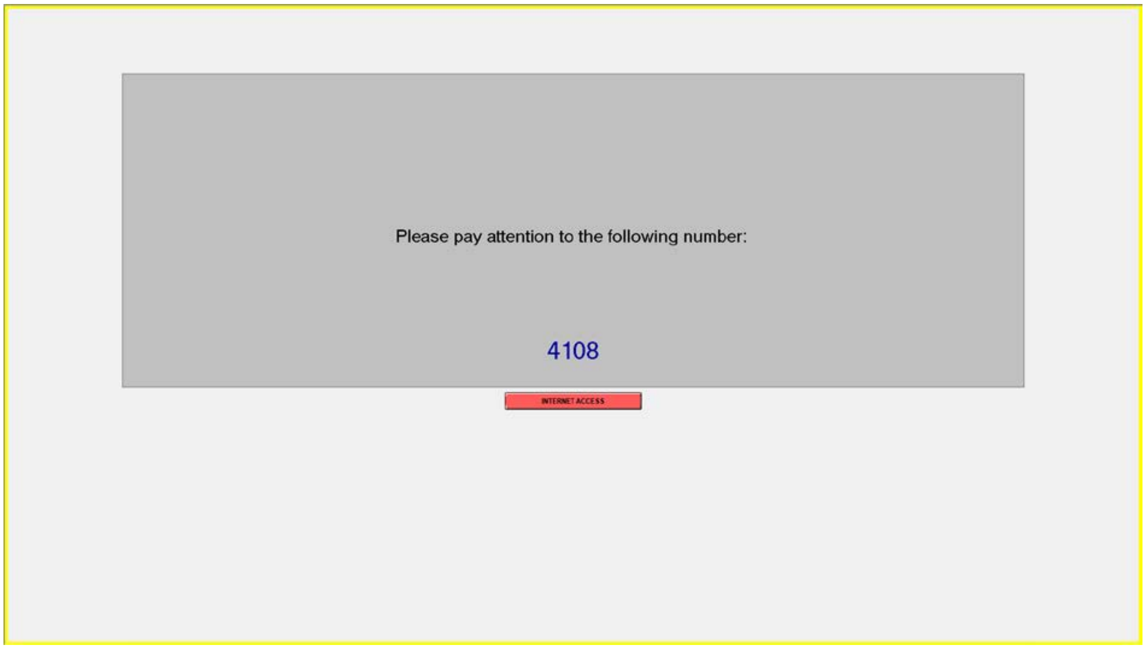
[After stage 1]

In Stage 2, you will repeat the same attention task with a small modification. We will explain this modification in more detail soon, and afterwards you will be asked to answer some questions regarding the new attention task.

The modification

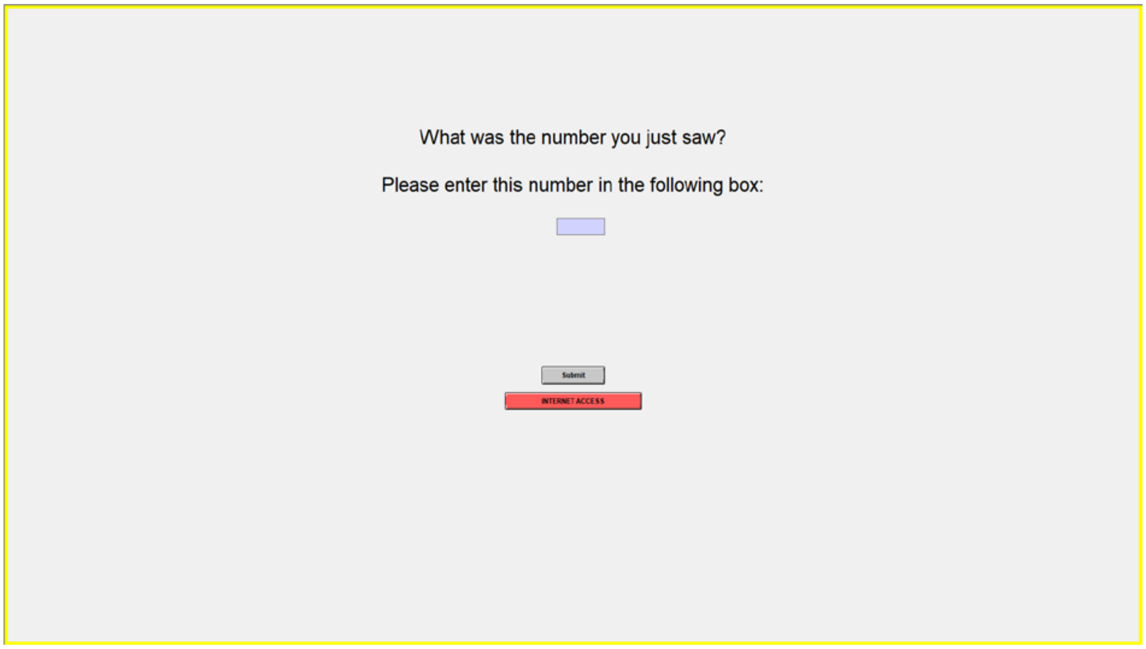
The attention task in Stage 2 is similar to the one you completed in Stage 1: your main task is to focus attentively on a four-digit number that will appear on your computer screen for a period of up to 30 minutes. There will again be 5 prompts to enter the last number you saw on the screen, and you will again earn 120 tokens per correct answer.

However, below the four digits, you will now see a button labeled "Internet Access", as shown in the screenshots below. You can click on this button at any point during the attention task.



Please pay attention to the following number:

4108



What was the number you just saw?

Please enter this number in the following box:

If you click this button, you will be able to surf the internet for the remainder of the 30-minute period instead of continuing with the attention task. Once you click the button, you will not be able to return to the task, so clicking it means that you forfeit the chance to earn more money through answering any future prompts correctly. You will, however, earn money from all correct answers up to the point of clicking the button.

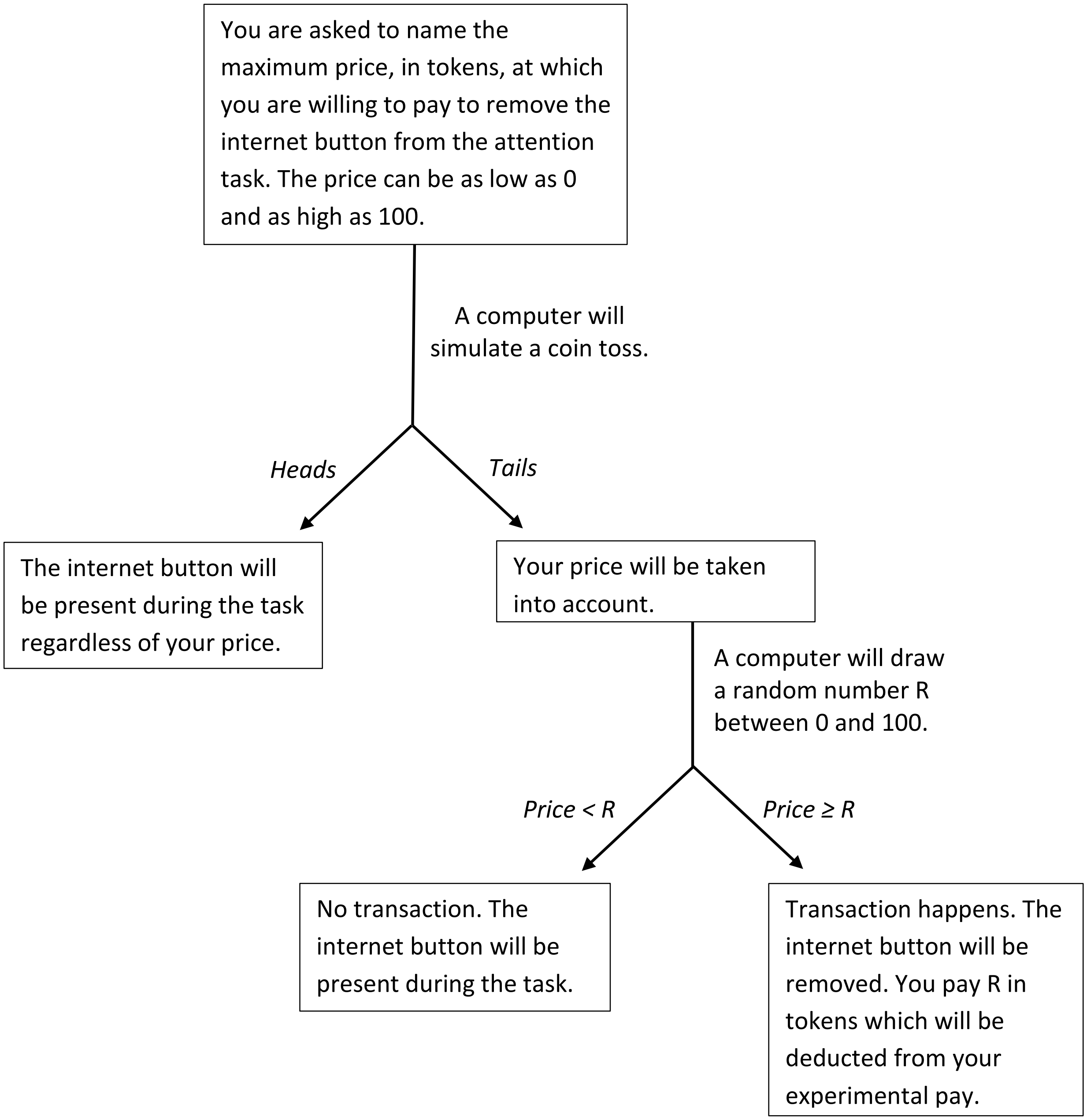
If you do not click the button, you will simply continue with the attention task. The button will continue to be present for the remainder of the 30-minute period.

Removing the “Internet Access” button

If you would like to remove the “Internet Access” button, for example because you think you might be able to concentrate better without it, you have the possibility to do so. We will now describe how this works.

You can pay tokens to remove the “Internet Access” button. Removing the button means that there will be no possibility for internet access during the attention task and you will therefore be participating in the attention task for the whole 30 minutes. The screen displayed will exactly be the same as in Stage 1, without the “Internet Access” button.

Starting from the top, determining whether to remove the button or not involves the following steps:



You are asked to name the maximum price, in tokens, at which you are willing to pay to remove the internet button from the attention task. The price can be as low as 0 and as high as 100.

A computer will simulate a coin toss.

Heads

The internet button will be present during the task regardless of your price.

Tails

Your price will be taken into account.

A computer will draw a random number R between 0 and 100.

Price < R

No transaction. The internet button will be present during the task.

Price ≥ R

Transaction happens. The internet button will be removed. You pay R in tokens which will be deducted from your experimental pay.

As you can see, the possibility of successfully removing the internet button increases the higher your stated price. Note in particular that:

- Your chance of removing the internet button is maximized (but is not guaranteed) if you state a price of 100.
- If you are not willing to pay anything to remove the internet button, you should enter a price of 0.

Your decision is final and cannot be changed.

No matter if you pay to remove the internet button, keep the button but never click it, or keep the button and click it, you will spend the same amount of time (up to 30 minutes) in Stage 2.

You will now have a practice round to ensure you understand how the process of removing the internet button works. When you have finished the practice round, you will be asked to state your **actual** price for removing the button, as well as answer a few questions about the new attention task.

Finally, the computer will toss the coin and (if TAILS comes up) draw the random price to determine the outcome of the transaction. You will be informed about whether the "Internet Access" button will be present or not, and subsequently you will start the attention task in Stage 2.

Stage 2

On-screen instructions.

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP18/21 Cormac Ó Gráda: The Next World and the New World: Relief, Migration, and the Great Irish Famine
WP18/22 Lisa Ryan, Sarah La Monaca, Linda Mastrandrea and Petr Spodniak: 'Harnessing Electricity Retail Tariffs to Support Climate Change Policy' December 2018
WP18/23 Ciarán Mac Domhnaill and Lisa Ryan: 'Towards Renewable Electricity in Europe: An Empirical Analysis of the Determinants of Renewable Electricity Development in the European Union' December 2018
WP19/01 Ellen Ryan and Karl Whelan: 'Quantitative Easing and the Hot Potato Effect: Evidence from Euro Area Banks' January 2019
WP19/02 Kevin Denny: 'Upper Bounds on Risk Aversion under Mean-variance Utility' February 2019
WP19/03 Kanika Kapur: 'Private Health Insurance in Ireland: Trends and Determinants' February 2019
WP19/04 Sandra E Black, Paul J Devereux, Petter Lundborg and Kaveh Majlesi: 'Understanding Intergenerational Mobility: The Role of Nature versus Nurture in Wealth and Other Economic Outcomes and Behaviors' February 2019
WP19/05 Judith M Delaney and Paul J Devereux: 'It's not just for boys! Understanding Gender Differences in STEM' February 2019
WP19/06 Enoch Cheng and Clemens Struck: 'Time-Series Momentum: A Monte-Carlo Approach' March 2019
WP19/07 Matteo Gomellini and Cormac Ó Gráda: 'Brain Drain and Brain Gain in Italy and Ireland in the Age of Mass Migration' March 2019
WP19/08 Anna Aizer, Paul J Devereux and Kjell G Salvanes: 'Grandparents, Mothers, or Fathers? - Why Children of Teen Mothers do Worse in Life' March 2019
WP19/09 Clemens Struck, Adnan Velic: 'Competing Gains From Trade' March 2019
WP19/10 Kevin Devereux, Mona Balesh Abadi, Farah Omran: 'Correcting for Transitory Effects in RCTs: Application to the RAND Health Insurance Experiment' April 2019
WP19/11 Bernardo S Buarque, Ronald B Davies, Dieter F Kogler and Ryan M Hynes: 'OK Computer: The Creation and Integration of AI in Europe' May 2019
WP19/12 Clemens C Struck and Adnan Velic: 'Automation, New Technology and Non-Homothetic Preferences' May 2019
WP19/13 Morgan Kelly: 'The Standard Errors of Persistence' June 2019
WP19/14 Karl Whelan: 'The Euro at 20: Successes, Problems, Progress and Threats' June 2019
WP19/15 David Madden: 'The Base of Party Political Support in Ireland: An Update' July 2019
WP19/16 Cormac Ó Gráda: 'Fifty Years a-Growing: Economic History and Demography in the ESR' August 2019
WP19/17 David Madden: 'The ESR at 50: A Review Article on Fiscal Policy Papers' August 2019
WP19/18 Jonathan Briody, Orla Doyle and Cecily Kelleher: 'The Effect of the Great Recession on Health: A longitudinal study of Irish Mothers 2001-2011' August 2019
WP19/19 Martina Lawless and Zuzanna Studnicka: 'Old Firms and New Export Flows: Does Experience Increase Survival?' September 2019
WP19/20 Sarah Parlane and Lisa Ryan: 'Optimal Contracts for Renewable Electricity' September 2019

UCD Centre for Economic Research Email economics@ucd.ie