UCD CENTRE FOR ECONOMIC RESEARCH

WORKING PAPER SERIES

2022

Pessimism and Overcommitment: An Online Experiment with Tempting YouTube Content

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

WP22/03

January 2022

UCD SCHOOL OF ECONOMICS
UNIVERSITY COLLEGE DUBLIN
BELFIELD DUBLIN 4

Pessimism and Overcommitment: An Online Experiment with Tempting YouTube Content*

Claes Ek† Margaret Samahita‡

January 24, 2022

Abstract

This paper explores the possibility that demand for costly commitment may prove unnecessary and thus excessive. In an online experiment, subjects face a tedious productivity task where tempting YouTube videos invite procrastination. Subjects can pay for a commitment device that removes the videos with some probability less than one, allowing us to compare their willingness to pay with realized material and psychological costs of temptation. A significant share of subjects overestimate their commitment demand, being overly pessimistic about their performance when tempted. However, the total realized *ex-post* disutility from undercommitment is greater than that from overcommitment.

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

*This draft supersedes an earlier draft circulated under the title "Pessimism and Overcommitment". We thank Giuseppe Attanasi, Pol Campos-Mercade, Christina Gravert, Olof Johansson-Stenman, Leonhard Lades, Lisa Norrgren, Eva Ranehill, Séverine Toussaert and participants at the SABE ECR Workshop (Dublin), UCD Behavioural Science and Policy seminar, University of Gothenburg Behavioural and Experimental Economics seminar, ESA Junior Faculty webinar and UCD Behavioural Science Workshop on "The Economics of Self-Control" for helpful comments and suggestions. We also thank Miloš Fišar and the team at Masaryk University Experimental Economics Laboratory (MUEEL) for their assistance with running the experimental sessions. Financial support from the Torsten Söderberg Foundation [grant no. E25/18] and Formas [grant no. 2017-00225] is gratefully acknowledged. Our analysis plan is pre-registered at osf.io/dz3fp. The experiment was approved by University College Dublin Office of Research Ethics (reference no. HS-E-21-52-Samahita). Declarations of interest: none.

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

‡School of Economics and Geary Institute for Public Policy, University College Dublin. E-mail: margaret.samahita@ucd.ie

1 Introduction

Commitment devices have been proposed as a solution to self-control problems in a variety of settings, including health, personal finance and work.[1] The existing literature has largely focused on the underdemand of commitment, i.e. overoptimistic decision makers not demanding sufficient levels of commitment to completely eliminate the self-control problem (DellaVigna and Malmendier, 2006; Heidhues and Kőszegi, 2009; John, 2020; Bai et al., 2021); the policy implication being that commitment take-up should increase where possible (for example, Acland and Chow, 2018; Sadoff and Samek, 2019; Hoong, 2021). However, recent experimental research suggests that, possibly driven by errors (Carrera et al., forth.), commitment demand is often present among those who do not seem to need it, seemingly reducing welfare. Royer et al. (2015) find that agents with a high frequency of pre-study gym visits are relatively more likely to demand commitment to go to the gym. Similarly, in a home grocery delivery program, Sadoff et al. (2020) show that commitment demand is higher among agents who are less likely to exchange healthy items for unhealthy items. This implies that "overcommitment" could be a real problem among pessimistic decision makers, and suggests that increased take-up of costly commitment may sometimes be detrimental.

However, although such findings are suggestive, they are not direct proof of over-commitment. The recent studies use either costless (Sadoff et al., 2020) or refundable commitment (Royer et al., 2015), thus do not determine whether observed commitment demand is truly excessive and lead to an *ex-post* utility loss to the decision maker. In this paper, therefore, we present an online experiment designed to provide exactly such evidence. Our design allows us to evaluate, within a single sample, the existence and magnitude of both excess and insufficient commitment demand. We are also able to directly compare the severity of each error, and thus infer whether or not overcommitment is an economically significant phenomenon in comparison with undercommitment.

In our experiment, the commitment device removes YouTube video pop-ups and thumbnails that would otherwise be shown during a tedious online work task. Watching these YouTube videos means that the subject spends less time on the work task, thus

[1] For example, smoking cessation (Giné et al., 2010), grocery shopping (Schwartz et al., 2014; Sadoff et al., 2020), gym attendance (Milkman et al., 2013; Royer et al., 2015; Carrera et al., forth.), alcohol consumption (Schilbach, 2019), gaming (Acland and Chow, 2018), savings (Ashraf et al., 2006; Beshears et al., 2015; John, 2020), work tasks (Kaur et al., 2010; Augenblick et al., 2015; Toussaert, 2018; Houser et al., 2018) and screen time (Allcott et al., 2021; Hoong, 2021; Marotta and Acquisti, 2018; Zimmermann and Sobolev, 2020).

reducing productivity and earnings. We picked this setting because of its obvious relevance—since the start of the COVID-19 pandemic, an increasing number of individuals are working from home and experience the need to ignore distractions online (and offline). As a temptation, YouTube videos are ideal, being highly familiar to our US sample.[2] Along with other social media, YouTube use is something individuals often wish to restrict (Allcott et al., 2021).

We elicit subjects' willingness-to-pay (WTP) for the non-refundable commitment device that removes the videos. Commitment is then allocated with some probability strictly less than one, allowing us in an incentive-compatible way to observe the behavior of subjects who demand commitment but have to face temptation. We assume that subjects' valuation of the commitment device consists of two components: i) material loss from being exposed to temptation, and ii) any non-material 'psychological costs', such as the mental burden of maintaining self-control while facing temptation. For this reason, we elicit subjective beliefs regarding expected productivity, as well as the *ex-ante* expected and *ex-post* actual experienced difficulty of resisting temptation.

On the extensive margin, the commitment take-up rate (subject share stating positive WTP) is 40%—somewhat larger than, for example, 28% in Ashraf et al. (2006) and 25% in Acland and Chow (2018) but in line with most previous studies where agents who demand commitment are typically in the minority.[3] Since the allocation of commitment incorporates a random implementation rule, a large majority of subjects (94%) still have to face temptation, as intended. Of these exposed subjects, most but not all manage to resist being tempted by the videos (clicking on a video or spending extra time watching a pop-up).

We find that 19% of exposed subjects overestimate their need for commitment when compared (only) to their actual material loss from facing the temptation. By the same measure, a much larger 48% act according to standard accounts of naïve decision makers and underestimate their demand for commitment. We obtain similar results when comparing against both material and psychological costs of temptation: 17% of subjects

[2] According to Pew Research Center (2021), YouTube is the most popular social media platform in the US: it is used by 81% of Americans with 54% of users using it daily and 36% visiting the site multiple times a day.

[3] Given the extent of self-control problems in domains such as personal finance and health, this may be consistent with underdemand by naïve present-biased agents. On the other hand, Laibson (2015) shows that in the presence of commitment cost, relatively low demand may in fact also be consistent with overcommitment.

seem to state a WTP higher than seems justified from the sum of these costs. By the same measure, 24% of subjects underestimate the material and psychological costs of facing temptation and understate their WTP. In our data, overcommitment appears to be associated with excessive pessimism with regards to task productivity under temptation. Compared to non-overestimators, WTP overestimators display both greater *expected* productivity loss and smaller *actual* loss from temptation.

Finally, for each subject where demand for commitment appears either too large or too small, we calculate their realized disutility from the error. This is done under the (admittedly strong) assumption that all purely psychological costs of being tempted, such as effort costs from maintaining self-control, are zero. Thus, this calculation effectively yields the total monetary losses from being tempted. Summing these losses across our sample, we find that the total utility loss from undercommitment is nearly five times larger than that from overcommitment. Thus, although we observe substantial overcommitment in our sample, undercommitment seems relatively more severe, vindicating the traditional focus on the latter bias in the literature.

Because we use YouTube content to tempt and distract subjects, our paper also relates to a rapidly expanding literature on the effects of social media. Much of this work has focused on social media's impact on well-being, documenting negative effects including loneliness, depression, lower life satisfaction and physiological stress (Allcott et al., 2020; Kross et al., 2013; Mosquera et al., 2020; Vanman et al., 2018)—primarily due to social media's role in facilitating social comparison (Braghieri et al., 2021; Tromholt, 2016; Hunt et al., 2018) or a feeling of having wasted time (Sagioglou and Greitemeyer, 2014).[4] However, a small number of studies have also examined the impact of restricting social media use, and screen time more generally, on productivity as measured by academic performance, finding no effects (Collis and Eggers, 2019; Zimmermann and Sobolev, 2020). Bjerre-Nielsen et al. (2020) suggest that the negative effect of smartphone use on academic performance may be overestimated. The only study that has looked at the effect of implementing a screen time commitment device on productivity, and thus closest to our paper, is Marotta and Acquisti (2018). They randomize participants into either exogenous (forced) or endogenous (optional) use of an app that blocks certain social media websites.

[4]Another strand of the recent literature on social media has examined its effects on political behavior (see Zhuravskaya et al. (2020) for a review). Social media can influence political attitudes (Fujiwara et al., 2021), with potentially negative consequences including the propagation of extreme viewpoints (Müller and Schwarz, 2021, 2020; Bursztyn et al., 2019) and political polarization (Levy, 2021).

While productivity and earnings significantly increase in the former condition, only a minority use the app and no treatment effect is observed in the latter condition. Our study suggests that offering such tools at a cost can be potentially detrimental to a subset of individuals who, ex-post, do not need them as much as they were willing to pay.

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

We seek to identify subjects' excess demand for commitment during the following online transcription task adapted from Augenblick et al. (2015). A series of 35 blurry Greek characters are shown in one row on the screen. Subjects reproduce them by pointing and clicking on the corresponding buttons below the characters. For a submission to be considered correct, it must be 90 percent accurate, with a maximum of 3 errors. Subjects are asked to complete as many tasks as they can within 15 minutes. Each correct submission is paid USD 0.50. A screenshot of the task is provided in Figure 2.

The set-up of the experiment is shown in Figure 1; the complete instructions are provided in Appendix C.[5] There are two main stages, which we call Task 1 and 2; subjects are informed before Task 1 that only one of the stages will be randomized for payment. In Task 1, subjects perform the transcription tasks without temptation. Before each transcription task, subjects are shown an empty pop-up window which they will have to close to proceed to the next task. These windows are included as a control for the time taken to close similar pop-up windows that are later used with the temptation, to be described below. As in that later temptation stage, the task clock continues running while each pop-up (including the very first one) is shown, and each pop-up window's "Close" button appears two seconds after the pop-up itself. Immediately before the start of Task 1, we familiarize subjects with the setting by having them complete a single practice transcription task. They are then told how many characters they got right and whether this was enough for their submission to be counted as correct.

After Task 1, we reveal to each subject their total number of correct submissions. We then inform them that the next stage involves another 15 minutes of transcription tasks

[5]The online experiment is available at https://transcriptiontask-f46b6.web.app and the source code is available at https://github.com/msamahita/overcommitment.

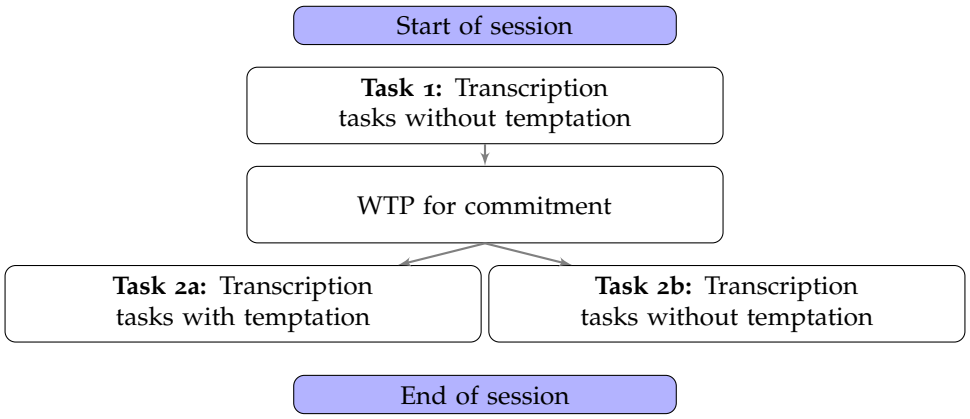


Figure 1: Experiment set-up.

Figure 2: Transcription task 1.

(Task 2), but that this time they will be faced with a temptation in the form of 10 YouTube thumbnails appearing at the bottom of the screen below each transcription task, as in Figure 3. The ideal temptation would be a personalized set of recommended videos as determined by the YouTube algorithm; however, for privacy reasons, YouTube does not allow recommendations to be scraped by any application programming interface (API). Instead, we use the following approach to provide partial personalization. At the start of the study, we list the 13 video categories for which trending videos (in the US) can be pulled from the YouTube API, and ask subjects to select the 3 to 5 categories that interest them the most.[6] For each subject, we then pull 200 videos per selected category, combine

Figure 3: Transcription task 2 with YouTube thumbnails.

them into a set of 600-1000 videos, and draw 10 (without replacement) to be presented below each new transcription task. Figure 3 shows an example screen with categories "People & Blogs, News & Politics, Howto & Style" selected.

Additionally, of the 10 videos shown with each transcription task, the first (top-left) is *automatically* played in the immediately preceding pop-up window; see Figure 4. Subjects will thus need to resist being distracted by the video and close the pop-up window to proceed to the task. The task clock will continue running while each pop-up (including the very first one) is shown. To make sure that each autoplaying video has sufficient loading time to start playing, each pop-up window's "Close" button only appears two seconds after the pop-up itself.

Subjects can click on any YouTube video at any point during the task, using the pop-up window or any of the thumbnails appearing below the transcription task. If the subject clicks on a video, a new browser tab will open where they will be able to view the corresponding video on YouTube. They can watch the video for as long as they like (and potentially open other tabs) and come back to the study tab at any time, and they can subsequently click on another video if they want. However, the task clock will keep running while they are watching the videos. We record i) the total time each subject spends watching the pop-ups, ii) how many times each subject accesses YouTube

[6]The categories are: Film & animation, Autos & vehicles, Music, Pets & animals, Sports, Gaming, People & blogs, Comedy, Entertainment, News & politics, Howto & style, Educational, and Science & technology.

Figure 4: Example of pop-up appearing before each transcription task, with "Close" button in bottom right.

by clicking on a pop-up or a video thumbnail, iii) the total time each subject spends on a YouTube tab accessed through the above methods and iv) the total time each subject spends on any other tab. If 15 minutes elapse while they are viewing a video, the video tab will automatically close and the subject will be taken back to the study tab. Thus, subjects are aware that to get the highest possible monetary pay-off they would have to exercise willpower to overcome the temptation.

Since subjects are able to freely access YouTube at any point during the experiment simply by opening a different browser tab, our temptation should not be understood as exploiting the desire to view interesting videos per se, but rather as having videos present *while working on the transcription tasks*. The intention is to capture the distraction individuals typically experience when working online, where engaging videos or links may appear in sidebar thumbnails and pop-ups. Similarly, that videos autoplay in the pop-up mimics social-media settings where videos play automatically as a user scrolls through their newsfeed. The external validity of our temptation has only increased since the start of COVID-19 as more individuals are working from home and experience the need to ignore distractions online and offline. As a result, our form of temptation is highly familiar to subjects. It also has immediate appeal given the personalization of the video categories and the monotonous transcription tasks, and should be perceived to be bad since subjects choosing this option forfeit the possibility to earn more money from

the work task. Qualitative feedback elicited at the end of the experiment suggests that subjects do consider the YouTube videos as a temptation to be avoided.

Subjects again complete an initial practice round prior to Task 2 itself. Given that the new setting is significantly more complex than than Task 1, the practice round now includes two transcription tasks with pop-up windows and videos. However, we then offer subjects a commitment device: the possibility of paying to fully remove all YouTube videos from the task environment for the entire 15-minute period. As in Task 1, subjects will still encounter a pop-up before each task (and the clock will continue running while each pop-up is open), but these pop-ups will not contain a YouTube video. The presence of the pop-ups even when subjects choose to remove the videos ensures that we control for the time taken to click the "close" button on the pop-up which autoplays a YouTube video. This enables us to isolate the effect of the temptation from the effect of having a pop-up shown.

WTP for this commitment device is elicited using the Becker-DeGroot-Marschak (BDM) mechanism combined with a coin toss as depicted in Figure 5, which is taken from our instructions to participants. Subjects state a price between 0 and 100 cents and are told that the possibility of successfully removing the YouTube videos increases the higher their stated price (this probability is equal to $WTP/200$). To guard against potential experimenter demand effects in either direction, subjects are informed both that this possibility is maximized (but not guaranteed) if they state a price of 100, and that they can ensure that the videos will *not* be removed by stating a price of 0. We do not allow for negative WTP since, as noted, subjects are already able to view trending videos at zero cost simply by accessing YouTube in a different browser tab. Payment for the commitment device is taken out of subjects' total earnings regardless of whether Task 1 or 2 is randomly picked for payment.

To ensure that subjects understand the procedure, we add a practice exercise specifically to illustrate how the BDM mechanism works. WTP is then elicited twice. The second elicitation (denoted WTP_2) occurs after subjects have been asked to reflect on their own productivity (see the next section), which may promote more accurate preferences. WTP_2 is framed as an option to revise the initial measure (denoted WTP_1), which is not used in our main analysis, though results are robust to using it in place of WTP_2. Finally, subjects proceed to Task 2 itself, where they do the transcription tasks with or without the videos,

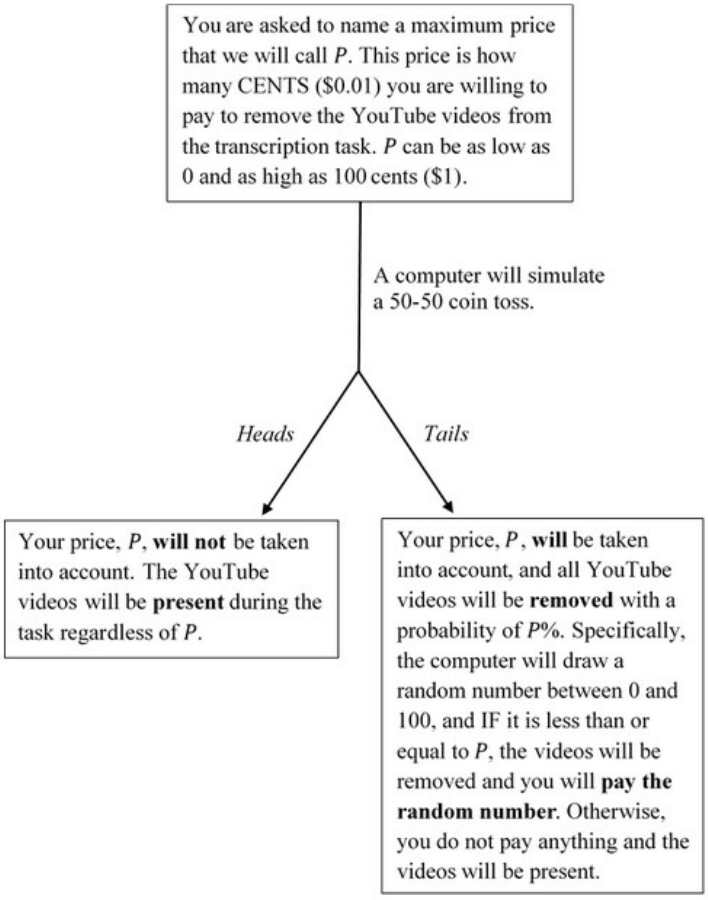


Figure 5: Elicitation of WTP for commitment.

as determined by the outcome of the coin toss and BDM mechanism.[7]

2.1 Measures of subjects' productivity beliefs

The subject's actual productivity is measured as y_1 and y_2, the number of correct submissions in Tasks 1 and 2 respectively. We also elicit unincentivized subject beliefs about their productivity. Directly after the conclusion of Task 1 and after subjects are told about their performance y_1, we ask subjects how many correct submissions they expect to get if they were to redo Task 1, where there was *n*o *t*emptation ($\hat{y}^{nt}$);[8] this is followed by

[7]The use of within-subject design is motivated by wanting all subjects to indicate their WTP for the commitment device with all information about the work task available to them, including their own experience of it.

[8]We consistently apply "hat notation" to all belief variables elicited in the experiment. Note that belief elicitation is not without issues. Although we ask subjects to state a single expected figure, it may be that

information about the second task and the elicitation of WTP_1. Next, we ask how many correct submissions they expect to get if the *t*emptation is present in Task 2 ($\hat{y}^t$).[9] These 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.[10] After the above variables have been measured, we elicit WTP_2 (henceforth WTP), explaining that it will determine whether the videos are displayed in Task 2.

2.2 Other variables

To better understand subjects' estimation of the psychological cost of temptation, we elicit a measure of expected temptation strength before Task 2 with the question "How tempted do you think you would be to click on any of the YouTube videos?". Subjects respond on a scale from 1 to 4 (not at all tempted, not that tempted, quite tempted, very tempted), denoted θ. Subjects' actual experience of the temptation, compared to *expectation*, 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 videos and concentrating on the transcription tasks was (i) higher than expected, (ii) lower than expected, or (iii) as difficult as expected (or N/A if the videos were removed). We do this to investigate whether subjects' WTP accurately captures their expectation of the psychological costs of temptation. We define the variable v to be equal to 1 in case (i) where psychological costs of temptation were overestimated; -1 in case (ii); and zero in case (iii).

In the post-experiment questionnaire we also elicit subjects' perception of their willpower using the brief self-control measure (Tangney et al., 2004). This question-

they instead consider a distribution of cases and report some other statistic, such as a modal outcome; or round values up or down. If so, their choices may appear less rational in our analysis than they actually are. In this paper, we make the assumption that subjects do state a single expected figure when asked to do so, and that it can be treated as certain conditional on the outcome of the coin toss and the BDM mechanism.

[9]Subjects also state the percentage likelihood ($\hat{p}^c$) that they will *c*lick on a video, if present, and also their expected productivity assuming that they *c*lick ($\hat{y}^c$) or do *n*ot *c*lick a video ($\hat{y}^{nc}$). These values allow us to construct a second, *inf*erred measure of expected productivity under temptation, $\hat{y}^{t,inf} = \hat{p}^c\hat{y}^c + (1-\hat{p}^c)\hat{y}^{nc}$. This measure excludes succumbing by simply keeping video pop-ups open, thus likely overestimating expected productivity under temptation. Indeed, on average $\hat{y}^{t,inf} > \hat{y}^t$ (9.85 vs 9.55, t-test, $p = 0.0009$), consistent with subjects expecting that being able to view videos in the pop-up will lower productivity.

[10]Using an incentivized measure of productivity in a previous experiment (see Section 4.4) yielded similar results as in our main analysis.

naire consists of 13 statements, to each of which the subject indicates their agreement on a five-point scale, reverse-coded where necessary (Cronbach's $\alpha = 0.89$). Some example statements include: "I am good at resisting temptation" and "I often act without thinking through all the alternatives". These values are aggregated to give ω, the perceived general level of willpower. We collect data on demographic variables such as age, gender, and time spent on YouTube daily on a scale from 1 to 4 (less than 30 minutes, from 30 minutes to 1 hour, from 1 to 2 hours, more than 2 hours).

3 Hypotheses

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

The decomposition of WTP can be shown within a simple expected-utility model. Recall that each Task is paid with probability $1/2$ and that the BDM outcome is only used with probability $1/2$. Conditional on the BDM being used, with probability $WTP/100$ the random number $R \in \{1, 2, ..., 100\}$ drawn by the computer is no larger than the stated price, in which case the videos are removed and the subject pays the amount R. Note that conditional on the BDM being used, R is paid regardless of the task selected for payment. We assume that Bernoulli utility takes as argument the sum of state-dependent costs and benefits, so $u = u(x + PC)$, with wealth x and psychological costs PC. For some fixed (possibly accurate) expectations on earnings with and without temptation (y^t, y^{nt}), subjects are taken to maximize expected utility as

$$\begin{aligned} U(WTP) = \frac{1}{2}\Bigg[&\frac{1}{2}\cdot u(50y_1 - PC) \\ &+\frac{1}{2}\left(\frac{100 - WTP}{100}\cdot u(50y_1 - PC) + \frac{1}{100}\int_0^{WTP} u(50y_1 - R)dR\right)\Bigg] \\ &+\frac{1}{2}\Bigg[\frac{1}{2}\cdot u(50y^t - PC) \\ &+\frac{1}{2}\left(\frac{100 - WTP}{100}\cdot u(50y^t - PC) + \frac{1}{100}\int_0^{WTP} u(50y^{nt} - R)dR\right)\Bigg] \end{aligned} \tag{1}$$

The last line, for example, is the case when Task 2 is used for payment (with probability 1/2) and the BDM outcome is taken into account (with probability 1/2). Then, with probability $(100 - WTP)/100$, the subject does not get the commitment and payoff is given by 50 times the number of correct answers when exposed to temptation less any psychological cost from the temptation. The final integral term means that with probability $WTP/100$ the subject does get the commitment and payoff is given by 50 times the number of correct answers when not exposed to temptation less the payment of R, the random number drawn.[11]

We solve for maximum WTP under the assumption of risk neutrality, which is reasonable given the small payments at stake (Rabin, 2000); Online Appendix A provides a robustness check assuming risk aversion using lab data from a previous experiment (see Section 4.4). Risk neutrality yields

$$WTP = 25(y^{nt} - y^{t}) + PC \tag{2}$$

The first term (expected material loss) is the subject's expected productivity without temptation less their expected productivity with temptation. 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 videos are present, even if they are never watched. For example, in an exploratory analysis of the effort task of Toussaert (2018), subjects who were exposed to unwanted temptation were found less productive, suggesting that self-control is indeed costly.

The second term (expected psychological cost) is not directly measured by our design. It may reflect at least two different motivations: (i) the disutility of expending effort to resist temptation (whether successfully or not); and (ii) any self-image loss should the subject succumb to temptation. In principle, psychological cost could also include the value (if any) of being shown thumbnails and videos within the experiment. Since subjects can freely access YouTube through other browser tabs, such value would not reflect the utility of watching videos per se, but rather that the mere presence of the videos might add novelty or variation to the transcription tasks. However, we expect this effect to be

[11] Thus, for simplicity we treat R as a continuous variable in $[0, WTP]$. The solution to the discrete problem corresponding to equation (1) adds a term $-1/2$ to equation (2), reflecting the fact that $\int_0^{WTP} R\,dR - \sum_{R=0}^{WTP} R = WTP^2/2 - WTP(WTP+1)/2 = -WTP/2$. Our results are (unsurprisingly) robust when comparing subjects' stated WTP with this slightly lower value of optimal WTP.

quite small compared to the psychological negatives of being distracted.

The analysis proceeds with the following steps.[12]

1. We begin by comparing WTP with **actual** performance in the transcription tasks. As suggested above, subjects might overestimate WTP relative to actual material loss because they (i) inaccurately estimate future material payoffs and/or (ii) anticipate non-material psychological costs of facing temptation. Although the latter explanation remains consistent with "correct" demand for commitment as given by equation (2), examining whether commitment exceeds material losses is arguably interesting in its own right.

   Note that, for the sub-sample of subjects who obtain commitment, we are unable to infer realized material losses because we never observe how these subjects perform when exposed to temptation. Thus, these subjects cannot be used to compare WTP with material losses. By contrast, for the sub-sample of subjects who do face temptation in Task 2, we may be able to infer material losses by comparing those subjects' Task 2 performance with that in Task 1, where they did not face temptation. In effect, we would use performance in Task 1 as the counterfactual, setting $y^{nt} = y_1$ (while $y^t = y_2$). However, if there are learning or fatigue effects across the Tasks, it should be clear that Task 1 performance cannot directly substitute for counterfactual Task 2 performance. Thus, we will first check whether performance differed across Task 1 and 2 among those who faced temptation in neither Task, i.e., obtained commitment. Of course, since there may be selection into this group, the test can only provide circumstantial evidence that our approach is valid.

   In any case, supposing that no adjustment for learning or fatigue is necessary, we can then classify each subject exposed to temptation as *material loss* (M) overestimators, M accurate estimators or M underestimators, depending on whether their WTP is above, below, or equal to what would maximize utility when only material loss is considered. In principle, the three classes are defined by $WTP > 25(y_1 - y_2)$, $WTP < 25(y_1 - y_2)$, and $WTP = 25(y_1 - y_2)$, respectively. However, since WTP is bounded at zero, we add all subjects who state $WTP = 0$ but have $y_1 - y_2 < 0$ to the last group rather than the first group of M overestimators. According to equation (2), these subjects' WTP should have been negative, but the experiment

[12] A pre-analysis plan is available at https://osf.io/dz3fp/.

only allows for values greater than or equal to zero. Similarly, all subjects for whom $25(y_1 - y_2) > 100$ and $WTP = 100$ are treated as M accurate estimators rather than M underestimators.

We then test the following null hypothesis. All tests of proportions are based on the standard normal approximation of binomial parameters and assuming a threshold share of 10%.[13]

Hypothesis 1. *Among the subjects who face temptation in Task 2, no more than 10% have* $WTP > 25(y_1 - y_2)$.

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

2. Next, we compare WTP with **expected** performance in the transcription tasks, as given by $\hat{y}^t$. We test whether or not WTP is overestimated relative to expected material losses, as captured by the difference between the expected number of correct answers without temptation less the expected number when exposed to temptation.

   Hypothesis 2. *Among the subjects who face temptation in Task 2, no more than 10% have* $WTP > 25(\hat{y}^{nt} - \hat{y}^t)$.

3. A WTP even higher than expected material losses may be driven by anticipation of psychological costs, but also by "true" overestimation (of both material losses and psychological costs). Because the actual PC is unknown and not elicited in our design, we cannot test directly whether stated WTP is larger than the entire realized right-hand side of equation (2). However, an indirect test is possible. Starting from equation (2) and denoting expected quantities by subscript e and actual values (i.e.

[13] This threshold can be interpreted as the maximum proportion attributable to subject confusion or demand effects. While seemingly conservative, this threshold is picked given a lack of existing studies measuring such drivers in this setting. De Quidt et al. (2018) suggest that typical demand effects are modest in experiments. Our setting involves little uncertainty regarding the task, and additionally WTP is elicited twice, which should minimize any confusion. Nevertheless, it is conceivable that confusion and experimenter demand would affect a larger share of subjects than 10%. Yet even then, we note that experimenter demand effects are not unlike the impact of a nudge designed to increase commitment take-up.

accurate expectations) by subscript a, true overestimation in the expected-utility model with risk neutrality would be characterized by

$$\begin{aligned} WTP\left(\cdot_e\right) > WTP\left(\cdot_a\right) &\iff 25(y_e^{nt} - y_e^t) + PC_e > 25(y_a^{nt} - y_a^t) + PC_a \\ &\iff 25\left((y_e^{nt} - y_e^t) - (y_a^{nt} - y_a^t)\right) > -(PC_e - PC_a) \end{aligned} \tag{3}$$

Thus the overestimation of material losses needs to exceed any *under*estimation of psychological costs. Our approach is thus to ask the subjects whether they think resisting temptation was easier than expected, yielding v. If $v = 1$, our interpretation is that $PC_e > PC_a$, and the RHS of the inequality is negative. This in turn implies that a sufficient condition for overestimation is that the LHS is greater than or equal to 0. Furthermore, for subjects who find that ignoring the temptation was just as easy or difficult as expected ($v = 0$), the RHS of the above inequality is taken to be zero, implying that a sufficient condition for overestimation is that the LHS is strictly greater than 0. Note that because these conditions are not necessary, they imply a lower bound on the number of overestimators.

Testing the pair of conditions 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}. As in testing Hypothesis 1, we suggest to use $y_a^{nt} = y_1$, in which case the LHS gives $(y_e^{nt} - y_e^t) - (y_1 - y_2)$. We then use elicited beliefs such that $y_e^{nt} = \hat{y}^{nt}$ and $y_e^t = \hat{y}^t$. Additionally, as before, subjects cannot state negative WTP, so $WTP > 0$ is a necessary condition for strict overestimation.

Hypothesis 3. *Among the subjects who face temptation in Task 2, no more than 10% have* $(\hat{y}^{nt} - \hat{y}^t) - (y_1 - y_2) \geq 0$ *and* $v \geq 0$*, with at least one strict inequality, and* $WTP > 0$*.*

Rejection of this hypothesis suggests that a substantial share of subjects have overestimated WTP.

4 Results

4.1 Descriptive results

We recruited subjects from Prolific, a platform that provides a pool of participants for online research. We started by recruiting an initial sample of 350 US participants, nationally representative in terms of age, gender and ethnicity. After paying out their experimental earnings, our budget allowed us to recruit a further 75 participants, which we proceeded to do as pre-registered. Since the minimum number of subjects required by Prolific for nationally representative sampling is higher than 75, the second wave pre-screens participants using the following criteria: US country of residence, balanced in terms of gender, and excluding participants who have participated in the previous study. We pool the samples in our analysis and, after accounting for incomplete responses, end up with a total of 409 participants. Results are similar if the first sample is analyzed separately; we do not have enough power to analyze the second sample separately.

Summary statistics from the experiment are presented in Table 1. Despite the fact that the average subject expects material losses from temptation ($\hat{y}^{nt} - \hat{y}^{t}$) corresponding to about 1.13 correct submissions, most subjects have little willingness to pay for the commitment device, consistent with previous studies such as Augenblick et al. (2015). 247 subjects (60.4%) state $WTP = 0$. The average WTP is 10.62 cents, or 26.80 for those with positive WTP. The conditional-on-positive WTP distribution is given in Figure 6. In total, 26 subjects are successful in getting the temptation removed after stating a price greater than the random number drawn and given that the BDM outcome is used.

4.2 Evidence of temptation

Recall that there are two main ways that subjects may succumb to temptation in our experiment. First, they may click a thumbnail or pop-up video to open a new YouTube browser tab where the corresponding video can be watched. Second, they may spend time watching videos directly in a pop-up window within the study tab. Our data suggests that while most subjects are able to resist both types of temptation, a non-negligible share of tempted subjects indeed succumb to at least one of them.

For the first type of temptation, 60 of the 383 subjects who are exposed to videos (15.7%) click on a pop-up or thumbnail (or both) and subsequently spend an average of 86

Table 1: Summary statistics.

Variable	Mean	SD	Min	Max	N
WTP, final stated maximum price (in cents) for removing videos	10.62	21.78	0	100	409
Behavior of those who do NOT face temptation					
y_1, correct submissions in Task 1	8.42	4.47	0	15	26
y_2, correct submissions in Task 2	11.08	6.11	0	20	26
d_{p1}, seconds elapsed during pop-up in Task 1	57.12	17.92	21	109	26
d_{p2}, seconds elapsed during pop-up in Task 2	64.50	35.07	36	208	26
Behavior of those who face temptation					
y_1, correct submissions in Task 1	9.36	5.05	0	22	383
y_2, correct submissions in Task 2	11.80	5.63	0	27	383
d_{p1}, seconds elapsed during pop-up in Task 1	56.15	34.99	14	321	383
d_{p2}, seconds elapsed during pop-up in Task 2	77.34	102.19	4	900	383
n_{click}, times subject clicks on a YouTube pop-up or thumbnail	0.49	1.87	0	19	383
d_{YT}, seconds spent on YouTube accessed through pop-up or thumbnail	85.50	188.20	0	907	60
Beliefs about self-productivity in Task 2					
$\hat{y}^{nt}$, predicted correct submissions in Task 2 if temptation NOT present	10.69	6.66	0	90	409
$\hat{y}^{t}$, predicted correct submissions in Task 2 if temptation present	9.56	5.55	0	35	409
$\hat{y}^{nc}$, predicted correct submissions in Task 2 if video NOT clicked	10.08	5.41	0	35	409
$\hat{y}^{c}$, predicted correct submissions in Task 2 if video clicked	7.37	4.75	0	33	409
$\hat{p}^{c}$, predicted likelihood of clicking a video	0.13	0.23	0	1	409
Psychological measures					
θ, how tempted subject expects to be	1.69	0.79	1	4	409
v, ignoring temptation was easier than expected	0.30	0.73	-1	1	383
ω, score on brief self-control scale (Tangney et al., 2004)	42.51	10.23	17	65	409
Subject characteristics					
Age	40.67	16.30	18	81	409
Male	0.49	0.50	0	1	409
Daily time spent on YouTube	2.25	1.10	1	4	409
1: $<$30 minutes					
2: $>$30 minutes & $\leq$1 hour					
3: $>$1 hour & $\leq$2 hours					
4: $>$2 hours					

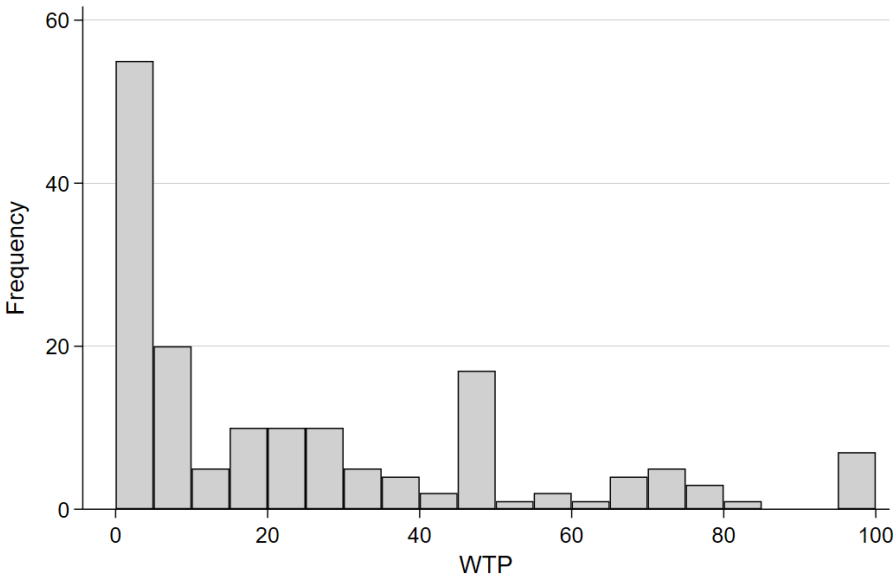


Figure 6: WTP for commitment device, for the 40% of subjects stating $WTP > 0$. Each bin represents an interval of length 5, i.e. the first bin is $0 < WTP \leq 5$, the second $5 < WTP \leq 10$ and so on.

seconds on YouTube, about the time it takes most subjects to complete one transcription task. Five subjects spend more than half of the total Task 2 duration (900 seconds) on YouTube. On the other hand, about half of the 60 subjects spend less than 10 seconds on YouTube after clicking on a video. For these subjects, it appears that the video pop-up or thumbnail is sufficiently tempting to induce them to leave the work task, but not enough for them to continue watching.

For the second type of temptation, the median extra time spent in pop-up windows in Task 2 compared to Task 1 is exactly zero seconds, yet the distribution of such pop-up time differentials is quite asymmetric, with a long right tail where subjects spend a great deal more time in pop-ups in Task 2. For example, while the 5th percentile of extra viewing time is -37 seconds, the 95th percentile is 162 seconds. As a result, the average extra time is about 21 seconds. Taken together, the end result of both temptation mechanisms is that, for example, 19 of the 383 tempted subjects (5.0%) spend one-third or more of total Task 2 time either in pop-ups or in YouTube browser tabs; yet only a single subject spent at least a third of the total Task 1 time in pop-ups. Hence, we confirm that the temptation worked

as intended.[14]

4.3 Main results

We begin by comparing subjects' WTP for removing temptation with their *actual* productivity loss when exposed to temptation. As discussed in Section 3, we do so for the 383 subjects who do not obtain commitment.[15] Taking each tempted subject's Task 1 performance as their non-tempted counterfactual, we then classify subjects as *material loss* (M) overestimators, M accurate estimators or M underestimators according to whether their WTP is above, equal to, or below what would maximize utility when only material loss is considered.

Table 2 summarizes these classifications. While around 54.8% of subjects are classified as M accurate estimators, the majority have $y_1 < y_2$ but $WTP = 0$. Around 29.2% of subjects are M overestimators, greater than the 10% attributed to confusion ($p < 0.0001$), thus rejecting Hypothesis 1. The proportion of M overestimators is also significantly greater than that of M underestimators (29.2% vs 16.0%, two-sample test of proportion, $p < 0.0001$), contrary to the classical narrative of underdemand for commitment devices.

As seen in the column $y_1 - y_2$, the average M overestimator actually performs better in Task 2 than in Task 1, seemingly making their positive WTP for commitment device (on average 21.09) larger than justified. However, Table 1 also shows that those *not* facing temptation in Task 2 clearly exhibit a learning effect: subjects do better in Task 2 than Task 1 ($\bar{y}_2 = 11.08$ vs $\bar{y}_1 = 8.42$, t-test, $p = 0.0017$). This casts substantial doubt on the validity of the estimates just reported. On the other hand, we can adjust for learning among those who do face temptation by subtracting from their y_2 the average learning effect among those who do not face temptation (2.65); this is the approach we take in the

[14]To control for other types of online procrastination, we also record how much time subjects spend on browser tabs other than the work task and YouTube. Only 18% of tempted subjects opened such tabs during Task 1 and/or Task 2. As for the duration of viewing these tabs, due to a coding error, this variable is only updated when the subject returns to the task tab, meaning that it is understated for subjects who are still on the non-task tab when the 15 minutes end. Nevertheless, this duration is not significantly different in Task 1 vs Task 2, both for the group with and without temptation.

[15]This is a selected sample that tends to exclude subjects with very high WTP and thus high probability of getting commitment. We perform two robustness tests to address concerns about selection bias. First, we use just the half of the (exposed) sample where the coin flip decides against commitment, regardless of WTP. Second, we apply frequency weights with respect to WTP to the 383 subjects to account for the 26 "missing" subjects. Both methods do not change our conclusions.

rest of the paper.[16] Table 2 shows that, with the adjustment, around 19.3% of subjects are M overestimators, again greater than the 10% attributed to confusion ($p < 0.0001$), and rejecting Hypothesis 1. We now observe more M underestimators than overestimators (47.8% vs 19.3%, two-sample test of proportion, $p < 0.0001$).

Table 2: Classification of subjects who face temptation.

	Original data				Adjusted for learning effect			
Classification	N	WTP	$y_1 - y_2$	Frequency	N	WTP	$y_1 - y_2$	Frequency
M overestimators	112	21.09	-3.41	29.24%	74	26.24	-1.97	19.32%
$WTP > 0$	112	21.09	-3.41		74	26.24	-1.97	
M accurate estimators	210	0.24	-3.55	54.83%	126	0	-2.49	32.90%
$WTP > 0$	1	50	2					
$WTP = 0$	209	0	-3.57		126	0	-2.49	
M underestimators	61	7.59	3.16	15.93%	183	5.10	2.96	47.78%
$WTP > 0$	23	20.13	3.22		62	15.05	3.23	
$WTP = 0$	38	0	3.13		121	0	2.82	
Total	383	7.51	-2.44	100%	383	7.51	0.22	100%

Notes: Average WTP and productivity loss of subjects. In columns 2-5 we use the number of correct submissions in Task 2 as y_2, in columns 6-9 we adjust for learning effects by subtracting 2.65 from the number of correct submissions in Task 2. *WTP* and $y_1 - y_2$ show the average values within each type. *M overestimators* are those for whom $WTP > 25(y_1 - y_2)$, *M accurate estimators* are those for whom $WTP = 25(y_1 - y_2)$ (or $25(y_1 - y_2) < 0 = WTP$ or $25(y_1 - y_2) > 100 = WTP$), and *M underestimators* are those for whom $WTP < 25(y_1 - y_2)$.

In summary, we state our first result:

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

Given that both M overestimators and M underestimators are present in our data, how should we judge the relative severity of these two biases? One natural approach is to calculate the welfare (summed utility) loss for each type; since subjects are assumed risk neutral, utils will correspond to cents in this exercise. Because we are unable to directly measure the psychological cost of temptation, the calculation assumes that these equal zero for all subjects. Had they been included, optimal WTP would likely increase, and

[16] Adjusting for learning was not pre-specified in our analysis plan because such effects were found non-significant in a previous experiment (see Section 4.4). However, note that performing the adjustment lowers y_2, making WTP overestimation less likely. In the tempted group, learning is confounded with temptation in Task 2, so the identifying assumption is that average learning effects are comparable across tempted and non-tempted subjects. Gross productivity gains from Task 1 to Task 2 are similar across the groups (2.44 vs 2.65, t-test, $p = 0.7660$).

welfare losses for overestimators (underestimators) would be lower (higher). We first substitute each subject's stated WTP, along with y_1 and y_2, into expected utility function (1). Then, we compare the result with expected utility evaluated at the optimal feasible WTP, which will be either $25(y_1 - y_2)$, zero, or 100. Summing over all M overestimators yields a total welfare loss of \$6.42; lower than that for M underestimators at \$30.75.[17]

In general, it is possible that some subjects overestimate WTP compared to realized material losses because they have mispredicted those losses. However, if subjects also state WTP higher than their *expected* material losses, it appears some other factor, such as costs from having to maintain self-control, must be in play as well. Thus, we next compare WTP with subjects' *expected* material losses as elicited in the experiment, $25(\hat{y}^{nt} - \hat{y}^{t})$. Among subjects who face temptation, 18.8% have WTP greater than this quantity, a proportion significantly greater than the 10% threshold ($p < 0.0001$). Note that, since this test does not consider realized outcomes and thus does not require a counterfactual, it may also be performed on the full sample of subjects, including those who do not face temptation. Then, the proportion of subjects who ex-ante overestimate their WTP is 21.5%.

In any case, we conclude the following:

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

A possible explanation for such a discrepancy between WTP and expected material losses is that subjects expect to experience psychological discomfort from temptation, and seek to avoid it by stating a higher WTP. We therefore check whether WTP is still overestimated even when accounting for subjects' expectation of the psychological cost of temptation. As derived from equation (3) above, a sufficient condition, given $WTP > 0$, is that subjects' overestimation of material losses exceeds their underestimation of psychological cost, i.e. that $(\hat{y}^{nt} - \hat{y}^{t}) - (y_1 - y_2) \geq 0$ and $v \geq 0$, with at least one strict inequality. Without adjusting for learning effects, the proportion of overcommitters is significantly higher than that of undercommitters for whom $(\hat{y}^{nt} - \hat{y}^{t}) - (y_1 - y_2) \leq 0$ and $v \leq 0$, with at least one strict inequality (22.2% vs 8.9%, two-sample test of proportion, $p < 0.0001$). With the adjustment for learning, the proportion of subjects who overestimate WTP using this measure is 17.5%, less than that of undercommitters at 24.3% but significantly greater than the 10% threshold ($p < 0.0001$). We conclude that:[18]

[17] Without adjusting for learning effects, the total welfare loss from all M overestimators is \$13.30; larger than that for M underestimators at \$11.31.

[18] Recall from footnote 6 that we also construct an inferred measure of expected productivity under

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

4.3.1 Mechanisms behind overcommitment[19]

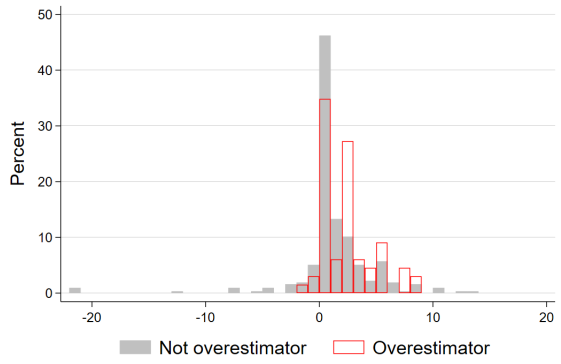
For a subject to be classified as an overestimator in Hypothesis 3, we require that $(\hat{y}^{nt} - \hat{y}^{t}) - (y_1 - y_2) \geq 0$. From this condition, it is not clear which LHS term(s) are the main mechanism for overcommitment. For example, overestimators might do equally well with and without temptation ($y_1 = y_2$) but believe they do worse when exposed ($\hat{y}^{nt} > \hat{y}^{t}$); or, they might think they will do equally well while in fact performing *better* when exposed. To check this, we compare outcomes and expectations between the 17.5% of subjects who overestimate their demand for commitment (as defined in Hypothesis 3), and the remaining 82.5% who do not, within the group of 383 subjects who face temptation.

We find support for both aspects of overcommitment: WTP overestimation is associated with both high expected productivity loss *and* low actual productivity loss (and indeed high actual productivity gain). As shown in panel (a) of Figure 7, overestimators expect relatively greater productivity losses when facing temptation; the difference in distribution is significant (Kolmogorov-Smirnov test, $D = 0.2674$, $p = 0.001$). Moreover, while the majority of subjects do better in Task 2 than in Task 1, panel (b) shows that actual productivity gains among overestimators are greater than for non-overestimators. The distributions are significantly different (Kolmogorov-Smirnov test, $D = 0.3661$ and $p < 0.001$). Comparing the individual components ($\hat{y}^{nt}, \hat{y}^{t}, y_1, y_2$), the only one for which distributions differ between the two groups is y_2: overestimators do significantly better than non-overestimators (Kolmogorov-Smirnov test, $D = 0.1867$ and $p = 0.042$). Thus, overcommitment appears to reflect a kind of unfounded pessimism: overestimators expect to do worse when exposed to temptation, when in fact they do quite well—not only relative to the first task, but to other subjects as well.

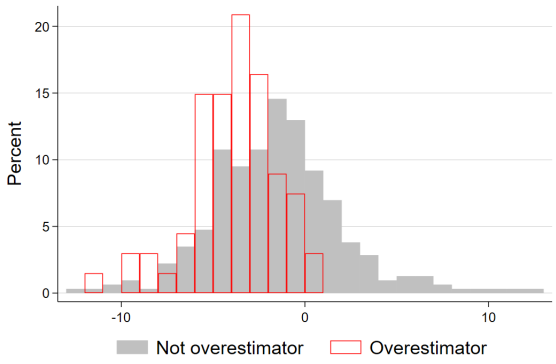
Do overcommitters differ from others in terms of their behavior when facing temptation? For example, one might imagine that one reason that overcommitters do particularly well in Task 2 is because they turn out to be especially good at resisting temptation.

temptation, $\hat{y}^{t,inf}$. Using this measure in place of $\hat{y}^{t}$ (as pre-registered) yields 19.3% of WTP overestimators for Result 2, and 17.0% for Result 3. Both results survive the Holm-Bonferroni correction for multiple hypotheses testing.

[19] While the analyses here were not-preregistered, we believe they provide further insights into the factors associated with overcommitment.

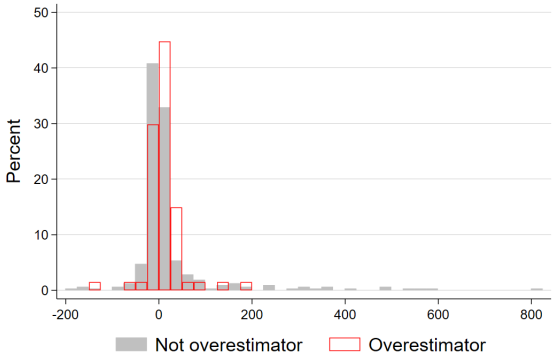


(a) $\hat{y}^{nt} - \hat{y}^{t}$: expected productivity loss

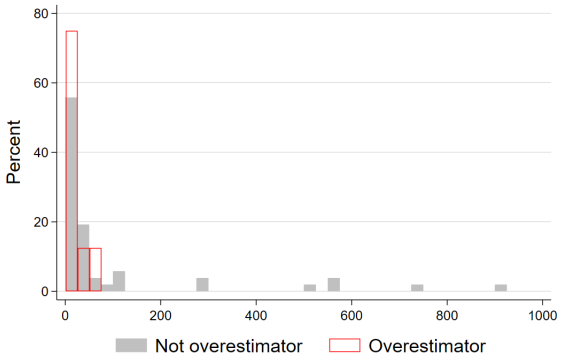


(b) $y_1 - y_2$: actual productivity loss

Figure 7: Distribution of realized outcomes and beliefs about productivity. One subject stated $\hat{y}^{nt} = 90$, panel (b) excludes this outlier.



(a) Extra time on pop-ups in Task 2



(b) Total time on YouTube in Task 2 for those who click on a video

Figure 8: Distribution of time spent on temptation

A natural starting point is to examine whether overcommitters themselves think they are especially likely to struggle under temptation. Recall that we elicit θ, a measure of how tempted subjects expect to be to click on any of the YouTube videos. However, its distribution is not significantly different between overcommitters and others (Kolmogorov-Smirnov test, $D = 0.1183$, $p = 0.422$), indicating that overcommitters do not *expect* to succumb more by clicking on a video. Neither is overcommitting associated with lower perceived self-control as elicited using ω, the aggregate score on the brief self-control scale (Tangney et al., 2004) (Kolmogorov-Smirnov test, $D = 0.0929$, $p = 0.727$).

As for *actual* behavior under temptation, when we compare the extra time spent on pop-ups (Task 2 pop-up duration minus Task 1 pop-up duration), overcommitters appear to spend longer watching these windows than non-overcommitters (Kolmogorov-Smirnov test, $D = 0.1865$, $p = 0.043$); see panel (a) of Figure 8—though the distribution for non-overcommitters displays a fat right tail.[20] Also, while a smaller fraction of overcommitters click on a video, the difference is not significant (11.9% vs 16.5%, two-sample test of proportions, $p = 0.3557$). Finally, focusing only on those who do click a video at least once, the total time spent in YouTube tabs is again not significantly different across the two groups (Kolmogorov-Smirnov test, $D = 0.2692$, $p = 0.696$); see panel (b) of Figure 8. Hence, there is at best weak evidence that overcommitters succumb more by watching the pop-ups longer than non-overcommitters. Overall, our data point toward task productivity being the main characteristic that differs between overcommitters and non-overcommitters.[21]

4.4 Overcommitment in the lab

Our results are also confirmed in an earlier lab experiment with 289 subjects conducted at Masaryk University Experimental Economics Laboratory (MUEEL) in Brno, Czech

[20] It is also worth noting that overcommitment negatively correlates with YouTube use ($\rho = -0.13, p = 0.0092$), though not with age and gender.

[21] We also examine whether and how commitment demand more generally, as captured by WTP values, vary with expected and actual behavior when facing temptation. Among the 383 tempted subjects, WTP is positively correlated with θ ($\rho = 0.12, p = 0.0218$). Consistent with expectation, WTP is also positively correlated with the likelihood of succumbing by clicking on a video ($\rho = 0.14, p = 0.0081$). Conditional on clicking a video, however, higher WTP is not associated with a longer time spent in YouTube tabs ($\rho = -0.08, p = 0.5477$), nor with longer pop-up duration in Task 2 net of pop-up duration in Task 1 ($\rho = 0.08, p = 0.1262$). WTP also does not significantly correlate with ω (Tangney et al., 2004) ($\rho = -0.08, p = 0.1279$), YouTube use, age, or gender.

Republic during the period 27-30 May 2019.[22] There, we use a work task similar to that used in Toussaert (2018): for a period of up to 30 minutes, subjects are asked to pay attention to their computer screen where a four-digit number increments every three seconds. At five random (subject-specific) times, they are prompted to enter the last number they saw, after which the number is reinitialized; the last prompt always occurs after the 25-minute mark, ending the task. Each correct answer is paid 120 CZK, about $5.20 at the time of the experiment. The potential earnings from this task are set to be relatively high to induce subjects to be interested in completing the task.

The temptation takes the form of an additional button on the screen which allows internet access. 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. Demand for a commitment device that removes the internet access button is elicited in a similar manner to the current paper.

In the experiment, 27.0% of subjects have a positive demand for commitment, and, given the random allocation, all but 12 of our 289 subjects still have to face temptation with four succumbing and clicking on the internet button. We find that 22.0% of exposed subjects overestimate their demand for commitment when compared to their actual material loss from facing the temptation. By the same measure, and perhaps surprisingly, only around 5.8% act according to standard accounts of naïve decision makers and underestimate their demand for commitment. Compared to *expected* material loss, 17.7% of subjects are still classified as overestimators. We obtain similar results when checking whether subjects overestimate both material and psychological costs of temptation: 14.8% of subjects seem to state a WTP higher than seems justified from the sum of these costs. By the same measure, only 7.2% of subjects underestimate the material and psychological costs of facing temptation and understate their WTP.

[22] More details are available in an earlier version of this paper, see https://papers.ssrn.com/abstract=3980519. The experiment was pre-registered at https://osf.io/rn8uc/ and complete results of the pre-analysis plan is available at https://osf.io/4dgm7/.

5 Concluding Remarks

When facing self-control problems, people may misestimate future outcomes or their own preferences, leading to sub-optimal commitment demand. Previous research has focused on overoptimistic beliefs resulting in underdemand for commitment. With a few exceptions, excess commitment demand has so far been understudied in the literature. We present the first clear evidence of this possibility online and in the lab and make the following contributions.

First, we are the first to rationalize the demand for commitment by quantifying the different components of temptation costs. We show that a significant share of subjects overdemand a commitment device with real and non-refundable costs, thus incurring welfare losses. This is true when we compare WTP with material loss from temptation, but also when we take into account psychological cost—which has so far eluded existing studies using field settings. Although it is conceivable that these estimates are specific to our setting, we use a design where, shortly before the productivity task with temptation, subjects do the exact same task without temptation. Since they also have high familiarity with the temptation we use, participants should have a good idea about the difficulty of the task—yet even so, some subjects overestimate their WTP.

Second, we are able to calculate the welfare loss due to excessive as well as insufficient commitment demand. Results indicate that, although there is substantial overcommitment in our sample, total welfare loss from undercommitment is much larger than from excess commitment. While this suggests that the existing literature is right to focus on undercommitment, policy should consider the risk that costly commitment may yield lower utility than facing temptation and thus be harmful.

Third, we explore the factors associated with overcommitment. We do not find strong evidence that overcommitters expect to succumb, or indeed do succumb, to a different extent than non-overcommitters. Instead, overdemand for the commitment device is associated with subjects' pessimism in material loss, predicting lower-than-realized performance due to costly exercise of self-control in resisting the temptation. This contrasts with recent findings in Carrera et al. (forth.), where excessive commitment demand appears driven by mean-zero random noise in a setting with high uncertainty about the future.

The above conclusions do come with some caveats; mainly, that other motives besides

material losses and psychological costs of temptation may drive behavior. For example, if an agent derives signals about what kind of person he is from his actions, then succumbing to temptation may lead to psychological self-image costs beyond those considered in this paper. Similarly, obtaining commitment may have additional value as a signal of being capable of sophisticated reasoning. It is conceivable that such motivations lead to higher WTP and even that subjects construct motivated beliefs about subsequent material losses to justify their decision. However, most subjects do resist temptation, and we would expect that achieving such an outcome would provide *higher* self-image, self-confidence, or signalling value than obtained when the agent achieves the same goal using the commitment device. Moreover, facing temptation also allows learning to occur (see, e.g., Ali, 2011). If so, a high WTP for commitment should appear even less rational than it does in our analysis. Further research should explore the implications of these issues for welfare and public policy.

We see our findings as a first step towards showing the existence of excess demand for commitment. The empirical applications and their policy implications will need to be investigated further in future research.

References

Acland, D. and Chow, V. (2018). Self-control and demand for commitment in online game playing: Evidence from a field experiment. *Journal of the Economic Science Association*, 4(1):46–62.

Ali, S. N. (2011). Learning self-control. *The Quarterly Journal of Economics*, 126(2):857–893.

Allcott, H., Braghieri, L., Eichmeyer, S., and Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, 110(3):629–76.

Allcott, H., Gentzkow, M., and Song, L. (2021). Digital addiction. NBER Working Paper 28936.

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.

Bai, L., Handel, B., Miguel, E., and Rao, G. (2021). Self-control and demand for preventive health: Evidence from hypertension in India. *The Review of Economics and Statistics*, pages 1–22.

Beshears, J., Choi, J. J., Harris, C., Laibson, D., Madrian, B. C., and Sakong, J. (2015). Self control and commitment: Can decreasing the liquidity of a savings account increase deposits? NBER Working Paper 21474.

Bjerre-Nielsen, A., Andersen, A., Minor, K., and Lassen, D. D. (2020). The negative effect of smartphone use on academic performance may be overestimated: Evidence from a 2-year panel study. *Psychological Science*, 31(11):1351–1362.

Braghieri, L., Levy, R., and Makarin, A. (2021). Social media and mental health. Available at SSRN: https://ssrn.com/abstract=3919760.

Bursztyn, L., Egorov, G., Enikolopov, R., and Petrova, M. (2019). Social media and xenophobia: Evidence from Russia. NBER Working paper 26567.

Carrera, M., Royer, H., Stehr, M., Sydnor, J., and Taubinsky, D. (forth.). Who chooses commitment? Evidence and welfare implications. *The Review of Economic Studies*.

Collis, A. and Eggers, F. (2019). Effects of restricting social media usage. Available at SSRN: https://ssrn.com/abstract=3518744.

De Quidt, J., Haushofer, J., and Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, 108(11):3266–3302.

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

Fujiwara, T., Müller, K., and Schwarz, C. (2021). The effect of social media on elections: Evidence from the United States. NBER Working paper 28849.

Giné, X., Karlan, D., and Zinman, J. (2010). Put your money where your butt is: A commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4):213–35.

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. *American Economic Review*, 92(5):1644–1655.

Hoong, R. (2021). Self control and smartphone use: An experimental study of soft commitment devices. *European Economic Review*, 140:103924.

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

Hunt, M. G., Marx, R., Lipson, C., and Young, J. (2018). No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology*, 37(10):751–768.

John, A. (2020). When commitment fails: Evidence from a field experiment. *Management Science*, 66(2):503–1004.

Kaur, S., Kremer, M., and Mullainathan, S. (2010). Self-control and the development of work arrangements. *American Economic Review*, 100(2):624–28.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., Shablack, H., Jonides, J., and Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. *PloS ONE*, 8(8):e69841.

Laibson, D. (2015). Why don't present-biased agents make commitments? *American Economic Review*, 105(5):267–72.

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3):831–70.

Marotta, V. and Acquisti, A. (2018). Interrupting interruptions: A digital experiment on social media and performance. *Available at SSRN: https://ssrn.com/abstract=3283951*.

Milkman, K. L., Minson, J. A., and Volpp, K. G. (2013). Holding the hunger games hostage at the gym: An evaluation of temptation bundling. *Management Science*, 60(2):283–299.

Mosquera, R., Odunowo, M., McNamara, T., Guo, X., and Petrie, R. (2020). The economic effects of Facebook. *Experimental Economics*, 23(2):575–602.

Müller, K. and Schwarz, C. (2020). From hashtag to hate crime: Twitter and anti-minority sentiment. Available at SSRN: https://ssrn.com/abstract=3149103.

Müller, K. and Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 19(4):2131–2167.

Pew Research Center (2021). Social media use in 2021. Available at: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5):1281–1292.

Royer, H., Stehr, M., and Sydnor, J. (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(3):51–84.

Sadoff, S. and Samek, A. (2019). Can interventions affect commitment demand? A field experiment on food choice. *Journal of Economic Behavior & Organization*, 158:90–109.

Sadoff, S., Samek, A., and Sprenger, C. (2020). Dynamic inconsistency in food choice: Experimental evidence from two food deserts. *The Review of Economic Studies*, 87(4):1954–1988.

Sagioglou, C. and Greitemeyer, T. (2014). Facebook's emotional consequences: Why Facebook causes a decrease in mood and why people still use it. *Computers in Human Behavior*, 35:359–363.

Schilbach, F. (2019). Alcohol and self-control: A field experiment in india. *American Economic Review*, 109(4):1290–1322.

Schwartz, J., Mochon, D., Wyper, L., Maroba, J., Patel, D., and Ariely, D. (2014). Healthier by precommitment. *Psychological Science*, 25(2):538–546.

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.

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

Tromholt, M. (2016). The Facebook experiment: Quitting Facebook leads to higher levels of well-being. *Cyberpsychology, Behavior, and Social Networking*, 19(11):661–666.

Vanman, E. J., Baker, R., and Tobin, S. J. (2018). The burden of online friends: The effects of giving up Facebook on stress and well-being. *The Journal of Social Psychology*, 158(4):496–508.

Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12:415–438.

Zimmermann, L. and Sobolev, M. (2020). Digital nudges for screen time reduction: A randomized control trial with performance and wellbeing outcomes. Working Paper.

Appendices

A Assuming aversion to uncertainty in the lab

In our experiment, temptation may have been perceived as a 'risk' or uncertainty which subjects would like to avoid. Such uncertainty aversion need not be the same thing as 'risk aversion' in the traditional sense given the small stakes in the lab, over which subjects should be risk-neutral (Rabin, 2000). Nevertheless, what looks like an overestimated WTP for commitment compared to the optimal WTP of a risk-neutral agent may become rationalizable or even underestimated when compared to the optimal choice under uncertainty aversion. In the absence of a better way to parametrize such aversion to uncertain lab payments, in this section we check whether WTP remains overestimated if subjects are assumed to be (strongly) risk-averse with CRRA utility.

Recall that in Section 3 of the main text, subjects maximize a utility function which 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. However, expectations about succumbing are not elicited in our experiment. While we do elicit the subjective probability that a video will be clicked ($\hat{p}^c$; see footnote 8 in the main text), our design also allows subjects to procrastinate without clicking: specifically, by watching videos directly in the pop-up windows. Hence, aversion to uncertainty cannot be fully accounted for in our experiment. As an alternative, below we reproduce an analysis related to the earlier experiment summarized in Section 4.4.

In that experiment, commitment removes a button that allows the subject to surf the internet. Accounting for the participation fee and experimental earning in the lab, we therefore change the temptation-related utility terms to obtain

$$\begin{aligned} U(WTP) = \frac{1}{2}\Bigg[&\frac{1}{2}\cdot u(100+120y_1-PC) \\ &+\frac{1}{2}\left(\frac{100-WTP}{100}\cdot u(100+120y_1-PC)+\frac{1}{100}\int_0^{WTP}u(100+120y_1-R)dR\right)\Bigg] \\ &+\frac{1}{2}\Bigg[\left(1-\frac{WTP}{200}\right)\left(p^c u(100+120y^c-PC)+(1-p^c)u(100+120y^{nc}-PC)\right) \\ &+\frac{1}{2}\left(\frac{1}{100}\int_0^{WTP}u(100+120y^{nt}-R)dR\right)\Bigg] \end{aligned}$$

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, and nor is there utility from internet surfing. The solution under risk neutrality, denoted WTP_{RN}, is:

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

where $y^t = p^c y^c + (1 - p^c)\, y^{nc}$.

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

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

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

$$\begin{aligned} \frac{dU}{dWTP} = \frac{1}{200\,(1-\eta)} \Big\{ & (100 + 120y_1 - WTP)^{1-\eta} + \left(100 + 120y^{nt} - WTP\right)^{1-\eta} \\ & - (100 + 120y_1 - PC)^{1-\eta} - p^c\,(100 + 120y^c - PC)^{1-\eta} \\ & - (1 - p^c)\,(100 + 120y^{nc} - PC)^{1-\eta} \Big\} = 0 \end{aligned} \tag{A.1}$$

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 (A.1) with the relevant y_1, y^{nt}, y^c, y^{nc} and p^c values inserted for each individual subject. η, the coefficient of relative risk aversion, has been estimated in different studies to be around 1.[23] To be conservative, we present results for several values of η up to $\eta = 3$, 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 1 in the risk-neutral case). In equation (A.1), y^{nt} is thus interpreted as the *actual* number of correct answers when the subject is not exposed to temptation; as per above, we use y_1 as the counterfactual. (There

[23] 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 0.06.

are no significant learning effects in the group that obtains commitment.) p^c is obtained using the percentage of subjects who succumb out of all subjects exposed to temptation; it equals 1.4%. For subjects who do not succumb, $y^{nc} = y_2$, while y^c, the counterfactual had they succumbed, is obtained using the average productivity of subjects who do succumb, which is $y^c = 2$. In the same way, for subjects who succumb, $y^c = y_2$ while the counterfactual $y^{nc} = 4.93$, the average productivity for those who do not succumb. Comparing the resulting WTP_{RA} with the WTP stated by each subject, the proportion of M overestimators under different values of η are given in the first row of Table A.1. Around 17% of subjects are still considered to be M overestimators under conventional levels of risk aversion, 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 (79% under $\eta = 1$ or 1.5), though the magnitude of underestimation compared to material loss is much smaller (around 7 tokens under $\eta = 1$ or 1.5) compared to the risk-neutral case (82 tokens). Nevertheless, our first result is robust to assuming CRRA with $\eta \leq 3$.

Table A.1: Proportion of overestimators under risk aversion.

Relative to	$\eta = 0.5$	$\eta = 1$	$\eta = 1.5$	$\eta = 2$	$\eta = 3$
(1) Actual material loss	17.33%	16.97%	16.97%	15.52%	15.52%
(2) Expected material loss	11.55%	11.55%	11.55%	11.19%	9.75%
(3) Actual material loss and psychological cost	14.08%	14.08%	14.08%	14.08%	12.64%

We next turn to subjects' WTP considering expected material loss (corresponding to Hypothesis 2). We proceed as above, except that we now use each subject's predictions of their own performance $\hat{y}^{nt}$, $\hat{y}^c$, $\hat{y}^{nc}$ and $\hat{p}^c$. As shown in the second row of Table A.1, 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 less than 12% for all values of η and no longer significant. Putting aside psychological cost, 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.

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 3 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 Online Appendix B. Given this fact, we may proceed as follows.

For all subjects with $WTP > 0$ and $v \geq 0$, and for all expected PC values consistent with $0 < WTP < 100$, we plug in appropriate outcome variables in (A.1) 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 $0 < WTP_e < 100$.[24]

As shown in row (3) of Table A.1, we find that such subjects make up between about 13-14% of all subjects who face temptation ($p < 0.1$ for all values of η, $p < 0.05$ for conventional values of $\eta \leq 2$). Hence, our result of overestimation relative to both actual material and psychological costs is also robust to assuming CRRA with $\eta \leq 3$.

Overall, even assuming a very strong degree of risk aversion, our conclusion that a significant share of subjects overestimate 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.

[24] In principle, since both stated WTP and all parameters related to expected material losses are known, we might use them in (A.1) 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 it is sometimes negative, which we interpret as there being some random error in subjects' WTP responses.

B Proof that WTP under risk aversion is increasing in psychological cost

Assuming CRRA with $\eta > 1$, the first-order condition is restated below:

$$\begin{aligned}\frac{dU}{dWTP} = \frac{1}{200\,(1-\eta)} \Big\{ &(100 + 120y_1 - WTP)^{1-\eta} + \left(100 + 120y^{nt} - WTP\right)^{1-\eta} \\ &- (100 + 120y_1 - PC)^{1-\eta} - p^c\,(100 + 120y^c - PC)^{1-\eta} \\ &- (1-p^c)\,(100 + 120y^{nc} - PC)^{1-\eta} \Big\} = 0\end{aligned}$$

The second derivative is

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

The partial derivative of the first-order condition with respect to PC is

$$\begin{aligned}\frac{\partial^2 U}{\partial WTP \partial PC} &= \frac{1}{200}\left[\frac{1}{(100 + 120y_1 - PC)^{\eta}} + \frac{p^c}{(100 + 120y^c - PC)^{\eta}} + \frac{1-p^c}{(100 + 120y^{nc} - PC)^{\eta}}\right] \\ &> 0\end{aligned}$$

Using the implicit function theorem,

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

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

C Instructions

Begin on next page.

Welcome!

This is a study about decision-making conducted by researchers at University College Dublin and the University of Gothenburg. The study has been given ethical approval by the university's ethics committee. You must be at least 18 years of age to participate in this study.

The study will take around 45 minutes to complete. Participation involves completing a simple task and answering an online questionnaire. All of your responses will remain anonymous throughout and only aggregate results will be published.

The study does not involve any risk of harm. In addition to your participation fee, you may receive additional bonus payment depending on the decisions you make. If you wish to withdraw at any point during the study, you can simply close your internet browser. Please note that you need to complete the entire session to receive payment for your participation.

If you have any questions regarding this study, please email margaret.samahita@ucd.ie.

I have read and understood the above and want to participate in this study. (radio button)
-Yes
-No

What is your Prolific ID? (text entry)
_____

NEXT BUTTON

----------PAGE BREAK----------

The following are **YouTube** video categories. Please **select the 3 to 5 categories** that interest you the most.

[Tickboxes:]

- Film & animation
- Autos & vehicles
- Music
- Pets & animals
- Sports
- Gaming
- People & blogs
- Comedy
- Entertainment
- News & politics
- Howto & style

- Educational
- Science & technology

NEXT BUTTON

----------PAGE BREAK----------

As a reminder, after the session you will receive your participation fee. Additionally, you may receive **additional bonus payment depending on your decisions during the study.**
The study consists of 2 stages. 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 study. Because it is uncertain which stage will be chosen for payment, you should carefully consider all decisions.

You are about to begin with Stage 1.

NEXT BUTTON

----------PAGE BREAK----------

STAGE 1

During Stage 1, you will be asked to transcribe a line of blurry letters from a Greek text, as shown in the example below. Each task will be shown on a new screen and consists of a row of blurry Greek text that will appear at the top of the screen. For each letter, you will need to find and select the corresponding letter from the alternatives presented below the text. For your task submission to be considered correct, your submission must be 90% accurate.

You are asked to complete as many transcription tasks as possible in 15 minutes. Each correct submission will earn you $0.50 (50 cents), should Stage 1 be randomly chosen for payment. After the 15 minutes, you will automatically progress to the next screen.

Additionally, immediately before each transcription task, there will be a pop-up window which you will have to close to proceed to the next task. The task clock will continue running while each pop-up (including the very first one) is shown.

You will now have one practice task before moving on to the real tasks. There is no time limit to the practice task. Please note that the practice task will end once you click Submit.

SUBMIT

NEXT BUTTON

----------PAGE BREAK----------

Practice task

**********ROW OF BLURRY GREEK TEXT**********

CLEAR GREEK LETTERS TO SELECT

SUBMIT BUTTON

----------PAGE BREAK----------

You transcribed X characters accurately out of 35 and thus your submission is considered CORRECT/INCORRECT.

You will now begin with a real task. Your **15 minutes** will start as soon as you click the Next button.

NEXT BUTTON

----------PAGE BREAK----------

Actual task

*****15 minutes of the transcription task WITHOUT videos*****

NEXT BUTTON

----------PAGE BREAK----------

Your total number of correct submissions is

X

NEXT BUTTON

----------PAGE BREAK----------

Suppose that now you are to repeat the transcription task again.

This second time, how many correct submissions would you expect to get in 15 minutes?

_____

NEXT BUTTON

----------PAGE BREAK----------

STAGE 2

We would like you to pay attention to the following information, therefore you will not be able to proceed and the NEXT button will not appear for 2 minutes.

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

<u>The modification</u>

The task in Stage 2 is similar to the one you completed in Stage 1: you will be asked to transcribe as many lines of blurry Greek letters as possible within 15 minutes. You will again earn $0.50 (50 cents) per correct (90% accuracy) submission, should Stage 2 be randomly selected for payment. After the 15 minutes, you will automatically progress to the next screen.

However, below the transcription task, there will now be a series of 10 YouTube videos, as shown on the next screen. The videos on your screen will be personalised for you by importing YouTube videos from your chosen categories that are currently trending in the US. A new set of 10 videos from your chosen categories will be shown for each new transcription task. You can click on any YouTube video at any point during the task.

Immediately before each transcription task, there will again be a pop-up window which now **automatically plays one of the videos that will appear below the task.** You will have to close the pop-up window to proceed to the next task. The task clock will continue running while each pop-up (including the very first one) is shown.

We do NOT record which YouTube videos you see or view. We only record <u>whether clicks are made</u> on any video.

If you click on a video, a new tab will open where you will be able to view it. You can watch the video for as long as you like and come back to the transcription task tab at any time, and you

can subsequently click on another video if you like. However, the task clock will keep running while you are watching any video. If 15 minutes elapsed while you are viewing a video, the video tab will automatically close and you will be taken back to the study tab.

Watching videos means that you spend less time on the transcription task. Hence, you may potentially have fewer correct submissions in the 15-minute period, thus lowering your bonus payment. If you do not click on any video, you will simply continue with the transcription task.

In the next screen, you will practice the transcription task with the new modification: the addition of the YouTube videos. To help you to familiarize yourself with the new setup, you will now face **two** practice tasks in a row.

NEXT BUTTON

----------PAGE BREAK----------

Practice task

**********ROW OF BLURRY GREEK TEXT**********

CLEAR GREEK LETTERS TO SELECT

SUBMIT BUTTON

----------PAGE BREAK----------

Practice task 1:
You transcribed X characters accurately out of 35 and thus your submission is considered CORRECT/INCORRECT.

Practice task 2:
You transcribed X characters accurately out of 35 and thus your submission is considered CORRECT/INCORRECT.

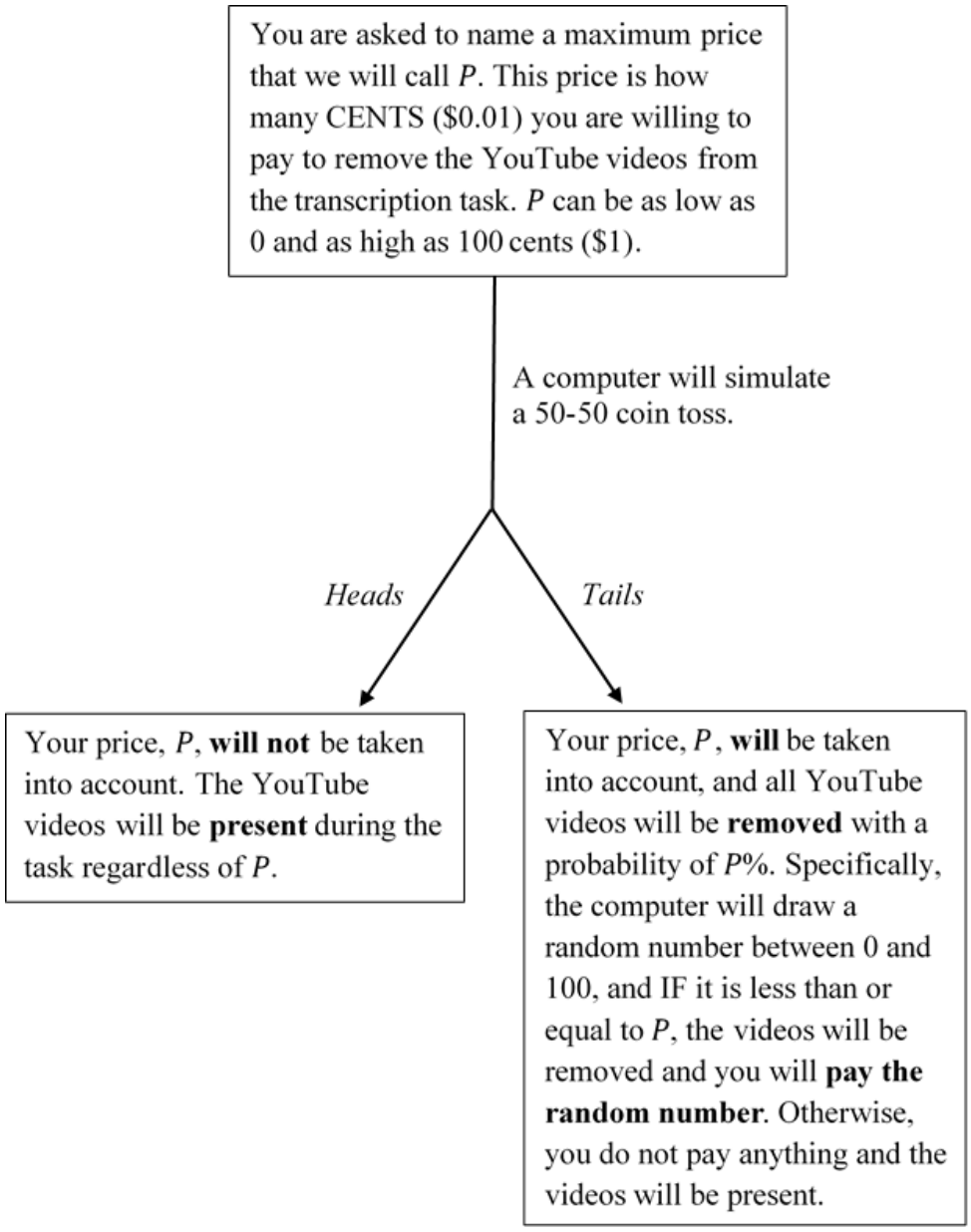
NEXT BUTTON

----------PAGE BREAK----------

We would like you to pay attention to the following information, therefore you will not be able to proceed and the NEXT button will not appear for 2 minutes.

<u>Removing the YouTube videos</u>
If you would like to remove the YouTube videos, for example because you think you might be able to better perform the task without them, you have the possibility to do so. The screen displayed throughout the 15 minutes of Stage 2 will be exactly the same as in Stage 1, without the YouTube videos. You will still encounter a pop-up before each task (and the clock will continue running while each pop-up is open), but as in Stage 1 these pop-ups will not contain a YouTube video.

We will now describe how to remove the YouTube videos. It is done by paying money taken out of your bonus payment. **Note that there is no right or wrong choice: you are entirely free to state a maximum price that seems best to you.** Starting from the top, determining whether to remove the videos or not involves the following steps:



Example 1. Participant A stated a price of 60 cents and the simulated coin flip came up Heads. Because the coin flip had that result, the videos will be **present**, and Participant A will not have to pay anything.

Example 2. Participant B stated a price of 40 cents and the simulated coin flip came up Tails. Because the coin flip had that result, the computer will now draw a random number between 0 and 100, removing the videos **with 40% probability**. The random number drawn was 32 and this is less than 40, so the videos will indeed be removed, and 32 cents will be deducted from Participant B's bonus payment.

As you can see, the probability of removing the YouTube videos increases, the higher your stated price. Note in particular that:

- Your chance of removing the YouTube videos is maximized (but, due to the simulated coin toss, is not guaranteed) if you state a price of 100 cents. **Note that this maximum price corresponds to the earnings from 2 correctly submitted tasks.**
- If the videos are removed, you will pay the random number drawn, NOT your price. You will never pay a higher price than the one you state, P. Any amount you pay will be taken out of your bonus payment regardless of whether Stage 1 or 2 is chosen for payment.
- Your total bonus is your earnings from the randomly chosen stage minus any payment to remove the YouTube videos. Hence, your total bonus may be negative though not lower than -$1. A negative bonus will be taken out of your participation fee.
- If you are not willing to pay anything to remove the YouTube videos, you should enter a price of 0. **This will ensure that you will do the transcription task WITH YouTube videos.**

NEXT BUTTON

----------PAGE BREAK----------

No matter if you pay to remove the YouTube videos, keep the videos but never watch them, or keep the videos and watch any one of them, you will spend the same amount of time (15 minutes) in Stage 2.

You will now have a practice round to ensure you understand the process of paying to remove the YouTube videos. When you have finished the practice round, you will be asked to state your actual price for removing the videos, as well as answer a few questions about the new transcription task.

Finally, the computer will toss the coin. If TAILS comes up, it will then draw a random number to determine the outcome. You will be informed about whether the YouTube videos will be present or not, and subsequently you will start the transcription task in Stage 2.

NEXT BUTTON

----------PAGE BREAK----------

PRACTICE round

Maximum TEST price for removing YouTube videos

State a TEST price, in cents, you are willing to pay to remove the YouTube videos. Please enter a number between 0 (cents) and 100 (cents) (inclusive)

_____

NEXT BUTTON

----------PAGE BREAK----------

The test price you stated is
X

The coin toss resulted in

HEADS/TAILS

Therefore, your test price will/will NOT be taken into account

(If tails:) Given that the coin flip came up TAILS, your chance of removing the YouTube videos is now
X%

(If tails:) The random number drawn is
Y

The outcome for the next stage would have been
Transcription task WITH/WITHOUT Youtube videos

END OF PRACTICE ROUND

NEXT BUTTON

----------PAGE BREAK----------

Before we move on to your actual price, please take a moment to consider what you would prefer for the upcoming transcription task. (radio button)

- I prefer to do the upcoming transcription task with the YouTube videos **present**.
- I prefer to do the upcoming transcription task with the YouTube videos **not present**.

Think about what your choice implies for your price to remove the button on the next screen.

NEXT BUTTON

----------PAGE BREAK----------

ACTUAL round

Maximum ACTUAL price for removing YouTube videos

What is the highest price, in cents, you are willing to pay to remove the YouTube videos? Please enter a number between 0 (cents) and 100 (cents) (inclusive)

_____

NEXT BUTTON

----------PAGE BREAK----------

You will soon learn if the YouTube videos are removed or not in Stage 2.

Suppose the videos are NOT removed, and you continue to have the option of clicking on a video and viewing it during the transcription task.

In this situation, how many correct submissions do you expect to get in 15 minutes?

_____

NEXT BUTTON

----------PAGE BREAK----------

You will soon learn if the YouTube videos are removed or not in Stage 2.

Suppose the videos are NOT removed, and you continue to have the option of clicking on a video and viewing it during the transcription task.

How likely (in %) would you say you are to click on a video? _____

Suppose that you DO click on a video.
In this situation, how many correct submissions do you expect to get in 15 minutes?

_____

Suppose that you DO NOT click on a video.
In this situation, how many correct submissions do you expect to get in 15 minutes?

_____

NEXT BUTTON

----------PAGE BREAK----------

How tempted do you think you would be to click on any of the YouTube videos? (radio button)

- Not at all tempted
- Not that tempted
- Quite tempted
- Very tempted

NEXT BUTTON

----------PAGE BREAK----------

You have previously stated a price of X cents for removing the YouTube videos.

You have since answered questions and thought more about the potential outcomes in Stage 2. Now you have a chance to revise your stated price, if you wish.

Would you like to revise your stated price?
If yes, please enter a new price between 0 (cents) and 100 (cents) (inclusive). If not, simply enter the same price of X. **This decision is final and cannot be changed!**

_____

NEXT BUTTON

----------PAGE BREAK----------

Now we will tell you the result of your actual transaction.

The price you stated is
X

The coin toss resulted in

HEADS/TAILS

Therefore, your price will/will NOT be taken into account

(If tails:) Given that the coin flip came up TAILS, your chance of removing the YouTube videos is now
X%

(If tails:) The random number drawn is
Y

The outcome for the next stage is
Transcription task WITH/WITHOUT Youtube videos

CLICK NEXT TO START STAGE 2

NEXT BUTTON

----------PAGE BREAK----------

Actual task

*****15 minutes of the transcription task WITH videos*****

NEXT BUTTON

----------PAGE BREAK----------

End questionnaire

What is your age (in years)?

_____

What is your gender? (radio button)

- Male
- Female

Do you think the difficulty of ignoring the YouTube videos and concentrating on the transcription task was higher or lower than expected (when you chose your price for removing the videos)?

- Ignoring the YouTube videos and concentrating on the transcription task was more difficult than expected
- Ignoring the YouTube videos and concentrating on the transcription task was neither easier nor more difficult than expected
- Ignoring the YouTube videos and concentrating on the transcription task was easier than expected
- N/A - I was not shown the YouTube videos

How much time per day do you spend on YouTube? (radio button)

- Less than 30 minutes
- From 30 minutes to 1 hour
- From 1 to 2 hours
- More than 2 hours

Which of the following best applies to you?

- I was not interested in the YouTube videos at all because I did not care about them
- I was not interested in the YouTube videos at all because I was concentrating on the transcription task
- At first I was not interested in the YouTube videos, but as time passed, I got bored and started thinking about them
- At first I thought a lot about the YouTube videos, but as time passed, I managed to start focusing more on the transcription task
- I kept thinking about the YouTube videos and this prevented me from staying focused on the transcription task
- I chose to view a YouTube video almost immediately
- N/A - I was not shown the YouTube videos

NEXT BUTTON

----------PAGE BREAK----------

Using a 5-point scale, please indicate how much each of the following statements reflects how you typically are. (5-point Likert scale, from “Not at all” to “very much”)

I am good at resisting temptation.
I have a hard time breaking bad habits.
I am lazy.
I say inappropriate things.
I refuse things that are bad for me.
I wish I had more self-discipline.
I do certain things that are bad for me, if they are fun.
Pleasure and fun sometimes keep me from getting work done.
I have trouble concentrating.
I am able to work effectively toward long-term goals.
People would say that I have iron self-discipline.
Sometimes I can't stop myself from doing something, even if I know it is wrong.
I often act without thinking through all the alternatives.

NEXT BUTTON

----------PAGE BREAK----------

We would be very interested in hearing about your experience in this study. Please write any comment in the textbox below.

[5-line textbox]

NEXT BUTTON

----------PAGE BREAK----------

Stage chosen for payment:
Stage 1 / 2

Number of correct submissions in that Stage:
X

Earnings from correct submissions:
X

Cost of removing YouTube videos
X

Total bonus payment:
X

Please click RETURN TO PROLIFIC to record the completion of this study.

RETURN TO PROLIFIC button (https://app.prolific.co/submissions/complete?cc=81E5FDC9)

UCD CENTRE FOR ECONOMIC RESEARCH – RECENT WORKING PAPERS

WP21/06 Margaret Samahita and Leonhard K Lades: 'The Unintended Side Effects of Regulating Charities: Donors Penalise Administrative Burden Almost as Much as Overheads' February 2021
WP21/07 Ellen Ryan and Karl Whelan: 'A Model of QE, Reserve Demand and the Money Multiplier' February 2021
WP21/08 Cormac Ó Gráda and Kevin Hjortshøj O!Rourke: 'The Irish Economy During the Century After Partition' April 2021
WP21/09 Ronald B Davies, Dieter F Kogler and Ryan Hynes: 'Patent Boxes and the Success Rate of Applications' April 2021
WP21/10 Benjamin Elsner, Ingo E Isphording and Ulf Zölitz: 'Achievement Rank Affects Performance and Major Choices in College' April 2021
WP21/11 Vincent Hogan and Patrick Massey: 'Soccer Clubs and Diminishing Returns: The Case of Paris Saint-Germain' April 2021
WP21/12 Demid Getik, Marco Islam and Margaret Samahita: 'The Inelastic Demand for Affirmative Action' May 2021
WP21/13 Emmanuel P de Albuquerque: 'The Creation and Diffusion of Knowledge - an Agent Based Modelling Approach' May 2021
WP21/14 Tyler Anbinder, Dylan Connor, Cormac Ó Gráda and Simone Wegge: 'The Problem of False Positives in Automated Census Linking: Evidence from Nineteenth-Century New York's Irish Immigrants' June 2021
WP21/15 Morgan Kelly: 'Devotion or Deprivation: Did Catholicism Retard French Development?' June 2021
WP21/16 Bénédicte Apouey and David Madden: 'Health Poverty' July 2021
WP21/17 David Madden: 'The Dynamics of Multidimensional Poverty in a Cohort of Irish Children' August 2021
WP21/18 Vessela Daskalova and Nicolaas J Vriend: 'Learning Frames' August 2021
WP21/19 Sanghamitra Chattopadhyay Mukherjee: 'A Framework to Measure Regional Disparities in Battery Electric Vehicle Diffusion in Ireland' August 2021
WP21/20 Karl Whelan: 'Central Banks and Inflation: Where Do We Stand and How Did We Get Here?' August 2021
WP21/21 Tyler Anbinder, Cormac Ó Gráda and Simone Wegge: '"The Best Country in the World": The Surprising Social Mobility of New York's Irish Famine Immigrants' August 2021
WP21/22 Jane Dooley and David Madden: 'Ireland's Post Crisis Recovery, 2012-2019: Was It Pro-Poor?' September 2021
WP21/23 Matthew Shannon: 'The Impact of Victimisation on Subjective Well-Being' September 2021
WP21/24 Morgan Kelly: 'Persistence, Randomization, and Spatial Noise' October 2021 *(For revised version of this paper see WP21/25)*
WP21/25 Morgan Kelly: 'Persistence, Randomization, and Spatial Noise' November 2021
WP21/26 Eliane Badaoui and Frank Walsh: 'Productivity, Non-Compliance and the Minimum Wage' November 2021
WP21/27 Annette Broocks and Zuzanna Studnicka: 'Gravity and Trade in Video on Demand Services' December 2021
WP21/28 Linda Mastrandrea: 'Linking Retail Electricity Pricing and the Decarbonisation of the Energy Sector: a Microeconomic Approach' December 2021
WP22/01 Doina Caragea, Theodor Cojoianu, Mihai Dobri, Andreas Hoepner and Oana Peia 'Competition and Innovation in the Financial Sector: Evidence from the Rise of FinTech Start-ups' January 2022
WP22/02 Sandra E Black, Paul J Devereux, Fanny Landaud and Kjell G Salvanes: 'The (Un)Importance of Inheritance' January 2022