Appendices

A 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.

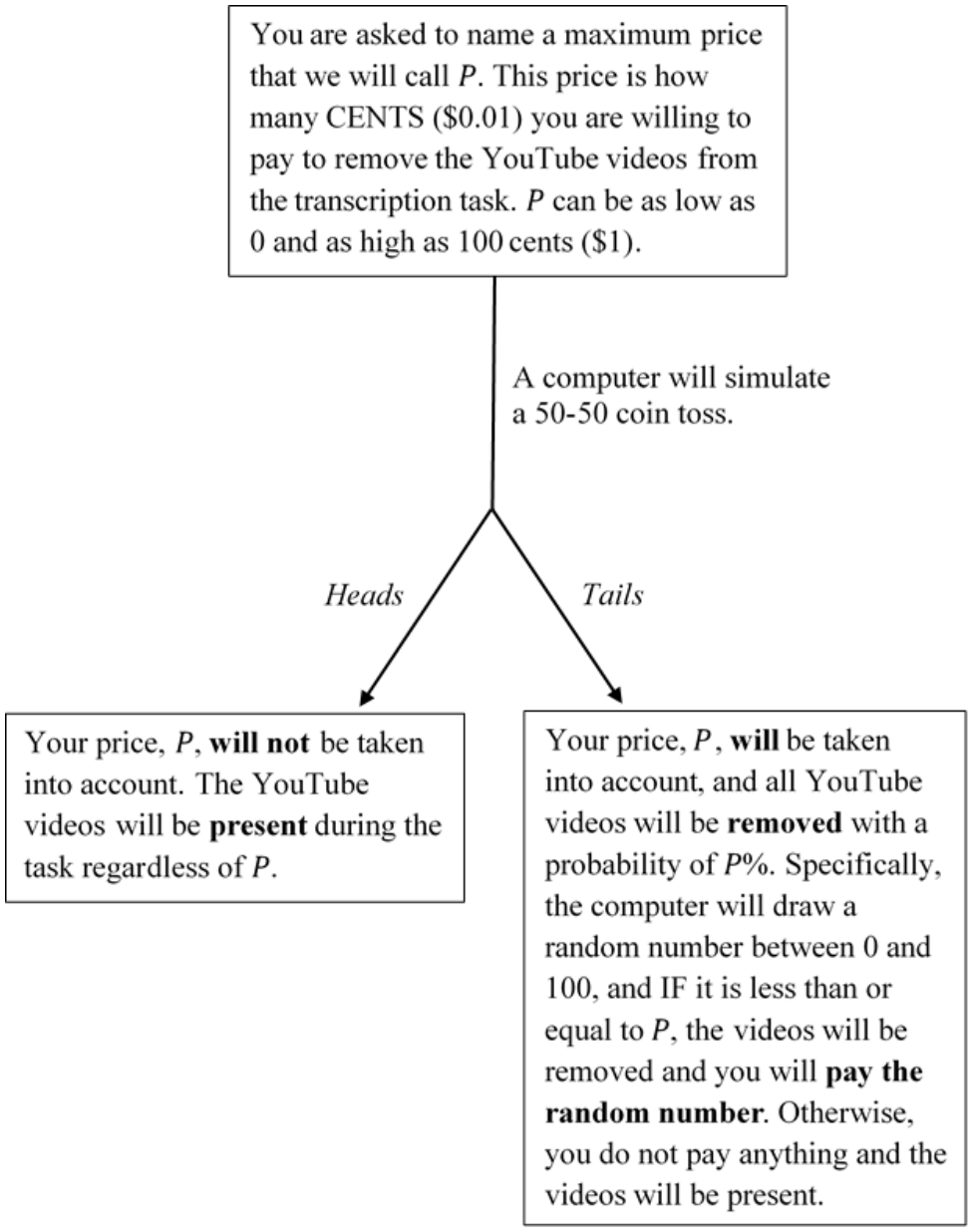
NEXT BUTTON

----------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)

B Hypothesis tests

In the pre-analysis plan, we pre-registered three null hypotheses based on the three definitions of overestimators in Definitions 1-3 respectively. These tests of proportions are based on the standard normal approximation of binomial parameters and assuming a threshold share of 10%.[29] The hypotheses are:

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

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

Hypothesis B.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* $WTP > 0$.

In Table B.1 we provide all hypothesis test results, including robustness checks.

Table B.1: Hypothesis tests and robustness checks.

Hypothesis	Main specification	Using WTP_1	Discrete R	Coin flip Heads	Frequency weights
H1	19.3%	19.8%	19.3%	19.9%	23.8%
H2	18.8%	19.8%	19.6%	18.6%	22.8%
H3	17.5%	18.3%	17.5%	19.0%	21.6%

Notes: Proportion of overestimators. All p-values < 0.0001 from tests of proportions based on the standard normal approximation of binomial parameters and assuming a threshold share of 10%. The robustness tests (last four columns) are: (i) using WTP_1 in place of WTP_2; (ii) using a discrete rather than continuous distribution for R to calculate optimal WTP; (iii) using only the half of the (exposed) sample where the coin flip results in Heads; and (iv) using frequency weights with respect to WTP to account for the 26 "missing" subjects who obtain commitment.

[29] This threshold can be interpreted as the maximum proportion attributable to subject confusion, demand effects, or random error ("noise"). The use of this threshold reflects 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.

C Additional tables and figures

Table C.1: Correlates of WTP residual

	(1) WTP residual, Definition 1	(2) WTP residual, Definition 2
θ	0.233	-13.056*
	(6.912)	(6.802)
ω	0.125	-0.785
	(0.474)	(0.499)
Male	-19.602**	-16.513*
	(9.345)	(9.670)
Age	0.001	0.597**
	(0.294)	(0.302)
YouTube use	-8.171*	-6.673
	(4.682)	(6.034)
Constant	27.814	36.363
	(29.113)	(30.353)
N	409	409
R-sq	0.026	0.034

Notes: OLS regressions of WTP residual as per Definitions 1 and 2. θ: how tempted subject expects to be (1-4). ω: score on brief self-control scale (13-65). Male: dummy variable which equals 1 if subject is male. Age: subject age in years. YouTube use: daily time spent on YouTube where 1: $<$30 minutes, 2: $>$30 minutes & $\leq$1 hour, 3: $>$1 hour & $\leq$2 hours, 4: $>$2 hours. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

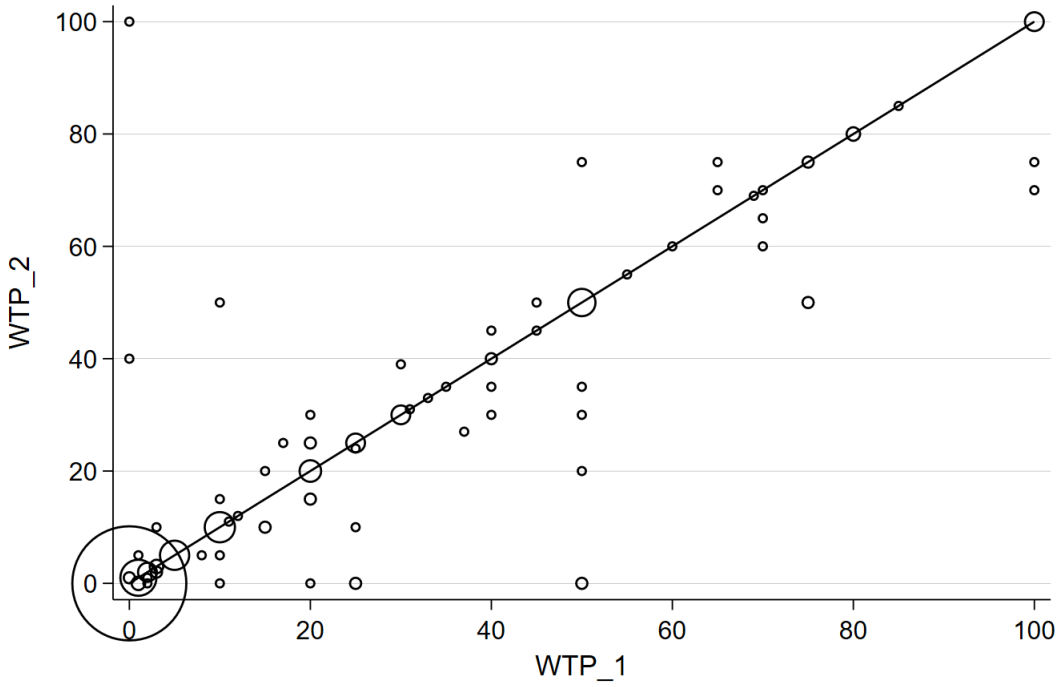


Figure C.1: Scatterplot of WTP_1 and WTP_2, circle size is proportional to frequency weight

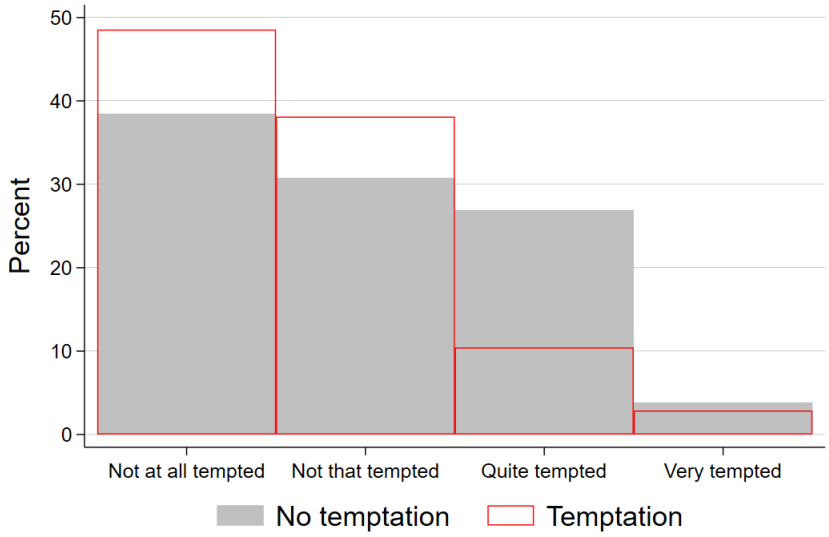


Figure C.2: How tempted subject thinks they would be to click on a video

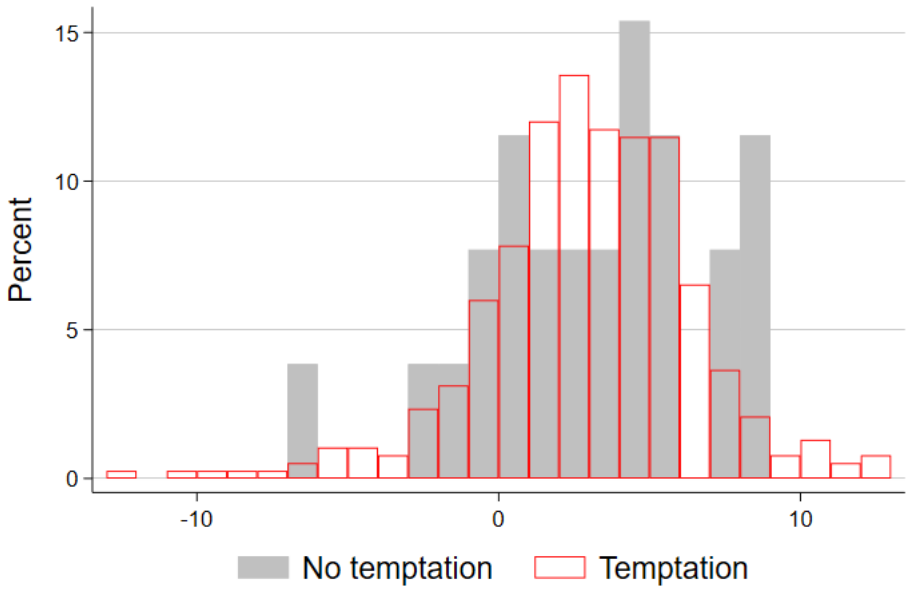


Figure C.3: $y_2 - y_1$: learning effect

D Additional robustness checks

D.1 Assuming WTP is paid conditional on Task 2 being chosen for payment

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

$$\begin{aligned} U(WTP) = \frac{1}{2}\Bigg[&\frac{1}{2}\cdot u(50y_1 - PC) \\ &+\frac{1}{2}\left(\frac{100-WTP}{100}\cdot u(50y_1 - PC) + \frac{1}{100}\int_0^{WTP} u(50y_1 - 0)dR\right)\Bigg] \\ &+\frac{1}{2}\Bigg[\frac{1}{2}\cdot u(50y^t - PC) \\ &+\frac{1}{2}\left(\frac{100-WTP}{100}\cdot u(50y^t - PC) + \frac{1}{100}\int_0^{WTP} u(50y^{nt} - R)dR\right)\Bigg] \end{aligned}$$

and the solution under risk neutrality is given by

$$WTP = 50(y^{nt} - y^t) + 2PC$$

Clearly, the analysis for Definition 3 is equivalent. For Definition 1, the proportion of overestimator using the new WTP threshold given above is 18.02% ($p < 0.0001$). For Definition 2, the proportion of overestimators is 17.23% ($p < 0.0001$).

D.2 Robustness of definitions 1-3 in the presence of effort costs

In this section, we explore the implication of adding linear or quadratic effort costs in our assumed utility function.

D.2.1 Linear effort

Assuming constant effort cost $k^l > 0$ per correct submission, we write the effort-cost function as $e(y) = k^l y$. With this effort function, expected utility function U can be

rewritten as

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

The rational WTP under risk neutrality is then

$$WTP = (25 - k^l)(y^{nt} - y^t) + PC$$

Note that WTP is somewhat lower than in Equation 2 of the main text. Nevertheless, Definitions 1 and 2 are clearly unchanged since these are meant to only take into account effects on subjects' *material* payoffs. Thus, effort cost will only affect our third definition, which takes into account non-material components of the utility function. Assuming that expectations about future effort costs are correct ($k_e^l = k_a^l = k^l$) and using the same approach as in the main text, true overestimation of WTP would be characterized by

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

Definition 3 in the main text thus remains a sufficient condition for overestimation as long as $k^l < 25$. But it is straightforward to show that a subject will choose $y_1 > 0$ only if $k^l < 25$.[30] Hence, for any subject with nonzero effort, Definition 3 can still be used to flag overestimation.

D.2.2 Quadratic effort

Now assume instead that the effort function equals zero at zero correct submissions and otherwise exhibits linearly increasing marginal effort such that $e(y) = k^q y^2$, with $k^q > 0$.

[30] We would then assume that y_1 is subject to an upper constraint—due to the time limit—at which the agent would place themselves. This model may be a reasonable approximation of an effort function which increases quite slowly at low-to-medium effort and then exhibits an asymptotic "spike" near the limits of human ability to perform the task.

With this function, the risk-neutral expected utility function becomes

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

such that rational WTP is now given by

$$WTP = (y^{nt} - y^t)(25 - k^q(y^{nt} + y^t)) + PC$$

This is again clearly somewhat smaller than in Equation 2 of the main text; and again, Definitions 1 and 2 are not affected. For Definition 3, we again assume that expectations about future effort costs are correct ($k_e^q = k_a^q = k^q$), in which case

$$
\begin{aligned}
&WTP\,(\cdot_e) > WTP\,(\cdot_a) \\
&\iff 25\left((y_e^{nt} - y_e^t) - (y_a^{nt} - y_a^t)\right) - k^q\left[(y_e^{nt})^2 - (y_e^t)^2 - ((y_a^{nt})^2 - (y_a^t)^2)\right] > -(PC_e - PC_a)
\end{aligned}
$$

This condition implies that Definition 3 can no longer be used to identify overestimators. Indeed, since k^q is not elicited in the experiment, the magnitude of the LHS as a whole is unknown. Nevertheless, since $k^q > 0$, a more restrictive sufficient condition than Definition 3 can now be used to identify overestimators, namely the following:

Definition D.1. *Overestimator$_{3.1}$: a subject with:*

1. $(\hat{y}^{nt} - \hat{y}^t) - (y_1 - y_2) \geq 0$,
2. $(y^{nt})^2 - (y^t)^2 - (y_1^2 - y_2^2) \leq 0$, *and*
3. $v \geq 0$

with at least one strict inequality, and $WTP > 0$.

An analogous definition for underestimators reverses the direction of inequalities 1-

3. Within this smaller set of subjects, we again find a significantly higher proportion of undercommitters compared to overcommitters (4.2% vs. 0.2% of subjects, $p = 0.0001$).

D.3 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 *over*estimated 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 constant relative risk aversion (CRRA) utility.

Recall that in Section 3 of the main text, subjects maximize a utility function which only captures the BDM "lottery" and misses the second "lottery" faced by the subject: the risk of earning a lower amount if they succumb 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 14 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 an earlier but very similar experimental design (Ek and Samahita, 2020). As in the current study, we found a substantial proportion of overcommitters; however, there were more over- than undercommitters. The relative proportions of over- and undercommitters are, respectively, 22.02% and 5.78% by Definition 1, 17.69% and 14.80% by Definition 2, and 14.80% and 7.22% by Definition 3. Nevertheless, in line with the current experiment, the average WTP residual is -47.33 (t-test, $p < 0.0001$) by Definition 1 and -49.26 (t-test, $p < 0.0001$) by Definition 2—indicating that undercommitment may pose a larger problem, despite affecting fewer subjects.

In Ek and Samahita (2020), commitment removes a button that allows the subject to surf the internet during a (different) tedious lab task. 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:

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

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

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

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 (D.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.[31] 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 Definition 1 in the risk-neutral case). In equation (D.1), y^{nt} is thus interpreted as the *actual* number of correct answers when the subject is not exposed to temptation; as in the main text, we use y_1 as this counterfactual. (There 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; in Ek and Samahita (2020), this equals 1.4%. For subjects who do not click the button, $y^{nc} = y_2$, while y^c, the counterfactual had they done so, is obtained using the average productivity of subjects who do click, which is $y^c = 2$. In the same way, for subjects who succumb and click the button, $y^c = y_2$ 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 overestimators under different values of η are given in the first row of Table D.1. Around 17% of subjects are still considered to be 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). Nevertheless, our results on overestimation are robust to assuming CRRA with $\eta \leq 3$.

Table D.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 in relation to expected material loss (corresponding to Definition 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 D.1, we

[31] 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.

find a somewhat lower number of risk-averse subjects overestimate their WTP, compared to the case with actual material loss above.

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 Definition 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 Appendix D.4. 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 (D.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 $PC_e \geq PC_a$ by assumption), 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 $0 < WTP_e < 100$.[32]

As shown in row (3) of Table D.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, it should be clear that risk aversion runs in the direction of rationalizing what would otherwise look like *over*estimation of WTP. Thus, the overall conclusion of the present paper (i.e., that undercommitment dominates) is further supported by adding such concerns. However, based on the results just reported, the share of overestimators also appears unlikely to decrease very much.

[32] In principle, since both stated WTP and all parameters related to expected material losses are known, we might use them in (D.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.

D.4 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.