Excuse-Driven Present Bias*

Luca Flóra Drucker and Marc Kaufmann

Central European University

May 21, 2020

Preliminary draft, do not circulate

Click here for latest version

Abstract

We test whether people behave in a more present-biased way when they can excuse such behavior. In an experiment on Amazon Mechanical Turk, we elicit workers' willingness to work (WTW) today and to work two days later. We elicit this WTW when we provide low-excuse alternatives or high-excuse alternatives: while the low-excuse alternative always requires to work, the high-excuse alternative provides a plausibly higher degree for excuse-making by adding a 10% chance of not having to do extra work. We find that the WTW today drops by \$0.11 more than the WTW in two days when we move from the low-excuse to the high-excuse alternative, as if the high-excuse alternative was worth more when it allows avoiding working today. Our design rules out risk and time preferences that do not depend on other alternatives present. We describe a planned follow-up study with the goal of replicating our finding with excuses not based on risk.

1 Introduction

There are many reasons why people rightly postpone unpleasant activities. Take a software engineer who has a list of bugs to fix and features to implement. She may wait to hear back

*We thank Jonathan de Quidt, Lorenz Goette, Botond Kőszegi and Bertrand Verheyden for helpful comments and discussions. We thank CEU and Botond Kőszegi for funding part of the project.

from a client to see if the bug persists, check if her colleague still cares about the feature, or decide to leave work for tomorrow when her schedule is free. When applied coherently, these reasons apply with equal strength no matter whether they justify putting off *immediate* work or *future* work. We explore experimentally whether people additionally apply such reasons *asymmetrically* by acting as if these reasons provided stronger justifications when they allow putting off immediate rather than future work. Such asymmetric use of reasons is what we mean by excuse-making. Present bias, with its time-inconsistent nature, is a prime candidate for excuse-making, since excuses may allow people to leave free rein to their impulsive behavior.

This description highlights the main challenge: any excuse worth its name is built atop a reason that is perceived to be genuine, no matter how flimsy. Therefore when we compare the willingness to work for people with and without excuses, we should find that the former work less, because they have reasons to work less. The same reason should lead to the same change in willingness to work. Instead, we find that the willingness for immediate work drops by \$0.11 ($N = 147$, p-value 0.011) more than the willingness for future work, which we interpret as excuse-driven present bias.

We describe our within-subjects design to identify excuse-making based on this change in willingness to work in Section 2 and the implementation on Amazon Mechanical Turk in Section 3. We provide two types of alternatives for work: *low-excuse* and *high-excuse* alternatives. The low-excuse alternatives require subjects to do extra future work for sure, while the high-excuse alternative provides a 10% chance of not having to do the extra work. We took the idea of using a chance of no work as a possible excuse from Exley [2015], who finds that people use risk in charitable giving as an excuse to donate less to charity.

If, as our results in Section 4 suggest, people excuse their present bias, this suggests strongly that present bias – or present-oriented behavior – is context-sensitive: the same person may behave in more or less present-biased ways in superficially similar situations. For instance, the more dimensions there are to a choice, the more present-biased a person may act. In addition, having excuses for present bias may be what keeps people from learning that they are present-biased. This in turn could ensure that they remain *naive* about their own present bias, which is in line with several studies that find that most subjects are predominantly naive

(Augenblick and Rabin [2019], Fedyk [2018]). At the same time, it leaves room for learning in situations with clear feedback, such as in Le Yaouanq and Schwardmann [2019].[1] Most situations provide more excuses than our experimental setup: life offers more important and urgent tasks to do, colleagues more requests for help, and Netflix more movies to chill than any lab study can ever hope for. If excuses increase people's present bias, daily life where excuses abound will exacerbate this beyond what we find in the lab.

In Section 5, we summarize our plans for a follow-up lab study, which aims at replicating our current results and strengthening our identification on two fronts. The first issue is to complement risk as an excuse with other, cleaner excuses, to rule out all state-dependent preferences with risk and time interactions. The other excuse we are going to test in the follow-up study is *a different type of task*. The idea is that subjects may use the fact that the high-excuse alternative offers them a different task in the future as an excuse to choose future work instead of present work.

The second issue is that our MTurk subjects report working roughly 20 hours per week on MTurk. Suppose, to make the point most clearly, that subjects find our task exactly as unpleasant as other MTurk tasks, and can earn a fixed hourly wage from these other tasks. When deciding how much to work today rather than two days from now, these subjects would base their decisions on which day offers a higher payment per task. Their choices would thus be driven by maximizing earnings, and not involve any time preferences. Moving to a student population will alleviate this concern as students are more likely to substitute our task for unpaid leisure or studying, which are less direct substitutes (and less likely to be equally unpleasant).

2 Identification of excuse-driven present bias

Before diving into our experimental design in Section 3, let us explore how to identify excuses conceptually – in particular, how to separate out the excuse from the reason.

As mentioned in the introduction, we consider the degree to which one and the same reason

[1] O'Donoghue and Rabin [1999] highlight how important it is whether people are aware of their present bias (sophistication) or unaware of it (naivete).

is applied asymmetrically as excuse-making. The challenge is in ensuring that we keep the reason constant, while bringing out the asymmetry. Let us observe that people only make excuses if (i) there is a motivation – conscious or subconscious – to distort the choice; and (ii) there is a potential excuse available. Thus when there is no motivation to distort or no possibility to distort, we won't see excuse-making. Excuse-making will only happen when we create situations where we expect people to have a strong motivation to distort their choices – ones where they can avoid immediate effort – and where we expect people to have the possibility to distort.

The basic idea is therefore to offer people 4 choices, for each combination of motivation to distort ('yes' and 'no') and possibility of distorting ('yes' and 'no'). The hypothesis we are testing is whether present bias is something for which people have a motivation to distort their choices, which is why have 2 choices involving tedious tasks today as the high-motivation case, and 2 choices involving tedious tasks in the future as the low-motivation case. For now we assume that we have two alternatives, a *low-excuse alternative* and a *high-excuse alternative* that allow for small and large choice distortions respectively.

Concretely, we want to find the following indifference points X, $X^{'}$, Y, and $Y^{'}$:

Low level of excuses:

$$\begin{aligned} \text{work in the future} + \$X &\sim \text{low-excuse alternative} \\ \text{work in the present} + \$X' &\sim \text{low-excuse alternative} \end{aligned}$$

High level of excuses:

$$\begin{aligned} \text{work in the future} + \$Y &\sim \text{high-excuse alternative} \\ \text{work in the present} + \$Y' &\sim \text{high-excuse alternative} \end{aligned}$$

Then $\Delta_C = X' - X$ is the difference between willingness to work in the future and in the present when the possibilities for making excuses are low (control), while $\Delta_T = Y' - Y$ is the same difference when the possibilities for making excuses are high (treatment).

The low- and high-excuse alternatives we provide in our experiment are having to do extra

future tasks for sure; and having a 10% chance of not needing to do the extra future tasks. This is based on Exley [2015], who finds that people distort risky choices in charitable giving in a way that is beneficial to them. The main concern of the design is to take care of the difference in utility between the low-excuse and the high-excuse alternatives, since a drop in the probability of having to do work should lead to a change in willingness to work. To highlight how the four choices in our design allow us to take care of this concern, let us rewrite the indifference conditions in utility terms. Denote by $d(x)$ the disutility of doing x tasks, by β the present bias parameter in the low-excuse condition, and by β_E the present bias parameter in the high-excuse condition. Let x_1 be the amount of tasks people have to do in the left-hand side options, and x_2 the amount of tasks in the right-hand side (low-excuse or high-excuse) options. We do not think that people's utility literally changes. Rather we think of this as a reduced-form way to capture the idea that people behave as though they were more present biased in the high-excuse setting.

Low-excuse (control):

$$\begin{aligned} X &= \beta\, d(x_1) - \beta\, d(x_2) \\ X' &= d(x_1) - \beta\, d(x_2) \end{aligned}$$

High-excuse (treatment):

$$\begin{aligned} Y &= \beta_E\, d(x_1) - 0.90 \cdot \beta_E\, d(x_2) \\ Y' &= d(x_1) - 0.90 \cdot \beta_E\, d(x_2) \end{aligned}$$

where we assume for simplicity that people have expected utility preferences, although this is not necessary. Taking the difference between willingness for future work over present work in the two conditions, we get the following:

$$\begin{aligned} \Delta_C &= d(x_1) \cdot (1 - \beta) \\ \Delta_T &= d(x_1) \cdot (1 - \beta_E) \end{aligned}$$

We finally obtain the difference of these differences,

$$\Delta\Delta = \Delta_T - \Delta_C = d(x_1) \cdot (\beta - \beta_E)$$

which equals zero when $\beta_E = \beta$. This holds for any risk preferences that are state independent, since the first differences wash out any state independent effect.

More is true. Even when we have state-*dependent* preferences, such as when the foregone option affects the choice, we can only get non-zero effects if this state-dependent effect is stronger when the alternative is work today rather than work in the future.[2] If, however, the high-excuse alternative affects the two dates differently, making work in the future more attractive, then $\beta_E < \beta$, and $\Delta\Delta$ will be positive. Our primary hypothesis is therefore that $\Delta\Delta > 0$, which we interpret as excuse-driven present bias. An additional way we plan to measure the size of the distortion is to consider how much it distorts the magnitude of the β-parameter an econometrician would estimate. Although excuses might be due to distortions along many dimensions – from distorting the probabilities, the disutility of doing tasks, or present bias – we capture the degree of distortion by considering how much it affects the estimate of the present bias parameter β. Next, we describe the experimental setting, using risk as an excuse, and in Section 5, we talk about other possible candidates for an excuse.

3 Experiment design

3.1 Technical details

We implemented the experiment in oTree [Chen et al., 2016] and ran it on the Amazon Mechanical Turk (MTurk) platform. All details were pre-registered in the AEA RCT Registry [Drucker and Kaufmann, 2019]. We ran the first pilot of the actual experiment in August 2019, with 43 participants who completed the experiment. We ran the main experiment in September 2019 with 147 participants completing it. Subjects started with a tutorial and

[2]While we are unaware of any such preferences that would explain our results, we think that excuses that circumvent the whole issue of risk by leveraging non-risk based excuses in the first place is cleaner. See our discussion in Section 5.

description to introduce subjects to the task. After the tutorial, they had the opportunity to sign up. Every participant who signed up completed one session on that same day and a second session two days later.

3.2 Payments

On average participants earned $\approx$ \$15, broken down as follows:

- \$1.50 for the tutorial
- \$1.00 for completing session 1
- \$10.00 for coming back and completing session 2 (and thus the study)
- ~\$3.00 in bonus payments based on choices to do additional work

We paid all subjects after session 2, within 2-3 days, providing all payments at once, even if subjects dropped out early. Thus dropping out of the study did not lead to early payments. For similar reasons, subjects who completed extra work in the first session did not receive the extra bonus payment unless they also completed session 2. Instead they they received a flat bonus of \$1.00 if they completed the first session, independent of the choices they made. In this way, we rule out that subjects who think that they may not come back for session 2 choose to do the extra work today in order to get extra money they would not get if they chose to work in the future and failed to come back.

3.3 Timeline

Table 1 shows the timeline of the experiment.

Table 1: Timeline of the experiment

	Session 0	**Session 1**	**Session 2**	**Payment**
Tasks	warm-up rounds of effort task	(right after Session 0) effort choices	(2 days after session 1) effort task	(within 2 days after Session 2)

Session 0	Session 1	Session 2	Payment
consent debrief survey	effort task		

3.3.1 The Tutorial

Session 0 is the tutorial[3] which described the study and required subjects to do 6 of our tasks to familiarize them with it. As is common on MTurk, there are many Workers who sign up but don't complete the study. By only signing subjects up after the tutorial, we reduced attrition in the crucial part of our experiment substantially. The experimental task consists of correctly counting the 1's in a matrix of 1's and 0's, and takes between 30-60 seconds for most people. When someone gives a wrong answer, we provide a new matrix, to avoid guessing repeatedly to get the right answer.[4]

There are two sizes of matrices in the experiment, a small one (7x12 cells), and a large one (10x15 cells), and subjects completed 3 of each in the tutorial. Figure 1 shows an example of a large matrix. At the end of the HIT, Workers completed a debrief survey about clarity of our instructions, how unpleasant they find both types of matrices, and how much time the spend working and how much they earn on MTurk per week. We also elicited a survey measure of patience, used by Falk et al. [2018][5]. Every Worker who completed the HIT, received $1.50 even if they didn't sign up for the study. Those who signed up could start the experiment right away.

3.3.2 The Main Sessions

If MTurk Workers decided to join the study, they received detailed instructions, followed by a comprehension check. Participants could only move on after giving correct answers. [6] Conditional on this, they started the actual study. In both sessions 1 and 2, participants had to count 25 matrices. The choices they made determined how many of these 25 matrices were small matrices (7 by 12) and how many of them were large ones (10 by 15). Figure 2 shows an example of a price list where

[3]This was the MTurk HIT advertised on MTurk. HIT is the acronym for Human Intelligence Task, which is one job to complete by a Worker on the MTurk platform

[4]Additionally to this, we allowed subjects to only get a certain number of matrices wrong, to avoid repeatedly entering the same number until a matrix pops up that has this number of 1's. We observed 1 subject in our pilot who we think followed this strategy and we wanted to avoid it.

[5]The question is "How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future? 0 means not willing at all, 10 means very willing"

[6]We highlighted the wrong answers, so all subjects could get it right with enough tries.

0	1	1	1	0	1	00	11	1	1	11	0	0	0	1
1	11	1	1	1	11	1	0	1	0	0	1	0	0	10
1	1	1	1	11	1	1	0	1	00	11	1	1	1	0
0	1	0	1	11	1	10	1	0	1	1	1	1	00	10
0	0	00	10	11	00	1	10	1	0	1	1	0	00	11
00	11	10	1	00	1	00	0	0	0	11	1	0	1	00
10	10	1	10	1	1	1	0	1	10	1	0	0	11	11
0	0	1	1	1	0	10	0	1	1	11	10	1	1	10
1	1	1	1	0	0	0	0	11	1	0	1	11	0	0
11	0	1	0	1	1	1	1	1	1	1	1	1	0	1

Figure 1: Example of a large matrix

subjects choose between 22 large matrices today for a $3.00 bonus or 20 large matrices today for a bonus ranging from $0.00 to $3.40, increasing in $0.20 increments.

Participants faced the same 15 price lists in a random order (randomized for every participant) and they knew that we'd pick one choice from one of the pages at random. If the implemented choice involved uncertainty (e.g., "20 large matrices in Session 2 with 90% probability"), then we resolved the uncertainty on the day the work was potentially due, right before subjects had to do the work. Participants then completed the work for session 1, after which we gave them the link to come back in 2 days to complete session 2. We also sent a reminder email on the day of session 2. In session 1, we additionally asked participants after they made their choices how they went about making these choices.

3.3.3 Choosing large matrices rather than extra matrices

We decided to let participants choose the number of hard tasks, rather than the number of extra tasks to avoid making subjects choose primarily based on the extra time taken and get them to think more about the difference in unpleasantness. The primary reason for doing this is that our subjects spend a lot of time working on MTurk (roughly 20 hours per week), and may have developed heuristics based on time it takes to do a task. Even more problematic is the case where MTurkers consider our task to be roughly as tedious as other tasks on MTurk, in which case they would primarily decide based on whether we pay more or less per hour. In that case, a choice to work less today might be driven not by present bias but by the fact that that the hourly wage offered for today is lower than the hourly wage offered in the future, given that the worker may already have decided to work several hours each

Add $3.00 to bonus; 22 large matrices today	Add $0.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $0.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $1.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.40 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.60 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $2.80 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.00 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.20 to bonus; 20 large matrices today
Add $3.00 to bonus; 22 large matrices today	Add $3.40 to bonus; 20 large matrices today

Figure 2: Example of a price list

of those days. In order to make such thinking less likely, we decided to let subjects choose the number of large tasks, rather than extra tasks. We have no proof that this worked, which is part of our reason for planning a follow-up study in the lab (see Section 5 for details), since there is less room for such task substitution in a lab setting.

3.4 Implementation of risk as an excuse

We test for excuse-driven present bias in two batches of four choices.[7] In each batch, we use the switching point in a price list as the indifference point, which gives us 4 inferred indifference points corresponding to X, X', Y, and Y' as described in section 2. Our pre-registered hypothesis was that $\Delta\Delta = (Y' - Y) - (X' - X) > 0$. For example, the $\Delta\Delta$ we get in batch 1 is the following:

$$\Delta\Delta = d(20) \cdot (\beta - \beta_E)$$

(see the exact choices in the Appendix).

Based on Exley [2015], our hypothesis is that in the high-excuse condition, people's willingness to work in the future decreases less than the willingness to work in the present, leading to a $\beta > \beta_E$. If this is the case, we will observe that $\Delta\Delta$ is positive. In the next section, we present the results for this hypothesis.

4 Experiment Results

We asked participants to rate the large and small tasks on a 10-point scale, comparing them to other tasks on MTurk. Participants on average rated the large matrix as significantly less pleasant (diff = 1.42 points) than the small matrix.[8] This shows that subjects do find the large matrix less pleasant, although some participants stated clearly that they didn't mind in their description of how they chose.

Out of the 154 participants who completed Session 1, 147 also completed session 2 and thus the experiment. Therefore, between the two sessions, attrition was only 4.5%. For the analysis of choices, we use the sample of those who completed all aspects of the experiment, to exclude those who, at

[7] 8 of the price lists correspond to the 2 batches, 4 choices correspond to a third batch – in which we later realized that we made a mistake in the options, so identification is not possible in that batch –, and the remaining 3 price lists provide extra data for identifying β and direct choices between low-excuse and high-excuse alternatives.

[8] We asked this question at the end of Session 0, in a debrief survey. See other debrief survey statistics in the Appendix.

the point of making choices, might have already known they would not finish it all. Only 1.7% of the choices had multiple switching points. We excluded these choices and the other choices in that batch from the analysis, as we could not infer an indifference point from them.[9] We present the results for our main hypothesis in this section. First, we show evidence for excuse-driven present bias in batches 1 and 2. Then, as a robustness check, we show that the effect we find cannot be driven by concave utility over money.

4.1 Main results

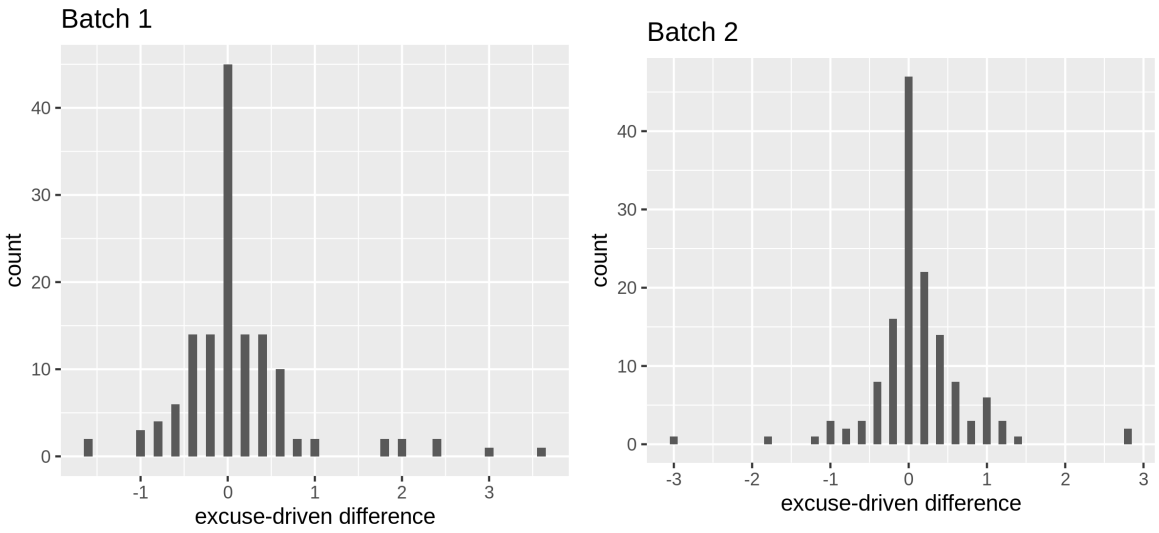

Figure 3: Excuse-driven difference in WTW in two days over WTW today

Figure 3 shows the distribution of the excuse-driven increase in the willingness to work in two days over willingness to work today ($\Delta\Delta$). Our main pre-registered test is a two-sided t-test on $\Delta\Delta$ for batches 1 and 2 jointly (although we report the individual t-tests too).[10] Specifically, we run the following regression, separately by batch, and then with batches 1 and 2 combined:

$$\Delta\Delta_{i,b} = \alpha + \varepsilon_{i,b}$$

Table 2 shows the results for this simple regression. Columns 1 and 2 correspond to the batches

[9] As our outcome of interest, $\Delta\Delta$, is a difference in the differences of such indifference points, we had to exclude all four choices in a batch if there was at least one choice with multiple switching points.

[10] Our hypothesis suggests a one-sided t-test, however we decided against pre-registering it as such since one-sided tests tend to be frowned upon. Alternatively our test can be interpreted as a one-sided t-test with 2.5% significance level.

separately, while in column 3 the two batches are pooled together. In column 3, standard errors are clustered at the participant level, since we have 2 observations per individual, one for each batch. The results show that there is an \$0.11 increase in the difference between willingness to work in two days vs today in the high-excuse condition. This is roughly 0.2 standard deviations of the difference between willingness to work in two days vs today. As described in Section 2, state-independent risk preferences cannot explain the result, nor are we aware of any existing theory about risk and time interactions that would explain it. One possible caveat could be concave utility over money. We rule out this explanation in the next section.

Table 2: Results: t-tests

	$\Delta\Delta_{i,b} = \alpha + \varepsilon_{i,b}$		
	Batch 1	Batch 2	Batches 1&2
	(1)	(2)	(3)
Constant	0.113*	0.106**	0.110**
	(0.062)	(0.053)	(0.043)
Observations	138	141	279
R^2	0.000	0.000	0.000
Note:			*p<0.1; **p<0.05; ***p<0.01
			Standard errors in parentheses,
			batches 1&2: clustered at the individual level

4.2 Ruling out concavity of money

Concave utility over money might also lead to $\Delta\Delta > 0$ depending on the choices offered.[11] Suppose that for a participant, $Y' = 9$, $Y = 6$, $X' = 2$, and $X = 0$. Then, $Y' - Y = 3$ and $X' - X = 2$, which would lead to $\Delta\Delta = 1 > 0$. However, with concave utility over money the increase from \$0 to \$2 and from \$6 to \$9 could be the same. To rule this out, we chose the payments such that for most participants we expected $X \geq Y$. In this way, concavity of money would if anything work against us finding an effect, pushing $\Delta\Delta$ down. Table 3 shows the results for a restricted sample of those for whom $X \geq Y$, so for whom the results cannot be explained by concave utility over money. The results for these participants are even stronger, for them, the excuse-driven increase in the willingness to work in two days over work today is \$0.19.[12]

[11] Over these small stakes, concave utility over money might be the result of loss aversion, or various types of framing.

[12] We have no explanation why the effect is stronger on this subset.

Table 3: T-test equivalents, ruling out concavity of money

	$\Delta\Delta_{i,b} = \alpha + \varepsilon_{i,b}$	
	Batches 1&2, all	Batches 1&2, $X \geq Y$
	(1)	(2)
Constant	0.110**	0.190***
	(0.043)	(0.049)
Observations	279	214
R^2	0.000	0.000
Note:		*p<0.1; **p<0.05; ***p<0.01
		Standard errors in parentheses,
		clustered at the individual level

Although we are not aware of any such theory, some form of risk and time interactions with state-dependence (other than excuse-making) might explain our results. To overcome these issues, we are planning a follow-up study in a laboratory with students. We turn to the description of this lab experiment, and to reasons why we chose to strengthen our resuls in the lab in the next section.

5 Design of follow-up lab study

We choose the lab due to several drawbacks of MTurk. To measure time preferences, we need a task or a consumption good that participants cannot substitute easily with other tasks they perform regularly. The MTurk Workers in our sample report working on average 20 hours weekly on MTurk, with tasks that may be close substitutes to our tasks. If our tasks are as unpleasant as other MTurk tasks, MTurkers may choose based on which tasks pay more, including whether we pay more per task in session 1 or session 2. Our participants find the small matrix similarly pleasant to other MTurk tasks (4.85 on a 0-10 scale)[13], while they considered the large matrix to be less pleasant (3.43), which suggests that we partly solved this issue with the two sizes of matrices. However the difference in pleasantness is not huge, and many subjects reported not caring differently about the small or large matrices. Students in the lab are less likely to substitute our tasks for equally tedious and paid tasks, which makes it a better subject pool for eliciting time preferences.

We additionally want to replicate our findings for other potential excuses, rather than only for risk. Our idea is that we introduce two different types of tasks that are commonly used as effort tasks in

[13] 0 was the least and 10 the most pleasant, while 5 was equally pleasant

the experimental literature – the matrix counting task, and a blurry Greek letter transcription task.[14] We then use a similar difference-in-differences strategy as with risk to identify excuse-making, using four choices. In the low-excuse condition people make choices about a *baseline* task, say, the matrix counting task. In the high-excuse condition, they have the option to choose another type of task – the blurry Greek letters task – for future work. We vary the baseline task between subjects, and across weeks within subjects.

Low-excuse condition:

$$\begin{aligned} x_1 \text{ matrices in 2 days} + \$X &\sim x_2 \text{ matrices in 2 days} \\ x_1 \text{ matrices today} + \$X' &\sim x_2 \text{ matrices in 2 days} \end{aligned}$$

High-excuse condition:

$$\begin{aligned} x_1 \text{ matrices in 2 days} + \$Y &\sim x_2 \text{ Greek tasks in 2 days} \\ x_1 \text{ matrices today} + \$Y' &\sim x_2 \text{ Greek tasks in 2 days} \end{aligned}$$

The excuse-driven change in the willingness to do 10 matrices in 2 days over today is then, analogously to our identification with risk in Section 2, denoting the disutility of doing x matrices with $d_M(x)$:

$$\Delta\Delta = \Delta_T - \Delta_C = d_M(x_1) \cdot (\beta - \beta_E)$$

The idea is that if people want to choose to work in the future, they can rationalize their choice by saying that they chose that option because it offered a different task. However, our design, by using the asymmetry in the choices, allows to distinguish just prefering the other type of task from indeed using the other task as an excuse to choose to work in the future.[15]

A final benefit of moving to the lab is the difficulty of tracking MTurk workers and following up with them, as the terms of service require us to only get in touch with workers through their platform. In the lab we can ask for phone numbers as well as emails and send text messages as reminders as well as emails. First, this allows us to have more days between sessions. This may matter, as Augenblick [2018] finds a few hours' β of 0.94, a daily β of 0.91 and a weekly one of 0.87. This suggests that we might have larger effects – and hence more statistical power – if we have more days between session

[14] This task was used e.g. in Augenblick and Rabin [2019].

[15] Hiding the choice of future tasks behind the choice of a different type of task is close to the implicit preferences framework presented in Cunningham and de Quidt [2016]. We plan to complement our identification with a test for implicit preferences for work in the future as described in their paper.

1 and session 2. To gain more data points, and to let participants become familiar with the setting, we plan to run the experiment for three to four weeks, with two sessions each week. Participants will make choices in the first session of each week, about work in the next week's two sessions and the current week's two sessions, to allow us to differentiate between present bias and exponential time discounting [as in Augenblick et al., 2015].

References

Ned Augenblick. Short-term time discounting of unpleasant tasks. July 2018. URL http://faculty.haas.berkeley.edu/ned/Augenblick_ShortTermDiscounting.pdf.

Ned Augenblick and Matthew Rabin. An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, 86(3):941–975, 2019.

Ned Augenblick, Muriel Niederle, and Charles Sprenger. Working over Time: Dynamic Inconsistency in Real Effort Tasks *. *The Quarterly Journal of Economics*, 130(3):1067–1115, 05 2015. ISSN 0033-5533. doi: 10.1093/qje/qjv020. URL https://doi.org/10.1093/qje/qjv020.

Daniel L Chen, Martin Schonger, and Chris Wickens. otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

Tom Cunningham and Jonathan de Quidt. Implicit preferences inferred from choice. *Available at SSRN 2709914*, 2016.

Luca Flóra Drucker and Marc Kaufmann. Excuse-driven present bias. *AEA RCT Registry*, September 2019. doi: 10.1257/rct.4010.

Christine L Exley. Excusing selfishness in charitable giving: The role of risk. *The Review of Economic Studies*, 83(2):587–628, 2015.

Armin Falk, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. Global Evidence on Economic Preferences*. *The Quarterly Journal of Economics*, 133(4):1645–1692, 05 2018. ISSN 0033-5533. doi: 10.1093/qje/qjy013. URL https://doi.org/10.1093/qje/qjy013.

Anastassia Fedyk. Asymmetric naivete: Beliefs about self-control. *Available at SSRN 2727499*, 2018.

Yves Le Yaouanq and Peter Schwardmann. Learning about one's self. 2019.

Ted O'Donoghue and Matthew Rabin. Doing It Now or Later. *The American Economic Review*, 89 (1):103–124, March 1999. ISSN 0002-8282.

A Appendix

A.1 Choices in batches 1 and 2

Batch 1:

1. 20 large matrices in Session 2 or 5 large matrices in Session 2 + \$0
2. 23 large matrices in Session 1 or 5 large matrices in Session 2 + \$0
3. 20 large matrices in Session 2 or 19 large matrices in Session 2 with $p = 0.9$ + \$0.20
4. 23 large matrices in Session 1 or 19 large matrices in Session 2 with $p = 0.9$ + \$0.20

Batch 2:

1. 22 large matrices in Session 2 or 15 large matrices in Session 2 + \$1.80
2. 25 large matrices in Session 1 or 15 large matrices in Session 2 + \$1.80
3. 22 large matrices in Session 2 or 15 large matrices in Session 2 with $p = 0.9$ + \$0.60
4. 25 large matrices in Session 1 or 15 large matrices in Session 2 with $p = 0.9$ + \$0.60

A.2 Debrief survey statistics

Table 4 shows summary statistics from the debrief survey for those who gave consent but not completed the experiment (attritors) and those who completed the whole experiment (completors), separately. Most participants found the instructions clear, and found the large matrix significantly less pleasant than the small matrix. Interestingly, completors hate the large matrix more than attritors, but the two groups are similar in other aspects. They work on average $\sim$ 20 hours weekly on MTurk, and earn $\sim$ 132 dollars per week from this work.

Table 4: Summary statistics from debrief survey

	(1) Attritors		(2) Completors		(3) Difference	
	mean	sd	mean	sd	diff	t
Instructions are clear	0.96	(0.20)	0.96	(0.20)	0.00	(0.00)
How pleasant is small matrix (0-10)	5.16	(2.83)	4.85	(2.15)	0.31	(0.71)
How pleasant is large matrix (0-10)	4.78	(3.14)	3.43	(2.33)	1.35***	(2.76)
How much worse is large matrix	0.39	(1.90)	1.42	(1.86)	-1.03***	(-3.31)
Patience (0-10)	7.37	(2.32)	7.22	(1.97)	0.14	(0.39)
Weekly working hours on MTurk	18.82	(15.19)	20.29	(13.93)	-1.47	(-0.60)
Weekly earnings on MTurk	118.43	(89.78)	135.63	(99.60)	-17.20	(-1.13)
Observations	49		147		196	

Note: Attritors are those who gave consent to participate, but did not finish the experiment, Completors are who completed all aspects of the experiment. In rows 2-3 ("How pleasant is small matrix" and "How pleasant is large matrix") 0 means very unpleasant and 10 means very pleasant compared to other tasks they do on MTurk. Row 4 is calculated as row 2 - row 3 individually. In row 5 ("Patience") 0 means very impatient and 10 means very patient. Weekly earnings on MTurk are in US dollars.
* p<0.1 ** p<0.05 *** p<0.01